International Encyclopedia of Ergonomics and Human Factors

Edited by
Waldemar Karwowski

VOLUME I

International Encyclopedia of Ergonomics and Human Factors

International Encyclopedia of Ergonomics and Human Factors

Edited by
Waldemar Karwowski

VOLUME I

Every effort has been made to ensure that the advice and information in this book is true and accurate at the time of going to press. However, neither the publisher nor the authors can accept any legal responsibility or liability for any errors or omissions that may be made. In the case of drug administration, any medical procedure or the use of technical equipment mentioned within this book, you are strongly advised to consult the manufacturer's guidelines.

To Bernardette,

My love and inspiration.

WK

Contents

Part 1 General Ergonomics

Part 2 Human Characteristics

Part 3 Performance Related Factors

Part 4 Information Presentation and Communication

Part 5 Display and Control Design

Part 6 Workplace and Equipment Design

Part 7 Environment

Part 8 System Characteristics

Part 9 Work Design and Organization

Part 10 Health and Safety

Part 11 Social and Economic Impact of the System

Part 12 Methods and Techniques

About the Editor

Waldemar Karwowski, Ph.D., P.E. is Professor of Industrial Engineering and Director of the Center for Industrial Ergonomics at the University of Louisville, Louisville, Kentucky, USA. He holds an M.S. (1978) in Production Engineering and Management from the Technical University of Wroclaw, Poland, and a Ph.D. (1982) in Industrial Engineering from Texas Tech University. He is also a Board Certified Professional Ergonomist (BCPE). His research, teaching and consulting activities focus on prevention of work-related musculoskeletal disorders, workplace and equipment design, human and safety aspects of advanced manufacturing, and theoretical aspects of ergonomics science.

Dr Karwowski is currently (2000-2003) President of the *International Ergonomics Association* (IEA), and is past President of the *International Foundation for Industrial Ergonomics and Safety Research*, as well as past Chair of the US TAG to the ISO TC159: Ergonomics/SC3 *Anthropometry and Biomechanics*. He is the author or co-author of over 200 scientific publications in the area of human factors engineering, industrial ergonomics and safety, mathematical modeling in human factors, and fuzzy systems, and has edited 25 books.

Dr Karwowski serves as Editor of the *Human Factors and Ergonomics in Manufacturing* international journal, Co-Editor of the *International Journal of Occupational Safety and Ergonomics*, and Consulting Editor of the *Ergonomics* journal. He is also a board member for *Human Factors, Applied Ergonomics, International Journal of Human-Computer Interaction, Universal Access to the Information Society: An International Interdisciplinary Journal, Occupational Ergonomics, Industrial Engineering Research: An International Journal of IE Theory and Application* (Hong Kong).

Dr Karwowski is also the Editor-in-Chief of the new journal *Theoretical Issues in Ergonomics Science* (TIES), launched in 2000, and designed to stimulate and develop a theoretical basis for the unique science of ergonomics.

Dr Karwowski is Fellow Member of the Human Factors and Ergonomics Society and recipient of the highest recognition in occupational safety and health in Poland, *Pro Labore Securo* (2000). He received the University of Louisville Presidential Award for Outstanding Scholarship, Research and Creative Activity in the Category of Basic and Applied Science (1995), and the W. Jastrzebowski Medal from the Polish Ergonomics Society (1995). He served as Fulbright Scholar, Tampere University of Technology, Finland (1990-1991), and was named an Outstanding Young Engineer of the Year by the Institute of Industrial Engineers (1989).

Board of Associate Editors

International Scientific Advisory Board

Erik Hollnagel, OECD Halden Reactor Project, Norway
Robin Herron, Colorado State University, USA
Jean-Michel Hoc, Université de Valenciennes et du
    Hainaut-Cambrésis (UVHC), France
William C. Howell, Arizona State University, USA
Andy Imada, University of Southern California, USA
Ed W. Israelski, Ameritech Co., USA
Toni Ivergard, Sweden
Steven L. Johnson, University of Arkansas, USA
Marilyn S. Joyce, Joyce Institute/Arthur D. Little, USA
Barry H. Kantowitz, Battelle Co., USA
Bronislaw Kapitaniak, Université Pierre et Marie Curie, France
Halimahtun Mohd Khalid, Universiti Malaysia Sarawak, Malaysia
Lise Busk Kofoed, Aalborg University, Denmark
Kazutak Kogi, Institute for Science of Labour, Japan
Ernst Koningsveld, Dutch Institute for Working Conditions, The
    Netherlands
Jefferson M. Koonce, University of Central Florida, USA
Danuta Koradecka, Central Institute for Labour Protection,
    Warsaw, Poland
Kenneth R. Laughery, Rice University, USA
Jerzy Lewandowski, Technical University of Lodz, Poland
Tom Leamon, Liberty Mutual Insurance Co, USA
Kee Yong Lim, Nanyang Technological University, Singapore
John Long, University College London, UK
Jacques Malchaire, Université Catholique de Louvain, Belgium
Thomas B. Malone, Carlow International, USA
Nicolas Marmaras, National Technical University of Athens, Greece
Markku Mattila, Tampere University of Technology, Finland
Najmedin Meshkati, University of Southern California, USA
Deborah A. Mitta, Georgia Tech Research Institute, USA
Pedro R. Mondelo, Universita Politecnica de Catalunya, Spain
Neville P. Moray, University of Surrey, UK
Ben B. Morgan Jr., University of Central Florida, USA
Mitsuo Nagamachi, Kure National College of Technology, Japan
Celestine A. Ntuen, North Carolina A&T State University, USA
Clas-Hakan Nygard, University of Tampere, Finland
Richard W. Pew, BBN Technologies, USA
Falzon Pierre, Conservatoire National des Arts et Métiers, France
Pieter Rookmaaker, AMG/Ergonomics, The Netherlands
Susumu Saito, NIOSH, Japan
Mark Sanders, California State University, USA
Steven L. Sauter, NIOSH, USA
Bente Schibye, National Institute of Occupational Health, Denmark
Pat Scott, Rhodes University, South Africa
Rabindra Nath Sen, Universiti Telekom, Malaysia
Joseph Sharit, University of Miami, USA
Barbara Silverstein, University of Washington, USA
Juraj Sinay, Technical University of Kosice, Slovakia
Gisela Sjogaard, Odense University, Denmark
James L. Smith, Texas Tech University, USA
Neville Stanton, Southampton University, UK
Leon Straker, Curtin University of Technology, Australia
Carol Stuart-Buttle, Stuart-Buttle Ergonomics, USA
Kazusige Suzuki, HQL, Japan
Jukka Takala, ILO, Switzerland
Harold Van Cott, Van Cott and Associates, USA
Kamiel Vanwoterghem, CERGO International, Belgium
Valery Venda, California School of Professional Psychology, USA

Max Vercruyssen, University of Hawaii, USA
Kim J. Vicente, University of Toronto, CA
Joel Warm, University of Cincinnati, USA
Richard Wells, University of Waterloo, Canada
Michael Wogalter, North Carolina State University, USA
Klaus J. Zink, University of Kaiserslautern, Germany

Foreword

Ergonomics is the discipline concerned with interactions among humans and other elements of a system (the term 'system' is intentionally vague and ranges from simple tools to complex sociotechnical structures). As a science, ergonomics is relatively young; however, the underlying ideal, human-centred technology, has played a pivotal role in the evolution of human society from its very start. The use of stone tools by *homo habilis*, over two million years ago, represents the dawn of ergonomics as an artform. Throughout the history of civilization, technological innovations were motivated by fundamental human aspirations for security, prevalence and self worth, and by problems arising from human-system interactions. In the 20th century, ergonomics emerged as a formal science in its own right, though it continues to be taught as part of other faculties such as engineering or psychology. Although it has been referred to as the science of work, ergonomics, in its broadest sense, is concerned with all forms of human activity. The application of the scientific method to the development of theory, principles and data relevant to design has had enormous impact on the reduction of user error, improved performance, reduction of occupational injuries and worker discomfort, increased usability and safety of computer systems and consumer products.

Today, profound changes are taking place that touch all aspects of our society: changes in the nature of work and play; changes in global commerce and communication; changes in science and technology; and changes in population migration and world demographics. These changes cannot but influence the future course of ergonomics since they relate to how people interact with technology in an increasingly dynamic and complex world. More importantly, they beckon ergonomics to play a more direct and vital role in shaping the society of the future.

An encyclopedia of ergonomics and human factors is crucial to the further development of the science and its applications. It serves to inform practitioners, educators, students and researchers in the field. As well, it is a useful resource for those not directly involved with ergonomics, but who want to understand one or more aspects of human–system interactions. It is a foundation of knowledge that serves all who have an interest in the field.

Preparing an encyclopedia in any discipline is a daunting task; preparing one in ergonomics and human factors is heroic. Part of the reason is that ergonomics is a very broad field, whose scope includes physical, cognitive, and organizational topics. Moreover, although the science of ergonomics is differentiated from other disciplines by the fact that it is concerned exclusively with the design of human–system interactions, it relies heavily on knowledge from related fields such as psychology, physiology, engineering, medicine, sociology, anthropology and kinesiology. That is, in addition to knowledge content unique to ergonomics, researchers and practitioners routinely apply knowledge from the relevant biological, behavioural and engineering sciences. It is clearly not possible for an encyclopedia of ergonomics to cover all of the related disciplines whose theories and data it borrows, adapts, and extends. Yet, it is similarly not practicable to exclude content from related sciences that is central to an understanding of human interactions with technology. The editors have, therefore, had to make difficult decisions about what to include and what to exclude in their effort to make the encyclopedia as comprehensive and informative as possible. They have clearly attained this goal in this unprecedented volume.

Another challenge that faced the editors was the fact the ergonomics is continually evolving. As an applied science, ergonomics evolves with technology and economic diversification. By necessity, an encyclopedia such as this reflects current directions and state-of-knowledge. Brian Shackle has identified the major thrusts of ergonomics applications in the 20th century as: military ergonomics, industrial ergonomics, consumer product ergonomics, human–computer interaction and software ergonomics, and cognitive ergonomics and organizational ergonomics. To this, Martin Helander has suggested adding eco-ergonomics as the main thrust of the first decade of the 21st century. The editors of this encyclopedia, have, therefore, had to strike a balance between covering the theoretical and methodological underpinnings of the field and elaborating topics closely aligned with major applications. Again, the balance achieved provides a rich scientific and technical resource yet one that retains strong relevance to practitioners.

Producing the first edition of an encyclopedia is unquestionably the most challenging for obvious reasons. Future editions will undoubtedly build on this work, update it and add significant new knowledge. However, the first edition represents a landmark, signifying the true coming of age of ergonomics. It is thus a timely and notable contribution to the field.

The editor, authors and the publisher are to be commended for undertaking the arduous task of producing such a fine resource. It will serve the discipline and profession well into the future.

Y. Ian Noy
President (1997–2000)

Acknowledgements

Part 1
General Ergonomics

An Annotated Review of Selected Military and Government Sources of Human Factors Design Related Criteria

W. F. Moroney† and J. A. Cameron‡

†Human Factors Program, University of Dayton, Dayton, OH, 45469-1430. USA

‡Crew System Ergonomics/Human Systems Technology Information Analysis Center (CSERIAC), Wright-Patterson AFB, OH, 45433-7022, USA

1. INTRODUCTION

The purpose of this chapter is to transfer the human factor design related criteria developed by US government agencies to non-government personnel. While many readers are familiar with ISO (International Standards Organisation), ANSI (American National Standards Institute) and SAE (Society of Automotive Engineers) standards, they are not familiar with design standards developed primarily by military agencies or agencies with a specific focus, such as the FAA (Federal Aviation Administration) and NASA (National Aeronautics and Aerospace Administration). Readers seeking design criteria, principles and practices for use in the design and development of systems, equipment and facilities will find these standards useful. Please note that since we have focused on design-related criteria we did not include guidelines or regulations developed by agencies such as NIOSH (National Institute of Occupational Health and Safety) or OSHA (Occupational Health and Safety Agency).

How are these criteria developed? Essentially, these standards were developed by committee and represent (1) the best of what was known at the time of their development and (2) the consensus arrived at by the committee. In some cases, an arbitrary decision is required. For example, luminance contrast (C) can be defined as either $(L_{max} - L_{min})/L_{min}$ or $(L_{max} - L_{min})/L_{max}$. The former formula produces whole numbers, while the latter formula produces values between 0 and 1.0. If luminance contrast is to be specified as design criteria then only one definition is acceptable, and DoD representatives elected to use the former definition.

The five major documents described here are:
- MIL-HDBK 46855: Human Engineering Program Processes and Procedures.
- MILSTD 1472: Human Engineering design criteria for military systems, equipment and facilities.
- Department of Defense (DoD) Human Computer Interface Style Guide.
- Federal Aviation Administration (FAA) Human Factors Design Guide (HFDG) for Acquisition of Commercial-off-the-Shelf-Subsystems, Non-developmental Items and Developmental Systems.
- NASA-STD-3000: Man–Systems Integration Standards.

Readers may also wish to consult the excellent British Defence Standard 00-25: Human Factors for Designers of Equipment. Defence Standard 00-25, is comprised of 13 parts and designed to "be viewed as a permissive guideline, rather than a mandatory piece of technological law" (vol. 1, p. 1). The text, written for designers with a variety of technical backgrounds, includes both general background information and human factors data. Most parts include definitions of terms relevant to that specific part, references and/or related documents, and sources for obtaining related documents. The 13 parts of this standard were issued over a period of more than ten years. Interested readers can download BDS STD 00-25 from the WWW site provided in the references.

2. MIL–HDBK 46855: HUMAN ENGINEERING PROGRAM PROCESSES AND PROCEDURES

MIL-HDBK 46855 is a DoD guidance document that describes the analysis, design and development, and test and evaluation considerations expected in a Human Engineering (HE) Program. It describes: (1) the program tasks, (2) the procedures and preferred practices useful in implementing the guidance and (3) selected HE methods. The handbook of approximately 300 pages consists of eight sections.

- 1. Scope: describes both the document in general and its applicability and tailoring (using selected portions in developing the HE program for a system under development).
- 2. Applicable Documents: lists relevant DoD standards, specifications, and handbooks as well as other pertinent government documents.
- 3. Definitions: defines the acronyms used in this HDBK.
- 4. Program Tasks: provides general program task guidance that the system developer can incorporate into the HE Program Plan. Issues addressed include, among others, risk management, non-duplication, analyses options, procedure development, and failure and error analysis.
- 5. Significance of HE for Program Acquisition: describes the role of HE in systems acquisition, including Human System Integration (HSI) and Manpower, Personnel and Training (MPT) implications. The range of HE activities are described, as well as the value of HE (descriptions of benefits and problems resulting from the lack of HE are provided).
- 6. HE Procedures for DoD Organizations: describes the HE responsibilities of the DoD organization acquiring the system. Descriptions of implementations unique to each service are provided. Details on program planning, budgeting and scheduling; preparing the Request for Proposal (RFP); proposal evaluation; and contract monitoring are also provided.
- 7. HE Procedures for Contractors: describes the HE responsibilities of the contractor developing the system. A listing of HE design standards and guidelines is provided. Descriptions of documentation (e.g. HE Test Plan, HE System Analysis Report) required by the acquiring agency are provided.
- 8. HE Methods and Tools: being of ~100 pages, provides details and references to a wide variety of time-tested methods. The methods described include mission, timeline, and workload analysis; diagrams (operational sequence, decision/action); and checklists, mockups, and mannequin usage. HE Test and Evaluation methodologies described include: HEDGE (Human Factors Engineering Design Guide

for Evaluation), interviewing and questionnaire techniques), physiological instrumentation, and physical measurement.

- Appendix A, Application and Tailoring.
- Appendix B, Task Analysis.
- Appendix C, Data Item Descriptors (DID).
- Appendix D, a matrix that cross-references sections of this HDBK with relevant DoD documents.

3. MIL-STD-1472 HUMAN ENGINEERING

MIL-STD-1472, provides design criteria, principles and practices for use in the design and development of military systems, equipment and facilities. It was originally developed in 1968 and has been periodically reviewed and updated since. It is currently undergoing technical revision, and the new version, revision F, should be available some time in 2000.

This standard is supported by two additional documents: MIL-HDBK-1908A, *Definitions of Human Factors Terms* (MIL-HDBK 1908 is currently undergoing revision) and MIL-HDBK-759C, *Human Engineering Design Guidelines*. These two documents provide important supplementary information for practitioners using MIL-STD-1472. MIL-HDBK-1908A provides standard definitions for human factors terms to supersede conflicting definitions in the literature and to eliminate unnecessary overlap. MIL-HDBK-759C provides additional guidelines and data that may be useful when designing military systems, equipment, and facilities. It is organized in sections that correspond with the top three levels of headings in MIL-STD-1472.

The body of MIL-STD-1472 is structured in five sections: (1) Scope, (2) Applicable Documents, (3) Definitions, (4) General Requirements and (5) Detailed Requirements. There is also a detailed index. Most of the substantive content of MIL-STD-1472 occurs in Sections 4 and 5.

- Section 4: provides high-level guidance about a variety of topics including functional allocation, human engineering design, fail safe design, simplicity of design, safety, ruggedness, nuclear/biological/ chemical (NBC) survivability and electromagnetic pulse (EMP) hardening. The 16 subsections of Section 5 in MIL-STD-1472E provide detailed requirements (principles, practices, criteria) concerning the topic areas described below.
- Section 5.1. Control/display integration: provides guidance concerning integration issues such as position and movement relationship, grouping, consistency, and control/display movement ratios.
- Section 5.2. Visual displays: much of the information about displays in MIL-STD-1472E concerns displays that are presented using technologies such as trans-illuminated displays, legend lights, and signal lights. Only limited information is provided about displays that are presented using computer technology. Human factors issues considered in Section 5.2 include display illumination; information content, precision, and format; location and arrangement; coding; scale types and designs; counters; and pointers.
- Section 5.3. Audio displays; includes information about topics including audio warning signals, verbal warning signals, speech transmission equipment, speech reception equipment, and speech intelligibility.
- Section 5.4. Controls: includes information about rotary and linear controls, about discrete and continuous controls, and

about keyboards, joysticks, trackballs, touch screens and mice. It also addresses human factors issues including direction of movement; arrangement and grouping; coding; and accidental actuation.

- Section 5.5. Labeling: includes both general information about labeling and specific information about label orientation and location, label content, qualities of the information presented (e.g. brevity, stated using familiar terms etc.), design of label characters and labeling of equipment.
- Section 5.6. Physical accommodation: generally based on criteria that are generally stated in terms of percentiles.
- Section 5.7. Workspace design: provides information about common working postures, about workspace design for seated and standing operators, and about specific workspace feature such as stairs, stair ladders, fixed ladders, ramps, doors, hatches and surface colors.
- Section 5.8. Environment: provides information about heating, ventilation, air conditioning, illumination, acoustical noise and vibration.
- Section 5.9. Design for maintainer: includes information about mounting, adjustment controls, accessibility, lubrication, cases, covers, access openings and covers, fasteners, conductors, connectors, test points, test equipment, failure indications, fuse requirements, printed circuit boards and designing for efficient handling.
- Section 5.10. Design of equipment for remote handling: includes information about the characteristics of the equipment to be handled remotely, feedback, manipulators, viewing equipment and illumination.
- Section 5.11. Small systems and equipment: provides information about portability and load carrying; tracking; and optical equipment and related equipment (e.g. sights, reticles, binoculars).
- Section 5.12. Operational and maintenance ground/ shipboard vehicles: includes basic information about vehicle design including information about seating; controls; operating instructions; visibility; heating and ventilation; trailers, vans, and intervehicular connections; cranes, materials handling and construction; and automotive subsystems.
- Section 5.13. Hazards and safety: includes information about safety labels and placards; pipe, hose, and tube line identification; general workspace hazards (e.g. emergency exits, illumination, thermal contact hazards); general equipment-related hazards; platforms, electrical, mechanical, fluid, toxic, and radiation hazards; trainers; and stealth and covert operations.
- Section 5.14. Aerospace vehicle compartments: contains limited information about the design of the crew station and passenger compartments, personnel entrance and exit, and emergency evacuation.
- Section 5.15. User–computer interface: include information about data entry, data display, interactive control, feedback, prompts, defaults, error management/data protection, data and message transmission, and system response time. It was not updated during the technical review of MIL-STD-1472E. For more current information, readers should consult the Department of Defense Human Computer Interface Style

Guide, Volume 8 of the Department of Defense Technical Architecture Framework for Information Management (TAFIM see below).

- Section 5.16. Visual display terminals (VDT): directs the user to ANSI/HFS 100 (Human Factors Society) for information about the use of VDT in office environments. However, VDT use in other environments is governed by requirements in this MIL-STD.

4. DEPARTMENT OF DEFENSE (DOD) HUMAN COMPUTER INTERFACE STYLE GUIDE

The DoD Human Computer Interface (HCI) Style Guide is Volume 8 of the DoD Technical Architecture Framework for Information Management (TAFIM). This document, of > 300 pages, is an excellent source for HCI guidelines, which can be traced back to references. Volume 8 consists of 14 sections that are described below:

- Section 1. Introduction: the goal of this document is to provide HCI standardization, i.e. a common framework for HCI design and implementation within the DoD. It is intended for use by individuals who determine system requirements, program managers, system managers, software developers and application HCI designers. The secondary audience includes users, software maintainers, and test and evaluation personnel. The document is applicable in both the operational and the business environments.
- Section 2. Interface Style: written for software developers, this section describes strategies for selecting a user interface style and for re-designing interfaces to improve usability. It also addresses application portability across platforms.
- Section 3. Hardware: focuses on input/output (I/O) devices, displays technologies other than CRT (e.g. liquid crystal, large screen, stereoscopic, etc.), and alternate I/O devices (Braille printers, large keycaps on keyboards, etc.) for individuals with disabilities.
- Section 4. Screen Design: provides guidance concerning screen design, log-on/off procedures and the use of color.
- Section 5. Windows: provides guidance on basic window design including appearance, message areas, scroll bars, labeling and navigation.
- Section 6. Menu Design: provides information about advantages and disadvantages of menus, pull-down and pop-up menus, hierarchical menus, menu labeling and dialog menus.
- Section 7. Direct Manipulation: provides information about screen arrangement by users, and about metaphors and icons (types, use, design, evaluation).
- Section 8. Common Features: provides guidance concerning interface features, functions, and formats that should be used consistently in all DoD applications. It also provides information about on-line Help, user computer dialogs (e.g. interrupts, error management, alarms), and the use of function keys.
- Section 9. Text: addresses the use of text within windows (labeling and updating fields, and the text cursor). It also provides guidelines for form completion (form layout, error management, etc.).
- Section 10. Graphics: provides guidelines for presenting data in graphical formats including tactical graphics (overlays,

symbology, and terrain representation), pictographic representations (digitized maps, pictures, etc.), and presentation graphics (graphs, pictures, and diagrams). Guidelines pertaining to graphical characteristics of the user interface (e.g. screen design, windows, icons, buttons, etc.) are also provided.

- Section 11. Decision Aids: describes when to use decision aids and expert systems, and offers guidance concerning requirements definition for decision aids, features of decision aids, decision aid interfaces and displays, and user training.
- Section 12. Query: deals with accessing data from Database Management Systems. This Section describes the types of database queries and database storage methods. It focuses on user-oriented database design and provides specific guidance on query screen designs, user requirements, user-friendliness, search options, and differing design requirements for novice and expert users.
- Section 13. Embedded Training: deals with *on-line training* that focuses on the learning process, as opposed to on-line help which provides assistance with specific functions, commands, etc. It provides guidance on embedded training including components of embedded training, instructional structure, and presentation. Guidance is also provided on screen design, navigation within embedded training, error feedback, and the ability to modify embedded training.
- Section 14. Emerging Technology: is divided into two sections: (1) personalization of the user interface to meet the skill levels and characteristics of different users (this includes adaptive modeling and workgroup situations) and (2) multimedia (including authoring systems and navigating within multimedia).
- Appendices: an 18-page glossary and 31 pages of references are provided.

5. FEDERAL AVIATION ADMINISTRATION HUMAN FACTORS DESIGN GUIDE FOR ACQUISITION OF COMMERCIAL–OFF–THE–SHELF–SUBSYSTEMS, NON–DEVELOPMENTAL ITEMS AND DEVELOPMENTAL SYSTEMS

The FAA Human Factors Design Guide (HFDG) provides referenced information to assist in the selection, analysis, design, development, and evaluation of new and modified FAA systems, facilities and equipment. This 1996, 2-inch loose-leaf binder document (also available on CD-ROM) combines guidance from other sources into one "human factored, user friendly" document. The document contains 14 sections, four appendices and an index that are described below:

- Section 1. Introduction: describes the purpose, scope and format of the document.
- Section 2. Complementary Documents: describes the sources from which data were integrated into this document. The sources cited include 32 Government Specifications, Handbooks and Orders, 10 federal regulations (including OSHA), and 20 non-government documents (ANSI/HFS, ASME (American Society of Mechanical Engineers), etc.).
- Section 3. Definitions: provides ~275 definitions of terms used in the text.
- Section 4. General Design Requirements: provides general

principles for designing or selecting systems and equipment, and discusses human performance and human–system interactions at the top level.

- Section 5. Maintenance Automation: primarily tutorial in nature. It provides general principles, as well as guidance, on human-centered automation; process control lessons; command, control and communications; systems engineering; monitoring; interfaces; remote maintenance; and maintenance management information.
- Section 6. Designing Equipment For Maintenance: provides criteria and guidelines related to designing equipment for handling; packaging, arrangement and mounting of equipment; access openings; covers and shields; cases; fasteners; connectors; lines and cables; packaging, layout, and mounting of internal components; adjustment controls; failure detection and isolation; fuses and circuit breakers; test points and service points; test equipment; and tools.
- Section 7. Human–Equipment Interfaces: addresses display–control integration. It also provides specific guidance concerning the following topics:
  - Visual displays (principles, trans-illuminated displays, scale indicators, CRT, large screen displays, light emitting diodes, flatpanel displays, liquid crystal displays, plasma, electroluminescent displays, stereoscopic displays, and touch panels);
  - Audio Displays (warnings and signals, controls, and voice communication);
  - Controls (selection of, movement, arrangement and grouping, coding, compatibility with hardware, accidental activation, foot operated, hand operated, keys, thumbwheels, knobs, cranks, pushbuttons. keyboards, levers, joysticks, ball controls, stylus devices, etc.);
  - Labeling and Marking (general guidance, location and orientation, typographic matters, designing label characteristics, wording, and information); and
  - Accommodating People with Disabilities (controls and displays for people with disabilities, telecommunications, and safety for people with disabilities).
- Section 8. Human–Computer Interfaces: provides guidance concerning user computer interaction, basic screen design and operation, windowing, data entry, data display, user guidance, data communication, input devices, and accommodating people with disabilities.
- Section 9. Workplace Design: describes workplace layout, designing of passageways, common working positions, standard console design, visual display terminals (VDT), and accommodating people with disabilities.
- Section 10. User Documentation: provides guidance on writing user documentation, layout and format, components of documents (cover pages, figures, tables, etc.), specific user document contents (proceduralized instructions, interactive electronic technical manuals), and accommodating people with disabilities.
- Section 11. System Security: describes general design practice, physical security and access control, identification and authentication, auditing, information and data protection, documentation of security safeguards, and security training.
- Section 12. Personnel Security: provides guidance on,

workspace safety, safety labels and placards. It also sets limits/specifies protection for the following types of hazards: liquid and gas, toxic, radiation, special chemicals, temperature, fire, noise, explosion and implosion, radiant energy and lasers.

- Section 13. Environment: provides guidelines on ventilation, temperature and humidity, illumination, and noise.
- Section 14. Anthropometry and Biomechanics: provides information about the application of anthropometric and biomechanical data and about anthropometric variability as well as anthropometric and biomechanical data on reach, human strength, and handling capacity. It concludes with a section on designing for physical comfort.
- Appendix A. References: 13 pages of references are provided.
- Appendix B. Sources: sources of the data contained in each section are specified.
- Appendix C. Standard Actions — Pushbuttons: provides definitions of functions that are performed with pushbuttons in windows (e.g. back, close, clear, cut, compile, etc.).
- Appendix D. Standard Verbs: provides definitions of verbs for use in task analysis and in writing procedural instruction (e.g. accomplish, align, find, clamp, etc.).
- Index: the 81-page index allows the reader to locate information that may be located in several sections.

6. NASA–STD–3000: MAN–SYSTEMS INTEGRATION STANDARDS

The NASA *Man–Systems Integration Standards* (NASA-STD-3000) is a multi-volume set of documents that specifies generic requirements for space facilities and equipment that interface directly with crewmembers is applicable to all manned space programs. Of primary interest are Volume 1, *Man Systems Integration Standards* (NASA-STD-3000, Revision B) and Volume 2, *Man–System Integration Standards: Appendices.*

Volume 1 is divided into 14 sections. In general, each section contains three, and sometimes four, kinds of information: (1) an overview of the section's content, (2) design considerations (background information that can help a user understand the rational behind specific requirements), (3) design requirements (contractually binding standards) and (4) design examples (sometimes included to illustrate important information). Although written for application in the space environment, with the exception of microgravity concerns, much of the information contained in the NASA-STD-3000 can be applied to comparable human interface/engineering problems in other environments.

- Section 1. Introduction: includes a statement of purpose; an overview of the entire set of Man–System Integration Standards; a statement of scope, precedence, and limitations; and general instructions on how to use the documents.
- Section 2. General Requirements: focuses on basic design information related to simplicity and standardization.
- Section 3. Anthropometry and Biomechanics: presents quantitative information about human body size, joint motion, reach, neutral body posture, body surface area, body volume, and body mass properties.
- Section 4. Human Performance Capabilities: documents the significant ways that the performance capabilities of humans may change when they go into space. It includes information about vision; the auditory system; olfaction and taste;

kinesthesia; reaction time; motor skills/coordination; strength; and physical workload.

- Section 5. Natural and Induced Environments: indicates the kinds of conditions to which a crewmember will be exposed during space flight including information about the effects of the composition of the atmosphere, microgravity, and acceleration, and specification of acceptable noise, vibrations, radiation, and thermal levels.
- Section 6. Crew Safety: deals with general safety concerns such as mechanical hazards; electrical hazards, fire protection and control, and decompression as they relate directly to crewmembers.
- Section 7. Health Management: discusses measures that must be taken to maintain crewmember health including both preventive care and medical care.
- Section 8. Architecture: provides information about the placement, arrangement and grouping of compartments and crew stations in space modules, including design data for concerning traffic flow, translation paths, location coding, orientation of workstations especially in microgravity, physical body envelopes for essential crew functions, hatches, doors, windows, lighting, and mobility aids and restraints.
- Section 9. Workstations: covers basic workstation design, including layout, controls, displays, labeling, coding and user/computer interaction design.
- Section 10. Activity Centers: discusses design and layout requirements for off-duty crew stations in the space module including facilities for personal hygiene, body waste management, trash management, crew quarters, recreation and meeting facilities, exercise and medical facilities, laundry facilities, and storage.
- Section 11. Hardware and Equipment: offers general equipment design guidance for tools; drawers and racks; closures and covers; mounting hardware; handles and grasp areas for portable items; restraints; mobility aids; fasteners; connectors; windows; packaging; crew personal equipment including clothing; and cable management.
- Section 12. Design for Maintainability: provides general guidance concerning maintainability and specific requirements concerning design; physical access; visual access; removal, replacement, and modularity; fault detection and isolation; test points; and requirements for a maintenance data management system.
- Section 13. Facility Management: addresses issues associated with housekeeping, inventory control, and information management.
- Section 14. Extravehicular Activity (EVA): establishes guidelines for extravehicular activity.

Volume 2, *Man–Systems Integration Standards*: *Appendices*, contains the appendices which pertain to the Man–System Integration Standards. These appendices include a bibliography, list of sources used to develop specific paragraphs, glossary, abbreviations and acronyms, units of measure and conversion factors and index/keywords listing. Volume 3, *Man–Systems Integration Standards: Design Handbook*, provides a condensed version of quantitative data from Volume 1. However, this volume currently reflects of contents of Revision A, rather than Revision B, of Volume 1.

7. OBTAINING THESE DOCUMENTS

The Internet has profoundly changed our ways of acquiring technical information — so much so that most of the documents described in this section can be viewed on-line and/or downloaded directly from the web.

The two US Department of Defense documents described in this section, MIL-STD-1472, *Human Engineering*, and MIL-HDBK-46855, *Human Engineering Program Processes and Procedures*, can be acquired from the Department of Defense Single Stock Point (DODSSP) using the following information: DODSSP Customer Service, Defense Printing Service Detachment Office, Building 4D, 700 Robbins Avenue, Philadelphia, PA 19111-5094, USA; Tel.: +1 (215) 697 2179; URL: http://www.dodssp.daps.mil/

The DODSSP URL provided above allows access to the Acquisition Streamlining and Standardization Information System (ASSIST) which provides electronic access to some military standards and handbooks.

The remaining documents described here can be accessed directly on the Web (either viewed on-line and/or downloaded) using the information the Reference section. In addition, the "NSSN: A National Resource for Global Standards" web site (URL: http://www.nssn.org/) allows one to search for standards documents produced by > 600 organizations. Their search engine is an excellent source for locating most international civilian and military standards.

ACKNOWLEDGEMENT

REFERENCES

Dates are not provided because these documents, or sections of the documents, are updated on an irregular basis and readers should consult the most current version. The documents reviewed here are those available in February 1999. Current versions may be obtained from the sources listed.

DoD, Human Computer Interface Style Guide. Vol. 8 of Technical Architecture Framework for Information Management (TAFIM). [http://www-library.itsi.disa.mil/tafim/ tafim3.0/pages/tafim.htm]

DoD, MIL-HDBK-1908A, *Definitions of Human Factors Terms* [http://www.dodssp.daps.mil/]

DoD, MIL-HDBK-759C, *Human Engineering Design Guidelines* [http://www.dodssp.daps.mil/]

DoD, MIL-HDBK 46855: Human Engineering Program Processes and Procedures [http://www.dodssp.daps.mil/]

DoD, MILSTD 1472: Human Engineering [http://www.dodssp.daps.mil/]

Federal Aviation Administration (FAA) Human Factors Design Guide (HFDG) for Acquisition of Commercial-off-the-Shelf-Subsystems, Non-developmental Items, and Developmental Systems [http://www.tc.faa.gov/hfbranch/hfdg/index.html]

Ministry of Defense, British Defence Standard 00-25: Human Factors for Designers of Equipment. [http://www-sa.jsc.nasa.gov/FCSD/CrewStationBranch/Msis/MSIS_home.htm]

NASA-STD-3000: Man–Systems Integration Standards. [http://www-sa.jsc.nasa.gov/FCSD/CrewStationBranch/ Msis/MSIS_home.htm]
National Resource for Global Standards [http://www.nssn.org/]

Australia: Ergonomics Society of Australia

M. I. Bullock

PO Box 371, Indooroopilly 4068, Brisbane, Australia

1. INTRODUCTION

The Ergonomics Society of Australia and New Zealand (ESANZ) was established in 1964 and separate Ergonomics Societies for New Zealand and Australia (New Zealand Ergonomics Society and the Ergonomics Society of Australia, ESA) were formed in 1986.

The ESA is an active and thriving society with representation from many disciplines. It has branches in each State and Territory, which report to the National Committee. The Society's business is directed by its Strategic Plan, the responsibilities for which are shared by branches. An annual national conference, a bimonthly newsletter and State symposia and workshops provide a means of communication for members. A major initiative of the ESA in 1996 was the organization of a conference entirely on the Internet: the Cyberg Conference, to offer conference opportunities for those unable to attend international conferences through travel. The Cyberg Conference will be held triennially. Several Special Interest Groups (such as in Computer–Human Interaction, Ergonomics Education and Product Ergonomics, Cognitive Ergonomics, Occupational Health and Safety, together with a group for the assistance of developing countries) offer a focus for specialist interests.

The Society has evolved since its formation, reflecting changing ergonomic interests and evolving responsibilities for ergonomists. The early history of ergonomics in Australia centers around individual and collaborative research projects concerned with the enhancement of operator comfort, safety and performance and a recognition of the importance of multidisciplinary activity. Only later, as knowledge of ergonomics grew and the value of its application was acknowledged, did opportunities become available for ergonomics consultants to analyze and solve work-related problems.

2. EARLY RESEARCH FOCUS ON ERGONOMICS

The importance of considering worker capacities and limitations in the design of work and working environments which would ensure effective integration of man and machine was recognized in Australia quite early in the 20th century, and a wide range of ergonomics-related research studies was undertaken.

In the 1930s the aviation industry was the focus of much of the early ergonomics-related research. Factors relating to visual standards, changes in atmospheric pressure with altitude, the problems of blackout in aircrew and the problems of noise in aircraft were areas for particular interest. Later, the Director of Aviation Medicine, Dr John Lane, was the catalyst for many developments, and through his efforts Australia led the world in many features of air safety. The design of air traffic control systems, navigational aids to assist aircraft landing, and visual displays to eliminate irrelevant information and assist the operator to organize incoming data all attracted attention.

Working in association with Flying Personnel Research Units were government instrumentalities such as the National Standards Laboratories, the CSIRO, the Acoustic Laboratory, the National Health and Medical Research Council and the Fatigue and Tropical Unit of the University of Queensland. This collaboration demonstrates an early concern about human performance by a range of scientific disciplines. A Human Engineering Research Group was established within the Aeronautical Research Laboratories of the Australian Defence Scientific Service in the Department of Supply, in about 1957. This represented the first formally constituted research group in ergonomics as such in Australia. Three of the principal researchers associated with this group, Dr Colin Cameron, Professor Ron Cumming and Lane, were to become instrumental in the later formation of the Ergonomics Society.

The effects of climate provided a special research interest. The extreme temperatures of the tropical and arid regions in Australia led to major studies being carried out within the Tropical and Fatigue Laboratory within the Department of Physiology at The University of Queensland into the physical and psychological effects of tropical service, and the design of clothing for flying in the tropics and at low temperatures. Other studies in the School of Public Health and Tropical Medicine at the University of Sydney investigated the effects of climatic extremes on comfort and performance of people of all ages, whether healthy or sick. Early work of psychologists, principally by Dr Provins in South Australia, was concerned with studying the relationship between environmental conditions, body temperature and the performance of skilled tasks. The results of these studies had application to work at Australia's station at Mawson, in Antarctica and in the mining industry in northern Australia. Engineers became actively involved in designing safety features in the design of load haul dump vehicles (LHD) and underground mining vehicles.

Collaborative research studies by the Departments of Mechanical Engineering and Physiotherapy at the University of Queensland investigated the skills demanded of LHD drivers in coping with vibration. Extensive studies were carried out by physiotherapist Professor Margaret Bullock to determine the optimal worker–pedal relationship to minimize spinal movements, in an attempt to control the prevalent back injuries. The physical stresses associated with manual sugarcane harvesting led engineers to develop an automated system of cane harvesting and bulk storage.

Large-scale surveys of school children allowed collection of data relevant to furniture design for schools, while in the 1960s the influence of seating design, posture and work place layout on the production of musculoskeletal injuries during process work was studied by Dr Peres within a human engineering group of the Division of Occupational Health in New South Wales. In 1963, the name of this group was changed to Ergonomics Group. Thus, an interest in the prevention of musculoskeletal injuries in industry was an important component of ergonomics in Australia from its earliest days.

Demonstrating an early interest in rehabilitation ergonomics, engineers and medical practitioners collaborated in the 1960s to examine upper limb stresses in process work and the design of prosthesis that would enable disabled persons to become productive workers. Dr Patkin also carried out considerable

research during the 1970s and 1980s, in the ergonomic design of surgical instruments, and by Professor David Ferguson in relation to the causation of repetition injuries.

The Australian Road Research Board was influential in supporting the collection of data relevant to Australian design rules for motor vehicle safety, and the driver's work-space. Ergonomics also had an influence on traffic engineering.

The development of ergonomics practice in Australia has been closely associated with interests in occupational health and safety. The excessive amount of lost time from work because of musculoskeletal injury and the subsequent costs forced employers to introduce measures of control. Positive changes in occupational health and safety practices and also in management style were introduced into many work places in Australia during the 1980s, due in part to the major contributions of Professor Ferguson.

The increasing use of computers as part of the new technology and the importance of developing effective user interfaces led to the formation of a Computer–Human Interaction Special Interest Group of the ESA, and this has proved to be of considerable value.

3. EDUCATION IN ERGONOMICS IN AUSTRALIA

Initially in Australia education in ergonomics was offered within relevant professional programs, including Engineering, Psychology, Physiology, Architectural Science, Physiotherapy, Occupational Health and Applied Arts and Industrial Engineering. Today, postgraduate qualifications in ergonomics (at Postgraduate Diploma or Masters Degree level) are offered within some tertiary Institutions. No undergraduate program totally devoted to the preparation of an ergonomist is offered within Australia.

4. PROFESSIONAL CERTIFICATION

The high incidence of work-related musculoskeletal injuries in the 1980s had implications for control of the quality of ergonomics practice in Australia and acted as a catalyst to establish a standard of practice for the ergonomist and to create a register of professional or certified ergonomists.

The criteria that could be used for a program of certification of professionally qualified ergonomists in Australia aroused much debate within the Society and in 1985 a proposal to proceed with developing a professional certification scheme was adopted by the Society. In 1990, 21 Society members were awarded professional certification status at the first ceremony of its kind in Australia. The Professional Affairs Board is active in updating its criteria for membership.

5. COMPETENCIES IN ERGONOMICS

In 1990 the Australian Government moved to introduce competency-based assessment in all occupations and professions. It was realized that definition of ergonomics competencies was vital for the comprehensive review of the certification procedure, as a basis for recertification, and as a resource in planning and accrediting education programs. The ESA has now defined and published its outline of core competencies of an ergonomist.

6. CURRENT ISSUES

An issue uppermost in the mind of the ESA today is quality of practice. The broadened scope of ergonomics, the core competencies required by any person working within the field and the importance of quality practice have led to concerns to define the professional ergonomist more specifically. The issues of professional certification of ergonomists, the specification of optimal educational requirements for an ergonomist and accreditation procedures have generated considerable interest in the ESA.

7. SPECIAL CONTRIBUTIONS

It is important to recognize the major contributions made by some eminent ergonomists in the formation and direction of the ESA. In particular, these are Cummings, Lane, Cameron, Professor David Ferguson and Dr Alan Howie. Each had a profound influence not only on the Society itself, but also on the achievements of ergonomics in Australian society. Through their efforts and that of many others, challenges have been met and human performance has been enhanced.

8. PROFESSOR RON CUMMING

The late Professor Cumming (1920–86) was one of the first graduates in aeronautical engineering in Australia and, as the leader of Human Engineering Group, carried out important research in operational aspects of aviation at the Aeronautical Research Laboratories. He later became the founding Chairman of the Australian Road Research Board and was involved in committees drafting safety design rules for motor vehicles. In a major shift of career, he became a Professor of Psychology at Monash University, successfully combining his two disciplines in major ergonomics research.

Cumming became the first President of the Ergonomics Society of Australia in 1964, he was the first Australian Fellow of the (American) Human Factors Society and was elected a Fellow of the Australian Academy of Technological Sciences and Engineering (FTSE).

9. DR JOHN LANE

The late Dr Lane (1918–99) is recognized as the father of aviation medicine in Australia and as a pioneer in road safety. In the 1950s he was involved in the development of the Aeronautical Research Laboratories and the 'T-vasis' visual safe aircraft landing system which is still in use today. The holder of a Master of Public Health Degree from Harvard University, he undertook extensive epidemiological research over 40 years. In 1960, trained as a space surgeon by NASA, he became Australian aeromedical monitor in the manned space program.

Lane was a founding member of the Aviation Medical Society of Australia and New Zealand, and of the Ergonomics Society of Australia, being its second President. As a foundation member of the Human Factors Committee of the Australian Road Research Board, he was active in promoting road safety research and had considerable influence on the formulation of vehicle safety standards. He was a member of the Traffic Inquiry Committee of the National Health and Medical Research Council.

Lane was an inspiration to many aspiring ergonomists in Australia and contributed to the development of the discipline not only through his own broad interests and activities, but also by his ability to lead and encourage others to undertake research and work in the ergonomics field.

Biosketch: Corwin Bennett

S. Konz

Department of Industrial and Manufacturing Systems Engineering, Kansas State University, Manhattan, KS 66506 USA

Corwin Bennett received a BS in Industrial Psychology from Iowa State in 1950 and an MS and PhD in Experimental Psychology from the University of Nebraska in 1951 and 1954. He worked first for IBM as a research psychologist from 1954 to 1966 (manager of human factors for IBM's space guidance division). He then shifted to academia, first as Associate Professor for Rensselaer Polytechnic in Troy, New York, from 1966 to 1970, then as Professor, Industrial Engineering at Kansas State University in Manhattan, Kansas, from 1970 until his death.

Corwin was a member of the Illuminating Engineering Society, the Human Factors Society, the Ergonomics Society, the American Psychological Association and the Society of Engineering Psychologists. He was a fellow of both the Illuminating Engineering Society and the Human Factors Society.

As a small example of his national professional activity, he was a member of the US National Committee TC 3.4 Discomfort Glare for 14 years, a member for 8 years of the Industrial Lighting Committee of the Illuminating Engineering Society and a member for 9 years of the Education Committee. He also was very active in state level activities of his professional societies as well as the many campus duties of a professor.

As can be seen, Corwin was interested in illumination and lighting. Specific projects ranged from the office (VDT workstation lighting; discomfort glare in offices; esthetics of interior spaces) to the factory (industrial inspection) and highway (illumination for railroad grade crossings, discomfort glare for roadway lighting).

Corwin had over 70 professional publications; his work on the esthetics of interior design is summarized in his *Spaces for People: Human Factors of Design* (Prentice-Hall, 1977).

Biosketch: EtienneGrandjean

M. Graf, J. Nemecek and H. Krueger
Swiss Society for Ergonomics
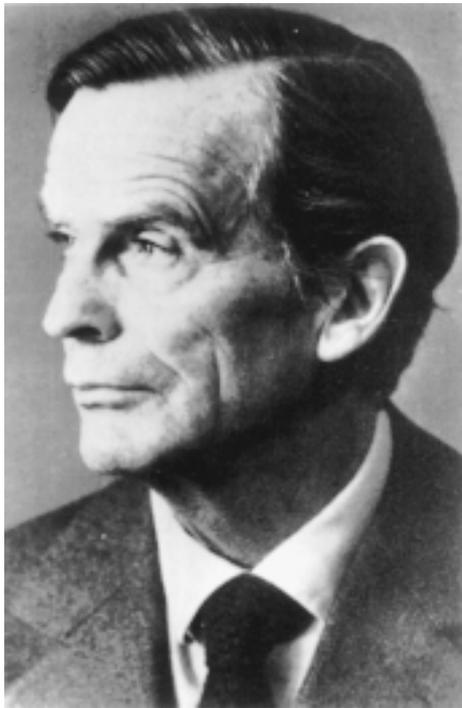
Figure 1. Etienne Grandjean.

The name Etienne Grandjean is known by students of ergonomics throughout the world as the author of *Fitting the Task to the Man*. He was a leading pioneer of the science of ergonomics, always emphasizing and ensuring its practical application. In Switzerland, his home country, his name was synonymous with ergonomics for several decades.

Grandjean was born in 1914 in Bern. After completing a medical degree he spent several years at the University Hospital in Lausanne gaining further experience in internal medicine, pathology and physiology. After completing a PhD in human physiology he took up the post of Laboratory Director for the firm Wander, Inc. In 1950, at 36 years of age, he was offered the position of Professor of Industrial Hygiene and Applied Physiology at the Swiss Federal Institute of Technology in Zurich. Before commencing in the new post he spent 6 months at Harvard University in the USA under a Rockerfeller Award.

The curriculum of his predecessor did not include ergonomics, so Grandjean chose to redefine his subject area, set up the new discipline and adjust the curriculum, firstly to suit the needs of his engineering students. He was exceptionally skilled at making the complex interactions between workplace conditions and the human organism comprehensible to the technically oriented student. Analogies and multiple examples were his favored tools; however, he was always careful not to oversimplify the underlying medical processes.

For 33 years Grandjean imparted the principles of humanization of the workplace to young mechanical and electrical engineers, architects, civil engineers and later to information technology students. But the total number of his students was much higher, as he was often invited to give special lectures and courses to a variety professional groups and associations. In this way he spread ergonomic concepts to personal managers, senior managers of industrial concerns, banks and insurance companies, public utilities, trade unions, and many other groups.

Although the interaction between the health and well-being of the working person and his or her work performance are clearly related, Grandjean never allowed performance improvements to be the primary focus of the ergonomist. He always kept preventive healthcare in the foreground of all his work.

His initial scientific work concentrated on physiological research into the effects of physical loading at work. Early studies were with forestry workers, agricultural workers, telephone exchange employees and orchestra players. The production process in the workplaces of Europe and the USA in the 1950s was characterized by heavy manual work with high energy consumption. The demands on the circulatory system often lay at the borders of human endurance. By way of example, lumberjacks at that time were required to cut through forests with handsaws, work that often stressed the cardiovascular system to its limits. With the introduction of the chainsaw their physical workload was substantially decreased.

In the later post-war decades the production process underwent a rapid change that resulted in an overall reduction in the number of worker exposed to heavy muscular work. Workers now used machines. They controlled and supervised complicated equipment, eventually fully preprogrammed production units. As a result of these changes the demands on the worker changed. Jobs increasingly required skilled manual activity, perception and interpretation of information, rapid and correct decision-making and increased vigilance. Grandjean noted that the new working methods and tools carried new types of stresses and strains for the worker, such as noise, vibration, exhaust fumes, increased accident risk, unnatural body postures and increased performance demands. Grandjean changed his area of scientific interest to reflect these new conditions.

He was sensitive to the effects on the workers of the new production and management technology and almost always responded rapidly to developments with his scientific investigations, often to preempt them. A well-cited example is his sensitivity to the advent of the VDU workplace. Although close to retirement, he projected, ahead of the physicists and engineers, the effects of developments in computer hardware on operators. Keeping the human in mind, he appealed to several manufacturers to change their still-conceptual plans better to suit the users. His main topics of interest were the prevention of musculoskeletal disorders and visual complaints.

The subject of seating interested him for many years. Even today many students and workers sit on chairs whose design is based on his specifications. He was one of the first ergonomists to recognize that no seat design could ensure health, and to emphasize the importance of adequate physical movement at the workplace.

The results of his investigations on the effects of modern technology on the human were included in his first book, published in German in 1963 as *Physiologische Arbeitsgestaltung, Leitfaden der Ergonomie* (translated roughly as Physiological Workplace Design, Guidelines from Ergonomics). The book was

written not only with students in mind, but also other practitioners such as engineers, planners, managers and administrators. The text was translated into English and published in 1980 as *Fitting the Task to the Man – An Ergonomic Approach*. It rapidly achieved worldwide popularity and has been translated into over ten languages. It has become a standard introductory text for ergonomics students in many countries. Technological advances were reflected in later editions. The fourth, completely revised edition was published in 1991.

The second area of Grandjean's research and teaching covered the wider living space of humans. Work at home, behavior in the home, the space available for people to move, the dimensions and design of rooms and living spaces, as well as the special situation and needs of handicapped and invalid persons all contained important issues and problems of interest to him. The home is also exposed to disturbances and annoyances that originate outside its borders, environmental factors which impinge on the home, particularly noise from roads, railways and aviation, and environmental chemical pollutants. As in ergonomics, Grandjean identified the important problems for the home environment, systematically investigated them and defined guidelines. Here too explained the problems clearly to architects and civic planners and he influenced their future decisions to make people's home-life more comfortable. The multiple themes of his living area research were laid down in his second book, *Wohnphysiologie — Grundlagen gesunden Wohnens* (The Physiology of Living — Principles of Healthy Homes). It was aimed at sensitizing architects and interior designers to ergonomic considerations.

To reflect the broader scientific interest the Institute was renamed in 1973 as the Institute for Hygiene and Applied Physiology. During Grandjean's directorship of the Institute, over 900 articles were published, covering not only ergonomics themes, but also such diverse subjects as environmental hygiene, industrial hygiene, experimental toxicology, psychophysiology, psychotoxicology, misuse of drugs and stimulants at the workplace, etc. The content of his work ranged from purely experimental natural science research, through epidemiological studies, the implementation of scientific findings, legislation on workplace hygiene, health politics and product design. His advice was sought by numerous national and international journals.

Apart from his well-known scientific publications, Grandjean also published widely in more popular media, being well aware that ergonomics is an applied discipline and its success also depends on the application of its principles to real work places and situations.

Aside from his teaching of undergraduate students of the Swiss Federal Institute for Technology, Grandjean did not neglect the further education of his own collaborators. Under his supervision dozens of young scientists were prepared for academic carriers or completed postgraduate studies. He imparted a systematic scientific approach to his co-workers in a tolerant professional manner, encouraging critical evaluation of work without injuring personal feelings. For most of his colleagues he was not only the boss and supervisor, but also a model of collegiality. His charm was appreciated by all that met him.

The effects of his work were soon recognized far outside the borders of Switzerland. He initiated and collaborated in countless international congresses and edited several books from these meetings. He was granted multiple honors, distinctions and prizes by domestic as well as foreign organizations and universities, including three honorary doctorates from the Universities of Surrey (1970), Stuttgart (1976) and Geneva (1984). He was awarded the "Prix René Barth" from the University of Paris (1953), the American Industrial Hygiene Association's "Yant Memorial Award" (1970), the International Prize for Occupational Medicine "Buccheri la Ferla" (1976) and was the first recipient of the Award of the International Ergonomics Association (1982). He became a Fellow of the Human Factors Society in 1978.

In 1956 Grandjean was a foundation member of the Swiss Society for Social and Preventive Medicine. He was its President for 10 years and was later an honorary member. He regarded the introduction of subspecialties within ergonomics with some concern and felt that ergonomics had to be an interdisciplinary science, not just a topic for specialists solely in psychology, medicine, engineering or design. It was important for him to get practitioners from various disciplines to contribute their knowledge and approach to solving problems. He loved discussion with his peers as much as with younger scientists and sought through constructive questioning and praise always to learn from others.

Although an all-rounder in the best sense of the word, Grandjean never forgot his roots as a doctor of medicine and the importance of the individual. He was continually interested in special cases and was wary of norms and standards, which he felt too often formed psychological cupboards for active thought and lead to a stagnation of progress. They did not correspond well to his medically oriented way of thinking or his artistic leaning, both of which placed most emphasis on the individual and the essential variety in humans and their experiences. He continually sought practical solutions to the problems he found, and he thought about what practitioners needed to be taught to be most effective. Science was not a hypothetical endeavor to Grandjean but an aid to decision-making, with direct effects on the people concerned.

He was a passionate skier and mountaineer as well as an enthusiast of modern art, a humanist, full of the joy of life, with little concern for himself. He spoke French, German, Italian and English fluently. His cultural sensitivity gained him acceptance all over the world.

In the Fall of 1983 Grandjean retired as Director of the Institute for Hygiene and Applied Physiology. He remained a frequent visitor to the Institute for many years, using his retirement to undertake a full revision of his popular textbook. He also continued to work in an advisory capacity to multiple organizations and to regularly attend conferences.

On the 11 November 1991 Grandjean lost his fight against a cancer and passed away. The foundations of ergonomics had been laid during his career and his role in the development of this new scientific discipline was substantial. His impact on it can still be seen in many areas today.

Biosketch: Jean-Marie Faverge

Jean-Marie Faverge was born in 1912 in the French Jura. He was educated in mathematics, before obtaining a degree at the Institut de Psychologie de l'Université de Paris and a state diploma from the Psychotechnicien. In 1945 he was elected Chairman of the "Groupe de psychométrie pédagogique" at the Centre de Recherches et d'Etudes pédagogiques of the French Education Ministry.

He was then invited by Dr André Ombredane in 1947 to join the Centre d'Etudes et de Recherches Psychotechniques in Paris, and soon afterwards he became Professor at the Institut de Psychologie de l'Université de Paris and at the Institut National d'Orientation Professionnelle. His teaching about the adaptation of work to man was based on statistical methods and on job analysis and it led him publish a series of now classic books: *Méthodes statistiques en psychologie appliquée*; *L'analyse du travail* (with A. Ombredane) and *L'adaptation de la machine à l'homme*. These publications contributed significantly to the launch of the ergonomics movement in the French-speaking world.

In Paris, he soon became a member of the "Human Factors — Safety" commission of the European Community for Coal and Steel (ECSC) and civil councilor of the Scientific Action Committee for National Defence. His works on the accident process in coalmines, and in the steel industry started in earnest.

In 1959, he was appointed head of the Psycholocal Laboratory at the Université Libre de Bruxelles, where he developed a teaching in Industrial Psychology and in Ergonomics; at the same time he contributed to the creation of a Institut du Travail modeled on the proposal of the International Labour Organization in Geneva.

Between 1960 and 1975, he carried out a set of researches for the Belgian Productivity Centre (now Institute for Research on Improving Working Conditions), leading him to distinguish four steps or procedures in ergonomics: postural, information, systems and cognitive (or heuristic). During this period, he investigated the concepts of safety and fiability (or reliability) in industrial processes. In 1976–77, Faverge conducted a follow-up study devoted to the difficulties of achieving breakthroughs in ergonomics in industrial settings.

Illness forced his retirement in 1980, leaving one of the richest legacies in the French-speaking world: as a scientist, he was in a position to build up and nurture two different schools of thought: first in Paris, with J. Leplat or M. de Montmollin and the building up of ergonomics; afterwards in Brussels, with G. Karnas on the work analysis, and with V. De Keyser (who teaches at Liège University) and the reliability research in the field of psychological ergonomics.

He was member or chairman of several scientific societies such as the French Psychological Society, the International Association of Applied Psychology and the Psychometric Society. He was one of the French and Belgian founders of the Société d'Ergonomie de Langue Française, which is his major achievement for ergonomic development.

Biosketch: Longin Paluszkiewicz, 1925–89

D. Koradecka

Central Institute for Labour Protection, Czerniakowska Street 16, 00712, Warsaw, Poland

Professor Longin Paluszkiewicz, one of the originators of the Polish science of ergonomics, belonged to the most renowned specialists in ergonomics-related psychology.

He was born on 28 December 1925 in Mstów near Czestochowa. In 1952 he graduated from the University of Warsaw with an MA in Philosophy. His speciality was psychology. In 1968 a PhD was conferred on him and in 1980 a professor's title in the field of Organization and Management.

Paluszkiewicz's career was almost entirely connected with the Central Institute for Labour Protection where he was at first a junior researcher, then Head of Engineering Psychology Department and, finally, Head of the Ergonomics Department. However, he began — before entering university — as a manual worker (fitter and electrician) and next as an administrative worker. During his studies he worked in a traditional domain of psychology counseling. He counseled juveniles on adapting to, sometimes difficult, living conditions, as well as to adults in the field of mental hygiene.

In research work he was originally occupied with an influence of various sources of light and eye and face protectors on both eyesight and the quality and efficiency of work. Yet, his main scientific interest was perception of signals. Paluszkiewicz's related research dealt with ergonomic properties of display and actuator devices (scale graduation, the length of the pointer, its distance from a scale, the color of the background, etc.), the influence of the probability of appearing signals on their perception.

The results of the research on working habits and on acquiring proficiency in using display devices were, in addition to the above-mentioned research, a significant contribution by Paluszkiewicz to the knowledge of ergonomics. The results of the research also included factors that lower this proficiency and making the use of such devices difficult for the perception of information.

Further research concerned various systems of display and actuator devices and the structure of actuator units for vehicles and multifunctional building machinery, as well as designing these units according to the rules and requirements of ergonomics.

In his scientific activity Paluszkiewicz was clear when raising and analyzing problems, and inquiring, and he showed a specific ability to consider the examined problems from different angles. This gathered around him representatives of the humanities and the sciences, who created a harmonious group of ergonomists. Paluszkiewicz's role was to inspire, program and supervise the work of the group.

Paluszkiewicz attached importance to research methodology. In this respect he was a perfectionist and he inculcated such an approach in his co-workers. He also showed an understanding for the relationship between theoretical solutions and their implementation. This understanding was, to a great extent, a result of his experience in industry. Additionally, the findings of his research as well of the research in leading countries, which he published, provided designers and constructors with indispensable data for designing products according to the rules and requirements of ergonomics.

Among his first publications, *Engineering Psychology* deserves consideration. Published first in 1963 (2nd ed. 1966), it was prepared together with Jan Okon, a master under whose wing Paluszkiewicz originally developed. *Engineering Psychology* was the first original Polish scientific publication on ergonomics. An extensive chapter, "Engineering Psychology," included in the university textbook *Psychology of Work* edited by Okon in 1971; is also an important publication in the output of Paluszkiewicz.

Paluszkiewicz included the findings of both his own and world research concerning display and actuator devices in his *Ergonomic Properties of Display and Actuator Devices* (1975). Being an author or co-author of Polish and foreign publications, he contributed to the development of psychology of work and ergonomics as well as to the popularization of knowledge about them. These publications included textbooks for university students and training guides used by the Polish Academy of Sciences, the Chief Technical Organization and companies, as well as 80 entries for Polish reference publications, such as *Encyclopaedia of Technology*. He also prepared documentation for the International Labour Office (ILO) in Geneva and the Council for Mutual Aid.

Paluszkiewicz was a member of the Editorial Committee of *Ergonomia* and also co-operated with *Polish Psychological Bulletin*. He was also a respected lecturer of ergonomics and psychology of work at the University of Warsaw and several other universities, and at training for staff currently responsible for working conditions, i.e. directors, designers, technologists and production organizers. Clarity of style, precision of expression and the ability to present clearly a given subject made his lectures and various presentations always interesting and drew the audience's attention. This derived not only from his knowledge, skill and aptitude, but also from the adjustment of the form and content of lectures and presentations to the listeners' specialization and intellectual level. He inculcated the need for such preparation in all his co-workers.

He was a patriot and he worked for the benefit of society. During World War II he took part in the Warsaw resistance movement and was an active participant in the Polish scout organization "Szare Szeregi." In the Warsaw Uprising he fought in Mokotów in the regiment 'Baszta' and was seriously wounded. Professor Paluszkiewicz presented this period in the third volume of his *Harcerze 1939–45: Szare Szeregi*.

In 1977 Paluszkiewicz co-founded the Polish Ergonomics Society, a member of the International Ergonomics Society (IEA). For his contribution to the development of the Polish Ergonomics Society he was awarded the Medal of Wojciech Jastrzebowski and the title of an honorary member.

He also took part in the meetings of many committees and societies, e.g. the Ergonomics Committee of the Polish Academy of Sciences, the Ergonomics and Labour Protection Committee of the Chief Technical Organization and the Polish Psychological Society. For his merits in scientific and research, organizational, educational and social work he was awarded a number of state distinctions, among others, Zloty Krzyz Zaslugi in 1974 and Krzyze Orderu Odrodzenia Polski (Kawalerski in 1978 and Oficerski in 1986).

Professor Longin Paluszkiewicz died on 26 September 1989 in Warsaw.

Biosketch: Paul Branton

D. J. Oborne

Department of Psychology, University of Wales Swansea, Singleton Park, Swansea SA2 8PP, UK

1. INTRODUCTION

Being an eclectic science, ergonomics attracts its practitioners from a range of disciplines. Although Paul Branton (1916–90) was formally trained as a psychologist, his roots before his relatively late university career were in commerce, industry and the Royal Navy (1942–48). Each environment played a significant role in the development of his views of ergonomics, from both philosophical and psychological perspectives. Each enabled him to become the founder of a predominantly person-centered view of ergonomics.

2. PHILOSOPHICAL INPUTS

Branton's emotional roots were firmly founded in philosophy. Before his training at the University of Reading (1959–62) he was particularly concerned with understanding the bases of personal value judgements and studied critically the philosophical foundations of ethical behavior in general, and morals and politics in particular. He was inspired by the teachings of the German philosopher at Göttingen University, Leonard Nelson (1882–1927), whose ideas were part of a German philosophical tradition that extends back to the critical reasoning of Immanual Kant (1724–1804). This interest in the philosophical bases for understanding behavior, particularly in the tradition of Critical Philosophy, remained with him throughout his life, and helped him to develop and maintain his person-centered view of ergonomics.

Kant's main preoccupation was with moral or ethical behavior. Thus, when he spoke of (rational) action he meant moral action, including such aspects as good deeds, resistance to temptation,

and a virtuous disposition. Taking this view further, Branton argued for a set of metaphysical principles of work that are not at all separable from moral considerations. On the contrary, he felt that moral considerations intrinsically belong to ordinary human work. In this way he spent much of his life attempting to "marry" critical philosophy to psychology. With his colleague Fernando Leal he developed a view that he called the "New Science of Inner Life," the ultimate aim of which was to show that there is "a greatness lying inside the human person" that needs to be understood, nurtured and liberated.

3. PSYCHOLOGICAL AND PHYSIOLOGICAL INPUTS

During his undergraduate training at the University of Reading, Branton developed a strong interest in the physiological and anatomical bases of psychological events. Throughout his future careers his view developed that we should understand person-centered variables from a psychophysiological perspective. For example, boredom and monotony can be understood in terms of physiological adaptation and stimulus inhibition. His study of anesthetist's stress suggested that the stress could be considered to result from an accumulation of "mini panics" arising from a realization that the anesthetists, as well as their patients, had drifted off to sleep during long operations. Also seating comfort, which he studied extensively during his period at Furniture Industry Research Association (FIRA) and then at British Rail, was viewed to be the result of physiological adaptations to postural instabilities.

The recurrent psychophysiological theme that runs through much of Branton's work was stimulated by his undergraduate studies. He frequently considered the importance of rhythmical variations in bodily functioning. He was considerably influenced by the fact that bodies exhibit such rhythms in almost all of their functions — consciously and unconsciously. These fluctuations, he argued, can be perceived as being major influencers of our behavior when interacting with working systems.

On leaving university Branton moved to FIRA (1962–65) with a brief to study the comfort of seating and seats as part of an ergonomics service to that industry. During this time he further developed his thoughts about the nature of ergonomics and the centrality of the individual within the system. In particular, he began to develop views about ways of integrating observable behaviors, such as movement and posture with the less observable and more subjective behaviors to understand better, and fully meet, the needs of the user.

It was during his time at FIRA that Branton developed his theory of postural homeostasis to explain individual comfort-seeking behavior. This theory again emphasizes the interface between psychophysiology and behavior, particularly stressing the fluctuating nature of many psychophysiological processes — in this case the forcing system that maintains a posture when sitting or standing.

Branton argued that the upright or sitting body is an inherently unstable system and postural activity is one of dynamic, fluctuating, body states that maintain the desired posture — "a two-way traffic across the borders between body and mind." Thus, the homeostatic theory of seating comfort considers comfort as being the optimal state between two conflicting body states or

requirements: the need for stability on the one hand and the need to reduce pressure points created by stabilization on the other. Using this conceptualization, Branton developed a seating comfort metric that analyzed postural instability and shifts ("fidgets"), rather than relying on the traditional subjective rating scale approach.

Following his career at FIRA, Branton moved to the MRC Industrial Psychology Research Unit (1966–69) where his interests shifted more towards global industrial problems and how psychological interventions and investigations may help to explain them. It was here that his interests in shiftwork and accidents began to take shape, leading particularly to developing his understanding of stress and decision-making within individuals. His later work on anesthetist stress, particularly resulting from fluctuating levels of awareness, is an example work that had its critical formation during this part of his career.

Indeed Branton returned to this area of work in 1987, as an Honorary Research Fellow at Birkbeck College, University of London. At this time he collaborated with his colleague Pat Shipley on studies relating to stressful experiences of working people to problems of individual control, and being in positions of responsibility.

Moving from FIRA to become Chief Ergonomist in British Rail in 1969, Branton became free to develop his ideas in a specifically applied setting. The list of issues in which he became interested is long and includes studies of train drivers (emotional and physiological states), driver cabs, signaling, driver behavior, passenger environments, and physical environmental influences. Such issues were studied from the viewpoints of skilled behavior, stress, comfort and decision making.

To each area Branton brought his person-centered view of ergonomics: that one should consider the whole person within the system, including the individual's psychophysiological and mental models of the system with which s/he is interacting.

4. SUMMARY

Paul Branton's view of ergonomics was formed by almost all facets of his varied life — from his early life as a tailor in Austria before World War II to his work as an Honorary Fellow at Birkbeck College at the time of his death. His view, and his pivotal role in the foundation of a new facet of ergonomics — person-centred ergonomics — was forged within the realms of Kant's Critical Philosophy and Branton's sense of innate goodness of people.

These views were strengthened during the persecution years in Germany and Austria; in 1937 he fled from Vienna and took refuge in Palestine. In 1948 he moved to the UK and became interested in politics. He studied philosophy and social science in the 1950s, before undertaking full-time undergraduate training in psychology. From this point on, philosophy became married to psychology and different facets of person-centred ergonomics emerged with his employment at FIRA, MRC, British Rail, and then in "retirement" working as an independent consultant and as an Honorary Research Fellow at Birkbeck College.

REFERENCES

BRANTON, P., 1985, The social importance of ergonomics. *Ergonomist*, 183, 2–3.

BRANTON, P. and SHIPLEY, P., 1986, VDU stress: is Houston man addicted, bored or mystic? In B. Knave and P. Wildbaeck (eds), *Proceedings of International Scientific Conference on Work with Display Units* (Stockholm).

BRANTON, P., 1987, In praise of ergonomics — a personal perspective. *International Reviews of Ergonomics*, 1, 1–20.

LEAL, F., 1993, Paul Branton as a philosopher. In D.J. Oborne, R. Branton, F. Leal, P. Shipley and T. Stewart (eds), *Person-centered Ergonomics: A Brantonian View of Human Factors* (London: Taylor & Francis).

OBORNE, D.J., BRANTON, R., LEAL, F., SHIPLEY, P. and STEWART, T., 1993, *Person-centered Ergonomics: A Brantonian View of Human Factors* (London: Taylor & Francis).

Biosketch: Paul Morris Fitts

R. W. Pew

BBN Technologies, 10 Moulton Street, Cambridge, MA 02138, USA

Paul M. Fitts was a pioneer in the field of Human Factors Engineering, which he preferred to call Engineering Psychology. He was the first Director of the Psychology Branch of the AeroMedical Laboratory at Wright Patterson Air Force Base, which is now called, after several metamorphoses, the Human Effectiveness Division of the Air Force Research Laboratory. He was a Founding Member of the Human Factors Society of America, now the Human Factors and Ergonomics Society. He was also a prominent Experimental Psychologist who saw only the fuzziest of boundaries between theoretical research and applied system development. His chapter "Engineering Psychology" in the prodigious *Handbook of Experimental Psychology* (edited by S.S. Stevens and published in 1951) was a landmark definition of the field and provided exposure of this new discipline to a generation of psychology graduate students. Regrettably, his career was cut short by his untimely death from a heart attack in 1965 at 52 years of age.

Paul Morris Fitts was born in 1912 in Martin, Tennessee. He received his Bachelor of Science degree in Psychology at the University of Tennessee in 1934, his Master of Arts at Brown University in 1936 and his PhD in Psychology at the University of Rochester in 1938, under Dr Leonard Carmichael. His thesis was on motivation in animals. He immediately returned to the University of Tennessee as an Assistant Professor of Psychology. However, World War II interrupted his career and he moved to Washington in April 1941 to work with a team of psychologists in the US Army Air Forces Office of the Air Surgeon on problems of Air Force pilot selection and training. He served the Air Forces for more than 8 years, first as a civilian, then for 4.5 years as an Air Force officer, rising from first Lieutenant to Lieutenant Colonel by the time of his discharge in October 1946, and then for another 3 years as a civil servant. It was during this time that the US military began to see that psychologists could play a role in determining how to design equipment to adapt it to human requirements and Dr Fitts identified strongly with this new role.

On 1 July 1945, a Psychology Branch of the Aero Medical Laboratory was established at Wright Field in Dayton, Ohio. Dr Fitts, then aged 33, was appointed its first Director and it soon became the unit responsible for all aspects of engineering psychology for the Army Air Forces.

At the Psychology Branch Dr Fitts became involved in setting the research agenda for this new field of "Engineering Psychology" in the military aviation world. He also become involved in several research studies himself. He is probably best known for the pioneering analysis of pilot error experiences in reading and interpreting aircraft instruments and in operating aircraft controls (Fitts and Jones 1947a, b). They cataloged those mistakes on the basis of the design features to which they could be attributed. In addition to several design impacts that could be addressed directly, these studies stimulated research in the Air Force for many years to come. The instrument-reading study was the first to document problems with reading the three-pointer altimeter. The Aviation Safety Reporting System (ASRS) is an extensive NASA data base of commercial aircraft human error incidents anonymously reported and widely used for research that may be regarded as a contemporary manifestation of this early work.

In 1949 Dr Fitts returned to Ohio State University and reactivated the Laboratory of Aviation Psychology. It was later renamed the Systems Research Laboratory when it ceased to be dominated by aviation research. He remained at Ohio State, active in research, teaching and administration, and on the national scene until 1958 when he moved to the University of Michigan where he remained until his untimely death. At Michigan he broadened his interest in the theoretical aspects of human information processing and, together with three colleagues from the Psychology Department, founded the Human Performance Center, a unit within the Psychology Department that became a focus for faculty and graduate students interested in human performance and human information processing.

Throughout his career, Fitts championed the application of psychological research to equipment or systems design. He accomplished this initially through his own research, but after he left the Air Force, his work took a more academic turn, emphasizing research questions that had a payoff in applied work. However, his direct impact on engineering design continued to be felt through his consulting and his participation in significant engineering psychology developments nationally. He was instrumental in helping to organize Division 21 of the American Psychological Association, the division concerned with applied experimental and engineering psychology. He was a founding member and later President of the Human Factors Society, now the Human Factors and Ergonomics Society. He was chairman of several psychology and social science panels, boards and committees sponsored by the Air Force, the Department of Defense and the National Research Council that were influential in establishing a role for engineering psychology.

It was at Ohio State where he conducted the early research and published what is now perhaps his most famous paper in *Journal of Experimental Psychology* in 1954 entitled "The information capacity of the human motor system in controlling the amplitude of movement." Here he showed that an index of difficulty of a movement could be defined in information theoretic terms that accurately predicted movement time as a function of the accuracy required of a movement and the distance moved. The greater the accuracy required, the slower the movement, and the longer the distance moved, the slower the movement. Equation (1) is currently used to describe this relationship, which has come to be called Fitts' Law:

$$\text{MT} = a + b\ [\log_2\ (2A/W)], \quad (1)$$

where A = distance from the starting point to the center of the target; W = target width, and a and b are constants reflecting the specific movement circumstances. In the original paper and in subsequent work he showed that only the constants had to be adjusted to describe (1) reciprocal tapping where the size and separation of the target plates was varied; (2) disk transfer where the difference in the diameter of the disk hole and the peg onto which it was placed reflected the accuracy constraint; (3) a peg transfer task in which pegs were transferred to holes of different sizes; (4) the effect of adding a weight to the hand; and (5) for discrete or continuous movements and under various instructions to be rapid or accurate. Numerous investigators have examined

the predictions in great detail, most notably Welford (1960, 1968), and while marginal gains in predictive accuracy have been proposed by adjustments in the equation, the fundamental relationship and the robustness of the predictions have never been challenged. In subsequent work it has been shown to work for foot movements (Drury 1975), movements of the head (Soede *et al.* 1973), and even movements made with tweezers while looking through a binocular microscope (Langolf *et al.* 1976).

At The University of Michigan, Dr Fitts most important work was in human information processing, more specifically, he used a variety of research paradigms in which the dependent variable was human reaction time or response time. An important contribution was the quantification of the relationship between speed and accuracy of performance. He showed in the laboratory that given quantitative payoffs for performance that favored either speed and accuracy, the human operator would adjust behavior appropriately to maximize that payoff, which meant either responding faster and making more errors or slowing down and being more accurate, depending on the relevant monetary payoff at the time. He went on to develop a predictive sequential sampling statistical decision model that accounted for these trade-offs (Fitts 1966).

It seems likely that Dr Fitts introduced the term "Engineering Psychology." If he did not actually introduce it, surely he was the first to popularize it and to influence its adoption to represent psychology's interests in the human's role in system design and development. He used it in his introduction to a series of Psychology Branch research papers in 1947 to refer to the emerging area of psychological research on equipment design (Fitts 1947). Then in 1951 it became the centerpiece of the title of his chapter in Stevens' *Handbook of Experimental Psychology* (Fitts 1951).

The most interesting thing about Fitts's work is that while the scope of his theoretical and applied interests were very broad, his research had a thread of continuity and specificity that gave it focus. His original interests were stimulated by his introduction to issues from the applied perspective at the Psychology Branch of the Aero Medical Laboratory, and he sustained work on them virtually throughout his career, generalizing and broadening their importance and impact at each new opportunity. All these topics fell under the general rubric that he would come to call *human skilled performance*.

For more detail on Dr Fitts' life, see Pew (1994).

REFERENCES

DRURY, C.G., 1975, Application of Fitts' Law to foot-pedal design. *Human Factors*, 17, 368–73.

FITTS, P.M., 1947, Introduction to psychological research on equipment design. In P.M. Fitts (ed.), *Psychological Research on Equipment Design*. Report 19 of the Army Air Forces Aviation Psychology Program Research Reports (Washington, DC: US Government Printing Office), 1–19.

FITTS, P.M., 1951, Engineering psychology and equipment design. In S.S. Stevens (ed.), *Handbook of Experimental Psychology* (New York: John Wiley), 1287–40.

FITTS, P.M., 1954, The information capacity of the human motor system in controlling the amplitude of movement. *Journal of Experimental Psychology*, 47, 381–91.

FITTS, P.M., 1966, Cognitive aspects of information processing III. Set for speed versus accuracy. *Journal of Experimental Psychology*, 71, 849–57.

FITTS, P.M. and JONES, R.E., 1947a, *Analysis of Factors Contributing to 460 "Pilot-Error" Experiences in Operating Aircraft Controls*. USAF Air Materiel Command Memorandum Report No. TSEAA-694-12.

FITTS, P.M. and JONES, R.E., 1947b, *Psychological Aspects of Instrument Display. I. Analysis of 270 "Pilot-Error" Experiences in Reading and Interpreting Aircraft Instruments*. USAF Air Materiel Command Memorandum Report No. TSEAA-694-12A.

LANGOLF, G.D., CHAFFIN, D.B. and FOULKE, J.A., 1976, An investigation of Fitts' Law using a wide range of movement amplitudes. *Journal of Motor Behavior*, 8, 113–28.

Pew, R.W., 1994, Paul Morris Fitts, 1912–1965. In H.L. Taylor (ed.), *Division 21 Members Who Made Distinguished Contributions to Engineering Psychology* (Washington, DC: American Psychological Association), 23–44.

SOEDE, M., STASSEN, H.G., VAN LUNTEREN, A. and LUITSE, W.J., 1973, A light spot operated typewriter for physically handicapped patients. *Ergonomics*, 16, 829–44.

WELFORD, A.T., 1960, The measurement of sensory-motor performance: survey and reappraisal of twelve year' progress. *Ergonomics*, 3, 189–230.

WELFORD, A.T., 1968, *Fundamentals of Skill* (London: Methuen).

Biosketch: Ross A. McFarland

S. H. Snook

Department of Environmental Health, Harvard School of Public Health, Boston, MA, USA

Ross A. McFarland was one of the founding fathers of Human Factors and Ergonomics in the USA. He was also a founder of the Human Factors Society and served as its president from 1969 to 1971.

Dr McFarland was born in Denver, Colorado, in 1901, and was raised in Missouri. He received a bachelor's degree from the University of Michigan in 1923, and a PhD from Harvard University in 1928 in Physiological Psychology. He pursued an academic career of research and teaching. As a Research Fellow at the University of Cambridge in England from 1927 and 1928, Dr McFarland showed how the lack of oxygen during simulated flights could impair the behavior of RAF student pilots, giving rise to lack of insight and loss of judgment.

From 1929 to 1937 Dr McFarland was an Instructor of Psychology in the Department of Psychology at the College of Physicians and Surgeons, Columbia University, New York City. He investigated the central nervous system's complete dependence upon a normal supply of oxygen, glucose and other organic constituents. In studies using visual tests of differential light sensitivity, he demonstrated for the first time that impairments may be present at altitudes as low as 4000 feet in unacclimatized subjects. These and other findings led to regulations for the use of oxygen by pilots in civil aviation and by military pilots during night combat.

In 1935 Dr McFarland was a member of the International High Altitude Expedition to South America where he continued his studies on oxygen deprivation in acclimatized subjects. During the 4-month expedition members of the party were compared with natives living at altitudes up to 20 000 feet in Chile and Peru.

Dr McFarland joined the Harvard Fatigue Laboratory in 1937 as an Assistant Professor of Industrial Research. At that time the laboratory was housed in the basement of the Harvard Business School. Working with the sociologist Elton Mayo, he investigated some of the physiological problems of industrial workers. His studies centered around mental tests made under various environmental stresses (heat, cold, noise, vibration, etc.).

When Pan American Airways opened air routes over the Pacific and Atlantic Oceans in 1937 and 1939, Dr McFarland was asked to study the fatiguing effects of long distance flights on pilots. He was also an advisor to the airline in 1940 when air routes were openend across Africa. During 1939 and 1940 he became interested in pilot selection and the development of better tests in predicting success or failure. Because of the high failure rates in the military services, Dr McFarland and a group of colleagues at Harvard were asked to make a comprehensive analysis of this problem on 1000 naval aviators. Their results were published as *The Pensacola Study of Naval Aviatlors*. Thirty years later, when follow-up studies were made of the same aviators, the results were not only validated, but also they led to a greater understanding of problems related to heart disease and aging.

With the outbreak of World War II, it was apparent that many older persons would be required to work in industry. Through experimental studies of the aging process, Dr McFarland demonstrated that, with proper placement and supervision, men and women could work productively much longer than originally believed. He stressed the importance of functional rather than chronological age in judging an individual's ability to perform. He also pointed out the close relationship between the oxidative processes and certain functions of the central nervous system. He demonstrated these close relationships in tests of light sensitivity, immediate memory and the loss of insight. His theory, relating the sensory and mental changes seen in aging to alterations in the oxygenation of the body's tissues, has gained widespread acceptance.

In 1943 and 1944 he served as an Operations Analyst for the 13th Air Force in the Solomon Islands campaign, studying combat fatigue in air and ground forces. It was here that he became interested in the problems of designing equipment to meet human capabilities and limitations. He wrote one of the first textbooks in human factors, *Human Factors in Air Transport Design* (McGraw-Hill, 1946), intended primarily for engineers.

Dr McFarland became a member of the faculty of the Harvard School of Public Health, when the Harvard Fatigue Laboratory moved there in 1947. It was here that he developed a new approach to some of the difficult problems of health and safety, not only in the air, but also on the ground and in space. He was one of the first to emphasize the multiple causation of accidents, and to apply the methods of biostatistics and epidemiology in the study of highway injuries and fatalities. Many of the lessons learned in aviation and aircraft design were now applied to automotive design and safety. Research programs included the application of anthropometric and biomechanical data to vehicle design and operation, and the effects of alcohol and toxic agents (e.g. carbon monoxide) on driver performance. Many of the resulting principles have been incorporated into federal regulations relating to air and highway safety. Dr McFarland's new approach attracted many young physicians and engineers to the Harvard School of Public Health for advanced study. In 1953 he published his second book entitled *Human Factors in Air Transportation – Occupational Health and Safety* (McGraw-Hill), primarily for physicians and safety engineers.

In 1957 the Guggenheim Foundation sponsored a teaching and research center at Harvard, and in 1962 the center was endowed. Dr McFarland became the first Director, and the first Daniel and Florence Guggenheim Professor of Aerospace Health and Safety. He and his colleagues at the Harvard School of Public Health trained more than 200 young scientists. Many of his students assisted in the medical aspects of the space program, and others achieved leadership positions in the fields of aerospace medicine, occupational health and highway safety. In 1966, he published his third book, *The Human Body in Equipment Design*, with Drs Albert Damon and Howard W. Stoudt (Harvard University Press). It has been widely used by engineers and industrial designers.

In addition to his three books, Dr McFarland contributed over 200 publications to the scientific literature. Among the many

awards recognizing his professional achievements are the Longacre Award of the Aeromedical Association in 1947 for the safe utilization of aircraft, the Flight Safety Foundation Award in 1953, the John Jefferies Award of the Institute of Aeronautical Sciences in 1956 for contributions to aeronautics through medical research, the Walter M. Boothby Award from the Aerospace Medical Association in 1962 for research in aviation medicine, the Exceptional Service Award of the US Air Force in 1969, and the Distinguished Civilian Service Award of the Department of the Army in 1971. In 1963 he was the first American invited to present the Ergonomics Research Society Lecture, "In Search of a Theory of Ageing", in the UK. He was awarded the Honorary Doctor of Science Degree by Park College, Rutgers University, Trinity College and the University of Denver.

Dr McFarland had also been a consultant or technical advisor to many federal agencies, including the Department of Health, Education and Welfare, the Federal Aviation Administration, and the National Aeronautics and Space Administration. He served on numerous task forces and study groups dealing with various aspects of health and safety. He was also active in numerous scientific and technical organizations, including the Human Factors Society, the Ergonomics Research Society, the American Psychological Association, and the Aerospace Medical Association.

Dr McFarland died in 1976 at the age of 75. He left behind a legacy of outstanding scholarship and warm friendship. His contributions to the field of human factors and ergonomics were enormous, and he was valued highly as a teacher and advisor by his students. He was truly a pioneer of human factors and ergonomics in the USA, and a professional giant among his peers.

REFERENCES

FORBES, W.H., MOELLER, D.W., REED, R.B., THORN, G.W. and FERRIS, B.G. Jr, 1978, *Harvard Public Health Alumni Bulletin*, 35, 4–6.

STOUDT, H.W., 1976, *Human Factors Society Bulletin*, 19, 1–2.

Biosketch: Wojciech Bogumil Jastrzebowski

D. Koradecka (editor)
A. Golebiowska (translator)

In the treatise entitled "An Outline of Ergonomics, or the Science of Work Based Upon the Truths Drawn from the Science of Nature", published in *Nature and Industry: a weekly devoted to accessible presentation of all branches of natural sciences, their practical application to living, and the latest discoveries and inventions*, Poznan, Poland (1857), W. B. Jastrzebowski created the foundations of ergonomics.

Wojciech Bogumil Jastrzebowski, the inventor, scientist, educator, and naturalist, was born on 19 April 1799, in Gierwat, Poland. His father died before Wojciech was born, and his mother died when he was nine years old. He attended a primary school in Janów. In 1816, he began attending a secondary school, but due to the poverty and ill health his education was often interrupted. In December 1820, Jastrzebowski began studying at the Department of Building Engineering and Surveying of the University of Warsaw. His knowledge, enthusiasm, and ingenuity made his professors assign him several supporting research projects. In September 1822, he also began studying at the Department of Philosophy (Natural History Unit). During his studies he assisted in the research of several biologists, astronomers and zoologists.

At that time he was commissioned to set up the famous sun dial in the Warsaw Lazienki Park (the sun dial is still there today). As setting up a sun dial required individual measurements for each location, he designed a special piece of equipment "for determining compasses in any space and in every location". The Government Commission called it "Jastrzebowski's Compass", and the inventor was accepted into the ranks of the Warsaw Friends of Science Society.

When the (Polish) November Uprising against the Russian invaders broke out in 1830, Jastrzebowski fought at the battle of Olszynka Grochowska. At that time Jastrzebowski also developed a proposal for the creation of a League of Nations, which opened with the words: "Peace in Europe is permanent and everlasting." According to this proposal, a European Congress should publish a proclamation calling upon all the nations to form a brotherly alliance. All disputes between states would be defused by a commonwealth of nations, whose decisions would be objective and just, and when there was no more unhealthy armaments rivalry between friendly nations, all their effort should be directed towards improving the education of the youth, laws, science, agriculture and industry.

After the fall of the (Polish) November Uprising, Jastrzebowski could not go back to his job at the University of Warsaw. In 1836, he became Professor of Botany, Physics, Zoology and Horticulture in the Institute of Agriculture and Forestry in Marymont. Thanks to him, the Marymont gardens became true protected areas of rare shrubs and trees. Jastrzebowski's students were taught to be methodical and consistent, restrained and patient in the laborious task of fact collecting. At the same time, however, he cautioned them not to become simple archivists of facts; they were to try to unravel the mystery of their creation. Jastrzebowski devoted a lot of time to field trips with his students across all regions of Poland. Those expeditions resulted in a study entitled "Forecasting fair weather, sleet, wind and other changes of the air" (1847), which was also translated into Russian. Thanks to the study's reputation, Jastrzebowski was admitted to several Polish scientific societies, including the Krakow Science Society, the Kielce Agriculture Society, and the Agriculture Society in Lvov.

In addition to the historic treatise: "An Outline of Ergonomics, Or the Science of Work based upon the truths Drawn from the Science of Nature" (1857), Jastrzebowski published "Stichology" and "Mineralogy" (1852). Upon leaving the Institute in Marymont, Jastrzebowski was given the post of works inspector, preserving the dunes of Czerwony Bór and planting trees on them. He settled in Feliksówka, where he created model gardens, nature rooms, and a dormitory for trainees. During the next (Polish) uprising of January 1863 the dormitory emptied. His students and his two sons joined the uprising.

The "Florae Polonicae Prodomus" — even today one of the basic textbooks of plant geography in Poland — was published by Józef Rostafinski in Vienna in 1872. In it, he describes 1550 plants, of which 1090 are labelled with Jastrzebowski's name since their names had been taken from his herbarium. In 1874, Jastrzebowski left his study in Czerwony Bór and tried to recreate a garden with his favourite yews and larches on a small plot of land in the Warsaw suburb of Czyste. At the same time he accepted the offer of the Warsaw–Vienna Railway to plant hedges round its stations and stops. The last time he went out into a field for a rendezvous with nature was in 1879. Jastrzebowski died on December 30, 1882, and was buried at the Powazki Cemetery in Warsaw.

First edition in English: Wojciech Jastrzębowski "An outline of Ergonomics, or the Science of Work based upon the truths drawn from the Science of Nature". Editor: Danuta Koradecka, Warszawa, Central Institute for Labour Protection 1997, translation of Wojciech Jastrzębowski's text: Teresa Baluk-Ulewiczowa, translation of the biographical note: Aleksandra Gołębiowska.

Cognitive Engineering

T. Lambie

Ergonomics and HCI Unit, University College, 26 Bedford Way, London, WC1H 0AP, UK

1. INTRODUCTION

The concept of cognitive engineering is not fixed and unequivocal, but in its various expressions are found certain common features. The fluidity of the idea is not surprising since its arrival is relatively recent, and its formation into something more determinate awaits the outcome of its practices as all disciplines do. There is probably a consensus on what constitutes cognition or cognitive behavior — reasoning, remembering, planning, etc. However, agreeing on what we mean by engineering — particularly of cognition or cognitive behavior — is not so straightforward. One of the reasons that this is so is that engineering is not adequately understood in its relationship with science, and this has perhaps implications for how both science and engineering are understood. The question as it affects technology and engineering is addressed, for example, by the aeronautical engineer Vincenti (1990).

Understanding cognitive engineering is therefore understanding what engineering is by itself and in connection with cognition or cognitive behavior, and there are perhaps two sources for the introduction of the idea into the currency: Simon (1969) and Norman (1981).

2. ORIGINS

Although Simon did not originally give the name engineering to his "artificial science" of the mind, much of his argument suggested that, in general terms, there were two distinct disciplines — pure science and systematic design — that traditional engineering was stultified and, by implication, that the practice of design should be reconceived to get rid of this stultification. This reluctance to call his artificial science of the mind engineering was partly because he saw cognitive science — the investigation of how cognition worked — as dependent partly on cognitive design (e.g. the uses of computing models — *designing* them — to elucidate aspects of cognition). In the context of Artificial Intelligence, for example, he later (Simon 1980) writes that we should perhaps make an explicit recognition that there are two disciplines "by speaking of cognitive engineering as well as cognitive science": that the study of artificial intelligence "is interested, in its applied aspect, not only in understanding but [also] in improving it", and that if we make this recognition then he hopes "that the two ventures will keep in the closest relation with each other, as they have done through the past quarter century. The dangers of confusing the normative with the positive are slight compared with the losses that would be suffered from isolating the science from its engineering applications". The wording of the last sentence strongly implies that it is cognitive science that will suffer if they fail to communicate, not cognitive engineering. In Simon's view, in any case, the discipline of systematic cognitive design is a new and distinct kind of knowledge that has at least a troublesome relationship with science.

Norman (1987), more typically a member of the Cognitive Ergonomics (including HCI) community, writes, however, that he invented the term "cognitive engineering" "to *emphasise* the cognitive aspects of human–machine interaction. … Cognitive engineering is meant to combine with the applied disciplines not to replace them". It represents for Norman a "new approach … more than just psychology … more than psychology coupled with engineering. We need all the disciplines of cognitive science, plus engineering." So for Norman the adoption of the term means that we should broaden our vision of what is involved in the work of HCI; more importantly, that it should lead to principled design that takes us beyond the solution of particular design problems as a craft discipline, with each solution a new one unrelated, except tacitly, to any other. Norman, however, appears to have a more pragmatic interest than Simon in the adoption of an engineering view of cognition. Although he wants to institute a novel approach which provides a greater guarantee so that design knowledge acquired in one project may be carried forward with confidence to support the solution of similar design problems, he makes little of the gaps or the connections between scientific and engineering knowledge.

3. IN PRACTICE

Because it is more precise, methodical, etc. engineering is widely regarded as a more responsible and, consequently, a more politically and socially respectable mode of design. It is associated with operationalization and implementation of design and as such is more practical, out-in-the-open and down-to-earth. So, cognitive systems engineering as practiced by Hollnagel, Rasmussen, etc. was at least partly adopted and developed because it was concerned with systems that were safety critical and complex. As the work matured it attracted attention because it offered better means to design. Rasmussen's work is particularly notable in this respect. Representations and models were offered of a general nature, so that a language of representation might allow generalization across design problems (also Woods and Roth 1988); e.g. formal expressions of behavior at different levels (from tasks to physical operations) corresponding with different levels of knowledge description (from symbols to signs), these to be consolidated by empirical work. The question of the origins of these models and representations is not raised, but the detail and care with which they are expressed and employed means that engineering cognitive systems is probably, both ostensibly and actually, more precise, methodical, etc. It is called engineering therefore to distinguish it from design as art or intuition-based, but it is not just cognitive science either. This version of cognitive engineering is forward-looking to precision and the testing of models and representations rather than backward-looking to the its epistemological roots. The expression of this approach is refined and recapitulated with particular relevance to Cognitive Ergonomics and HCI by Woods and Roth (1988).

For the Cognitive Ergonomics and HCI community, engineering cognition (or cognitive engineering) is partly the expression of a desire to design and build as rigorously and carefully, and, arguably, therefore as explicitly and publicly as possible. The motivation for engineering is, thus, to achieve the aims of a design which answers the requirements of the users and satisfies them of, for example, the safety or financial economy of the design, since the design would not persuade us that it would continue doing what it was claimed to do were it not methodical, precise and explicit. Engineering is supposed to bring

more than faith to the business of design. Nevertheless, it is not enough to aim for design which has precision, rigor, and is public and open to criticism. The grounds for, or principles of, this type of knowledge, which is different from other given kinds, must be examined, and its distinct properties established, or we might be being persuaded of the precision, rigor, etc. of scientific truths not design certitudes.

4. IN PRINCIPLE

The determination of the boundary between science and engineering, as well as being one of the *consequences* of the practical advantages of an open and systematic design activity, might also be one of the *motivations* for the inception of the practice of a distinct discipline of cognitive design such as is involved in Human Factors (HF), HCI, Computer Supported Cooperative Work (CSCW), etc. Conventional engineering disciplines, e.g. aeronautical or electrical engineering, have formed into professions distinct from that of their cognate scientific disciplines (Vincenti 1990). They have an independent existence drawn on by the need to solve specific technical problems which are not soluble by reference to the universal models or theories of science but rather by the need for the construction of models contrived for the particular context set by the design requirements.

It has been found increasingly that addressing design problems in HF, HCI, CSCW, etc. also necessitates turning attention away from the psychological, computational or sociological research of a scientific nature (which aims at universality) and focusing on the problem posed by the target artifact, including the constraint that it meet the given requirements (which aim at the eccentricities of the particular). For Dowell and Long (1998) HF is largely a craft, "the heuristics it possesses being either 'rules of thumb' derived from experience or guidelines derived informally from psychological theories and findings", with the latter representing the science applied, i.e. the more or less *simple* laying of the scientific template onto the design problem to generate the design solution.

However, even if applied science were quite different from pure science and not its *simple* application but were endowed with techniques, methods and practices developed with ingenuity and rigor, in other words, a discipline whose practices resulted in general features which became the ground for addressing novel problems, nevertheless, it would still fall short of fully fledged systematic design or engineering. This is so because systematic design or engineering is not only concerned with the solution of particular problems by the development of particular models, contrasted with the general theories of pure science, but also must engage with the idea of the fabrication of artifacts — the making of something new or the changing of the physical world in some way to lead to an improvement in the performance of some system or other. What arises, then, is a triad — of pure science, applied science and engineering — with engineering principally oriented towards normative or prescriptive knowledge, while science (in its distinct forms as pure and applied) is aimed at positive or descriptive knowledge. This contrast between science and engineering is addressed expressly by Simon, as mentioned above, and by others more casually.

5. FROM DEFINITION TO DISCIPLINE

Without such a journey into the epistemology of science and technology, starting with a definition of engineering takes us a good way forward. Whiteside *et al.* (1988) claim, "Engineering almost always involves construction but it differs from merely 'building' something … design for a purpose (as contrasted with tinkering and puttering) is always an essential part of the process," and, "engineering always operates against a background of scarce resources: talent, materials, time, and money."

Whiteside *et al.* have a particular interest in 'usability engineering', but carrying through their reasonable definition takes us quite a way towards a characterization of the discipline of cognitive engineering more generally — from the inside out rather than from the outside in, as offered by the epistemological approaches. The scarceness of the resources — both physical and cognitive (Whiteside *et al.* include the cognitive resource "talent") — mean that the solution of design problems is dominated by the aim of arriving at *effective* solutions. The designer need not chase the vain goal of optimality or greatest economy except in relation to the requirements, but this property of effectiveness must be pursued with as much explicitness and rigor as possible to fulfill the aims of engineering as defined.

In particular, the effectiveness sought, according to the above, bears on, if carried out conscientiously, the parsimonious use of physical as well as cognitive resources, both consistent with the requirements. So, to focus on (1) the best way (with respect to the requirements) of carrying out the tasks (the specification of the work to be done) and (2) the most economical (with respect to the requirements) employment of cognitive resources results in a division of the design problem into two partitions — that of the work to be carried out, and that of the agency (cognitive or joint cognitive systems) to carry it out. Such a description of the general design problem for cognitive engineering is to be found in Dowell and Long (1998): the duality comprising, respectively, the Domain of Work (the domain) and the Interactive Work System (IWS). The procedure of design, in this conception of cognitive engineering, is to determine the domain and design the IWS to reflect this domain, taking into account the constraints on the cognitive system/s. Further, consistent with the definition of cognitive engineering and with its troublesome relationship with scientific knowledge, the domain should be determined in as *unscientific* (while remaining systematic) a manner as possible. The domain's determination should be "ecological" (also Woods and Roth 1988), i.e. the designer should examine the work as done in its normal setting, not some (probably) unrepresentative model of the work: *in vivo*, therefore, not *in vitro*. This does not exclude resorting to scientific knowledge, but *as such* the input must be recognized to be as conjectural as any other inspiration. The knowledge, whatever its source, becomes engineering knowledge through the practices devised and generalized in the course of solving the design problem.

6. CURRENT AND FUTURE VIEW

Whether considered from a practical or principled angle, or from an analysis of its terminological components, there are remarkable similarities between the various versions of cognitive engineering, and, in large part, the important aspects of the emerging discipline are summed up by Woods and Roth (1988). They describe, in round terms, its attributes as exhibited by Rasmussen (1986), Norman (1987), Dowell and Long (1988) and others. However, Woods and Roth's analysis does not dwell on the manner in which

technological knowledge is defined with respect to scientific knowledge, and, for some, this is what needs addressing in particular.

There is undoubtedly a problem with the unclear status of design knowledge for reasons to do with reliability or guarantee. If design knowledge is loosely connected to scientific knowledge it is important to know how loosely in order to monitor and control the connection, but once this is achieved or believed to be achievable we have something like a distinct applied science with its own rules and practices. And if this transition is acknowledged we cannot easily avoid the same reflection on the transition from this distinct discipline of applied science to that of engineering, since plainly we must account for the move from the positive to the normative which poses different requirements and results in a different domain specification — this transition bringing with it new dangers of degraded reliability. Addressing this problematic connection between the applied scientific knowledge and the putative engineering knowledge leads, consequently, to the definition of engineering — in this case, cognitive engineering. Given the many common features of the different strands of cognitive engineering it might be profitable to establish a common denominator in the form of an analytic or epistemological approach to the delineation of the discipline, thereby helping to bring unity to the fragmented area of cognitive ergonomics and HCI.

Cognitive engineering is a natural development in the face of increasing technical complexity and as a result of the need to control that complexity. Together they force a re-examination of the kind of knowledge that supports systematic design. We look

in general to science to provide guarantees of our knowledge of the natural world. However, perhaps our *understanding* of the natural world needs also the knowledge of how we manipulate it — creating the artificial — and that this knowledge must be allowed to establish its own guarantees; Simon's fear that cognitive engineering might become estranged from cognitive science suggests that cognitive engineering can only complement and consolidate cognitive science if it is truly independent. However, contrary perhaps to his untroubled attitude to the confusion of the positive and the normative, this independence requires constant vigilance in the discrimination of the kind of knowledge exploited by cognitive engineers.

REFERENCES

DOWELL, J. and LONG, J., 1998, Conception of the cognitive engineering design problem. *Ergonomics*, 41, 126–38.

NORMAN, D., 1987, Cognitive engineering — cognitive engineering. In Carroll (ed.), *Interfacing Thought* (Cambridge, MA: MIT Press).

RASMUSSEN, J., 1986, *Information Processing and Human–Machine Interaction: An Approach to Cognitive Engineering* (New York: North Holland).

SIMON, H., 1980, Cognitive science: the newest science of the artificial. *Cognitive Science*, 4, 33–46.

VINCENTI, W., 1990, *What Engineers Know and How They Know It* (John Hopkins University Press).

WOODS and ROTH, 1988, Cognitive systems engineering. In M. Helander (ed.), *Handbook of Human–Computer Interaction* (Amsterdam: Elsevier).

Core Competencies in Ergonomics

International Ergonomics Association
Human Factors and Ergonomics Society, PO Box 1369, Santa Monica, CA 90406, USA

1. INTRODUCTION

Any mature discipline and profession requires understanding of its core competencies. Also, the exercise of defining core competencies is itself well worth while, because it prompts a profession to look closely at itself, its goals and its perceived contribution to society. Once complete, it provides a record of standards by which the profession can ensure quality of performance.

2. COMPETENCY STANDARDS

Competency standards do not themselves represent an outline of certification requirements, although they may be a resource for the certification process. Nor do they represent a curriculum document, although they may help direct the development of a curriculum.

3. DEFINITION OF COMPETENCY

A competency is a combination of attributes underlying some aspect of successful professional performance. An outline of core ergonomics competencies should describe what it is that ergonomists can do in practice.

4. TERMS

Ergonomics competency standards have been developed in terms of units, elements and performance criteria, which is the accepted format.

- *Units of competency* reflect the significant major functions of the profession or occupation.
- *Elements of competency* describe the identifiable components of ergonomics performance which contribute to and build a unit of competency.
- *Performance criteria* describe the standards expected of performance in the ergonomist's work. Expressed in terms of outcomes and professional ergonomics performance, they provide the basis on which an expert assessor could judge whether the performance of the ergonomist reached the standard acceptable for professional practice.

5. SCOPE OF ERGONOMICS

The scope of ergonomics is broad, across many domains. Ergonomists can be involved in both proactive and retrospective problem solving. The contexts for ergonomics practice are diverse and ergonomics must relate to the workplace, transport, the home or to leisure activities, or to the use of a variety of products. The IEA Core Competencies must acknowledge this diversity and should be interpreted with this breadth of scope in mind.

6. USES OF CORE COMPETENCIES

Ergonomics core competencies could be used in a variety of ways. These include:

- Development or review of curricula in ergonomics.
- Accreditation of new and existing ergonomics educational programs.
- Development of comprehensive and equitable assessment processes for the evaluation of a person's professional competence.
- Recognition by ergonomics certification authorities of the competency of graduates holding qualifications in ergonomics conferred by recognized institutions.
- Assessment of competence of eligible overseas qualified ergonomists seeking to practice in another country.
- Assessment of eligible ergonomists who have not practiced for a defined period of time and who are seeking to re-enter the profession or to be re-certified.
- Development of continuing education programs offered by the federated societies.
- Determination of need for continuing professional education by employers.
- Preparation of public information defining ergonomics roles and responsibilities.

7. BENEFITS OF NATIONAL (AND INTERNATIONAL) COMPETENCY STANDARDS

Those who have been involved with the application of Competency Standards have found them of benefit in the following ways:

- National consistency.
- Chance to examine the profession and its scope.
- Better definition of the profession.
- Basis for communication at a national (and international) level.
- A resource for education establishments and curricula.
- Provision of a more equitable basis for certification.
- Quality assurance.

8. REVIEW OF COMPETENCY STANDARDS

Any set of competencies has a limited life and this IEA document will be reviewed on a regular basis.

9. PRESENTATION

The core competencies have been presented in two formats "summary" and "full".

The *summary* version presents the units and elements of ergonomics competency as a summary, for those who require a concise overview. It is expected that any assessment of an individual or program would benchmark against this summary.

The *full version* presents a complete set of units, elements and performance criteria to illustrate the standards of performance required. This full version would be used to illustrate and give more detail on examples of, and criteria for, professional performance against which judgement can be made.

10. VERSION 2, MAY 1999, PPE COMMITTEE: SUMMARY OF CORE COMPETENCIES IN ERGONOMICS: UNITS AND ELEMENTS OF COMPETENCY

Unit 1. Investigates and Analyses the Demands for Ergonomics Design to Ensure Appropriate Interaction Between Work, Product or Environment and Human Capacities and Limitations

1.1 Understands the theoretical bases for ergonomics planning and review.

1.2 Applies a systems approach to analysis.

1.3 Understands the requirements for safety, the concepts of risk, risk assessment and risk management.

1.4 Understands and can cope with the diversity of factors influencing human performance and quality of life, and their interrelationships.

1.5 Demonstrates an understanding of methods of measurement and interpretation relevant to ergonomics appraisal and design.

1.6 Recognizes the extent and limitations to own professional competence

Unit 2. Analyses and Interprets Findings of Ergonomics Investigations

2.1 Evaluates products or work situations in relation to expectations for safe and effective performance.

2.2 Appreciates the effect of factors influencing health and human performance.

2.3 Analyses and interprets research data accurately and without bias, consulting appropriately where required.

2.4 Understand relevant current guidelines, standards and legislation.

2.5 Makes and can justify decisions regarding relevant criteria which would influence a new design or a solution to a specified problem.

Unit 3. Documents Ergonomics Findings Appropriately

3.1 Provides a succinct report in terms understandable by the client and appropriate to the project or problem.

3.2 Communicates clearly to the relevant workforce or general public, and if feasible to the scientific community.

Unit 4. Determines the Compatibility of Human Capacity with Planned or Existing Demands

4.1 Appreciates the extent of human variability influencing design.

4.2 Determines the quality of match and the interaction between a person's characteristics, abilities, capacities and motivation, and the organization, the planned or existing environment, the products used, equipment, work systems, machines and tasks.

4.3 Identifies potential or existing high-risk areas and high-risk tasks, where risk is to health and safety of the individual completing the task or any others affected.

4.4 Determines whether the source of a problem is amenable to ergonomics intervention.

4.5 Justifies decisions on ergonomics interventions or implementations.

Unit 5. Develops a Plan for Ergonomic Design or Intervention

5.1 Adopts a holistic view of ergonomics.

5.2 Incorporates approaches that would improve quality of life as well as performance.

5.3 Develops strategies to introduce a new design

5.4 Considers alternatives for improvement of the match between the person and the product, the task or the environment.

5.5 Develops a balanced plan for risk control, with understanding of prioritization and costs and benefits involved.

5.6 Communicates effectively with the client, any stakeholders, the public and professional colleagues.

Unit 6. Makes Appropriate Recommendations for Ergonomics Changes

6.1 Makes and justifies appropriate recommendations for design-based changes

6.2 Makes and justifies appropriate recommendations for organizational planning-based changes

6.3 Makes and justifies appropriate recommendations for personnel selection, education and training

Unit 7. Implements Recommendations to Improve Human Performance

7.1 Relates effectively to clients and all stakeholders, at all levels of personnel.

7.2 Supervises the application of the ergonomics plan.

7.3 Manages change effectively and sympathetically

Unit 8. Evaluates Outcome of Implementing Ergonomics Recommendations

8.1 Monitors effectively the results of ergonomics change implementation

8.2 Carries out evaluative research relevant to ergonomics

8.3 Makes sound judgements on the quality and effectiveness of ergonomics change implementation

8.4 Modifies a design or program in accordance with the results of evaluation, where necessary.

8.5 Understands the principles of cost–benefit analysis for any ergonomics change.

Unit 9. Demonstrates Professional Behavior

9.1 Shows a commitment to ethical practice and high standards of performance and acts in accordance with legal requirements.

9.2 Recognizes personal and professional strengths and limitations and acknowledges the abilities of others.

9.3 Maintains up-to-date knowledge of national strategies and scientific state of the art, relevant to ergonomics practice.

9.4 Recognizes the impact of ergonomics on people's lives.

11. CORE COMPETENCIES IN ERGONOMICS: FULL OUTLINE

11.1. Units, Elements and Performance Criteria

11.1.1. Unit 1. Investigates and assesses the demands for ergonomic design to ensure the optimal interaction between work, product or environment and human capacities and limitations

11.1.1.1.   Element 1.1. Understands the theoretical bases for ergonomic planning and review of the workplace

Performance criteria:

1.1a Understands theoretical concepts and principles of physical and biological sciences relevant to ergonomics.

- Demonstrates a working knowledge of physics, chemistry, mathematics, anatomy, functional anatomy, physiology, pathophysiology, exercise physiology and environmental science as they apply to ergonomics practice.
- Can apply knowledge of biomechanics, anthropometry,

motor control, energy, forces applied as they relate to stresses and strains produced in the human body.

- Demonstrates an understanding of the pathology relating to environmentally or occupationally generated disorders or causes of human failure.

1.1b Understands the effects of the environment (acoustic, thermal, visual, vibration) on human health and performance.

1.1c Understands theoretical concepts and principles of social and behavioral sciences relevant to ergonomics.

- Demonstrates a working knowledge of sensory, cognitive and behavioral psychology and sociology, and recognizes psychological characteristics and responses and how these affect health, human performance and attitudes.
- Can apply knowledge of information intake, information handling and decision making; sensory motor skills, human development and motivation principles as they relate to human performance.
- Understands the principles of group functioning and socio-technical systems.

1.1d Understands basic engineering concepts, with a focus on design solutions.

- Demonstrates an understanding of design and operation of technologies in which they work.
- Appreciates hardware design problems.
- Understands and can apply the basics of industrial safety

1.1e Understands and can apply the basics of experimental design and statistics.

1.1f Understands the principles of organizational management.

- Demonstrates an understanding of individual and organizational change techniques, including training, work structuring and motivational strategies.

1.1g Demonstrates an understanding of the principles of ergonomics and human–machine interface technology.

11.1.1.2. Element 1.2. Applies a systems approach to analysis

Performance criteria:

1.2a Demonstrates a knowledge of the principles of systems theory and systems design and their application to ergonomics.

1.2b Demonstrates a knowledge of the principles of ergonomics analysis and planning in a variety of contexts, and the scope of information required to ensure quality of life.

1.2c Understands the determinants and organization of a person's activities in the field and plans the analysis according to the organization's strategy and purposes.

1.2d Can explain the scientific or empirical rationale for appraisals selected and has the expertise required to perform them.

1.2e Identifies the demands of the situation and accesses sources of appropriate information.

1.2f Develops action plans with those involved and identifies the critical factors of the ergonomic analysis.

1.2g Carries out a systematic, efficient and goal orientated review of demands appropriate to ergonomics, addressing the needs of the project.

11.1.1.3. Element 1.3. Understands the requirements for safety, the concepts of risk, risk assessment and risk management

Performance criteria:

1.3a Recognizes the importance of safety principles, guidelines and legislation in risk management

1.3b Understands the goals of risk management.

- Demonstrates ability to manage change.
- Understands how to gain commitment of management and participation of worker in risk management approaches.

11.1.1.4. Element 1.4. Understands and can cope with the diversity of factors influencing human performance and quality of life and their interrelationships

Performance criteria:

1.4a Understands the organizational, physical, psychosocial and environmental factors which could influence human performance, an activity, a task, or use of a product and knows how to cope with adverse conditions.

1.4b Understands the impact of individual factors on other possible factors and the implications for ergonomic assessment.

1.4c Recognizes those aspects of the environment that are flexible and changeable.

11.1.1.5. Element 1.5. Demonstrates an understanding of methods of measurement relevant to ergonomic appraisal and design

1.5a Understands the type of quantitative and qualitative data required to clarify the basis for ergonomic appraisal and design, and validates the measurements selected for data collection and/or application.

1.5b Demonstrates the ability to carry out appropriate surveillance of the nature and magnitude of risks.

1.5c Selects the appropriate form of measurement for the particular context.

1.5d Applies measurement procedures and uses measurement instruments effectively, or refers appropriately to other ergonomics team members, to quantify load on the person and human characteristics.

1.5e Understands the concepts and principles of computer modeling and simulation.

1.5f Understands the use of the computer for data acquisition, analysis and design development.

11.1.1.6. Element 1.6. Recognizes the scope of personal ability for ergonomic analysis

1.6a Appreciates when it is necessary to consult and collaborate with a person with different professional skills to ensure comprehensive measurement taking and analysis.

11.1.2. Unit 2. Analyses and interprets findings of ergonomics investigations

11.1.2.1. Element 2.1. Evaluates products or work situations in relation to expectations for error-free performance

Performance criteria:

2.1a Determines the demands placed on people by tools, machines, jobs and environments.

2.1b Evaluates user needs for safety efficiency, reliability and durability, and ease of use of products and equipment and how these are met.

11.1.2.2. Element 2.2. Appreciates the effect of factors influencing health and human performance

Performance criteria:

2.2a Has a basic understanding of the mechanisms by which work or prolonged exposure to environmental hazards may affect human performance or be manifested in injury, disorder or disease.

2.2b Defines efficiency, safety, health and comfort criteria.

2.2c Specifies the indicators of poor match between people and their tools, machines, jobs and environments.

11.1.2.3. Element 2.3. Consults appropriately regarding analysis and interpretation of research data

11.1.2.4. Element 2.4. Analyses current Guidelines, Standards and legislation, regarding the variables influencing the activity

Performance criteria:

2.4a Refers to and applies relevant scientific literature and national and international recommendations and standards appropriate to the project.

2.4b Matches measurements against identified Standards.

11.1.2.5. Element 2.5. Makes justifiable decisions regarding relevant criteria which would influence a new design or a solution to a specified problem

11.1.3. Unit 3. Documents ergonomic findings appropriately

11.1.3.1. Element 3.1. Provides a succinct report in terms understandable by the client and appropriate to the project or problem

11.1.4. Unit 4. Determines the compatibility of human capacity and planned or existing demands

11.1.4.1. Element 4.1. Appreciates the extent of human variability influencing design

Performance criteria:

4.1a Understands the influence of such factors as a user's body size, skill, cognitive abilities, age, sensory capacity, general health and experience on design features.

11.1.4.2. Element 4.2. Determines the match and the interaction between a person's characteristics, abilities, capacities and motivations, and the organization, the planned or existing environment, the products used, equipment, work systems, machines and tasks

11.1.4.3. Element 4.3. Identifies potential or existing high risk areas and high risk tasks

11.1.4.4. Element 4.4. Determines whether the source of a problem is amenable to ergonomic intervention

11.1.5. Unit 5. Develops a plan for ergonomic design or intervention

11.1.5.1. Element 5.1. Adopts a holistic view of ergonomics in developing solutions

Performance criteria:

5.1a Identifies the relative contribution of organizational, social, cognitive, perceptual, environmental, musculoskeletal or industrial factors to the total problem and develops solutions accordingly.

5.1b Considers the impact of legislation, codes of practice, Government Standards and industry-based standards on defined problems and possible solutions.

11.1.5.2. Element 5.2. Incorporates approaches which would improve quality of life in the working environment

Performance criteria:

5.2a Provides opportunities for self-development.

5.2b Considers factors influencing the person's sense of satisfaction with the workplace.

11.1.5.3. Element 5.3. Develops strategies to introduce a new design to achieve a healthy and safe work place

Performance criteria:

5.2a Understands the iterative nature of design development.

5.2b Recognizes the practicalities and limitations of applying ergonomics, including the introduction of change.

5.2c Prepares a design specification report based on the systematic analysis to meet the objectives of the project, for use by industrial designers, engineers, computer scientists, systems analysts, architects or other professionals.

11.1.5.4. Element 5.4. Considers alternatives for optimization of the match between the person and the product, the task or the environment and to achieve a good performance

Performance criteria:

5.3a Establishes appropriate short and long-term goals relevant to the defined problems, in consultation with the client.

5.3b Considers the options available and the balance of approaches to be applied, relevant to the objectives.

5.3d Considers the potential benefits and costs of each form of ergonomic solution.

11.1.5.5. Element 5.5. Develops a balanced plan for risk control

Performance criteria:

5.4a Appreciates the background information required for effective risk management.

5.4b Understands how to control adverse physical and chemical conditions and major pollutants.

5.4c Establishes priorities in relation to level of risks identified, and to their consequences for health safety.

5.4d Selects appropriate forms of risk control, based on theoretical knowledge and ergonomics practice and develops a

comprehensive, integrated and prioritized approach for realistic risk control.

5.4e Identifies where assistive devices and aids could enhance compatibility between the person and the environment.

5.4f Considers the needs of special groups (e.g. aging or disabled).

11.1.5.6. Element 5.6. Communicates effectively with the client and professional colleagues

Performance criteria:

5.5a Discusses with the client, users and management the design or intervention strategies available, their rationale, realistic expectations of outcome, limitations to achieving outcome, and the costs of the proposed ergonomics plan.

5.5b Establishes effective relationships and collaborates effectively with professional colleagues in other disciplines in the development of ergonomic design solutions.

11.1.6. Unit 6. Makes appropriate recommendations for ergonomic design or intervention

11.1.6.1. Element 6.1. Understands the hierarchies of control systems

6.1a Recognizes the safety hierarchy, application of primary and secondary controls and the order of introducing controls.

11.1.6.2. Element 6.2. Outlines appropriate recommendations for design or intervention

Performance criteria:

6.2a Utilizes the systems approach to human–workplace integrated design for new or modified systems and understands design methodology and its use in systems development.

6.2b Applies correct design principles to design of products, job aids, controls, displays, instrumentation and other aspects of the workplace, work and activities and considers human factors in the design of any utility.

6.2c Drafts systems concepts for a functional interaction of tasks/technological variants, work means/tools, work objects/materials, work places/work stations and the work environment.

6.2d Develops appropriate simulations to optimize and validate recommendations.

6.2e Outlines details of the appropriate concept and develops specific solutions for testing under realistic conditions.

6.2f Provides design specifications and guidelines for technological, organizational and ergonomic design or redesign of the work process, the activity and the environment which match the findings of ergonomic analysis.

6.2g Is able to justify recommendations.

11.1.6.3. Element 6.3. Outlines appropriate recommendations for organizational management

Performance criteria:

6.3a Understands the principles of total quality management.

6.3b Recognizes the need to design organizations for effective and efficient performance and good quality of work place.

6.3c Recommends changes to the organizational design appropriate to the problem identified.

6.3d Considers issues such as participation, role analysis, career development, autonomy, feedback and task redesign as appropriate to the client and defined problem.

11.1.6.4. Element 6.4. Makes recommendations regarding personnel selection

Performance criteria:

6.4a Recommends personnel selection where appropriate as part of a balanced solution to the defined problem.

6.4b Applies appropriate criteria for personnel selection, where relevant, according to the nature of the demands.

11.1.6.5. Element 6.5. Develops appropriate recommendations for education and training in relation to ergonomic principles

Performance criteria:

6.5a Understands current concepts of education and training relevant to application of ergonomic principles, including encouragement of learning.

6.5b Implements effective education programs relevant to understanding the introduction of ergonomic measures or to the control of potential risks in the workplace, home, public or leisure environments, and to achieve safe and comfortable and successful performance and productive output in new and/or changed activities.

11.1.7. Unit 7. Implements recommendations to optimize human performance

11.1.7.1. Element 7.1. Relates effectively to clients at all levels of personnel

Performance criteria:

7.1a Communicates with the users, management and other professional colleagues in relation to method of implementation of the new design or risk control measures.

7.1b Uses appropriate processes to motivate the client to participate in the recommended ergonomics program and to take responsibility for achieving defined goals.

7.1c Where appropriate, provides individual guidelines for personnel in a form understandable to the client.

11.1.7.2. Element 7.2. Supervises the application of the ergonomic plan

Performance criteria:

7.2a Implements appropriate design or modifications.

7.2b Facilitates the adaptation to new approaches to activity.

7.2c Provides appropriate feedback on progress to client.

7.2d Incorporates methods to allow continuous improvement.

11.1.7.3. Element 7.3. Manages change effectively

Performance criteria:

7.3a In a work environment, where necessary, overcomes resistance of workers, managers and labor unions to change, and gains their cooperation for implementing new approaches.

11.1.8. Unit 8. Evaluates outcome of implementing ergonomic recommendations

11.1.8.1.   Element 8.1. Monitors effectively the results of ergonomic design or intervention

Performance criteria:

8.1a Selects appropriate criteria for evaluation.

8.1b Assesses level of acceptance of and satisfaction with implemented ergonomic measures.

8.1c Produces clear, concise, accurate and meaningful records and reports.

11.1.8.2.   Element 8.2. Carries out evaluative research relevant to ergonomics

Performance criteria:

8.2a Demonstrates rational, critical, logical and conceptual thinking.

8.2b Critically evaluates new concepts and findings.

8.2c Demonstrates a knowledge of basic research methodology for ergonomics research in an area relevant to individual ergonomic expertise.

11.1.8.3.   Element 8.3. Makes sound judgements on the quality and effectiveness of ergonomics design or intervention

Performance criteria:

8.3a Considers the cost effectiveness of the program in terms of financial implication, improvement in productivity, product usability and human requirements for the enhancement of comfort and safety.

11.1.8.4.   Element 8.4. Modifies the program in accordance with results of evaluation, where necessary

11.1.9. Unit 9. Demonstrates professional behavior

11.1.9.1.   Element 9.1. Shows a commitment to ethical practice and high standards of performance and acts in accordance with legal requirements

Performance criteria:

9.1a Behaves in a manner consistent with accepted codes and standards of professional behavior.

11.1.9.2.   Element 9.2. Recognizes personal and professional strengths and limitations and acknowledges the abilities of others

Performance criteria:

9.2a Recognizes extent of own knowledge in ergonomics, appreciates areas where knowledge and skill are lacking and knows what to do and whom to contact to access missing expertise.

9.2b Demonstrates a desire for life long learning, regularly reviews and updates knowledge and skills relevant to current practice of ergonomics, to ensure appropriate breadth and depth of understanding.

9.2c Recognizes those areas of ergonomics where knowledge is limited and consults appropriately with professional colleagues to ensure application of relevant expertise to particular problems.

9.2d Recognizes the value of tem work between multidisciplinary experts.

11.1.9.3.   Element 9.3. Maintains up-to-date knowledge of national strategies relevant to ergonomics practice

Performance criteria:

9.3a Demonstrates knowledge of government legislation relating to occupational health, control of environmental hazards and other areas relevant to ergonomics practice.

9.3b Understands the industrial, legal and liability issues that impact upon professional ergonomics practice, and takes appropriate action regarding them.

11.1.9.4.   Element 9.4. Recognizes the impact of ergonomics on peoples' lives

Performance criteria:

9.4a Appreciates the social and psychological impact of ergonomics investigations.

9.4b Appreciates professional responsibilities and requirements.

Cultural Ergonomics

M. Kaplan, B. B. Morgan Jr,* and J. Kring‡

International Center for Cultural Ergonomics, *College of Arts and Sciences, and ‡Department of Psychology, University of Central Florida, Orlando, FL 32816, USA

1. THE CONCEPT OF CULTURAL ERGONOMICS

1.1. History

The term "cultural ergonomics" was first introduced in print by Kaplan (1991) while formally setting forth many of the field's features. Pioneering contributions by Chapanis (1975) and Wisner (1989) appeared earlier. The environment in which the new discipline began to emerge had been expanded by two relevant factors: (1) concerns for industrially developing countries had been introduced especially by Sen (1984), Shahnavaz (1984), and Kogi (1985); (2) Hendrick's groundbreaking work (Hendrick 1987) on macroergonomics had enlarged the scope of ergonomics by encouraging attention to ergonomically relevant human-human concerns. The concentration on cultural considerations fits into this milieu very well.

1.2. Focal Elements Encompassed by the Concept

1.2.1. The cultural milieux of ergonomics

The role of the cultural settings in which work activities occur has been emphasized by Kaplan (1998). Noting the great variety of human groupings, both large and small, in the world's approximately 250 countries (themselves milieus), he argued that "whenever we focus upon a particular example of work, we may speak of the cultural milieu in which it is occurring." Various milieus are shown in Figure 1.
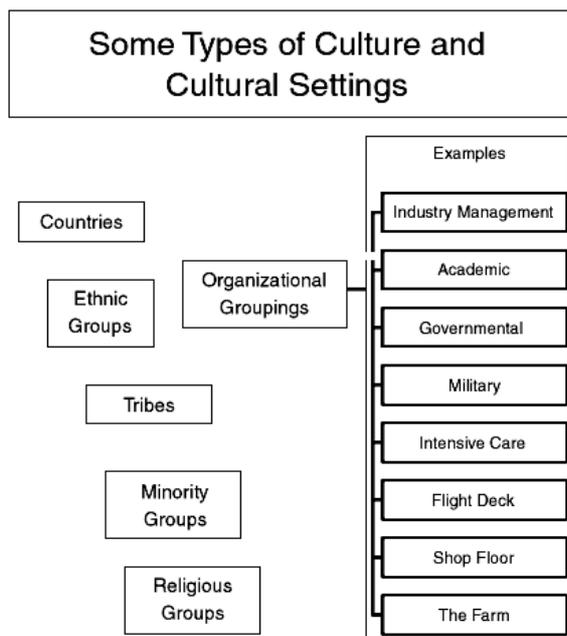
Figure 1. Types of Culture and Cultural Settings
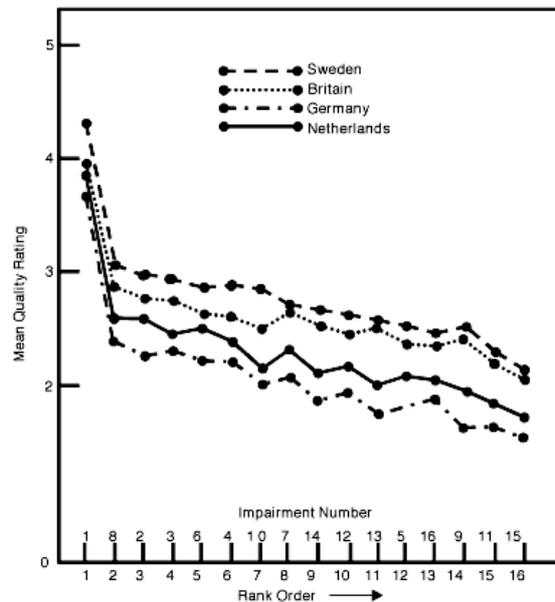
Figure 2. A performance-culture relationship: judgment of picture quality on a monitor as a function of equipment impairment, with country of judges as a parameter (adapted from Gleis and Renhall, 1986; reprinted with permission of Televerket, Farsta, Sweden).

The kinds of human work are many and various. But according to Kaplan, whatever the work, when it is a subject for ergonomic tasking, it calls for recognition of cultural considerations. It was recommended that, when approaching a task, the ergonomist should ask the following questions and, in its execution, take appropriate account of the answers:

- What are the characterisitcs of the cultural milieu in which the work is performed?
- What are the implications for conduct of the work?
- What are accustomed ways of behaving and performing the tasks by the individuals involved?
- What do individuals bring from *their* cultural backgrounds to *this* cultural/work environment?

1.2.2. Human performance and human interfacing

In work environments worldwide, the performance of tasks and, when applicable, the human interfacing that affects the performance are susceptible to cultural influence. This is a central feature of cultural ergonomics, and it applies both to individual performance and to the group performance that is integral to teams and sociotechnical systems. It also gives rise to the notion of culture as an independent variable and as a parameter of functions relating properties of performance to other independent variables that determine them (Figure 2).

More broadly, one may speak of performance-culture relationships, where the performance is occurring in a work environment. Such relationships and pertinent problems requiring solutions and research can be seen throughout topical areas of ergonomics and in relevant aspects of related disciplines such as experimental psychology, industrial psychology, learning psychology, sensory psychology, social psychology, industrial

engineering, business and management, anthropology, and political science.

1.2.3. Cross-national and cross-cultural comparisons

The necessity of developing workable approaches to ergonomic issues with wide-ranging and even global consequences has engendered a need for acquisition and codification of comparative national and cultural information for designing effective and desirable work procedures and programs. Typical areas of concern are (1) international standardization, (2) international variations in occupational health and safety regulations, as in the US National Institute of Occupational Safety and Health (NIOSH) and comparable agencies in other countries, and (3) well-being of workers (International Labour Office 1984).

1.2.4. Ergonomic involvement in critical global-level issues

Extending beyond the confines of discrete cultural settings are problems in which the world itself is the cultural milieu. As enumerated by Moray (1995), they include population pressure, pollution, water shortage, urbanization, food, energy, health and medicine, waste, violence and terrorism, migration, and the clash of cultures. They go beyond usual notions of ergonomics, but Moray sees a role for ergonomics. Arguing that solution of these problems "requires changing human behaviour," he suggests an interdisciplinary approach where "what ergonomics has to contribute to [these] problems ... is essentially a technology for changing behaviours to that which offsets the problems."

1.3. A Summary of Cultural Ergonomics

An integration of the foregoing focal elements, cultural ergonomics is currently regarded as a *coalescence of topical areas* belonging to ergonomics and related disciplines, where each topical area is *susceptible to some form of cultural influence*. It seeks to understand how cultural factors influence and interact with human performance and human interfacing in work environments worldwide. In relation to the cultural settings in which they occur, it explores relevant applied human factors problems and endeavors to design effective solutions for them.

2. TAXONOMY

Given the sizable number and variety of topical areas that are encompassed in the subject matter of cultural ergonomics, there has been a felt requirement for the organization of its content into a workable framework (Kaplan 1998). Here is a proposed taxonomy.

2.1. Categories of Cultural Concern and Culturally Influenced Topical Areas

Comprehensive categories of cultural concern in ergonomics are subdivided into culturally influenced topical areas. Here are the proposed categories followed by some of their topical areas:

- Interaction of cultural variables and standard areas of HFE
  Cultural factors in anthropometry, biodynamics, audition, voice communication, visual perception and displays, controls, human-computer interaction, lighting, temperature, humidity, ventilation, ability assessment, work layout, shift work and hours of work, safety, simulation and training.
- Industrially developing countries and technology transfer
  Analysis of cultural differences with regard to technology

design and utilization (Shahnavaz 1994), ergonomic problems, implementing technology transfer, adapting work procedures to local practices, needed training, adapting training to culture, applying anthropotechnology, using approaches from cognitive anthropology, language and communication problems, shift work.
- Comparative national and cultural approaches to safety issues
  Aviation and aerospace safety; railroad, maritime, and highway safety; local culture influence on the safety culture and power plants; regulation of chemical, gas, biological, and nuclear hazards; safety on the shop floor; industrial safety regulation; international variation in power plant design and safety (Parsons and Taylor 1995); interaction with time of day and performance; attention to age factors; role of abilities and problem-solving facility.
- Organizational design, functioning, and management
  Heterocultural teams and workplace diversity, multinational enterprises, international space station, cultural interactions with macroergonomics and participatory ergonomics, cultural variations in motivators, language and communication in business and industry, modifying products for other countries, modifying work and management procedures for other countries, influence of individualism or collectivism and related dimensions in varied work environments.
- Cultural variation and training needs
  Managing cultural influences affecting students' reactions to cultural or national variations in the teaching of knowledge and skills, adapting teaching to variations in cultural backgrounds, determining and dealing with variations in cultural reactions to simulation training, designing training for harmonious functioning of heterocultural teams, using training techniques from behavior analysis to *rapidly* overcome gaps between deficient skill levels in industrially underdeveloped countries and those required for implementing new technologies, using these techniques to produce the behavioral modifications necessary for resolution of global ergonomic problems (Moray 1995).

3. COMPLEXITY IN CULTURAL INFLUENCE: HETEROCULTURAL TEAMS

When examining performance-culture relationships, it is general to observe discrete cultural factors generating particular performance consequences, as in a foreign language or accent clouding accurate communication between pilot and tower in the English-language environment of air traffic control. Also there are multiple consequences from the heterocultural composition of teams (Morgan *et al.* 1994) and sociotechnical systems. They arise from the involvement of more than one individual and from the interactions generated within these groups.

Attention to cultural and ergonomic concerns in the functioning of teams and sociotechnical systems is of special importance, since they lie at the heart of so many work activities throughout the world, some of them critical for human survival, e.g., air transport, defence operations, and coping with disaster.

3.1. Interactions

Whatever the outcomes of their functioning, whatever the measures of their performance, teams and sociotechnical systems

reflect the contributions of individual members and components. And each individual is exposed to cultural factors. These cultural factors may influence or determine characteristics of the individuals that in turn affect their performance.

Among these characteristics, themselves at least partially *functions of culture*, one can cite aspects of personality; motivation and specifiable incentives to work; knowledge, skills, and abilities; and attitudes toward a variety of concerns, such as authority, other individuals and groups, work requirements, and one's self.

It is a complex undertaking to specify the cultural factors and the ways in which they relate to particular characteristics, and it is equally complex to spell out (a) how these characteristics interact to determine individual performance and (b) how individual performances interact to produce team and system outputs. Systematic research on these matters is indicated.

3.2. Comparative Effectiveness of Heterocultural and Homogeneous Teams

Ultimately, culturally oriented microanalysis may be useful in accounting for the effectiveness of heterocultural team and system outputs. Current examples of these outputs and some cultural aspects of the teams producing them are introduced in the next sections.

3.2.1. Positive consequences of cultural diversity for performance outcome

In some cases of group decision making, heterocultural groups can be more effective than their homogeneous counterparts in identifying problems and generating alternative solutions. Research by Watson *et al.* (1993) showed this effect appearing midway in a 17-week group decision-making task. Each of their heterocultural groups contained a White American, an African-American, a Hispanic-American, and a foreign national from Asia, Latin America, Africa, or the Middle East.

In actual work settings that bring together individuals from different cultures, "groupthink" and its consequences can be avoided in decision-making activities. The quality and diversity of new ideas, as in the management strategy of idea generation and implementation, can be significantly improved by drawing on team members' unique cultural qualities and strengths. A similar positive contribution of diversity is evident in the design of work procedures (e.g., in space walks) and in engineering plans for the international space station, which draw upon specialized experience and resources from 16 nations.

3.2.2. Negative consequences of cultural diversity for performance outcome

Hostility arising from cultural background issues can lead to errors and failures in the execution of team tasks. The international space station provides the setting for potentially serious problems and danger in the interfacing among its crew members. Here are some culturally generated issues that have caused difficulties on short international missions and which have been reported in surveys and interviews: personal hygiene, religious practices, food preferences, misinterpretation or lack of humor, approaches to decision making, refusal to accept alternative points of view (Santy *et al.* 1993, Lozano and Wong,1996).

Cultural variation in the perception of risk is another problem. Research by Vredenburgh and Cohen (1995) showed that com-

pared with Asian subjects, Hispanic, Caucasian, and African-American groups perceived a significantly higher risk in certain situations. Wherever joint participation is required to assure that a dangerous task is accomplished, it could be jeopardized by those team members who, perceiving too high a risk, may be loath to take part. So-called independent and interdependent views of one's self, a function of culture, when jointly represented in a team, may disrupt group performance.

In the independent case, typical of Western cultures, the self is seen as separate from the "social context" and bears little relationship to family, friends, and coworkers. The interdependent case, common to Asian cultures, entails deep involvement with social context (Markus and Kitayama 1991). It has been suggested that where both types are present, independent members might seek to stand out from other team members or promote their own goals over and above the team's goal, thus lessening team cohesiveness and possibly generating conflict.

4. PROGRESS IN CULTURAL ERGONOMICS

4.1. Infrastructure

An International Center of Cultural Ergonomics has been initiated at the University of Central Florida. A central place for channeling information and for communicating among workers and students in the field, it maintains the Cultural Ergonomics Clearinghouse and a website called Culturelink; it also fosters research and educational projects.

4.1.1. Data files

Two sets of files, the *topical area files* and the *regional/cultural area files* have been started within the Cultural Ergonomics Clearinghouse. Their purpose is to organize in a central location both new and hitherto scattered examples of performance-culture relationships associated with the numerous topical and regional/cultural areas. They are intended as a resource for practitioners, researchers, and students.

4.2. Extension of the Taxonomy

Here are some categories of cultural concern that may be added to the taxonomy or given increased attention: (1) human-computer and human-systems interfaces, (2) standardization, (3) ergonomic approaches to aging problems, (4) ergonomic approaches to disability problems, (5) ergonomics and technology of design.

4.3. University Curricula

The first formal course in cultural ergonomics has been introduced at the University of Central Florida along with supervised graduate study in the field. Arguments have been adduced for increasing graduate study and graduate research as part of ergonomics curricula (Kaplan 1998).

REFERENCES

CHAPANIS, A.C. (ed.), 1975, *Ethnic Variables in Human Factors Engineering* (Baltimore: Johns Hopkins University Press).

GLEIS, N. and RENHALL, P., 1986, Measurement of picture quality. *TELE*, 40, 21-25.

HENDRICK, H.W., 1987, Macroergonomics: a concept whose time has come. *Human Factors Society Bulletin*, 30, 1-3.

INTERNATIONAL LABOUR OFFICE, 1984, Improving working conditions and environment: an international programme (Geneva: ILO).

KAPLAN, M., 1991, Cultural ergonomics: an evolving focus for military human factors. In Gal and Mangelsdorff (eds) *Handbook of Military Psychology* (London: John Wiley), pp. 155-67.

KAPLAN, M., 1998, The cultural milieu of ergonomics. Paper presented at the Global Ergonomics Conference, 9-12 September, Cape Town, South Africa.

KOGI, K., 1985, Improving working conditions in small enterprises in developing Asia (Geneva: ILO).

LOZANO, M.L. and WONG, C.K., 1996, Multicultural human factors concerns aboard the International Space Station Alpha. *Proceedings of the Human Factors Society 39th Annual Meeting*, pp. 912-16.

MARKUS, H.R. and KITAYAMA, S., 1991, Culture and the self: implications for cognition, emotion, and moptivation. *Psychological Review*, 98, 224-53.

MORAY, N., 1995, Ergonomics and the global problems of the twenty-first century. *Ergonomics*, 38, 1691-707.

MORGAN, B.B. Jr, BOWERS, C.A. and JENTSCH, F., 1994, Issues in the performance of heterocultural teams. Paper presented at the International Congress of Applied Psychology, July, Madrid, Spain.

PARSONS, S. and TAYLOR, L., 1995, IEA guidelines for application of human factors/ergonomics principles in the design and operations of electric power systems, Parsons and Associates, Saratoga CA.

SANTY, P.A., HOLLAND, A.W., LOOPER, L. and MARCONDES-NORTH, R.,1993, Multicultural factors in the space environment: results of an international shuttle crew debrief. *Aviation, Space, and Environmental Medicine*, March, 196-200.

SEN, R.N., 1984, Application of ergonomics to industrially developing countries. *Ergonomics*, 27, 1021-32.

SHAHNAVAZ, H., 1984, The importance of ergonomics in developing countries. In Shahnavaz and Babri (eds) *Ergonomics in Developing Countries* (Lulea: Lulea University Press), pp. 4-27.

SHAHNAVAZ, H., 1994, Ergonomics, culture. and technology. Paper presented at the Symposium on Perspectives in Cultural Ergonomics, 12th Congress of the International Ergonomics Association, Toronto, Canada.

VREDENBURGH, A. G., and COHEN, H. H., 1995, *Proceedings of the Human Factors and Ergonomics Society 30th Annual Meeting*, pp. 1015-1019.

WATSON, W.E., KUMAR, K. and MICHAELSON, L.K., 1993, Cultural diversity's impact on interaction process and performance: comparing homogeneous and diverse task groups. *Academy of Management Journal*, 36, 590-602.

WISNER, A., 1989, *Vers une anthropotechnologie* (Paris: Conservatoire des Arts et Metiers).

Defining Ergonomics/ Human Factors

M. S. Wogalter
Department of Psychology, North Carolina State University, USA
P. G. Dempsey
Liberty Mutual Research Center for Safety and Health, Hopkinton, MA, USA
P. A. Hancock
Human Factors Research Laboratory, University of Minnesota, USA

The focus of this encyclopedia is ergonomics and human factors. The title *International Encyclopedia of Ergonomics and Human Factors* in some respects suggests that ergonomics and human factors are possibly two separate subject areas which are being covered. The conjunctive "and" strongly supports this interpretation. If ergonomics and human factors were synonymous, why not just use one name? Several years ago, the main professional organization of this area in the United States decided to change its name by adding the term ergonomics. They also decided to add the conjunctive "and." It is now called the Human Factors and Ergonomics Society (HFES). If ergonomics and human factors are the same, then why not use a slash instead of the "and" and call it Human Factors/Ergonomics Society? And for that matter why not call this encyclopedia, the *International Encyclopedia of Ergonomics/Human Factors*? It is difficult to pronounce a slash (so it is usually silent), but persons unfamiliar with the field would probably find the slash version even more strange sounding, hearing a string of three words and not knowing if one or two are adjectives. If you are going to keep both ergonomics and human factors in the name — 'and' sounds better. In fact, there is a journal that uses "and", *Human Factors and Ergonomics in Manufacturing.* This journal is edited by the same main editor of this encyclopedia (Professor W. Karwowski).

Many professionals consider the terms ergonomics and human factors synonymous, although others do not necessarily concur. To some, ergonomics has a traditional relationship with the physical aspects of work, while human factors has a greater relationship to cognitive involvement. Ergonomics evolved from studying the interactions between humans and their surrounding work environment (with environment defined broadly to include machines, tools, the ambient environment, tasks, etc.). Use of the term "human factors" tends to be a North American phenomenon with individuals who do work (research, teaching, practicing) that is most concerned with "above the neck" processing (perceptual and cognitive processes, etc.). The rest of the world more frequently use the term ergonomics to include "above the neck" processing as well as "below the neck" processing. In the latter, the areas of biomechanics and physical workplace design are emphasized. The use of the term ergonomics in the United States typically implies "below the neck" activities.

More recently in the United States, ergonomics has been added to names having the label human factors as ergonomics became better known (through mass market public media, such as advertising for cars and chairs). Also the superordinate organization, a level over the national and culture-specific organizations of the field, is called the International Ergonomics Association. Further, all of the worldwide societies use a form of the word ergonomics, not human factors.

Since ergonomics has been more closely allied with the physical aspects of the field, to distinguish it better from the more mental/cognitive part, there has been increasingly greater use of the label "cognitive ergonomics" versus "industrial or occupational ergonomics." Indeed a technical group in HFES has been formed using this name. There is a journal, *International Journal of Cognitive Ergonomics*. One could contrast the name with another journal called the *International Journal of Industrial Ergonomics*, although there is some overlap in subject matter covered in them. We expect to see greater use of the term ergonomics, but we also expect to see people distinguishing between the physical and cognitive sides of the field.

There appears to be a growing consensus that human factors and ergonomics (HF/E) refer essentially to a common body of knowledge. Despite this confluence, we still suffer from a lack of name recognition. The lay public, business, government, and academics generally do not have much of an idea what the field is all about. Most individuals have little problem understanding what established areas like physics, chemistry, mathematics, and astronomy deal with. Like similarly recognized subjects such as history and geography, these areas form the basis of school curricula. The relatively new field of psychology, which is about 100 years old, has become such a well-recognized area that many high schools now offer courses in it. HF/E has this not reached this level of exposure. In fact, exposure to the field is rather scant even for students in colleges and universities. Martin and Wogalter (1987) examined the availability of HF/E courses to college students in the United States. Fifty schools were selected randomly from each of four categories of universities and colleges (research, doctoral, masters, and baccalaureate/liberal arts) from a listing of four-year colleges and universities in the United States. Only 2% (one school) in the sample of liberal arts colleges and only 10% of the master's universities had a course in HF/E. Of the doctoral institutions, 62% had not a single HF/E course, and 44% of the research institutions had no HF/E courses. Other than a brief mention in a back chapter of an introductory psychology text book or of an industrial/organizational psychology text book, most college students have virtually no (or at best, scant) opportunity to learn about the field. This is particularly true if the university does not have an Industrial/Systems Engineering Department.

One obvious and crucial problem lies in the two predominant names that we have talked about above. Human factors is a general, indistinct term; one cannot derive from this name the content of the knowledge domain. A lay person might guess that the field deals with human beings, but they probably would not recognize that it deals with (among other things) people interfacing with technology. Rather, the lay person might expect that a human factors psychologist deals with some special form of therapy, or perhaps, person-to-person interaction (and interestingly, this is one of the few domains that human factors does not address). Also, an engineer who says their area is human factors will also have problems eliciting much understanding by lay persons either.

With the term ergonomics, the problems are different. One is that, unfortunately, the word ergonomics is very close to economics; the two can easily be confused by listeners and

readers. But considering this differently, this resemblance can turned into an advantage as Hendrick (1996a) did in his influential publication entitled "Good Ergonomics is Good Economics."

The "ergo" of ergonomics means work. The breadth of the field could be considered constrained by this prefix. Thus, how "work" is defined is critical. Many people may limit "work" to mean activities associated with employment. This frame of reference would not include leisure pursuits, an area certainly covered by the field's intent. Work can, however, be interpreted broadly, as in its meaning that it involves the general physical expenditure of energy to accomplish a goal. Thus, most of what of humans do (and their bodily processes) could be justifiably considered to be work, and thereby, ergo-related.

But what besides work and the involvement of humans define the field? Whatever the actual name, it should be asked how the area is bound, what is its unique knowledge content, what are its central theses, and how do we provide a concise, succinct statement that characterizes the area? Here, we address the definition question, not simply as another exercise in polemics, but rather as a fundamental evaluation of where our area stands at the start of a new millennium and to distill a way to advance our enterprise to a higher level of societal recognition and value.

One way to examine how an area embraces its domain is to see how it is being represented in various definitions. Definitions reflect how people specify some topic or concept using available language. Terms most frequently used to describe an area's scope

Table 1. Example definitions.

Brown, O., and Hendrick, H.W. (1986)
. . . the relations between man and his occupation, equipment, and the enviroment in the widest sense, including work, play, leisure, home, and travel situations.

Chapanis, A. (1995)
. . . is a body of knowledge about human abilities, human limitations and other human characteristics that are relevant to design

Hancock, P. A. (1997)
. . . is that branch of science which seeks to turn human-machine antagonism into human-machine synergy.

Mark, L.S. and Warm, J.S. (1987)
. . . attempts to optimize the fit between people and their enviroment.

Howell, W. and Dipboye, R. (1986)
Person-machine system design.

Meister, D. (1989)
. . . the application of behavioral principles to the design, development, testing and operation of equipment and systems.

Clark, T.S. and Corlett, E.N. (1984)
. . . study of human abilities and characteristics which affect the design of equipment, systems, and jobs and its aims are to improve efficiency, safety, and well being.

Sanders, M.S. and McCormick, E.G. (1993)
. . . designing for human use.

Wickens, C.D. (1992)
. . . is to apply knowledge in designing systems that work, accommodating the limits of human performance and exploiting the advantages of the human operator in the process.

Table 2. Moderate-length definitions formed from the most frequent terms

(a)	Designing and engineering human–machine systems.
(b)	Applying science to people performing in working enviroments.
(c)	Studying man's limited capabilities relate to safe job operation
(d)	Improving knowledge on the fit between users and tasks.
(e)	The interface between people and machines in systems.

can be a significant source of insight. In the present work, we extracted concise phrases describing HF/E from a previous work (Wogalter et al. 1998) that involved analysis of numerous definitions.

Previously, we took the language from a set of 134 definitions from 78 sources compiled by Licht, Polzella, and Boff (1990), and supplemented them with another 56 definitions from 35 sources of various kinds including HF/E textbooks and brochures, World Wide Web sites, introductory psychology, industrial/ organizational psychology and safety engineering textbooks (Wogalter et al. 1998). Definitions selected were intended to describe the field circumscribed by one or more of the following names: ergonomics, human factors, human factors engineering, and engineering psychology. Some were short, dictionary type definitions (e.g., "the study of work" and "human-machine interface"); other were much longer accounts giving the contents and goals of the field. Example definitions are given in table 1.

In the process of limiting the final list to the most frequently mentioned content words, Wogalter et al. (1998) first stripped the original set of definitions of certain elements, such as connecting words (e.g. the, and, to, which) that were unlikely to reveal meaningful interpretation. Additionally, the basic names designating the field were deleted, e.g. the term "ergonomics" was deleted if it appeared as part of the definition. The terms '"human factors", "human factors engineering", and "engineering psychology" were also deleted when they co-occurred in these specific sequences, but the terms themselves were retained if they occurred in other word contexts and sequences. The remaining terms were then sorted alphabetically. Words with identical prefix roots were combined when the ending/suffix did not change the basic meaning of the word. Words with high frequencies were then used to create definitions of moderate length that express the field of HF/E.

Wogalter et al. (1998) argued that the content words with the highest frequency of mention across the included 134 definitions suggests that they are meaningful components describing the field. By combining these high frequency terms, basic or core definitions of the field can be formed. Moderate frequency terms could be used elaborate the definitions with additional terms that include the methods, goals, and other details. On example is: HF/E involves the application of engineering design to the study and production of safe and efficient human–machine systems. Other examples include Chapanis' (1995) and Wicken's (1992) definitions in table 1. A few additional examples of moderate length appear in table 2.

In Table 3, we have categorized the set of terms in another way. Here there is a small set of categories under the headings who, what, how, when/where, and goal. A quick study of this

Table 3. A short list of terms assigned to a simple category structure describing human factors/ergonomics.

Who	What	How	When/Where	Goal
human	system	engineering	enviroment	safety
people	machine	designing	work	comfort
users	equipment	applying	life	efficiency
person	product	studying		
	technology	optimizing		

categorization undoubtedly inspires a well-suited definition or two. The table also provides a concise set of reference terms for describing our field to others.

Across the entire set of definitions examined in this exercise, the statements reflect a diversity in detail and purpose, varying in how much is given on the field's content, methods, and goals. Sometimes it was difficult to tell whether the wording was actually a definition. Wogalter *et al.* (1998) tended to be liberal in accepting wording as a definition. Under different criteria, some statements would not be considered to be a true definition, but rather a description of methods or goals.

A recent survey and a series of focus groups in the United States (Hendrick 1996b) revealed that one of the primary complaints of HFES members was that untold numbers of people outside the field know little, if anything, about our field. As we discussed at the outset, part of the problem has been our name, but also some of the problem may be that our definitions are not user-friendly. While we formed some of the word groupings and definitions ourselves (which undoubtedly reflect some our own personal biases), they were not produced considering the varied population groups to whom they may be proffered. Using the word lists, a different set of definitions could be formed to target different recipient groups (e.g. lay persons vs. engineering/science experts). In fact, we believe that you can tailor definitions to a specific audience with whom you are speaking to or working.

Technology is a powerful single force that is shaping human behavior. Too often, technology is "mindless" with respect to the individuals who either use it or are affected by it. A small but growing group of professionals seek to mediate between growing technical systems and their human users. As technology become complex, there needs to be even greater efforts in HF/E to enable synergistic relationships. Such an effort will be crucial to the path of true technological progress in the coming years. In order to play its role more effectively, the field needs a clear, concise unequivocal and usable term to describe our efforts. We suspect that the term ergonomics will take that role, along with adjectives of physical and cognitive.

REFERENCES

BROWN, O., and HENDRICK, H.W., 1986, *Human factors in organizational design and management — II* (Amsterdam: North-Holland).

CHAPANIS, A., 1995, Ergonomics in product development: A personal view. *Ergonomics*, 38, 1625–38.

CLARK, T.S., and CORLETT, E.N., 1984, *The ergonomics of workspaces and machines: A design manual* (London: Taylor & Francis).

HANCOCK, P.A., 1997, *Essays on the future of human-machine systems* (Minneapolis, MN: Banta).

HELANDER, M.G., 1997, The human factors profession (Chapter 1). In G. Salvendy (ed.) *Handbook of Human Factors and Ergonomics* (2nd ed.) (New York: John Wiley).

HENDRICK, H.W., 1996a, Good ergonomics is good economics. *Ergonomics in Design* (Santa Monica, CA: Human Factors and Ergonomics Society).

HENDRICK, H.W., 1996b. Road map to the future: Revised strategic plan. *HFES Bulletin*, 35(10), pp. 1, 5.

HOWELL, W., and DIPBOYE, R., 1986, *Essentials of industrial and organizational psychology* (3rd ed.) (Pacific Grove, CA: Brooks/Cole).

LICHT, D.M., POLZELLA, D.J., and BOFF, K.R., 1990, *Human factors, ergonomic, and human factors engineering: An analysis of definitions.* Paper presented at the American Psychological Association (APA), New York. (Manuscript copy available from CSERIAC, Dayton, Ohio).

MARK, L.S. and WARM, J.S., 1987, Editors' introduction. In *Ergonomics and Human Factors* (New York: Springer-Verlag).

MARTIN, D.W., and WOGALTER, M.S., 1997, The exposure of undergraduate students to human factors/ergonomics instruction. *Proceedings of the Human Factors and Ergonomics Society*, 41, 470–3.

MEISTER, D., 1989, *Conceptual aspects of human factors* (Baltimore, MD: Johns Hopkins University Press).

WICKENS, C.D., 1992, *Engineering psychology and human performance* (2nd ed.) (New York: Harper Collins).

WOGALTER, M.S., HANCOCK, P.A., and DEMPSEY, P.G. (1998). On the Description and Definition of Human Factors/Ergonomics. *Proceedings of the Human Factors and Ergonomics Society 42nd Annual Meeting* (Santa Monica, CA: Human Factors and Ergonomics Society), pp. 671–4.

Design

W. W. Gasparski

Institute of Philosophy and Sociology, Polish Academy of Science, Nowy Swiat 72, 00-330 Warsaw, Poland

1. INTRODUCTION

Design (1) is the kind of meta-action (an activity), the supreme purpose of which is a conceptual preparation of change, particularly of other action (actions or systems of them) or any of its elements; (2) the noun "a design" is a description (a pattern) of an artifact thought needed and thus is worthy to be implemented for the purpose of a change; and (3) the verb "to design" is to perform an action (a meta-action) aimed at formulating a design.

In different languages, i.e. in different cultural environments, words of different etymology are used as labels of a concept of design. Let us examine the most characteristic ones.

In English, it is *design* itself based on the Latin *designo* = "to define," "to point out," "to mark," "to form"; and *dissigno* = "to unseal," "to manage." In many languages (whether Anglo-Saxon or not) *design*, taken this time from the English not from the Latin, means "industrial design," i.e. design with an aesthetic flavor, e.g. *disseny* in Catalan, *diseno* in Castilian (Spanish), *estetique industerille* in French. Also, French equivalents for *design* are *dessein*, which means "intention," or *dessin*, which means "pattern" (Polish *deseñ* = decorative design).

Generally, the French ultimate equivalent to design is *conception* from the Latin *conceptio* = "concept," "conception." In Polish and other Slavic languages the label is *projektowanie* based on the Latin *proicio* = "to place something ahead" (like in English "to project a missile"), which is similar to German *projektirung* of the same Latin origin. In both languages "engineering design," especially mechanical one, is labeled *konstuowanie*, *konstruirung* respectively from the Latin *construo* = "to cast," "to arrange." In one group of languages different kinds of design are labeled through using one noun (e.g. *tervezes* in Hungarian) plus different adjectives, while in the other group of tongues different nouns are used. It reflects different beliefs in the unity or disunity of the different kinds of human activity in question.

All languages are unified, however, in one common question, namely whether *design* and *planning* are synonyms (like *suunnitella* in Finnish) or not like (*sheji* = "design" and *jihua* = "planning" in Chinese).

2. PLANNING AND DESIGN

According to Nadler (1981),

> Planning and design are classified together ... because their definitions overlap. The words are often used interchangeably as in "Planning a vacation," or "designing a health care delivery system." No purpose is served by saying that "planning" is open-ended while "design" is specific, or that the former has a longer time horizon, or that the latter is project-rather than program-oriented. Whether it be an architect's blue print, a five-year land-use map, or a family's financial plan, solution specifications are detailed, resource allocations are proposed, innovation is encouraged, and purposes are defined — and this is planning and design.

On the other hand, according to Bunge (1985), design and planning are different, though subsequent, phases of the technological process: "technology may be regarded as the field of knowledge concerned with designing artifacts and planning their realizations, operation, adjustment, maintenance, and monitoring in the light of scientific knowledge."

3. DESIGN AND PRAXIOLOGY

The praxiological point of designing and planning is similar to the Nadler (1981) one, for praxiology recognizes different names for the same kind of human action taken from a methodological point of view. Praxiology, however, accepts traditional names of design/planning-like professions, e.g. architectural design versus urban planning, organizational design versus economical planning, etc. (Gasparski 1993). That position is close to the one expressed by Simon (1981):

> The second state in decision making is to device or discover possible courses of action. This is the activity that in fields like engineering and architecture is called "design"; in military affairs "planning"; in chemistry "synthesis"; in other contexts "invention," "composition"; or that most approving of labels — "creation."

REFERENCES

BUNGE, M. 1985, *Treatise on Basic Philosophy*, vol. 6: *Epistemology and Methodology II: Understanding the World* (Dordrecht: Reidel).

GASPARSKI, W. 1993, Design, science, and philosophy: the praxiological perspective. In M.J. de Vries, N. Cross and D.P. Grant (eds), *Design Methodology and Relationships with Science* (Dordrecht: Kluwer), 165–89.

NADLER, G. 1981, *The Planning and Design Approach* (New York: Wiley).

SIMON, H.A. 1981, *The Sciences of the Artificial* (2nd ed.) (Cambridge, MA: MIT Press).

Ecological Approach

J. R. Hajdukiewicz and K. J. Vicente

Cognitive Engineering Laboratory, Department of Mechanical & Industrial Engineering, University of Toronto, Toronto, Canada

1. INTRODUCTION

Over the years many diverse approaches have evolved to study problems in human factors, typically with origins from the field of psychology. This diversification is due to the very large range of problems and multidisciplinary nature of human factors. A number of approaches (e.g. the information processing approach) focus predominantly on the behavior of human operators with less emphasis on the context or environment where work takes place. These approaches may be efficient and useful when work domains are relatively predictable, and tasks are routine and stable. However, many modern work systems are complex, have great flexibility in operation, and have unanticipated situations that require problem solving and appropriate intervention. For these systems, it is important that tasks and actions be considered in the context of the work situation. A relatively new approach based on the ecological perspective accounts for the capabilities and limitations of human operators as well as the context and environment where actions are performed. This approach provides a way of studying human operators, the work environment and their interactions within a single, coherent framework.

The purpose of this chapter is to describe this ecological approach to human factors. The first section presents an overview of the ecological approach, including its foundations and fundamental characteristics. The second section discusses the implications of the ecological approach to human factors in terms of the fundamental characteristics, as well as some examples of how the approach has been applied to human factors problems. The final section discusses the benefits and limitations of the ecological approach to human factors.

2. FOUNDATIONS OF THE ECOLOGICAL APPROACH

The ecological approach is a systems approach for studying the interactions between human organisms and their environment (Meister 1989, Flach 1990, Vicente 1990). The approach has its foundations with the research of Brunswik (1956) and Gibson (1979) in the field of ecological psychology. Each researcher differed in the details of their respective views, but there were a few aspects in common. As a key similarity, both shared the notion that the interaction between the human organism and its environment is at the root of psychology. In contrast, the prevalent view of psychology, referred to as the organismic approach, has been the study of human behavior in relative isolation from its environment (Meister 1989, Flach 1990, Vicente 1990).

Central to the ecological approach are the dual notions that the environment offers the human organism opportunities for action and the human organism adapts to the structure of the environment. These insights are important for studying human behavior, because characteristics of the environment both guide and constrain human behavior. Gibson (1979) referred to these opportunities for action as affordances. It is useful to describe these affordances in relation to system purposes. Thus, affordances are a functional description of the environment relevant to the human operator. As a result, an ecological approach to human factors must start by analyzing the environment and its constraints, relevant to the human operator, before studying human behavior in that environment. This tenet contrasts with the organismic approach, which does not emphasize the importance of the environment, and typically starts by studying human behavior independent of context (Flach 1990, Vicente 1990).

The next section describes the four fundamental characteristics of the ecological approach: reciprocity of person and environment, representative design, primacy of perception, and analyzing the environment first.

2.1. Reciprocity of Person and Environment

Key to the ecological approach is the notion that the human operator and work environment are reciprocally coupled and cannot be studied independently from one another. The coupled interactions between the two form the basis for understanding human behavior. The environment is functionally described with reference to the human operator in terms of affordances or opportunities for action. If the human operator directly or indirectly perceives the affordances, they can guide actions in the environment. In this way, the human operator's behavior is constrained by the environment.

A simple example to illustrate this point, based on a case study by DeSanctis (1984), is the difference between sales graphs and tables as appropriate displays for the human operator. Without knowing the context or task environment in which the information is used and by whom, it is difficult to determine the appropriate display. If the task is to determine a specific sales value on a particular date, tables may be more appropriate than graphs. If the task is to determine the general trend in sales over a range of dates, graphs may be more appropriate than tables. Without considering both the human operator and the work environment simultaneously, design requirements may be erroneous or incomplete. Thus, the reciprocal and coupled relations between the human operator and environment are important, and should form the basis for systems analysis and design.

2.2. Representative Design

The conditions under which one evaluates a design should be representative of the target of generalization in order to have meaningful results. This includes choosing appropriate environments, scenarios, tasks, and human operators. Note that not all aspects of the target environment have to be duplicated, only the ones that are relevant to the problem being studied. Also, careful attention must be paid to the confounding effects of the evaluation that are not representative of the target situation. These effects may produce misleading results that put into question any generalizations of the evaluation. A quote from Toda (1962: 165) gives special emphasis to the importance of representative design. "Man and rat are both incredibly stupid in an experimental room. On the other hand, psychology has paid little attention to the things they do in their normal habitats;

man drives a car, plays complicated games, designs computers, and organizes society, and rat is troublesomely cunning in the kitchen."

In analysis or design, just as much attention should be focussed on representing aspects of the actual environment, as is spent on the people working in that environment.

2.3. Primacy of Perception

Another characteristic of the ecological approach is that information should be presented in a form that allows human operators to exploit the power of their perceptual systems. These systems are more efficient, in terms of energy and time, and less error prone compared with the cognitive processes involved in deduction, analytical reasoning, and problem solving. They are the everyday skills that have been refined through evolution and that are developed in practice. A simple example to illustrate this point is how a person drives a motorcycle. One approach is to let the driver use the sensory information provided by their vestibular senses. This information can be used to directly perceive the dynamics of the motorcycle in the environment and consequently, to act appropriately. Since driving a motorcycle is functionally equivalent to solving a complex set of state equations describing the dynamics of the motorcycle, another approach is to provide the driver with these equations and the tools to solve them indirectly (i.e. via a calculator or other computational tool). Solving the state equations indirectly is a complex and effortful task when compared with perceiving the solutions to these equations directly. Clearly, much is to be gained by taking advantage of actors' perceptual skills.

2.4. Start with Analyzing Environment

The last characteristic of the ecological approach is that the problem should be understood first by starting with an analysis of the environment, before considering actual human behavior in that environment. This characteristic follows from one of the basic tenets of the ecological approach, that the environment has a strong influence on behavior. The parable of an ant's path on a beach by Simon (1981: 64) provides an example of the implications of this characteristic (Figure 1). "Viewed as a geometric figure, the ant's path is irregular, complex and hard to describe. But its complexity is really a complexity in the surface of the beach, not the complexity in the ant."

Simon's (1981) parable about an ant on the beach.
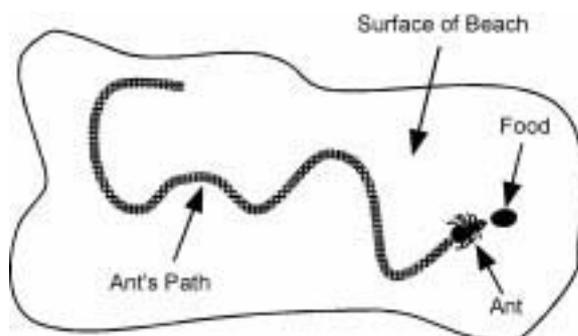
Figure 1. Simon's (1981) parable about an ant on the beach.

In the example the beach may be considered as the environment where the ant acts. The ant's cognitive processes may be simple (e.g. move in direction of food and lowest surface gradient). But if the contour of the beach is complex, the ant's behavior (e.g. trajectory) may be viewed as complex. This feature is predominantly the result of the complexity of the environment, not of the cognitive processes of the ant itself.

By studying the environment first, those external influences on behavior may be partitioned from internal influences (e.g. preferences) relevant to the human operator. This characteristic emphasizes the importance of first identifying the work domain landscape and opportunities for action with reference to the human operator, in order to begin to understand human behavior in that environment.

3. IMPLICATIONS OF THE ECOLOGICAL APPROACH FOR HUMAN FACTORS

The increased emphasis that the ecological approach places on the coupling between the human operator and the environment has important implications for the field of human factors. This section discusses some of these implications for analysis, design, and evaluation in human factors, in terms of the foundations and fundamental characteristics previously presented. In addition, the application areas that have benefited from this approach are briefly mentioned.

3.1. Analysis

In terms of analysis, the ecological approach provides a different perspective compared with the organismic approach that has typically been incorporated in human factors. These differences are discussed in terms of human performance modeling and task analysis.

3.1.1. Human performance modeling

With traditional approaches to human performance modeling, behavior is modeled in relative isolation from its context or environment. By not considering the environment and its influence on behavior deeply, analysts may attribute more information processing capabilities (e.g. elaborate mental constructs and processes) to the human operator than actually possessed (Vicente 1990). In Simon's (1981) parable of the ant, the complex trajectory might be explained as the ant having complex psychological mechanisms. In contrast, the ecological approach starts by analyzing the environment relative to the human operator. This tack assists in partitioning accurately those aspects of the environment that influence behavior and the internal mechanisms of the human operator.

Kirlik et al. (1993) provide an example of how an ecological description of the environment can lead to an account of skilled performance in a complex human–machine system. In the study, subjects performed a complex supervisory control task. Kirlik et al. found that it was possible to account for skilled performance in this domain with a model that relied almost exclusively on perception and action. This seems contrary to an organismic perspective that may predict a considerable amount of cognitive processing for the human operator for this task. Therefore, it is important for the analyst first to describe the environment, in order to partition accurately the influences on behavior of the environment and internal mechanisms of the human operator.

3.1.2. Task analysis

With traditional task analysis methods (i.e. organismic approach), analyses typically start with the human operator. The result is typically a description consisting of a single normative sequence of overt behaviors (Vicente 1999). This approach is efficient if the process is routine and predictable. However, there is minimal flexibility to deviate from this path for unanticipated situations. In contrast, the ecological approach starts with an analysis of the environment and narrows the action space by incorporating the capabilities of the human operator. The result is a functional description of the environment consisting of an envelope or field of constrained action possibilities. No single normative sequence of overt behaviors is prescribed. This captures the great flexibility in operation of the system, allowing the human operator to develop efficient strategies for operations, and cope with varying and unanticipated situations. However, this approach requires relatively more autonomy because normative task sequences are not prescribed beforehand.

The difference between the ecological and organismic approaches to task analysis may be illustrated by considering the set of action possibilities for a hypothetical system (Figure 2). With the ecological approach, the boundaries that constrain behavior are identified explicitly. Human operators adapt to these constraint boundaries so trajectories to reach the goal can be determined dynamically and implicitly from any starting position in the action space. In contrast, with the organismic approach, a predefined trajectory to reach the goal is generally identified and made explicit beforehand, incorporating past experience and expert behavior in the work environment. There may be significant limitations in diverting from this prescribed path because the constraint boundaries and other action possibilities may not be apparent to the human operator.

If these constraint boundaries are relatively dynamic, uncertain, or situation-dependent, then the ecological approach may be more appropriate because the human operator may be able to adapt to changing conditions. However, if the constraint boundaries are predictable and stable, the organismic approach may be more appropriate because there may be no need for the human operator to implicitly generate trajectories dynamically. Therefore, it is important to consider the characteristics of system that is analyzed to determine whether the ecological or organismic approaches are most appropriate.
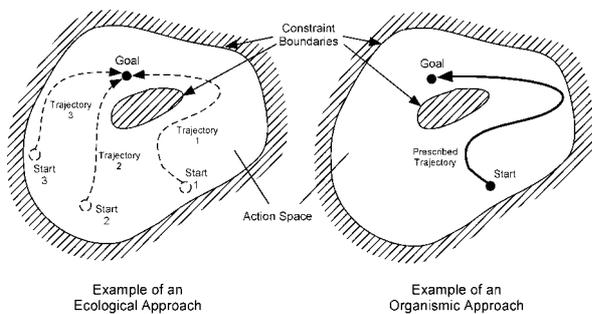
Figure 2. Example of the difference between ecological and organismic approaches to task analysis.

3.2. Design

The ecological approach also provides direction for human factors activities for designing computer-based systems. The goal of design is to develop representations that directly map actual situations in the environment with the human operator's understanding of these situations. The environment relevant to the human operator may be functionally described in terms of interface content and structure. The interface form, a direct or indirect representation of the content and structure, may then be specified to be compatible with the capabilities and limitations of the human operator (e.g. perceptual skills). The emphasis of starting with the environment is critical for design. Before one designs the displays of a system used by a human operator to control the environment, one must know the content and structure of the environment first. Otherwise, the representations developed may not specify the actual environment, potentially resulting in a mismatch between actual situations and the human operator's understanding of these situations. For a detailed account of an ecological approach to interface design for complex systems, refer to Torenvliet and Vicente (1999) on ecological interface design.

3.3. Evaluation

In terms of evaluation, the ecological approach emphasizes the requirement of representativeness to properly evaluate a design. The conditions under which one evaluates the design must be representative of the selected target situation. If not, the differences in the contexts in which behavior takes place (between the evaluation and target conditions) must be addressed. Otherwise, there is a potential for conflicting and confounding results.

3.4. Application Areas

The ecological approach has been applied to a number of very diverse research areas in human factors. One area of human factors has benefited from the ecological approach is the design of interfaces for complex human–machine systems. This area is characterized, in part, as having large problem spaces, dynamic events, a high degree and nested levels of coupling, uncertainty, and disturbances within the environment. The ecological approach is appropriate for this area because it accommodates flexibility in operation, essential for coping with complexity and unanticipated events. Other areas of human factors that have benefited from the ecological approach include human–computer interaction and physical ergonomics design. In addition, wide ranges of application domains have benefited from the ecological approach. These include aviation, command and control, computer programming, engineering design, information retrieval, medicine, process control, and workplace design. For a detailed discussion of a few of these application domains, refer to Chery and Vicente (1999).

4. BENEFITS AND LIMITATIONS OF THE ECOLOGICAL APPROACH

From the discussions above, a number of benefits of the ecological approach to human factors become evident. First, by considering the environment to be as important as the human operator, compatibility and interactions between the two may be realized. Aspects of the environment that influence behavior may be partitioned from the internal mechanisms of the human operator. Second, by analyzing the environment in terms of action

opportunities relative to the human operator, the ecological approach potentially accommodates flexibility to adapt to unforeseen contingencies and recover from errors. These action opportunities allow the human operator to solve problems and generate action paths dynamically and implicitly in the environment based on the actual situation. Third, the display of information is designed to specify the action opportunities in the environment and take advantage of the perceptual systems of the human operator. This potentially provides greater coupling between the human operator and the environment. Fourth, the ecological approach to human factors appears to have a broad scope of applicability across various application domains.

There are a number of limitations of the ecological approach to human factors that are also worth mentioning. First, much effort is required to analyze a system using the ecological approach, because the constraint boundaries and associated action opportunities relative to the human operator need to be determined before design. Second, since no particular sequence of actions is prescribed from the analysis, the human operator may have greater responsibility and autonomy in how the tasks are performed. While this aspect can be beneficial in managing dynamic and uncertain situations, human operators may require relatively more effort to perform routine tasks compared with invoking a prescribed sequence of actions. Third, ecological approaches have not been thoroughly tested compared with organismic approaches (e.g. information processing). Additional efforts are required to evaluate the approach, especially for larger scale systems.

5. CONCLUSIONS

The ecological approach provides a useful perspective for a broad range of human factors problems, especially in domains that have increased complexity, flexibility in operation, and the possibility for unanticipated situations. This chapter has outlined the fundamental characteristics of the ecological approach: reciprocity of person and environment, representative design, primacy of perception, and start with analyzing the environment. These characteristics provide a productive foundation for analyzing, designing and evaluating systems from a human factors point of view.

REFERENCES

BRUNSWIK, E., 1956, *Perception and the Representative Design of Experiments* (2nd ed.) (Berkley: University of California Press).

CHERY, S. and VICENTE, K.J. (1999), Ecological interface design — applications [in this volume].

DESANCTIS, G., 1984, Computer graphics as decision aids: directions for research. *Decision Sciences*, 15, 463–87.

FLACH, J.M., 1990, The ecology of human–machine systems I: Introduction. *Ecological Psychology*, 2, 191–205.

GIBSON, J.J., 1979, *The Ecological Approach to Visual Perception* (Boston: Houghton-Mifflin).

KIRLIK, A., MILLER, R.A. and JAGACINSKI, R.J., 1993, Supervisory control in a dynamic and uncertain environment: A process model of skilled human–environment interaction. *IEEE Transactions on Systems, Man, & Cybernetics*, SMC-23, 929–52.

MEISTER, D., 1989, *Conceptual Aspects of Human Factors* (Baltimore: Johns Hopkins University Press).

SIMON, H.A., 1981, *The Sciences of the Artificial* (2nd ed.) (Cambridge, MA: MIT Press).

TODA, M., 1962, The design of a fungus-eater: A model of human behavior in an unsophisticated environment. *Behavioral Science*, 7, 164–83.

TORENVLIET, G.L. and VICENTE, K. J. (1999), Ecological interface design — theory [in this volume].

VICENTE, K.J., 1990, A few implications of the ecological approach to human factors. *Human Factors Society Bulletin*, 33, 1–4.

VICENTE, K.J., 1999, *Cognitive Work Analysis: Toward Safe, Productive, and Healthy Computer-based Work* (Mahwah: Lawrence Erlbaum).

Epistemological Issues about Ergonomics and Human Factors

F. Daniellou

Laboratoire d'Ergonomie des Systèmes Complexes, Université Victor Segalen Bordeaux, France

1. INTRODUCTION

The reader who browses on the web through the sites of the IEA Federated Societies will find a number of alternative definitions either of ergonomics or of the aims of ergonomics and human factors. These definitions have in common:

- a reference to science
- a reference to "knowledge about the characteristics of human beings"
- a reference to application of that knowledge to the design of "systems and devices of all kinds".

Claims for such a cross-bred identity is not common among scientific disciplines. It may be taken as an invitation to an epistemological examination of the nature of ergonomics.

2. WHAT IS EPISTEMOLOGY ?

Epistemology is the philosophical activity that deals with knowledge and its truth. More practically, as far as science is concerned, the question is to determine the rules and tests that have to be utilized to allow for the use of the word "scientific" to qualify models or theories. There is a unanimous agreement among scientists that not all theories may be called "scientific", but almost none about the nature of the trials they have to undergo to gain this label.

2.1. Major Debates

Major debates are developing in this respect. Let us consider three of them that might be of some importance for ergonomics.

- Some scientists consider that the rules which govern sciences are steady, universal, and complete. They can be used to check the scientific validity of any theory in any discipline. Others dwell on the diversity of disciplines, some of which study "artificially reduced" phenomena in the laboratory (e.g. hydraulics), while others are confronted with irreducible complexity in the field (e.g. hydrology). They argue that epistemological rules have to be reworked within the disciplines, and that this epistemological reflection is one of the tasks of any disciplinary group.
- Some authors consider science as a globally cumulative process, each research adding new knowledge to an age-old construction. Others (Kuhn 1970) describe the history of sciences as a series of crises, turning points, ruptures. New theories do not always encompass former ones, they bring new light on certain topics as well as unexpected shade on others.
- Specific epistemological issues deal with human sciences. Those have to face the fact that human behavior resists prediction. human liberty is not considered as an absence of determinism, but as an ability to rearrange biological, cognitive, psychic, historical, social, cultural contradictory determinisms into new room for action.

2.2. Assessment of Scientific Models

Assessment and validation of scientific models refer to two different processes.

2.2.1 Assessment within the "reduction space"

Scientific models are usually produced in "worlds" the complexity of which has been artificially limited (e.g. the laboratory). In these worlds "perturbations" are avoided. Theories and laws are produced about "pure" phenomena (e.g. friction-free mechanics). The first assessment of theories validates the fact that models relevantly account for phenomena observed under these conditions of reduction. Congresses and journals rally specialists who assume the same "reduction spaces" and have competing theories about them (Latour 1993).

2.2.2 Relevance outside the "reduction space"

No car will ever run according to the laws of friction-free mechanics. This means that there is a need for a second assessment process of scientific theories. This one does not deal with the validity of the model to account for the artificial "world" but with its relevance to describe the real world.

This second assessment process is much less described in the scientific community than the former. One of the reasons for this is that the second assessment may not be carried out only by scientists. Competing scientific models do not only have theoretical differences, they also lead to different industrial or political decisions which have direct effects for the citizens. For instance, alternative models of shift work effects will insist on sleep deprivation or on social life alterations. If labor authorities consider that sleep deprivation models made in the laboratories are more "scientific" than surveys about social life alterations, they may produce laws about shift work that will minimize sleep deprivation, but may lead to severe family or social problems for the workers, who are not just sleepers.

Since competing theories lead to different effects for the citizens, the "second" assessment of theories may not be carried out only by scientists of the same discipline, probably not even by scientists of different disciplines. Processes involving representatives of categories of citizens likely to undergo the effects of theories (e.g. patients, consumers, workers, users) have been described (Latour 1993).

3. UNDERLYING ASSUMPTIONS

Any scientific theory comprises explicit assertions. But it is also based on implicit assumptions on a number of topics, which either the researcher considers as obvious, or the existence of which he/she is not aware. Such underlying assumptions are referred to as "backworld", "paradigm", "underlying assumptions", "beliefs" or "implicit vision". Epistemological approaches endeavor to detect and criticize those tacit assumptions, which are by nature not submitted to assessment.

As far as ergonomics and human factors theories are concerned, such tacit assumptions are numerous. Authors often refer to "work", "health", "improvement" or "technology" without much explicitation of their visions of what lies behind these concepts. Some alternative visions are listed below.

3.1. Models of the Human

Models of the human used by ergonomists implicitly refer to one to four of the following description levels:

- the biological level: the human is considered as an energy processing system (this level includes contributions from anatomy, physiology, biomechanics, biochemistry, toxicology, etc.);
- the cognitive level: the human is considered as an information processing system (this level includes the study of thought processes, representations, optimal decision-making, etc.);
- the psychic level: the human is the subject of a unique history, which has molded his or her personality, leading to a specific subjective processing of the situations he or she experiences;
- the social level: every single individual is a member of several social groups with different cultures, that will partly determine his or her values and habits.

No author would deny the existence of these different levels. In many cases, for the sake of rigor, a research will be framed within one or two of these description levels. In some cases, the authors will discuss what is left aside by not considering the other levels. In most cases they will simply do as if the levels they are using were the only relevant ones.

3.2. Models of Health

Implicit models of "health" can often be found "between the lines" in ergonomic publications.

- Health as the absence of acknowledged pathologies. This common approach does not consider fatigue, discomfort or ill-being as health-related issues, nor does it take into account social dimensions (e.g. poverty).
- Many authors would accept the WHO definition of health, which states that "health is a state of complete physical, mental, and social well-being and not merely the absence of disease or infirmity". This vision is much richer than the previous one. Still, it considers health as a "state": you have it or lose it, you are in health or out of health.
- Another model is to consider health as a process: "health is the possibility for every single individual to open up his/her own original way" to the target state described by WHO (Dejours 1995). Health is described here as an ongoing tentative construction (one gets nearer to health or drifts away from it) that can be hindered by situations encountered by the subject. Situations are favorable to health when they give the individuals opportunities to go forward in this construction. Negative situations for health are those which do not allow the individual a way out.

Such different starting points will lead to different descriptions of what should be done to improve occupational health.

3.3. Models of Work

The same diversity will be found about the concept of "work". A first group of differences concerns the levels of description. Most ergonomists would today not describe merely "physical" work — the reference to "cognitive" dimensions (including the so-called "manual" jobs) is now more or less generalized. However, some other dimensions are not commonly considered. For instance, fewer ergonomists will refer to the ethical dimensions of work, to its contribution to the construction of the person's identity, or to its role in the weaving of the social fabric.

A second group of differences is related to the way authors account for, or not, the difference between "what is demanded" of the workers and "what it demands" of them. In many cases, work descriptions are mostly task descriptions: what is considered as "work" is the objectives the worker has to meet (e.g. produce 300 parts a day, stick to quality requirements, adjust the equipment) in given surrounding environmental conditions (heat, noise, etc.). In this sense, work is the same for all workers in a given workplace. Another approach is to focus on work "activity", which is the mobilization of the workers' unique body, intelligence, and personality to fulfill the tasks. This approach will lead to stressing workers' individual and collective strategies to monitor and manage the situation, collect formal and informal information, anticipate incidents and events, cope with malfunctioning and variability, deal with their own variable capacities and limitations. This approach will highlight the fact that operators never only carry out the prescribed tasks, they bring (even apparently tiny) original contributions to the production. There would be no production if the workers stuck to prescribed tasks. Work is not given by formal prescriptions: it is an ongoing personal and collective creation that includes "coping with what is not given".

It appears that work or health may either be considered within one discipline description level, or be described as tentative compromises between resources and constraints falling within the province of different disciplines (see figure 1).

The construction of health and work activity make it necessary for the worker to find resources from different levels to face constraints stemming from different levels or from the outside. Any disciplinary approach makes a section in this transversal process. Work and health are, in a way, "perpendicular" to the fundamental disciplines which study them. One of the challenges for ergonomics is to produce models in which the necessary scientific reduction does not kill the nature of the processes.

3.3.1 Technology

Ergonomists' reference to "technology" may also be based on different backgrounds. In most cases, technology is referred to as the application of scientific knowledge to the design of artifacts. It has to do with physics, electricity, automation, etc. Another approach has been developed by anthropologists (e.g. Mauss, Haudricourt; see Lemonnier 1993) who consider technology in close connection to culture and to choices made by societies. Technology is that gearing through which the human cultural world meshes with the physical world. In this approach the
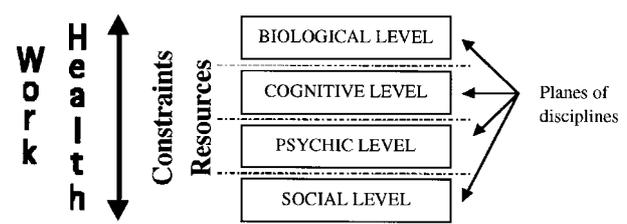
Figure 1: Health and work as transversal processes

keywords are not "technological progress" but "technological choices".

3.3.2 Organization

Organization is double-faced, but both faces do not have the same weight in ergonomic models. On one side is the organizational structure — charts, work division, procedures, formal rules (but also building lay-out, software passwords, etc.) — that will determine the frame within which the tasks will be accomplished. The second face is the "live" organization made up of the everyday social activity of all actors. There is a double determination between the organizational structure and the "live" organization: at a given instant, the structure largely determines the bounds of actors' social activity. In the long run, the structure can be regarded as an output of this activity. According to what the authors call "organization", they will have different approaches to organizational changes. Will the organization be changed only be designing a new organizational structure? Which kind of interactions between the actors does organizational change require? (Carballeda and Daniellou 1998).

3.3.3 Change

Models of change that can be found in ergonomic publications have different backgrounds. In many cases, the ergonomic contribution to change is an expert one: ergonomists give the scientific position about what is acceptable or not in the workplace. They will prescribe solutions, just like physicians prescribe a medicine: it is strong advice, the legitimacy of which is based on knowledge. It may be taken into account or not, but it is not supposed to be discussed.

Other authors refrain as much as possible from adopting expert positions in ergonomics. Their idea is that changes in workplaces are the results of social interactions between a number of actors (employers and employees, but also customers, suppliers, technicians, administrations, etc.). These ergonomists perceive that they contribute to changes in the workplaces through fostering the social debates and feeding them with specific information. They bring in facts and suggestions, but the actual change will be the results of negotiations between the actors. The ergonomist's contribution in this case is made of indisputable facts and disputable interpretations.

3.3.4 Describing, explaining and understanding

There may be several ways of describing and explaining what is happening in a workplace. In certain cases, the observer analyzes the situation "from outside", using measurements and his or her expert knowledge. In other researches, the workers' opinion is taken into account, by means of questionnaires or interviews.

Ergonomists have mixed feelings about the workers' word: on the one hand, they are usually aware of the importance of taking into account the operator's knowledge, and his or her subjective feelings about the workplace. On the other hand, they know that a number of factors will make it difficult for operators to express themselves about their work:

- a major part of the operator's knowledge is "tacit knowledge" (Polanyi 1983) that can be expressed through action, but that has never been put into words;
- there may be a linguistic difficulty, a lack of relevant words to express what happens in the workplace. In this case, the

workers often answer the observer's questions with something like "it is hard to say";
- workers themselves are caught in socially dominant representations of their own work. A worker may consider that the work is "purely manual work", even if the ergonomist demonstrates a high cognitive load, because the social description of the job refers to manual work;
- last, but not least, the harder the working conditions, the more the workers may build psychological defenses leading them to underestimate the difficulties and risks.

For all these reasons, a number of authors consider that work analysis has to be a joint production of the observer and the observed worker(s). The analyst produces a description and a first interpretation of the situation. The worker may be astonished ("I did not know that I was doing all that"), but he or she will be in a position to comment the description, suggest explanations of what has been observed, and put it in perspective. The "meaning" of work is neither given by the analyst nor by the worker, it comes out as the result of a joint construction.

Any scientific paper in ergonomics and human factors may therefore be questioned as to its underlying assumptions regarding at least the above listed topics.

4. WHICH KIND OF MODELS DOES ERGONOMIC RESEARCH PRODUCE ?

Coming back to the question "Under which conditions may a model be regarded as scientific?", the first point is to identify models of what ergonomic research produces. Two different kinds of models may be found in the results of ergonomic research.

4.1. Models of Operators' Work and its Results

Most of ergonomic research focuses on producing knowledge about the human at work. It may be limited within the frame of a "fundamental" discipline (e.g. physiology, cognitive psychology), or try to give evidence of "transversal" properties of work — for instance, the way a shift worker will organize his or her time takes into account factors such as job content, sleep, food, transportation, social life, family arrangements, life projects.

Knowledge about the human at work may be produced either in real work situations or in the laboratory, where there is an attempt to simulate some of the features of the workplace. In some cases, the research is an iterative process between field studies and laboratory studies.

Methodologies for producing knowledge about the human at work are thoroughly described in the articles of this encyclopedia.

4.2. Models of Change Processes

Another type of model is also dealt with in ergonomic publications. These relate to "work improvement", "change", "project management", or "action methodologies". They do not aim at producing knowledge about the work the ergonomists are observing, but rather about the ergonomists' own work. The question here is "Through which mechanisms can ergonomists improve workplaces?".

- The majority answer to this question in the international ergonomic community is "by applying scientific knowledge". Ergonomics, like engineering disciplines, is considered as an application of science.

- This approach has been thoroughly discussed by Schön (1983), not specifically concerning the ergonomists and human factor specialists, but about all kinds of "practitioners". He calls the first approach "technical rationality": it is dominant in engineering schools and universities: engineers, architects, physicians are supposed to *solve practical problems by applying scientific knowledge.* Schön states that the major challenge for practitioners is to predict the outcome of moves they are likely to make on reality in order to change it. Since we are in complex systems, any move will have expected effects as well as side effects. The practitioners' decision criterion is not "did I get what I wanted?" but "is the result — including the side effects — acceptable?". The practitioner's thought process is based on the availability of media (e.g. drawings) that make it possible to detect the side effects of moves without making them in the real world. This process is described as a "dialog with the situation".

Schön and Argyris plead for a "Theory of practice". Similarly, Latour studies "theorists' practice". Theories of practice are mostly the results of two methodologies:

- one is based on the observation of practitioners or theorists by a researcher, who studies their work as he or she would do with any other worker;

- the other one is "reflexive practice". Here, the researcher is involved in action, as an ordinary practitioner. He or she keeps records of the initial hypotheses, the events that occur, the decision making processes, the assumptions made throughout the intervention. Some of the final results are similar to the initial expectations, while some differ. An a posteriori reflexive position leads the researcher to question the underlying assumptions that have led to unexpected results, and enrich or modify the initial model. The assessment and validation of the new model, as to its predictability value, will be made in further projects by different researchers.

4.3. The Ergonomist as a Worker

In the model of "applied science" the ergonomist is the representative of knowledge about human work. He or she may have difficulties in the field, but their own activity is not considered an object of interest for scientific description.

In the models of "theories of practice", the ergonomist is regarded as a worker whose skills and difficulties have to be analyzed and modeled (in terms of new knowledge) in order to foster action in workplaces.

Such research programs are well known as regards theory of medical practice (Balint) or of ethnographical practice (Devereux). When analyzing the ergonomists practice, attention may be paid to their decision-making processes, their negotiation skills, their interactions with different actors, the kind of power and responsibilities they take on, the ethical problems they encounter, the human costs they go through, the psychic defenses they interpose to stand hard situations. Such models are extremely helpful to teach and train young ergonomists, not only through delivering them basic scientific knowledge.

5. KNOWLEDGE AND ACTION

The attempts of ergonomics to produce scientific knowledge encounter epistemological difficulties that are not usually found in one single discipline:

- ergonomics deals with technical devices and human beings in systems where the "laws" are of different natures;
- it is oriented towards action in real situations;
- it takes on both health and efficiency issues;
- it deals with social situations which raise all the questions of liberty and power struggle;
- it deals with complex systems characterized by non-linear answers, uncertainty about the initial state, variations in the context and the number of influencing factors;
- as a design discipline, it has to do with "things that do not yet exist", the existence of which is both a matter of technical feasibility and political will;
- it is continuously facing ethical dilemmas (what is an acceptable workload?);
- one of the way it uses to produce knowledge about the phenomena is by changing them (knowledge through action).

These features may be considered as weaknesses, that affect the possibility for ergonomics to be regarded as a scientific discipline. Or they may be taken as an opportunity for ergonomics to take part, with a head start over other disciplines, in epistemological debates that are developing about the relations between science and action in complex systems. Its strengths are the reciprocal of its weaknesses: work situations involve a number of constraints that no researcher would dare to impose in the laboratory. Ergonomists have learnt by and by to find ways out in extreme situations. The knowledge they have developed is continuously put to the test of real situations.

Two conditions are probably at stake in the possibility for ergonomics to be regarded as an "epistemologically leading discipline":

- the need for the ergonomic research community to take epistemological issues in earnest and tackle them in their congresses;
- the quality of the relations between ergonomics research and ergonomics practice, and their ability to act as *reciprocal* suppliers.

REFERENCES

CARBALLEDA, C., DANIELLOU, F., 1998, Ergonomic contribution to organisational changes. *Proceedings of the 13th IEA Congress,* 1, 34–6.

DEJOURS, C., 1995, Le facteur humain, Coll. *Que Sais-Je* (Paris: PUF).

KUHN, T.S., 1970, *The Structure of Scientific Revolutions* (2nd ed.) (Chicago: University of Chicago Press).

LATOUR, B., 1993, *We Have Never Been Modern* (Harvard University Press).

LEMONNIER, P., 1993, *Technological Choices* (New York: Routledge).

POLANYI, M., 1983, *The Tacit Dimension* (Gloucester, Mass.: Peter Smith).

SCHÖN, D., 1983, *The Reflexive Practitioner* (New York: Basic Books).

Ergonomics in the Nordic Countries

T. Ivergaard

Nordic Ergonomic Society (NES), Arbetslivsinstitutet, PO Box 610 (Frosostrand hus A), S-832 23 Froson-Ostersund, Sweden

Thirty years ago, the Nordic Ergonomic Society (NES) was established to represent Denmark, Finland, Norway and Sweden within the fields of Ergonomics and Human Factors. In recent years, the ranks of the NES have been enlarged through the inclusion of Iceland as a formal representative, as well as with the establishment of closer links with the Baltic countries, i.e. Estonia, Latvia and Lithuania. Over the years, national ergonomic societies have been established in all of the NES's member countries — Danish Society of the Work Environment (SAM), Finnish Ergonomic Society (ERY), Icelandic Ergonomic Society (VINNIS), Norwegian Ergonomic Society (NEF), and Ergonomic Society of Sweden (ESS).

The NES has, to a large extent, traditionally been oriented towards work-physiology and a large number of its members have backgrounds within such areas and work within physiology, physiotherapy and rehabilitation. However, even from its inception, the society has had members who are experienced within such fields as work psychology, design and engineering, and occupational health and safety.

Over the past decade, the NES also has had new members from such areas as work sociology, organisational psychology, leadership and training. Members have mainly been concerned with the application and practice of ergonomics. The number of members involved in research has also increased over the years. The principal areas of application for such research have been within industry and government.

As regards membership figures for each respective society, these are:

- Denmark — 908 members.
- Finland — 230 members.
- Iceland — 82 members.
- Norway — 177 members.
- Sweden — 226 members.

In total, the NES has 1623 members.

In conclusion, it should also be noted that in accordance with European Ergonomic Standards, a professional licensing function is attached to the NES.

The European Union's Policy in the Occupational Safety and Health Sector

P. Batra and A. Hatzopoulou
National Technical University of Athens (NTUA), Department of Humanities, Social Sciences and Law, Zografou–Athens, Greece

1. INTRODUCTION

Nowadays, the European Union plays a dominant role in the formation of Occupational Safety and Health (OSH) law frame and policy worldwide. The Community's policy in this sector has come through several stages that almost coincide with its stages of evolution. The purpose of this article is to record this evolution and investigate its impact on the improvement of health and safety conditions in European work sites.

2. THE TREATIES ESTABLISHING THE EUROPEAN ECONOMIC COMMUNITY

The forerunner of the European Union (EU), the European Economic Community (EEC), was established in the 1950s with three treaties:

- The European Coal and Steel Community (ESCS) Treaty, signed in Paris in 1951
- The European Community Atomic Energy Authority (EURATOM) Treaty, signed in Rome in 1957
- The European Economic Community (EEC) Treaty, also signed in Rome in 1957

The European Union (EU) was established by the Treaty of Maastricht in 1992, and modified by the Treaty of Amsterdam in 1997.

Historically, the European Coal and Steel Community Treaty was the first international economic agreement to include occupational safety and health provisions. The reason for this inclusion was the high number of employees in the coal and steel sector and the accompanying high number of occupational accidents and diseases in these industries. Following an accident in Belgian mines in 1956, which caused the death of 262 coal miners, the European Coal and Steel Community Council established a Permanent Committee for Safety in Mines. The main tasks of this tripartite body (2 governmental, 1 employer, and 1 employee representative from each member state) are the monitoring of safety conditions in mines, making proposals for safety improvement, and the dissemination of related information among the member states. The work of this committee over the years has been very beneficial and resulted in the reduction of the number of accidents in European mines.

The main remit of EURATOM was the development of the European nuclear industry along with safety conditions that would ensure the exclusion of hazards to the lives and health of Europeans and the promotion of works of peace. To achieve this scope, many restrictive processes were foreseen in the text of the treaty. The first Occupational Safety and Health directive adopted in 1959 concerned the protection of the public and workers from radioactive radiation.

3. THE PERIOD 1957–1986

In addition to the coal and steel industries, once established, the EEC turned its attention to OSH matters in other economic sectors. Although the main scope of the Treaty of Rome is the single European market, there are social policy references in Articles 117–128. Articles 100 and 118 were the legal base for the adoption of legislative and non-legislative measures in the OSH sector. These articles formed the most important base of the EEC Treaty for the adoption of legislation relative to this area. This provision has more commercial initiatives than social, since the differences in the laws and regulations among member states are factors that stop the free trade of goods and the mobility of people.

During the first 30 years of its existence, the EU aimed at the approach of the member states' legislation for the withdrawal of these barriers in a sequence of different issues affecting OSH. Article 100 refers to the approach of the clauses that directly affect the operation of the Common Market and provides the issuing of relative directives. Article 118 refers to working conditions and the prevention of occupational accidents. The limited authority in the social sector that was originally given to the bodies of the EU was aimed at limiting any misuse of the rules of competition within the Common Market. It was considered certain that the anticipated economic development would automatically result in social progress.

The EU's roles are very important, especially those of the Directorate General of Employment, Industrial Relations and Social Affairs (DG V), which is relevant to OSH matters. The DG V has initiated many directives, regulations, resolutions, and recommendations covering a wide range of OSH issues, always following the scientific and technological progress as well as the social and economic needs of the European workers and the evolution of the EC itself. The Committee, in performing its aims, is consulted by the Advisory Committee of Safety, Hygiene and Health Protection at Work, the Economic and Social Committee, and the European Foundation for the Improvement of Living and Working Conditions sited in Dublin. All these organizations have achieved a high scientific level of decisions through elaborate positions and social dialogue.

The Advisory Committee for Safety, Hygiene and Health Protection at Work was established in 1974 by a Council decision. The composition of the Advisory Committee consists of two governmental representatives, two employers' representatives, and two employees' representatives from each of the member states, and its responsibilities are:

- The exchange of information and experience in the field of OSH connected with existing legislation;
- The facilitation of cooperation among national management and syndicated organizations of employees and employers;

- The discussion of problems and the proposal of solutions applicable on a community scale;
- The examination of scientific progress and consultation with the European Committee.

Its common finding is that the Consultant became an active player in the issuing of documents of a scientific, technical, and political nature and he/she is vital in outlining policy in the field of OSH.

The European Institute for the Improvement of Living and Working Conditions, sited in Dublin, was established in 1975 as a specialized and autonomous organization of the Community. It is responsible for the applied research in the fields of social policy, the implementation of the new technologies, and the improvement and the protection of the environment. It has carried out programs concerning safety and hygiene, promotion of health, dangerous waste, sequences by the new technologies, participation of the employees, and shift work.

Until 1992, a landmark year for the realization of the internal market, the Community issued three action programs in the field of OSH. The first action program was about the period 1978–1982, during which the framework directive was issued for the protection of the employees from dangers related to their exposure to chemical, natural, and biological factors at work. In the same period, special directives were adopted which limited the exposure of employees to lead and asbestos, and where for the first time the role of the employees' representatives was recognized, and also a directive for taking measures against dangers from major industrial accidents. The second action program was about the period 1984–1988 and the main directives issued at that time included one concerning noise and another which forbids the production and use of four chemical substances particularly dangerous to the health of the workers.

4. THE SINGLE EUROPEAN ACT

The difficulties and delays for the completion of the EC policy in the field of OSH must be attributed to the inadequacy of the legislative base before 1986 — namely, the fact that the social policy clauses required unanimity for taking decisions. This gave minorities the opportunity to hold back the proposals of the Commission. By the Single European Act in 1986, and especially Articles 100 and 118, the Treaty of Rome was modified to what is related to OSH. The important change is the institution of minimum specifications for the working environment and the making of decisions in the Council by relative majority.

These changes directly affected the third action program 1988-1992, which was ruled by the philosophy that OSH cannot be considered in the light of financial terms alone, but must also take into account social terms. During this period, among the important legislative work that was completed, the framework for directive 89/391/EEC has a dominant position. This directive introduces the innovative elements of the "occupational risk self-assessment" and the adoption of social dialogue in OSH matters in enterprises. It also establishes a new legislative strategy, which is based on a general law frame with a wide field of application, completed by special arrangements in aiming at the special workplaces and working conditions be covered. The third action program was completed in 1992 with the celebration of the European year for the safety and health at work.

A field of community activities that is directly connected with OSH, but which is also an indirect way of restraining the free trade of goods, is the different national standards by which these goods are produced. In order to withdraw the barriers to transactions resulting from this, the Council established the procedure of information in the field of technical standards and specifications (Directive 83/189/EEC), which ensures the full information of the Committee for everything that happens in this field in every member state. Moreover, it instituted a procedure for the adaptation of fully compatible European and national standards, which are a result of mutual information and cooperation of the National Organizations of Standardization. This procedure is a new approach in the area of standardization and is based on the directives of technical approach, which led to harmonized European standards. In the context of the European Organizations of Standardization, technical committees are operating to examine groups of standards, many of which are related to OSH. Likewise, a great number of the directives of technical procedures and the standards, which are related to them, are connected with OSH issues (directives for low voltage machines and personal protection equipment).

The role of information is also important, not only to the social partners and to the greater population, but also to the members of the Commission. For the wider communication of the former, the Community has issued many informative editions in all the Communities' languages and also publishes the quarterly journal *Janus*. The information campaign reached its peak during the European year of OSH. There is a special committee composed of specialists in every member state for the dissemination of information to all members of the Commission about the progress of harmonization in the directives of minimum specifications. Additionally, special scientific committees are operating for the collection of national statistics on work accidents and occupational diseases.

5. THE SOCIAL CHARTER AND THE MAASTRICHT TREATY

In the 1980s, the "social space" concept was a sheer change from the "market" concepts that were dominant in 1970s. It was determined that the single market should not only be an area of free trade, but also a unified social area where the economic and social cohesion was to be safeguarded. These ideas were confirmed in the Social Charter, which covers twelve categories of fundamental social rights. Some of them are indirectly related to occupational safety and health issues: improvement of living and working conditions, social protection, informing and participation of employees, protection of children and teenagers, aged employees, disabled persons. There is also a special section referring to the safety and health of the employees, highlighting certain initiatives. In the early 1990s, the notion that employees' health and safety is an important part of the social dimension and an integrated aspect of the internal market's realization began to dominate.

The period from the institution of the European Community until the adoption of the Social Charter of the workers' social rights was characterized by the production of legislative work that aimed at the completion of the Common Market. In the 1980s, the notion was emerging that, if the social component was not developed in parallel, there was a risk of torpedoing economic development as well. The policy in the field of OSH

since the early 1990s is a reflection of the new social concepts that have slowly appeared on the European scene. This period is characterized by the undertaking of non-legislative measures aimed at the wide dissemination of information and the training of employers, employees and citizens in general, which are developing in parallel with the legislative action that is now focused on assuring the minimum acceptable working specifications.

In 1992, the treaty for the institution of the European Union, which changed the Treaty of Rome, was signed in Maastricht in the Netherlands. According to this, the twelve member states instituted a European Union, marking a new stage in the construction of the community. The changes introduced by the Maastricht Treaty in the field of OSH are more indirect than direct. They are expressed mainly in the "Protocol of Social Policy" that was appended to the treaty and is the commitment of the eleven signing countries (with the exception of the United Kingdom) to promote a more progressive policy in the social field. Some of the areas of responsibility that have been added by the main body of the Treaty are:

- Social protection, the raising of living standards, improvements to the quality of life, and protection of the environment;
- Policies in the social and environmental fields;
- Rules of the internal market that are related to the protection of OSH, the environment and consumers, who are dependent on a high level of protection.

In addition, articles that concern the protection of public health, the protection of consumers' health, security and economic interests, and the protection of the environment were embodied in the EEC Treaty. A small number of posterior directives are based on these articles, which are supplementary to the main OSH directives.

In 1993, the Commission produced the "Green Paper on the European Social Policy", which has explicit references to issues about the safety and the health of the employees. The issue of minimum common standards in working conditions is of great importance as they are directly connected to the issues of competition and social dumping. The will of the Commission to continue its action in the field of OSH, based on the existing directives, is ensured. The Community legislation continues to hold a dominant position in the securing of minimum specifications for health and safety to all European workers. However, at the same time, the notion has been realized that legislation is not the only way for the institution of social rules to be expressed, but that attention must be paid attention to the agreements arrived at among the social partners. Apart from the standards, the rules, the collective agreements, the definition of rights and responsibilities, oral and tacit behavioral rules are needed for the reduction of conflicts. The main thing contribution needed towards this direction is the unobstructed cooperation of the member states, the European Parliament, and the social partners. The idea that improvement of working conditions means increasing the production costs is considered played out, and, what is more, untrue. On the other hand, the Community shares the view that the improved safety and health of the workers leads to the reduction of production costs through reduction of accidents and diseases. In addition, statistics show that within the EC 10 000 000 work accidents and/or occupational diseases

occur every year, 8000 of which are fatal. The annual cost of these in compensation is estimated at 26 billion ecus, without including the indirect cost, which is not viable for the economy in general.

In Bilbao in 1994, the Community set up the European Organization for Safety and Health at Work as an additional means for the provision and distribution of technical and scientific information. The organization's aims are:

- The collection and distribution of OSH information among the member states;
- The diffusion of research results;
- The organization of educational seminars;
- The development of close bonds among specialized organizations of member states;
- The provision of technical and scientific help to the community bodies.

Other non-legislative actions that characterize the last decade of the twentieth century are the issuing and continuous updating of the European catalogue of occupational diseases and, among the member states, the exchange of public servants who are in charge of controlling the implementation of Directive 89/391/EC.

6. THE AMSTERDAM TREATY

By the end of the millenium, the Community aimed to complete the "General Framework of Action of the Commission of the European Communities in the Field of Safety, Hygiene and Health Protection at Work (1994–2000)". The legislative and non-legislative measures that the Community is going to take in the field of OSH by the end of the year 2000 are reported in this document. Matters that demanded special attention in this period were: transfer of issued directives to member states, small and medium-sized enterprises, social dialogue, information and training of employees, international cooperation and non-legislative measures. The program of non-legislative measures with a view to implementation is focused on eleven special actions the realization of which requires decisions from the Council and the Committee. Currently, decisions have been issued for the composition of the Superior Work Inspectors Committee, the scientific committee for the limits of exposure to chemical substances during work, the SAFE program, and the OSH European week which has been celebrated in the three successive years 1996–1998.

In June 1997, a revision to the Treaty of the European Union was made by the Treaty of Amsterdam. This revision takes into account the tremendous challenges which the Union has to deal with at the start of the twenty-first century. These can be epitomized in the globalization of the economy and its effects on occupation, emigration pressures, ecological imbalances and threats to public health, especially in the view of the widespread of drugs and AIDS. The Union is called on to deal with these issues by the provision of its proposed enlargement. With the Treaty of Amsterdam, substantial changes occurred, especially in social policy, but also to other policies that are related to OSH. After the signing of the Protocol for social policy by Great Britain in May 1997, this was embodied in the Treaty, thus establishing a coherent and effective legislative framework for the community to undertake action in this field. The more important changes, which are related directly or indirectly to the OSH field, are:

- Measures of harmonization, related to the protection of OSH,

the environment and consumers, which are recommended by the Committee, the Council or the Parliament, have to take into account the new scientific data.

- The Community and the member states must take into account the fundamental social rights and aim for the improvement of living and working conditions, social protection, social dialogue and the development of human resources.

The Community supports and complements the activities of the member states related to the improvement of the working environment. To this end, the Council, unanimously and after consulting with Economic and Social Committee and the Committee of the Regions, institutes through directives the minimum specifications of the working environment. Such directives shall avoid imposing administrative, financial, and legal constraints which would hold back the development of small and medium-sized enterprises. The member states can maintain or institute stricter national clauses, which reconcile with the other clauses of the treaty: namely, they do not impact the clauses of competition. They can, as well, assign the implementation of the relevant directives to the social partners, provided that is a mutual request and until the date of harmonization have taken the necessary measures through agreements.

The Commission shall have the task of promoting consultation among the social partners at a community level and shall take any relevant measure to facilitate their dialogue. To this end, before submitting proposals in the social policy field, the Commission shall consult social partners on the possible direction of Community action.

The social dialogue may lead to the establishment of conventional relations, as long as the social partners want so. The agreements made at a community level are implemented in the OSH field, when it is mutually requested, by the signing of parties and the Council's decision taken with qualified majority, and after the Commission's proposal.

The guarantee of a high level protection for the health of man must exist in all community policies. The reference to drugs is more intense. Indirect reference is made for the prevention of AIDS.

The demands for the protection of the consumers are taken into consideration for the determination and implementation of other community policies and activities.

The last few years have been characterized by the growth of harmonization among the member states of the Community directives, which have been issued according to Article 118 and by its modification of the fundamental directives in the light of scientific and technological progress. The path of the OSH structure will be concluded in a different environment, implied by the Amsterdam Treaty, the beginning of negotiations for the enlargement of the Community, and the changes in working conditions (for example, resulting from an increase in the number of aged and women workers, and new forms of occupation). In this light, efforts must be focused on the effectiveness of legislation, preparation for the EU's enlargement, the strengthening of the occupation and the focusing on new working hazards.

7. CONCLUSIONS

The experience acquired in the field of OSH to date can be utilized as a guide for the future. The basic principles have been founded and consolidated, eliminating regressions. Social dialogue is a practice that is going to be implemented more and more in the OSH field, since this was used in the first stages of the EU with great success. The institution of minimum specifications for the working environment with directives and the elaboration of European Standards compatible to them is another stabilized method. The new living and working conditions which may emerge in the Europe of the twenty-first century may change the traditional system of occupational health and safety that is currently practiced. The only sensible way forward, not only in order to sustain it but also to adopt new data, is by the creation and establishment of unified "social space" as a first priority of the European Union, and not as a supplementary to the Common Market.

REFERENCES

Commission of European Community, 1992, *Health and Safety at Work: A Challenge for Europe* (Luxembourg: Agency of Official Publications of European Community).

Commission of European Community — DG V, 1993, *European Social Policy: Options for the Union*, Green Paper of the Commission, COM(93) 551 final.

Commission of European Community — DG V, 1993, *Europe for Safety and Health at the Workplace: Social Europe*, 3/93.

Commission of European Community — DG V, 1994, *European Social Policy: The Route to the Future for the Union*. The White Paper (Luxembourg: Agency of Official Publications of European Community).

European Foundation for the Improvement of Working Life Conditions, 1992, *Four-Years Action Program 1993–1996* (Luxembourg: Agency of Official Publications of European Community).

KOUKIADIS, J. and BLANPAIN, R., 1993, *The Community Work Law Before and After Maastricht* (Thessaloniki: Paratiritis).

KOUTSIARAS, D. N. 1992, *The Community Social Policy: An Economic Point of View* (Athens: Hellenic Center of European Studies).

WRIGHT, F., 1998, Occupational health and safety: the European Union. In *Encyclopaedia of Occupational Health and Safety* (Geneva: ILO), vol. I, pp. 23.31–23.34).

http://europa.eu.int/abc/obj/treaties/el/eltoc.htm

http://europa.eu.int/comm/dg05/h&s/index_hs.htm

Exposure Assessment of Upper Limb Repetitive Movements: A Consensus Document

Developed by the Technical Committee on Musculoskeletal Disorders of the International Ergonomics Association (IEA) and endorsed by the International Commission on Occupational Health (ICOH)

D. Colombini, EPM – CEMOC, Milano, Italy

N. Delleman, TNO – Human Factors Research, Soesterberg, The Netherlands

N. Fallentin, National Institute of Occupational Health, Copenhaven, Denmark

A. Kilbom, National Institute of Occupational Health, Solna, Sweden

A. Grieco, EPM – Department of Occupational Health, University of Milan, Italy

Note: The views expressed in this document are those of the authors and do not necessarily reflect the views of the IEA.

WORKING GROUP

The following Group Members participated in the preparation of this document:

Dr Hudson de Araujo Couto, MG, Brazil

Prof. Thomas J. Armstrong, University of Michigan, USA

Prof. Gunnar Borg, Department of Psychology, Stockholm University, Sweden

Dr Peter Buckle and Dr David A. Stubbs, Robens Institute, University of Surrey, UK

Dr Alberto Colombi, PPG Industries, Pittsburg, Pennsylvania, USA

Dr Jan Dul, NIA-TNO, Hoofddorp, The Netherlands

Dr Ruddy C. Facci, Curitiba, Brazil

Dr Arun Garg, University of Wisconsin, Milwaukee, USA

Prof. Mats Hagberg, National Institute of Occupational Health, Solna, Sweden

Dr Christine M. Haslegrave, Institute for Occupational Ergonomics, University of Nottingham, UK

Dr Bronislav Kapitaniak, Laboratoire de Physiologie du Travail et du Sport, Universite Curie, Paris, France

Dr Waldemar Karwowski, Center for Industrial Ergonomics, University of Louisville, USA

Prof. Gianni Molteni, Department of Occupational Health, "Clinica del Lavoro Luigi Devoto", Milano, Italy

Dr V. Putz-Anderson, US Dept. Health & Human Services Cincinnati, Ohio, USA

Edoardo Santino, Sorocaba, Brazil

Dr Barbara Silverstein, Washington State Department of Labor & Industries, USA

Dr Bente Schibye, National Institute of Occupational Health, Copenhaven, Denmark

Prof. Gisela Sjogaard, Dept. Of Sports Science, University of Odense, Denmark

Dr Toru Itani, Department of Hygiene and Occupational Health, Nagoya City University Medical School, Japan

1. INTRODUCTION

1.1. Aim

This consensus document, while taking into consideration the most recent and significant contributions in the literature, intends to supply a set of definitions, criteria, and procedures useful to describe and, wherever possible, to assess the work conditions that can represent a physical overload for the different structures and segments of the upper limbs. The consequences of physical overload are represented by work-related musculoskeletal disorders (Hagberg *et al.* 1995).

The document is aimed at all the operators, i.e. occupational doctors but mainly technicians, who are, or will be, involved in the prevention of work-related upper limbs musculoskeletal disorders.

From this point of view, the document intends to provide methods and procedures (for an exposure assessment) easily applicable in the field, not requiring sophisticated instrumentation and, when possible, based on observation procedures.

The proposed methods shall be based as far as possible on knowledge and data from scientific literature: when these are contradictory or deficient, reference will be made to standards or pre-standards issued by national and international agencies and bodies, with the experience of researchers involved and common sense.

From time to time it shall be clarified on which (more or less consolidated) basis the different result analysis and interpretation procedures are proposed.

The "guiding principle" that brought about all the choices reported in the document can be summarized in a few lines of the International Code of Ethics of Occupational Health Professionals prepared by the International Commission on Occupation Health (ICOH 1992) "Special consideration should be given to the rapid application of simple preventive measures which are cost-effective, technically sound and easily implemented. Further investigations must check whether these measures are efficient and a more complete solution must be recommended, where necessary. When doubts exist about the severity of an occupational hazard, prudent precautionary action should be taken immediately."

Potential users increasingly demand an easily applicable method for description and assessment of work with repetitive movements, avoiding the risk of "trying to measure everything", "interpreting little" and "changing nothing".

In Europe this is a result of the new legislation (particularly, European Economic Community (EEC) directive 391/90/CEE) which requires employers to evaluate work hazards in their companies and reduce them if necessary. The European Union's "Machine Directive" also deals with this topic, by introducing the need to provide machine designers and manufacturers with easily applied methods with which to evaluate potential health risk factors connected with using these machines.

The group intends to give a response, even if there are still uncertainties from a strictly scientific standpoint; however, the group commits itself to perform subsequent validations especially of as yet unconsolidated issues. Therefore, what is proposed is not a "rigid pattern" but, once some reference points have been

set, it is intended as a dynamic tool able to seek in time the best point of equilibrium between knowledge from research and application requirements.

This document focuses specifically on identification of risk factors and describes some of the methods that have been developed for evaluating them. There is a rapidly developing body of literature on job analysis and as yet no agreement on a single best way to analyze jobs. It is the committee's belief that the appropriate methods depend on the reason for the analysis, e.g. walk through inspection, evaluation of a specific tool or work method, or analysis of a problem job. The appropriate methods will also vary from one work situation to another, e.g. office where workers are performing keyboard tasks or foundry where workers are using powered grinders. Professional judgement is required to select the appropriate methods. Analysis and design of jobs should be integrated into an ongoing ergonomics program that includes management commitment, employee involvement, hazard identification and control, training, health surveillance, and medical case management.

The aim of this section is not a complete overview of the literature. It simply intends to direct readers towards such studies that represent essential contributions for the operational choices which are then suggested by the authors of this study.

In 1987, C.G. Drury (Drury 1987) discussed a method for the biomechanical assessment of pathologies due to repetitive movements, and focused on three main factors (force, frequency, and posture); he suggested a description and assessment method which counts the daily number of "hazardous movements" for the body, and particularly for the wrist.

In 1988, V. Putz Anderson published an interesting book in which he systematically listed all the practical and theoretical knowledge, which was available at the time, on the control and management of cumulative trauma disorders (CTDs) (Putz Anderson 1988). Amongst other things, the book postulates a "risk model" for CTDs, based on the interaction of four main factors: repetitiveness, force, posture, recovery time.

In 1986—7, Silverstein, Fine and Armstrong (Silverstein *et al.* 1986, 1987) highlighted the connection between repetition and force risk factors and Cumulative Trauma Disorders (particularly Carpal Tunnel Syndrome). They also threw light on the fact that there is a synergistic mechanism between the two factors under consideration.

In 1992, Winkel and Westgard suggested some guidelines for the study of occupational risk factors — and of individual ones — related to alterations of the cervical and scapulohumeral region (Winkel and Westgaard 1992).

In 1993, a large group of authors who were part of an ICOH working group, mainly Scandinavian and American, presented a conceptual model for the interpretation of the development of occupational musculoskeletal disorders of the neck and upper limbs (Armstrong *et al.* 1993).

Again in 1993, Tanaka and McGlothin, two NIOSH researchers, presented a conceptual model for the study and prevention of the occupational carpal tunnel syndrome (CTS) (Tanaka and McGlothin 1993). In their method, the exposure limit required is determined by the repetitiveness of movements, by the force used, and by the postural deviations of the joint involved, the wrist in this case.

In 1993, McAtamney and Corlett proposed the Rapid Upper Limb Assessment (RULA) method (McAtamney and Corlett 1993) in which upper limb risk exposure was evaluated using a simple description of the posture, force, and repetitiveness of the muscular action. They also proposed a procedure to calculate a synthetic index.

Guidelines for "practitioners" were presented and discussed by Kilbom in 1994, for the analysis and assessment of repetitive tasks for the upper limbs (Kilbom 1994).

This is an extremely important review, both theoretically and practically, and supplies useful suggestions both for the definition of repetitive tasks, and for the classification of the different issues to consider during analysis.

Frequency of movement is pointed out as being of particular importance for the characterization of risk. For each body region (hands, wrist, elbow, shoulder), indications are given on the frequency limits of similar movements showing a high risk for upper limb injuries where such frequencies are exceeded.

With respect to action frequency, the existence of other overloading factors (high force, high static load, speed, extreme postures, duration of exposure) is considered as being an amplification of the risk level.

In 1995, the contributions of a qualified panel of authors were summarized in a volume devoted to work-related musculoskeletal disorders (WMSDS) (Hagberg *et al.* 1995). Starting from an analysis of the best designed studies on the subject, the book examines the different elements representing occupational risks which could be a cause of the various pathologies of the upper limbs, and possible measurement and analysis methods are indicated for each of the elements considered.

Moore and Garg suggested a model to analyze jobs for risk of distal upper extremity disorders. It is based on the measurement of six variables: force, action frequency, posture, recovery time within a single cycle, speed of work, daily duration (Moore and Garg 1995).

More recently, NIOSH's Center for Disease Control and Prevention published "Musculoskeletal Disorders and Workplace Factors: A Critical Review of Epidemiological Evidence for WMSDs of the Neck, Upper Extremity and Low Back." One of the things that this volume provides is a critical report of studies used to show the association between certain working risk factors (particularly repetitiveness, force, posture and vibration) and individual upper limb pathologies.

There are also various documents prepared by national bodies, institutes, and by International Standards agencies. These represent useful references for the definition of description and analysis procedures and criteria when dealing with tasks that present biomechanical overload for the upper limbs. Among these, the following deserve mention:

- the "Code of Practice Occupational Overuse Syndrome" issued by the Australian Health and Safety Commission in 1988 (Victorian Occupational HSC 1988);
- the "Ergonomic Program Management Guidelines for Meatpacking Plants", issued by the OSHA, USA, in 1991 (OSHA 1991);
- the standard plan for the control of CTDs which is being drafted by American National Standard Institute (ANSI), a USA standards agency (ANSI z—365 Draft Control of CTD—PONT. 1 upper extremity (ANSI 1995);

- the OSHA draft standard "Ergonomic Protection Standard" (Schneider, S. 1995).
- The draft standards pr-EN 1005/3 (Committee European Normative — CEN 1997) and pr-EN 1005/4 (CEN 1997) concerning, respectively, "Safety of Machinery, Human Physical Performance. Recommended force limits for machinery operation" and "Safety of Machinery, Human Physical Performance. Evaluation of Working Postures in Relation to Machinery."
- The draft standards ISO/DIS 11226 (1998): "Evaluation of working postures".
- Sjogaard G., Fallentin N. *"Exposure assessment in repetitive monotonous jobs: Danish national strategies for preventing work-related musculoskeletal disorders"*. Schriftenreihe der Bundesanstalt für Arbeitsschutz. Tagungsbericht 10. *"Problems and progress in assessing physical load and musculoskeletal disorders"*. Dortmund: Bundesanstalt für Arbeitsschutz: 73–81 1996.
- Circular letter No. 5/1994. "Monotonous repetitive work. A model for mapping and evaluating repetitive monotonous work". Danish Working Environment Service: 1–12 1994 (in Danish).

Beyond the specific objectives, these drafts continuously relate the evaluation of force and posture to the other factors. Thus, the force limits are related to action frequency, speed, and work duration. The acceptability of certain postures and movements of the various upper limb segments is established by the duration for which they are maintained and (in the case of movement) to the action frequency.

In summing up this report, room must be given to the checklists that are so often seen in the medical literature, although this is not the occasion to propose a detailed analytical review. In general terms, however, it can be said that checklists have proved useful, either in particular application contexts (e.g. industrial sectors with assembly lines in which the workers carry out a single task) or, in establishing the presence/absence of a job that is "problematic" for the upper limbs (without evaluating the merit of the individual risk factors and the possibility of reducing them).

This study takes into consideration several proposals derived from the most authoritative checklists. At the same time, it searches for models to describe and evaluate each of the main risk factors considered in those checklists.

2. ASSESSMENT GENERAL MODEL AND DEFINITIONS

The description and assessment general model, concerning all exposed workers in a given situation, is aimed at evaluating four main collective risk factors: repetitiveness, high force, awkward posture and movements, lack of proper recovery periods. Such factors shall be assessed as a function of time (mainly considering respective duration).

In addition to these factors, others, grouped under the term "additional factors", should be considered. These are mechanical factors (e.g. vibrations, localized mechanical compressions), environmental factors (e.g. exposure to cold) and organizational factors (e.g. pre-established and non-adjustable rhythms) for which there is evidence of causal or aggravating relationship with work-related musculoskeletal disorders of the upper limbs: the list of such factors is open.

Each identified risk factor is to be properly described and classified. This allows, on the one hand, the identification of possible requirements and preliminary preventive intervention for each factor and, on the other hand, eventually, to consider all the factors contributing to the overall "exposure" in a general and mutually integrated frame. From this viewpoint "numerical" or "categorical" classifications of results may be useful for an easier management of results, even if it is important to avoid the sense of an overt objectivity of methods whose classification criteria may still be empirical or experimental ones.

To this end, the following definitions are important:
- The work is composed of one or more *tasks*; these are definite activities (such as the stitching of clothing, the loading or unloading of pallets, etc), that can occur one or more times in one work shift;
- within a single task, several *cycles* can be identified. Cycles are sequences of technical actions, which are repeated over and over, always the same way. In general terms, cycles describe the completion of an operation on product or, sometimes, the completion of a product unit.
- within each cycle, several *technical actions* can be identified. These are elementary operations that enable the completion of the cycle operational requirements. They may imply a mechanical activity, such as turning, pushing, cutting, etc, or a control activity, e.g. checking for faults, etc. In this case, the action is not necessarily identified with the single body segment movement, but rather as a group of movements, by one or more body segments, which enable the completion of an elementary operation.

There is a terminological problem in defining a technical action: in fact, it is substantially attributable to the concept of "Therblig" (Barnes 1958, 1968) "work element" in some literature sources, or "micromotions and exertions" in the ANSI 2-365 (1995) draft. To solve this terminological problem, it was decided to keep the term "technical action" which is immediately understandable by worksite technical staff. Related examples (picking, shifting, moving, turning) should eliminate any doubtful interpretation. Table 1 lists the main terms used in this document, together with the definitions that best fit the authors operational choices for exposure assessment.

The suggested procedure for assessing the risk should follow the general phases listed hereunder:
- pinpointing the typical tasks of any job, and — among them — those which take place in repetitive and equal cycles for significant lengths of time;
- finding the sequence of technical actions in the representative cycles of each task;
- describing and classifying the risk factors within each cycle (repetitiveness, force, posture, additional factors);
- reassembling of the data concerning the cycles in each task during the whole work shift, taking into consideration the duration and sequences of the different tasks and of the recovery periods;
- brief and structured assessment of the risk factors for the job as a whole.

3. ORGANIZATIONAL ANALYSIS

Organizational analysis should come before the analysis of the

Table 1. Main definitions of recurring terms used in exposure assessment.

Organized work:
the organized grouping of work activities that are carried out within a single working shift; it may be composed of one or more tasks.

Task:
specific working activity aimed at obtaining a specific operational result
The following tasks are identified:
REPETITIVE TASKS: characterized by repeated cycles with mechanical actions
NON-REPETITIVE TASKS: characterized by the presence of non-cyclical mechanical actions

Cycle:
a sequence of technical, mainly mechanical, actions, that is repeated over and over, always in the same way.

Technical action (mechanical):
an action that implies a mechanical activity; not necessarily to be identified with the single joint movement, but rather with the complex movements of one or more body regions that enable the completion of an elementary operation

Potential risk factors

Repetitiveness:
the presence of events (i.e. cycles, technical actions) that are repeated in time, always in the same way.

Frequency:
number of technical (mechanical) actions per given time units (no. of actions per minute)
High frequency is the related risk factor.

Force:
the exerted force required by the worker for the execution of the technical actions.

Posture:
the whole postures and movements used by each of the main joints of the upper limb to execute the sequence of technical actions that characterise a cycle.
Awkward posture: hazardous postures for the main joints of the upper limbs.

Recovery period:
period of time between or within cycles during which no repetitive mechanical actions are carried out. It consists of relatively long pauses after a period of mechanical actions, during which the metabolic and mechanical recovery of the muscle can take place. Lack of recovery is the related risk factor.

Additional risk factors:
additional risk factors may be present in repetitive tasks but are neither necessary nor always present. Their type, intensity, and duration lead to an increased level of overall exposure.

-- work: 460 minutes -------------------------
-- pause: 20 minutes ------------------------ |480 minutes shift
-- daily pause time : ------ 60 (lunch) + 20 (pause)

Figure 1. Definition of repetitive task.

therefore without mechanical actions, it will be defined as a recovery for the upper limbs.

The tasks with non-repetitive mechanical actions remain and shall be defined as NON-REPETITIVE TASKS (but not ''recovery periods'').

The number of foreseen cycles within a REPETITIVE TASK, and the net duration of each cycle, must be counted at this point.

		MIN. IN SHIFT AND MAIN CHARACTERISTICS	
TASK A	PISTON ASSEMBLY	220	REPETITIVE TASK
TASK B	LINE MAINTENANCE OPERATIONS	60	NON-REPETITIVE TASK
TASK C	LINE VISUAL CONTROL	180	RECOVERY
PAUSE	2 x 10 MIN. EACH 60 MIN. (LUNCH)	80	RECOVERY
			TOTAL 480 MIN. WORK (+ 60 MIN. LUNCH)

Note: periods of "non-working" in the order of minutes.
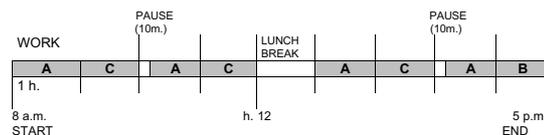
Figure 2. Definition of non-repetitive task.

The number of cycles often coincides with the number of pieces to be worked in each shift.

4. RISK FACTORS

	→ 440 pistons must be assembled in 220 min in the 8-hour shift
TASK A described in Example 2	→ the assembly of each piston is to be considered as a single cycle
	→ the cycle duration is 30 sec . (440 pistons in 220 min. net)

Figure 3. Definition of cycle time.

4.1. Repetitiveness: frequency

This is probably the risk factor of greatest importance in many jobs in industry, so much so that the syndromes of specific interest have often been defined through this element (repetitive strain injuries). The characterization of repetitiveness can be used to discriminate the tasks that must be assessed. To this end, the presence of a repetitive task for the upper limbs can be defined as the consecutive activity, lasting at least one hour, in which the subject carries out work cycles similar to each other and of relatively brief duration (a few minutes at maximum). The cycles require the carrying out of mechanical actions.

Once repetitive tasks are submitted to analysis, there is the more important problem of quantifying and assessing repetitiveness. In the literature, a definition of repetitiveness on

four main risk factors and of the additional factors. It is essential to focus on the *real task duration and repetitive tasks*, and on the existence and distribution of *recovery periods*.

The first phase of the analysis is finding the distribution of work times and pauses within the work shift(s).

In this way it is possible to find the duration and distribution of the macro-pauses (RECOVERY PERIODS).

The organized work shift may consist of one or more working tasks.

In turn, each task may be characterized by cycles or by other types of execution.

If the task is characterized by cycles with mechanical actions, it will be defined as a REPETITIVE TASK.

If it is characterized by check operations (examination, inspection) without movements or awkward postures, and

the basis of cycle duration can frequently be found. High repetitiveness, in particular, is postulated with cycles lasting less than 30 seconds (Silverstein *et al.* 1987).

This same proposal is completed by stating that high repetitiveness occurs when over 50% of the cycle time is spent in performing the same type of action. It is to be noticed, however, that it is possible that very short cycles do not require very frequent gestures and movements, and that longer cycles are carried out with a high frequency of actions. Since the development mechanism of tendon pathologies seems to be related to movement frequency, then action frequency is a more accurate estimation factor of this risk element.

Measuring the frequency of the single joint movements would be the best thing for assessment, as some authors suggest. In fact, the highest risk occurs when the same type of movement is frequently repeated by the same joint. Obviously, a direct measurement of joint movement frequency is not always feasible in the field. It would require the measurement of the frequency of each type of movement (flexion, extension, adduction, abduction, etc.) for each of the main joints, and for both upper limbs.

In this proposal repetitiveness is measured, on the one hand, by counting, within the cycle, the number of technical actions performed by the upper limbs and, on the other hand (see, hereafter, analysis of postures and movements), by identifying, for each action, how many times (or for how long) it involves a given posture or movement of each main segment/joint of the upper limbs (hand, wrist, elbow, shoulder).

A description of the technical actions often requires the filming of the job, which must then be reviewed in slow motion. When the cycles are very short (few seconds), a direct observation at the workplace could work as well as filming. If the task is technically complex, it is extremely useful, indeed often essential, to describe the action sequence with the help of company technical personnel experienced in the task itself. Often, the company already has records available in which the task is described and numbered, and the elements constituting successive technical actions are timed (methods—time measurements).

In order to analyze frequency, the following steps have to be implemented:

(a) Description of the technical actions: reviewing the film in slow motion, all technical actions carried out by the right and left upper limbs, must be listed in order of execution. When the same action is repeated more than once, such as drawing a screw, each repetition of a single action must be counted (i.e. turning four times). When a certain action is carried out repetitively, but not on all pieces (i.e. technical actions for quality control, to be carried out once every 4 pieces or every 20 pieces), then the action described is counted with the relevant fraction (one-fourth or one-twentieth). The technical action sometimes coincides with the joint movement. In this case, when the execution of a simple movement is aimed at the execution of a given technical action, it must be counted as a proper technical action (i.e. use finger flexion to push button). Table 2 gives a description of a brief but complex operating cycle recorded at a working post on an engine assembly line. The actions listed are ascribed to the left and/or right upper limbs in numerical terms. Actions not necessarily present in every cycle, are specially calculated.

(b) Calculation of action frequency; from the previously described work organization study, the following are already known: net time of repetitive task, number of cycles in repetitive task, duration of each cycle. From the technical action description it is possible to obtain: the number of actions per cycle, and therefore the action frequency in a given time unit: NO. OF ACTIONS PER MINUTE (Table 2). It is also possible to obtain the overall number of actions in the task/tasks, and consequently for the shift. Additional considerations should be also given to the "cyclicity" (i.e. the changing of frequency during the shift: regular, irregular or cumulated) of a job.

4.2. Force

Force more directly represents the biomechanical effort necessary to carry out a given action, or sequence of actions. Force may be described as being external (applied force) or internal (tension developed in the muscle, tendon, and joint tissues). The need to develop force during work-related actions may be related to the moving or the holding still of tools and objects, or to keeping a part of the body in a given position.

The use of force may be related to static actions (static contractions), or to dynamic actions (dynamic contractions). When the first situation occurs, it is generally described as static load, which some authors describe as a "distinct risk element" (Hagberg *et al.* 1995).

In the literature, the need for using force in a repetitive fashion

Table 2. Example of description and calculation of technical action frequency within a complete cycle.

	NUMBER OF ACTIONS IN A COMPLETE CYCLE	
	RIGHT	LEFT
- Pick up and position gasket in bushing lodging	2	
- Use pincers in right hand and hook 1st spring	2	1
- Use pincers in right hand and hook 2nd spring	2	1
- Shift calibre and push button	2	1
* Spring position correction (once every four pieces)		
- Pick up unscrewing tool and position it	2/4	3/4
- Unscrew or screw spring	1/4	1/4
- Put tool down		1/4
* Substitution of broken spring (once every 20 pieces)		
- Pick up unscrewing tool and position it	2/20	3/20
- Unscrew spring (4 times)	4/20	4/20
- Hold tool		1/20
- Pick out broken spring with pincers (2 times)	2/20	
- Pass tool into right hand	1/20	
- Pick up spring and put it on tool	1/20	2/20
- Position spring and screw it in (4 times)	5/20	5/20
- Put tool down		1/20
Total	9.5	5

No. of pieces per shift	=	2075
Cycle time	=	11.9 sec. per piece
No. of technical actions per cycle: on the right	=	9.5
on the left	=	5.0

Frequency:

Right $\dfrac{9.5 \times 60}{11.9} = 47.9 \cong 48$ (action/minute)

Left $\dfrac{5 \times 60}{11.9} = 25$ (action/minute)

is indicated as being a risk factor for tendons and muscles. Furthermore, a multiplying interaction has been shown to exist between force and action frequency (Silverstein *et al.* 1986), especially for tendon pathologies and nerve entrapment syndromes.

Force quantification in real life contexts is a problem. Some authors use a semi-quantitative estimation of external force via the weight of the objects being handled. In other cases, the use of mechanical or electronic dynamometers has been suggested. The internal force can be quantified by means of surface electromyography techniques.

Finally, other authors suggest using subjective evaluation methods through psychophysical rating scales. On account of in-field applicability problems, two different procedures are advised for the evaluation of the use of force associated with the technical actions present in one cycle.

4.2.1 Use of dynamometers to assess the force required to carry out a given technical action

This procedure is possible and recommendable for actions involving the use of levers, or components of machines and objects. In these cases it is possible to determine the force required to move a lever, or, if the dynamometer is equipped with the proper interface, to simulate the same working action by the workers involved (cross modality – matching). It is to be noticed however that not all technical actions requiring the use of force can be easily determined by means of dynamometers. Besides the observer should be able to establish which technical actions respectively do or do not require the use of force also by asking the operator for the forceful actions.

In order to evaluate the use of force, it is necessary to compare the results obtained on the field with the ability of a reference working population: relevant data are available either in the literature or can be provided by National or International Standards organizations (refer to Appendix C). Additional information on maximal action force may be obtained from the "force atlas" (Rohmert *et al.* 1994).

4.2.2 Use of psychophysical rating scales

In this case, worker's subjective evaluation is used to determine the physical effort associated with the cycle technical actions. Of the different psychophysical scales available in the literature, reference can be made to the "CR10 Borg scale" for perceived exertion (the category ratio scale for ratings of subjective somatic symptoms including perceived force, where 10 defines the relative maximum) (Borg 1982, 1998).

The use of subjective scales is not free from disadvantages likely to affect their reliability (e.g. non-acceptance of the "subjective" method by some employers, conflicting situations influence of motivation, presence of "pathological" subjects, wrong communication of the subjective evaluation goal).

Despite these objections, it is worth mentioning that this technique, if correctly used, allows researchers to evaluate the effort associated with any technical action. In terms of evaluation, reference values are provided by the scale itself. Besides which, according to some authors (Grant *et al.* 1994), the results of the implementation of Borg's Scale, when used for an adequate number of workers, have turned out to be comparable to those obtained with surface electromyography (value of Borg's Scale x

10 @ percentage value with respect to Maximum Voluntary Contraction MVC as obtained by EMG).

The quantification of the effort as perceived by the whole upper limb should theoretically take place for every single action that makes up a cycle. For practical reasons, the actions that require no, or minimal, muscle involvement could be identified respectively as 0 or 0.5 value in Borg's Scale; then the involvement description procedure could only describe those actions, or groups of actions, that require more force than the minimal amount, always using Borg's Scale.

This procedure, when applied to all workers involved, allows researchers to evaluate the average score among subjects for each technical action by asking for the use of force, as well as the weighed average score for all actions and the whole cycle time.

Finally it is to be emphasized that whatever the method used for the description and assessment of force, it is necessary to evaluate:

(1) the average level of force required by the whole cycle — referred to as the maximum force capability, it is defined by reference groups or the group of workers involved;
(2) whether there are in the cycle (and which and how many) technical actions requiring the development of force beyond given levels (peak force)?

It is useful to know also the presence of peaks because the knowledge of average level only can hide their presence.

4.3. Posture and Types of Movements

Upper limb postures and movements during repetitive tasks are of basic importance in contributing towards the risk of various musculoskeletal disorders. A definite agreement is found in literature as to the potential damage coming from extreme postures and movements of each joint, from postures maintained for a long time (even if not extreme) and from specific, highly repetitive movements of the various segments. Moreover, the description of postures and movements of each segment of upper limbs during technical actions of one cycle completes the description of "the repetitiveness" risk factor. The analysis of postures and movements shall be concerned with each single segment of upper limbs (hand, wrist, elbow, shoulder): it is aimed at checking the presence and time pattern in cycle (frequency, duration) of static postures and dynamic movements involving each segment/joint considered.

The description may be more or less analytical but has to be able to appreciate at least the following items:

(a) Technical actions requiring postures or movements of a single segment beyond a critical level of angular excursion (see below).
(b) Technical actions involving postures and/or movements which, even within acceptable angular excursion, are maintained or repeated in the same way.
(c) The duration expressed as a fraction of cycle/task time of each condition reported above.

Joint combination of such description factors (posture/time) will provide the classification of posture effort for each segment considered.

In order to identify the so-called angular excursion critical levels (point (a)), reference is to be made to data and proposals available in the literature.

Table 3. References to the medium/high involvement of joints during dynamic activity.

	RULA (Mc.Atamney et al., 1993) MEDIUM-HIGH INVOLVEMENT	CEN prEN 1005-4 ISO/CD 11226 INVOLVEMENT HIGH	MED.	OSHA (Schneider S., 1995) INVOLVEMENT	KEYSERLING (Keyserling W.M., 1993) INVOLVEMENT	DRURY (Drury C.G., 1987) HIGH INVOLVEMENT	OCCHIPINTI - COLOMBINI (Occhipinti et al., 1996) HIGH INVOLVEMENT	ANSI Z-365 (Ansi, 1995) HIGH INVOLVEMENT	MOORE - GARG (Moore et al., 1996) HIGH INVOLVEMENT
SHOULDER				UNSUPPORTED					
– Flexion	[45°- 90°] [>.90°]	> 60°	20-60°	UPPER LIMBS OR	UNSUPPORTED	+ 94°	+ 80°	UNSUPPORTED	-
– Extension	SUP.20°	IF PRESENT		ELBOW AT MEDIUM	ELBOW AT MEDIUM	+ 31°	+ 20°	ELBOW AT MEDIUM	-
– Abduction	IF PRESENT	> 60°	20-60°	CHEST HEIGHT	CHEST HEIGHT	+ 67°	+ 45°	CHEST HEIGHT	-
ELBOW									
– Flexion	[60°] [SUP.100]	IF EXTREME		-	-	+ 71°	+ 60°	IF EXTREME	-
– Supination	CLOSE TO MAX.RANGE	IF EXTREME		IF SPEEDY	IF PRESENT	+39°, + 57°	+ 60°	IF EXTREME	-
WRIST									
– Flexion	[15°] [SUP.15 °]	IF EXTREME		SUP.20°	IF	+ 45°	+ 45°	IF EXTREME	+ 30°
– Extension	[15°] [SUP.100 °]	IF EXTREME		SUP.30°	PRESENT	+ 50°	+ 45°	IF EXTREME	+ 40°
– Radial deviation	IF PRESENT	IF EXTREME		-	AND	+ 14°	+ 15°	IF PRESENT	-
– Ulnar deviation	IF PRESENT	IF EXTREME		-	EVIDENT	+ 24°	+ 20°	IF PRESENT	+ 20°
HAND									
– Pinch	·	·		IF > 2 POUND LOAD	IF PRESENT	IF PRESENT	DIFFERENT SCORES	IF PRESENT	·
– Grip	·	·		IF > 10 POUND LOAD	·	IF PRESENT	FOR DIFFERENT TYPES	·	·
– Keying	·	·		IF FOR LONG TIME	·	·	OF PINCH AND GRIP	·	·
– Other fingers movements	·	·		CLICK AND DRAG	TRIGGERING MOTION		SPEEDY FINGERS MOVEMENTS	PRESSING WITH THE TIP OF THE FINGERS	·
HOW TO CLASSIFY POSTURE EFFORT									
- STATIC POSTURE	YES	YES		·	·	·	YES	YES	·
- HIGH FREQUENCY	-	YES		·	·	YES	YES	YES	YES
- USE OF FORCE	YES	·		If grip with force	For objects > 2,7 Kg	YES	YES	YES	YES
- DURATION	-	YES		YES (time in the shift)	YES (> 1/3 of cicle time)	YES	YES	YES	YES
- SAME TYPE OF ACTIONS EVERY FEW SECONDS	YES (> 4 times/min.)	·		YES			YES	·	·

Table 3 summarizes some of the main description and classification methods of postures and upper limb movements during repetitive work.

It should be particularly noted that the international literature highlights the need both to describe the awkward postures and to study their duration and frequency (Putz-Anderson 1988; Hagberg *et al.* 1995; Moore and Garg 1995). It is, however, to be emphasized that at this stage it is not important to describe all the postures and movements of the different segments of upper limbs, but rather to focus on those that by typology or excursion level (as well as by duration) are the static postures and/or the movements involving greater effort and also requiring improvement.

On the other hand, the literature provides information on the risk of postures and movements maintained or repeated identically for prolonged times (point (b)). This holds true even if the excursion of such postures and movements does not reach the critical levels evidenced above. In this case, however, the "duration" factor (point (c)) becomes even more important in fixing criticalness of "stereotypy" of specific postures or motions.

4.3.1 Static postures

Static postures are considered critical when:
a) they near the extremes of the movement range — independently of the duration
b) they result in the body segment being held in an intermediate position within the joint range for a prolonged period of time. Table 4 shows an original suggestion (Rohmert 1973) of the

use of muscle force in an optimal balance between static load and recovery time.

Static postures may be present both in repetitive and non-repetitive tasks; an investigation is necessary as well. In this case the evaluation criteria are based on the type of 14 postures, continuative maintenance time, adequacy of recovery periods (see Table 4 and section 4.4). Appendix D1 contains the relevant parts of an easily applicable international consensus standard.

4.3.2 Movements

The literature often provides a definition associating movements with duration (or repetitiveness): work cycle less than 30 seconds or the same fundamental work actions performed during more than 50% of the cycle time are to be considered as critical (Silverstein *et al.* 1987). Using the latter criteria it is possible to assert that repetition of the same action in the same joint (stereotypy) is critical when it exceeds 50% of the cycle time, regardless of the size of the movement angle.

Further, the duration seems to be the major factor in view of determining the level of effort associated with movements with critical angular excursion. In this regard, the literature sources are less univocal but for practical purposes reference could be made to a fraction of the cycle equal or above 1/3 (Keyserling *et al.* 1993; Colombini 1998), or to the criteria for which the extreme movements of a single joint should not exceed a frequency of twice a minute; these can be the "rationale" with which to define a significant threshold of effort associated with critical angular excursions.

To obtain an exhaustive description of postural risk it is necessary to cover four operational phases:

Table 4. Rohmert's original suggestion concerning recovery periods (in seconds) for technical actions requiring isometric contractions (equal to or longer than 20 seconds) for applied times and forces.

FORCE (%MVC)	HOLDING DURATION (SEC.)	RECOVERY PERIOD	
	20	2	10%
	30	3	10%
	45	7	15%
up to	120	60	50%
20% MVC	180	180	100%
	240	480	200%
	300	1200	400%
	450	2700	600%
	20	10	50%
	40	40	100%
about	60	120	200%
30% MVC	90	360	400%
	120	720	600%
	150	1200	800%
	20	20	100%
about	30	60	200%
40% MVC	50	200	400%
	70	420	600%
about	20	40	200%
50% MVC	30	120	400%
	40	240	600%
	90	720	800%

(a) a separate description of postures and/or movements by each joint: shoulder, elbow, wrist, hand (type of grip and finger movements) and by type of effort (static, dynamic);

(b) static postures: observation of static postures close to extreme articular range, during the cycle/task time; observation of static postures in the medium of articular range held for a prolonged period of time; observation of grip positions during cycle/task time.

(c) joint movements: presence of articular movements, close to the limit of the range of motion during the cycle/task time; repetitive articular movements, due to presence of the same technical actions (independently of the articular range) for at least 50% of cycle time and subsequently of task time.

To allow the technician to make an easier description and classification of the postural effort (type of posture per time), a form is reported in Appendix A (Table 6) with an example of all the major items for each one of the four segments of the upper limb under consideration in order to obtain a postural score. Appendix D2 contains the relevant parts of an easily applicable international consensus standard.

For practical purposes, a significant cycle should be analyzed for each repetitive task. This is better achieved by videotaping.

The video could be reviewed in slow motion to describe and evaluate the effort of each joint segment, making a distinction between right and left side when the effort is asymmetrical.

4.4. Lack of Recovery Periods

The recovery period is a time during which one or more of the muscle groups that are usually involved in the work tasks are basically inactive (macro-pauses). The recovery period is one of the main factors for overall exposure assessment.

The following may be considered as recovery periods:

(a) work breaks, including the lunch break

(b) those periods during which tasks are carried out, which do not involve the usual muscle groups

(c) those periods within a cycle, with actions implying the total rest of the usually active muscle groups; to be defined as macro-pauses, these periods must be at least 15 consecutive seconds; micro-pauses of very short duration indicated as E.G. EMG gabs of at least 0.2 seconds seems to be of major functional significance but are difficult to detect without technical measurements.

The analysis of the recovery periods is first and foremost a check of their duration and distribution within the cycle, and a macroscopic examination of their presence, duration and frequency within the whole shift. With some exceptions (see later), represented by recovery periods for actions implying protracted static contractions, the description and the assessment of recovery periods should be based on the following:

(a) a description of the actual task sequences involving repetitive movements of the upper limbs, of "light" non-repetitive tasks, and of pauses;

(b) the frequency of the recovery periods with reference to the actual number of working hours per shift;

(c) a ratio between the total recovery time and the total working time, in a shift devoted to tasks involving repetitive movements.

The main problem encountered in analyzing recovery periods is the lack of criteria for an adequate assessment (duration, time scheduling).

In this connection, it is worth making the following considerations.

4.4.1 Static actions

As for static actions, classical muscular physiology studies (Rohmert 1973) provide criteria with which to assess the adequacy of recovery periods as an immediate consequence of a static effort (as a function of its intensity, mostly expressed in percentage of the MVC and duration of involvement).

It should be emphasized, however, that such data refer to effect such as performance or, at best, muscular fatigue but is not fully validated when considering major health effects.

While taking into account such deficiencies, the results of these studies are summarized in Table 4. It shows the various degrees of contraction force, as a percentage of MCV, and the various duration of contraction (in seconds); for each, the minimum necessary muscle recovery periods are indicated, both in seconds and as a percentage of the contraction time. The table is self-explanatory.

After each holding condition, and according to the force developed, an adequate recovery period must immediately follow. If such a period is either absent or inadequate, then there is a condition of risk, all the greater with the greater discrepancy between the actual situation and the optimal one. Table 4 is important as a prevention tool; it offers the optimal division between isometric contraction and recovery periods, to be alternated in strict succession. The corollary of the table, as a prevention tool, is that the force required during isometric contractions that last over 20 seconds (maintenance) must never exceed 50% of MCV.

As to using Table 4, if data about % MCV are not available, it is possible to estimate force using Borg CR10 scale.

Also within non-repetitive tasks characterized by the presence of static postures, it is necessary to use adequate RECOVERY TIMES as suggested in Table 4. Appendix D1 contains the relevant parts of an easily applicable, international consensus standard.

4.4.2   Dynamic actions

As for dynamic actions, no adequate studies are available for evaluating the optimum distribution between repetitive work time (with relative levels of muscular and tendinous effort) and recovery time. A partial exception is the case when so-called intermittent static actions are carried out. For this kind of action, reference should be made to the valuable contribution by Bystrom (1991) that established the maximum acceptable work time before it becomes necessary to have a recovery period, considering a given rate of muscular involvement (expressed in % MVC).

In most repetitive tasks, however, upper limbs actions are typically dynamic and, in the absence of consolidated scientific studies, concerning the optimal distribution of recovery periods, it becomes necessary to refer to "rough" and empirical data reported in the literature or in guide documents and standards (Victorian Occ. HSC Australia 1988; ISO TC 159 Draft 1993; Grandjean 1986).

Logically, if not strictly scientifically speaking, all these documents tend to state that:

(1) work involving repetitive movements of upper limbs cannot be continuously sustained for over one hour without a recovery period;

(2) the recovery period, within one hour of repetitive work, has to be in the region of 10–20% of working time (that is about 5–10 minutes per hour). These rough indications, still to be perfected, may guide description and assessment methods of RECOVERY TIMES with relation to "dynamic" activities of upper limbs.

An example of how to obtain a score for "lack of recovery period" is described in Appendix B.

4.5. Additional Risk Factors

There are other factors, apart from those already discussed, which are considered to be relevant in the development of WMSDs.

They always have their origin in work and must be taken into consideration whenever assessing exposure. They have been described as additional in this work, but not because they are of secondary importance — rather, because each of them can be either present or absent in the various occupational contexts. For a factor to be considered, it has to have an association with WMSDs' effects, as well as having a collective impact (that is, on the whole of exposed subjects or on significant groups of them) and not an individual impact (that is, on single subjects). In other terms, factors such as anthropometric measurements, psychological issues, extra work activities of single subjects are not to be considered at this stage of analysis. The additional risk factors may be mechanical, environmental, organizational ones. The list of factors mentioned here (Table 5) is only an indicative one and is not exhaustive: from time to time each operator will decide on the single factors of interest in view of assessing overall exposure.

Table 5. List of possible additional risk factors (not complete).

MECHANICAL
- Hand-arm vibrations
- Extreme precision in positioning objects
- Localized compression on upper limb structures
- Use of gloves
- Rapid or sudden wrenching movements of upper limbs
- Blows and shocks (such as hammering hard surfaces)

ENVIRONMENTAL
- Exposure to cold
- Exposure to heat

ORGANIZATIONAL
- Machine-paced task
- Incentive payment
- Routine overtime
- Working with tight deadlines
- Sudden peaks of high workload
- Lack of training

Mechanical and environmental factors can be described and assessed according to the corresponding time pattern (frequency, duration). This allows the definition of the amount of time (both with reference to cycle time and task time) spent with that factor. For assessment purposes, it will be considered that the optimal condition is represented by the absence, or a very limited presence, of each additional factor. Organizational factors can be described according to category classifications (at least as present/absent).

5. OVERALL EXPOSURE ASSESSMENT

An overall exposure assessment of upper limb WMSDs must account for different risk factors, described one by one and classified.

If, in fact, it is true that the simplest and most elementary prevention actions can be already undertaken after a good analysis of each risk factor, it is all the more true that more comprehensive prevention strategies (e.g. priority choices) must be based on the assessment of overall exposure, as determined by the different combination of the risk factors considered. In this regard, the literature even now provides data and convincing hypotheses on the interrelation between some of the considered factors.

The force–repetitiveness ratio was examined in relation to effects (Silverstein 1987) and muscular capability and physiology while taking activity times (duration) into consideration (CEN 1997).

On the other hand, the relationship between possible force development and some postures (or movements) of the hand—forearm segment (Eastman Kodak 1983) are known. In a recent CEN document (CEN – PrEN1005/4 1996) upper limb postures and movements are classified according to action frequency and overall task duration.

In another recent CEN document (CEN – PrEN1005/3 1997), the recommended force values to be developed in different kinds of manual action by upper limbs are provided in relation to variables such as action frequency, action speed, and overall task duration.

In spite of this, it should be stated that at the present state of knowledge there still is a lack of sufficient data to outline an accurate and parametric general model, combining all the risk factors considered, particularly when the issue is to fix the "specific

weight" of each factor in determining the overall exposure level. Accounting for this, we have to emphasize the necessity of having even partially empirical models for a synthetic assessment of overall exposure to the risk factors considered.

Methods and procedures for determining synthetic exposure scores are already available in the literature. Even when using simple checklists (Keyserling *et al.* 1993; Schneider 1995) the analytical process closes with a synthetic score classifying exposure.

When slightly more sophisticated exposure description and assessment methods are used (McAtamney and Corlett 1993; Moore and Garg 1995), models and procedures are, however, provided for calculating synthetic indices to account for risk factors.

A synthetic index has been recently proposed (Colombini 1998) providing a classification of the risk factors considered here (repetitiveness, force, posture, lack of recovery periods, and additional risk factors). This synthetic index model has been the object of positive preliminary tests, through epidemiological studies. It allows a classification of the results in a three-zone model, useful for implementing preventive actions following from the exposure assessment process.

Being well aware that the data supporting the above overall exposure assessment models are still deficient and often empirical, it is recommended that, if used, they should be adopted "critically" when studies are carried out for preventive action and/or for the active health surveillance of workers. In this respect, and with these goals, the following aspects should be considered:

(a) The exposure indices proposed at present have a methodological value, showing the concept of the integrated evaluation of risk factors.

(b) Such indices also have a practical value: even if they do not provide an absolute statement of the exposure (and thus of WMSD risk), at least they allow the ranking of exposure levels derived mainly from the combination of the different factors in the different work situations. This allows priority choices of action and intervention. Currently, an index may only be used in combination with health status monitoring (complaints, disorders) of the workers involved, in order to see whether the right action and intervention were chosen.

(c) The exposure indices proposed here should not be intended as standards or reference values to distinguish safe or hazardous conditions; this should be clearly emphasized to the potential users.

(d) The exposure indices proposed here, or in the future, need to be validated by laboratory studies, as well as by epidemiological studies (exposure/effect).

(e) The issue of exposure integrated assessment of upper limb repetitive movements is a crucial problem for future scientific and application developments of ergonomics.

(f) Finally, it should be emphasized that the exhaustive description and classification of the risk factors associated with a given repetitive task, the quantification of consequent exposure in a concise, albeit approximate, index, and the need to perform parallel studies on the clinical effects on exposed workers, all represent both an opportunity and a commitment to carry out further research and intervention in the field in the near future.

REFERENCES

ANSI (USA) — Draft ANSI Z-365, 1995, *Control of work-related cumulative trauma disorder.*

ARMSTRONG, T.J., BUCKLE, P., FINE, L.J., *et al.*, 1993, A conceptual model for work-related neck and upper-limb musculoskeletal disorders*, Scand. Work Environ. Health*, 19, 73–84.

BARNES, R.M., 1958, *Motion and Time Study: Design and Measurement of Work*, 6th ed. (New York: John Wiley).

BARNES, R.M., 1968, *Motion and Time Study: Design and Measurement of Work,* 7th ed. (New York: John Wiley).

BORG, G.A.V., 1982, A category scale with ratio properties for intermodal comparison. In H.G. Geissler and P. Petzold (eds), *Psychophysical Judgement and Process of Perception* (Berlin: VEB Deutscher Verlag der Wissenschaften), 25–34.

BORG, G.A.V., 1998, *Perceived Exertion and Pain Scales*, Human Kinetic Europe, Leeds (England).

BYSTROM, S., 1991, ''Physiological response and acceptability of isometric intermittent handgrip contractions'' *Arbete och alsa* , NIOSH (Sweden) – 38.

CEN — Draft PrEN 1005/4, 1996, *Safety of Machinery: Human Physical Performance. Part 4: Evaluation of working postures in relation to machinery.*

CEN — Draft PrEN 1005/3, 1997, *Safety of Machinery:. Human Physical Performance. Part 3: Recommended force limits for machinery operation*.

COLOMBINI, D., 1998, An observational method for classifying exposure to repetitive movements of the upper limbs, *Ergonomics*, 41(9), 1261–89.

DRURY, C.G., 1987, A biomechanical evaluation of the repetitive motion injury potential of industrial jobs. *Seminars on Occ. Medicine*, 2, 41–9.

Eastman Kodak Company, 1983, *Ergonomic design for people at work, Vol. 2. Van Nostrand Reinhold* (New York: Eastman Kodak).

GRANDJEAN, E., 1986, *Fitting the task to the Man* (London/Philadelphia: Taylor & Francis).

GRANT, A.K., HABES, D.J., and PUTZ ANDERSON, V., 1994, Psychophysical and EMG correlates of force exertion in manual work. *Intern. Journ. Industrial Ergonomics,* 13, 31–9.

HAGBERG, M., SILVERSTEIN, B., WELLS, R., SMITH, M.S., HENDRICK, H.W., CARAYON, P., and PERUSSE, M., 1995, *Work-related musculoskeletal disorders . A reference book for prevention*. I. Kuorinka and L. Forcier (eds) (London/Philadelphia: Taylor & Francis).

ISO CD 11226, 1995, *Ergonomics: Evaluation of Working Postures*.

ISO TC 159: Draft, 1993, *Human Physical Strength: Recommended Force Limits*.

KEYSERLING, W.M., STETSON, D.S., SILVERSTEIN, B., and BROWER, M.L., 1993, A checklist for evaluating ergonomic risk factors associated with upper extremity cumulative trauma disorders. *Ergonomics*, 36, 807–31.

KILBOM, A., 1994, Repetitive work of the upper extremities. Part 1: Guidelines for the practitioner. Part 2: The scientific basis (knowledge base) for the guide. *Intern. J. Ind. Ergonomics*, 14, 51—86.

McATAMNEY, L., and CORLETT, E.N., 1993, RULA: A Survey Method for the Investigation of Work-Related Upper Limb Disorders. *Applied Ergonomics*, 24(2), 91–9.

MOORE, J. S., and GARG, A., 1995, The strain index: a proposed method to analyze jobs for risk of distal upper extremity disorders. *Am. Ind. Hyg. Assoc. Journal*, 56, 443–58.

NIOSH, Center for Diseases Control and Prevention, 1997, *Musculoskeletal Disorders and Workplace Factors. A critical review of Epidemiologic Evidence for WMSDs of the Neck, Upper Extremity and Low Back.* US Department of Health and Human Services, July 1997.

OCCHIPINTI, E., and COLOMBINI, D., 1996a, Proposed concise index for the assessment of exposure to repetitive movements of the upper limbs. *La Medicina del Lavoro,* 87(6), 526–48 (in Italian).

OCCHIPINTI, E., and COLOMBINI, D., 1996b, Musculoskeletal disorders of the upper limbs due to mechanical overload: methods and criteria for the description and assessment of occupational exposure. *La Medicina del Lavoro*, 87(6), 491–525 (in Italian).

OCCHIPINTI, E., 1998, OCRA: A concise index for the assessment of exposure to repetitive movements of the upper limbs, *Ergonomics*, 41(9), 1290–1311.

OSHA, 1991, *Ergonomics Program Management Guidelines for Meatpacking Plants*. US Department of Labor, OSHA, 3123.

PUTZ-ANDERSON, V., 1988, *Cumulative Trauma Disorders. A Manual for Musculoskeletal Disease of the Upper Limbs* (London: Taylor & Francis).

ROHMERT, W., 1973, Problems in determining rest allowances. Part 1: Use of modern methods to evaluate stress and strain in static muscular work. *Applied Ergonomics*, 4, 91–95.

ROHMERT, W., BERG, K., BRUDER, R., and SCHAUB K., 1994, Force atlas. Kräfteatlas. Teil 1. *Datenauswertung statischer Aktionskräfte* (Bremerhaven: Wirtschaftdverlag NW).

SCHNEIDER, S., 1995, OSHA's Draft standard for prevention of work-related musculoskeletal disorders. *Appl. Occup. Environ. TNG*, 10(8), 665–74.

SILVERSTEIN, B.A., FINE, L.J., and ARMSTRONG, T.J., 1986, Hand wrist cumulative trauma disorders in industry. *Brit. J. Industrial Medicine*, 43, 779–84.

SILVERSTEIN, B.A., FINE, L.J., and ARMSTRONG, T.J., 1987, Occupational factors and carpal tunnel syndrome. *American Journal of Industrial Medicine*, 11, 343–58.

TANAKA, J. and McGLOTHLIN, J.D., 1993, A conceptual quantitative model for prevention of work related carpal tunnel syndrome. *Intern. J. Ind. Ergonomics*, 11, 181–93.

Victorian Occupational HSC (Australia), 1988, *Draft code of practice: Occupational Overuse Syndrome*.

WINKEL, J. and WESTGAARD, R., 1992, Occupational and individual risk factors for shoulder-neck complaints: Guidelines for the practitioner. 2. The scientific basis for the guide. *Intern. J. Ind. Ergonomics*, 10, 179–204.

APPENDICES

In appendices A and B, some risk assessment examples are shown in order to obtain exposure scores. They represent only a hypothesis of simple methods easily applicable in the field. In appendices C and D, information about the assessment of force, posture, and movement in easily applicable, international consensus standards are presented.

The appendices contain examples of practical methods previously used for measuring exposure and assessing risk. The methods presented in appendices A and B have proved useful in field studies and illustrate possible approaches to the analysis of work place risk factors. In addition, force capacity values included in the preliminary CEN standard (PrEN 1005-3) and the evaluation procedure for postures and movements used in ISO/DIS 11226 and prEN 1005-4 are presented in appendices C and D.

Note: The proposed methods available for work place risk assessment purposes are numerous and the examples presented in this appendix are not especially endorsed or recommended by the authors of the consensus document, nor by the IEA Technical Committee.

Appendix A: An example for calculating postural exposure scores (Colombini 1998)

In order to allow the technician to make an easier description and classification of posture effort (type of posture per time), Table 6 reports a form with an example of all the major items for each one of the four segments of the upper limb under consideration.

There are four operational phases in the form:

(a) A separate description of postures and/or movements by each joint: shoulder, elbow, wrist, hand (type of hold and finger movements), and by type of effort (static, dynamic).

(b) Static postures: observation of static postures close to extreme articular range, during the cycle/task time (A_3, C_3, D_3); observation of static postures in medium leverage angle held for prolonged period (A_4, C_4); observation of grip positions (D_1) during cycle/task time.

(c) Joint movements: presence of articular movements, close to the extreme joint excursions during the cycle/task time (A_1, B_1, C_1); repetitive articular movements, due to presence of same technical actions (independently of the articular range) for at least 50% of cycle time and subsequently of task time (A_2, B_2, C_2, D_2).

(d) Calculation (for each joint) of the postural involvement score within the cycle/task time summing the scores written in the square boxes, checked during the posture analysis.

The posture involvement score is attributed to each joint, taking into account that the presence of a significant effort is given by either of the two minimum scenarios, one for static postures and the other for movements, respectively.

For practical purposes, a significant cycle should be analyzed (preferably by a video) for each repetitive task.

The video could be reviewed in slow motion to describe and evaluate the effort of each joint segment, making a distinction between right and left side when the effort is asymmetrical.

A possible example to obtain a risk score for the lack of recovery periods is shown below.

Example

With the help of Table 6, analyze the work of an operator who picks up a handful of screws with the right hand, and for two-thirds of the cycle fits the screws into their holes, always using the same hand and holding his arms off the table. Cycle time is 15 sec; shift duration is 8 hours.

Begin by observing the shoulders.

The operator first takes the screws from a container (abduction/adduction more than 60°) and then keeps the arm in flexion, in a risk area (>60°) for 2/3 of cycle time.

In Table 6, under shoulder sign in A1: 1/3 (score 4), in A3: 2/3 (score 8), in A4 the continuously keeping the arm raised (score 4).

The posture score for the shoulder is 16: in this way critical movements, critical static postures and continuously arm-raised

Table 6: Analysis of upper limb postures as a function of time: a simplified model with an example (example 4)

TASK INVOLVEMENT ....

UPPER LIMB : o RIGHT   o LEFT

SHOULDER POSITIONS AND MOVEMENTS
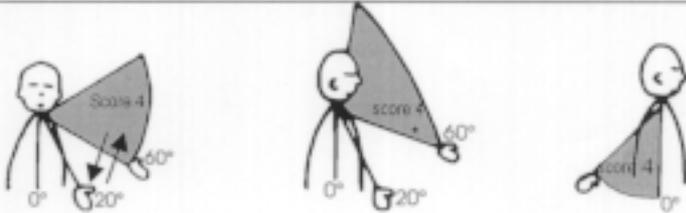
[A1] MOVEMENTS IN EXTREME ARTICULAR RANGE: THEY TAKE : [4] 1/3, [8] 2/3, [12] 3/3 of cycle/task time

[A2] LACK OF VARIATIONS:
performs work gestures of the same type involving the shoulder for at least 50% of the cycle/task time: [4]

[A3] KEEPS THE ARM RAISED (unsupported) by an angle more than 60° or in extension, for :
[4] 1/3, [8] 2/3, [12] 3/3 of cycle/task time.

[A4] KEEPS THE ARM RAISED (unsupported) continuously more than 20° or in extension for at least 50% of cycle task/time: [4]

RISK SCORE IN CYCLE: | 1, 6 |

ELBOW MOVEMENTS
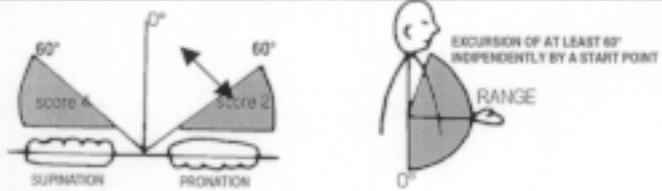
[B1] MOVEMENTS IN EXTREME ARTICULAR RANGE: [4] 1/3, [8] 2/3, [12] 3/3 supination
THEY TAKE : [2] 1/3, [4] 2/3, [6] 3/3 pronation } of cycle/task time
[2] 1/3, [4] 2/3, [6] 3/3 flexion

[B2] LACK OF VARIATION: Performs work actions of the same type involving the elbow for at least 50% of the cycle: [4]

RISK SCORE IN CYCLE: | 0, 8 |

WRIST POSITIONS AND MOVEMENTS

[C1] MOVEMENTS IN EXTREME ARTICULAR RANGE: [2] 1/3, [4] 2/3, [6] 3/3 r/u deviation
THEY TAKE : [4] 1/3, [8] 2/3, [12] 3/3 extension } of cycle/task time
[2] 1/3, [4] 2/3, [6] 3/3 flexion

[C2] LACK OF VARIATION: Performs work actions of the same type involving the wrist for at least 50% of the cycle/task time: [4]

[C3] STATIC POSTURES IN EXTREME ARTICULAR RANGE: THEY TAKE : [2] 1/3, [4] 2/3, [6] 3/3 of cycle/task time.

[C4] KEEPS THE WRIST FLEXED OR EXTENDED OR DEVIATED LATERALLY (ulna or radius)
continuously for at least 50% of cycle/task time: [4]

RISK SCORE IN CYCLE: | 0, 4 |

TYPE OF GRIP AND FINGER MOVEMENTS

[D1] TIME OF THE GRIP AND FINGER POSITION:

		1/3	2/3	3/3
[ ]	TIGH GRIP (1.5 CM)	[2] 1/3,	[4] 2/3,	[6] 3/3
[ ]	PINCH	[3] 1/3,	[6] 2/3,	[9] 3/3
[ ]	PALMAR GRIP	[4] 1/3,	[8] 2/3,	[12] 3/3
[ ]	HOOK GRIP	[4] 1/3,	[8] 2/3,	[12] 3/3
[ ]	KEYING	[4] 1/3,	[8] 2/3,	[12] 3/3
	....	[ ] 1/3,	[ ] 2/3,	[ ] 3/3
	....	[ ] 1/3,	[ ] 2/3,	[ ] 3/3
	....	[ ] 1/3,	[ ] 2/3,	[ ] 3/3

of cycle/task time

LACK OF VARIATIONS:
[D2] performs work actions of the same type involving the same finger(s) for at least 50% of the cycle: [4]
[D3] holds an object in a pinch or palmar (or hook) grip continuously for at least 50% of the cycle: [4]

RISK SCORE IN CYCLE: | 1, 3 |

positions are summed up considering time pattern during the cycle.

As for the elbow, the operator carries out pronation movements (and return from pronation) >60° for two-thirds of cycle time.

Under the elbow B1, enter two-thirds of cycle time for pronation movement (score 4), and in B2 enter "4" for movement stereotypy.

The elbow has an overall score of 8.

The wrist joint does light flexion (and return from the flexion) movements, but for gestures of the same type, and for two-thirds of cycle time. Under C2 fill in 4.

The overall score for the wrist is 4.

The hand is involved in concurrent precision grip (PINCH).

These gestures are always the same, and last for the whole of the cycle. Under D1 sign 3/3 (score 9); under D3 sign the box 4; the overall score the hand is 13.

When the cycle time is extremely short (e.g. shorter than 6 sec.), the stereotypy of the technical actions is always present.

Appendix B: An example for calculating lack of recovery period exposure scores (Colombini 1998)

A still empirical but effective way of performing such analysis is by examining individually the hours that make up the shift; for each hour it is necessary to verify whether repetitive tasks are carried out and whether there are adequate recovery periods (Colombini 1998).

According to the presence/absence of adequate recovery periods within each hour of the repetitive work under examination, each hour is considered as being either "risk-free", or "at risk" if there is a lack of adequate recovery periods.

The overall risk related to lack of recovery periods could be determined by the total number of hours of the shift in which recovery is insufficient.

A possible example to obtain a risk score for lack of recovery periods is shown in example 5 (Figure 4).

The risk due to a lack of recovery periods is classified with a score of 4. This expresses the number of hours in the shift in which the recovery is insufficient. In an eight-hour shift, with a lunch break but with no other pauses at all, the score will be 6; in fact, the hour of work followed by the lunch break, just as the last hour before the end of the shift, can be considered as risk-free, because they are followed by adequate recovery periods.

Appendix C: Recommended force limits for machinery operator (PrEN 1005–3 1996)

- Define relevant actions and force directions.
- Obtain distribution parameters (average and standard deviation) of the maximal isometric force for the relevant action in the general adult healthy European population.
- Decide if the machinery is intended for professional or domestic use.

In a working situation where a single repetitive task is carried out (task A), and where pauses are distributed as follows:
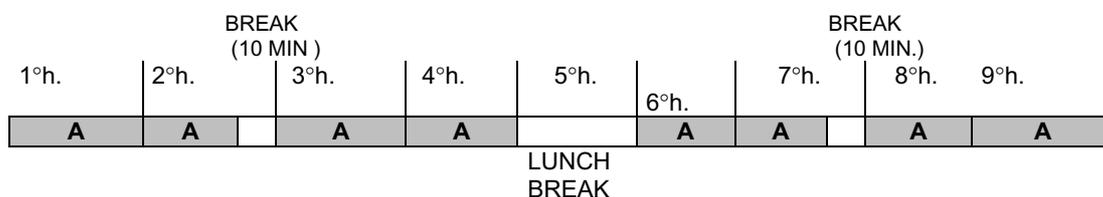
the following scheme (protocol) should be adopted:

1st HOUR	= 60 min. TASK (no rec.)	= RISK 1
2nd HOUR	= 50 min. TASK: 10 REC.	= RISK 0
3rd HOUR	= 60 min. TASK (no rec.)	= RISK 1
4th HOUR	= 60 min. TASK	
5th HOUR	= 60 min. REC.	= RISK 0
6th HOUR	= 60 min. TASK (no rec.)	= RISK 1
7th HOUR	= 50 min. TASK: 10 REC.	= RISK 0
8th HOUR	= 50 min. TASK (no rec.)	= RISK 1
9th HOUR	= 60 min. TASK + RECOVERY END OF SHIFT	= RISK 0

Figure 4. Protocol for work/rest schedule on repetitive tasks.

● Determine F_B, i.e. the 15th force percentile for professional use or the 1st percentile for domestic use.

Appendix D 1: Evaluation of working postures (ISO/DIS 11226 1998)

The holding time for upper arm elevation is evaluated using Figure 5.

It is recommended to provide adequate recovery time following the holding time for a certain upper arm elevation.

Appendix D 2: Evaluation of working postures in relation to machinery (CEN prEN 1005–4 1997)

Upper arm elevation

Step 1: refer to figure and table below

Step 2:

(a) Acceptable if there is full arm support; if there in no full arm support, acceptability depends on duration of the posture and period of recovery.

(b) Not acceptable if the machine may be used for long durations.

(c) Not acceptable if frequency ≥10 / minute and/or if the machine may be used for long durations.

The holding time for upper arm elevation is evaluated using the table below.

HOLDING TIME	ACCEPTABLE	NOT RECOMMENDED
> maximum acceptable holding time* ≤ maximum acceptable holding time*	X	X

● Taken from the figure below.
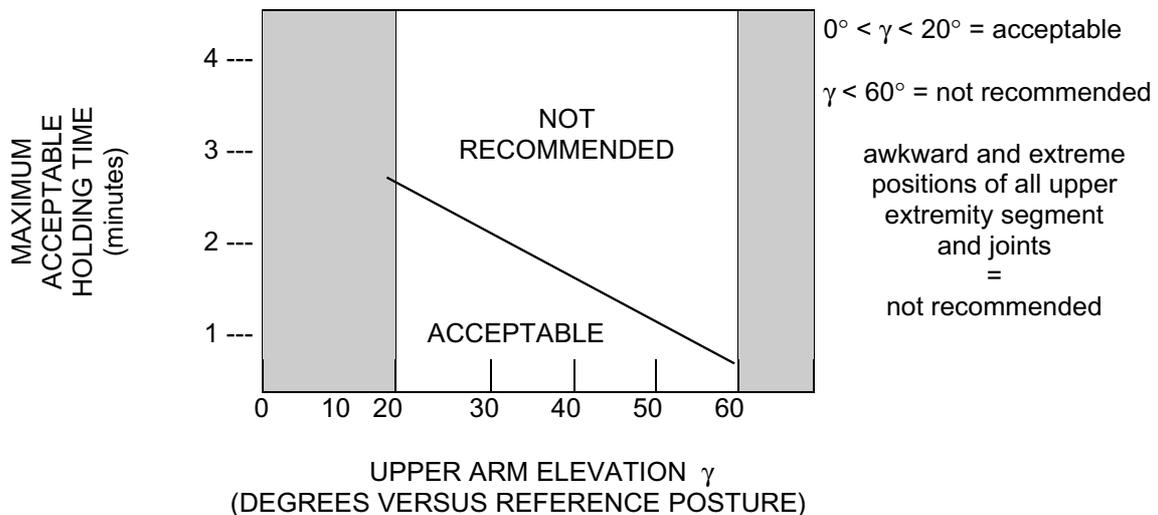
Figure 5. Maximum acceptable holding time.
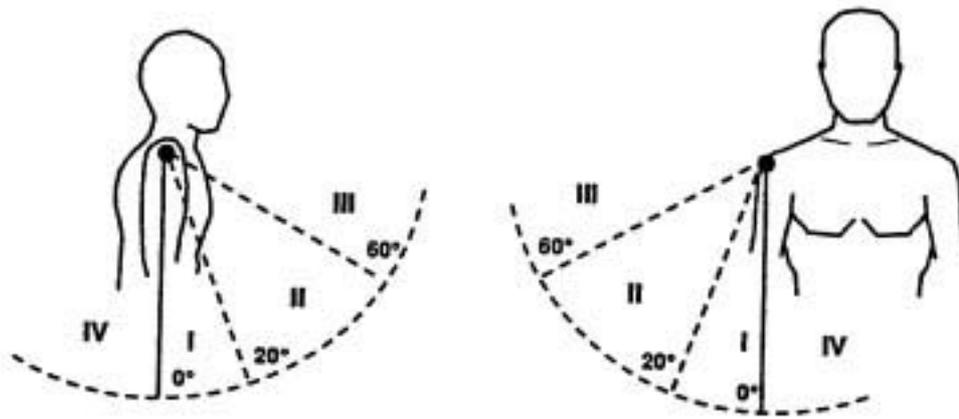
Upper arm elevation

Evaluation of upper arm elevation

	Static posture	Movement	
		low frequency (<2/minute)	high frequency (≥2/minute)
I*	acceptable	ACCEPTABLE	acceptable
II	conditionally acceptable (step 2A)	acceptable	conditionally acceptable (step 2C)
III	not acceptable	conditionally acceptable (step 2B)	not acceptable
IV	not acceptable	conditionally acceptable (step 2B)	not acceptable

Static posture and high frequency movements (≥ 2 / minute), awkward and extreme positions of all upper extremity segments and joints = not acceptable

* It is recommended to strive for working postures with the upper arms hanging down.

Figure 6.

Finnish Ergonomics Society

C.-H. Nygård

Tampere School of Public Health, University of Tampere, PO Box 607, 33101 Tampere, Finland

The Finnish Ergonomics Society (FES) was founded in 1985. Among the founders there were not only active ergonomists, but also institutions including the Finnish Institute of Occupational Health and the Center for Occupational Safety. The current number of members is around 230. The main objective of the FES is to promote the application of current knowledge of ergonomics in the design of safe and convenient, as well as productive and efficient human work systems. This includes improving machines and tools as well as work methods and the work environment. It is the adoption of work to man and of each man to his job. Participatory ergonomics is one of the most powerful processes currently used in Finland for improvements in work and environment — both in research and developments. The possibilities of ergonomics are widely admitted for the aging work force and the promotion of work ability. FES has the following main principals in the activities of the society:

- to place a special emphasis on the application of ergonomics in operation of both occupational health and safety specialists for the improvement of working conditions and environment;
- to advise on netted collaboration in providing information, training and education in the field of Ergonomics;
- to publish a newsletter for members and to hold a World-Wide Web homepage to guarantee the maintenance of the ergonomics information network;
- to support a national CREE committee for certifying European Ergonomists;
- to participate in preparing ergonomic standards;
- to organize an annual national meeting and, every 4 years, a Nordic conference on ergonomics;
- to cooperate in organizing with universities postdoctoral courses and congresses; and
- to work, in cooperation with other organizations, to create a basis for long-term development in education and research in ergonomics in the Baltic states (especially Estonia).

Fundamental Concepts of Human Factors

D. Meister

1111 Willow Avenue, San Diego, California 92109, USA

This article examines the concepts underlying the study of Human Factors Ergonomics (HFE), with particular emphasis on what makes HFE a distinctive discipline.

1. INTRODUCTION

Although the discipline has existed only since the end of World War II, its underlying concepts derive in part from its predecessor, psychology and in part from philosophical principles that go back to the early 19th century and even earlier. These concepts have developed a special meaning because of their application to technology, but it is important for HFE professionals to recognize that they are part of a much older scientific tradition.

2. HFE AS A DISTINCTIVE DISCIPLINE

All scientific disciplines have one critical factor, the need to *know*. The subject-matter of that knowledge determines the nature of the discipline. Physics examines the structure of the atom; biology, the nature of the cell; psychology describes how the human functions. HFE, which developed out of psychology, also seeks to discover the nature of human processes, but in a very special context: the technology the human creates to control the external world.

Many HFE professionals think that because of the close relationship between HFE and psychology, HFE is merely a specialized branch of psychology. Others, including the author of this section, feel that HFE is a distinctive discipline, despite its earlier and continuing ties to psychology. HFE represents humans in interaction with systems in the world around them. One finds little or no concern for such systems in psychology.

Some may feel that it makes little difference what one calls one's work. Labels do, however, make a difference, particularly if a theoretical viewpoint and a methodology are associated with the labels. If one thinks in the idiom of another discipline like psychology, the resultant concepts, and especially the way in which research is performed, will be those of that other discipline (Meister 1999). Although all comparisons are invidious, it is necessary to establish HFE distinctiveness simply to define the scope of its work. As an illustration, HFE is not interested primarily in physiological processes, although this is one of the factors affecting the human's interaction with technology. It is only when these physiological processes affect the way in which the human manipulates technology, that HFE examines those processes. The same is true for all other disciplines with which HFE has relationships.

The ties with psychology, physiology, safety, etc. still persist, just as psychology's ties with physics, chemistry and biology also persist, since the human is also and forever a physical, chemical and biological entity. But if one considers psychology an independent discipline, so is HFE. HFE professionals must always be aware that they are creating new knowledge. If they lack this awareness, they will create a knowledge that, whatever its other qualities, is essentially irrelevant to HFE.

3. HUMAN–TECHNOLOGY RELATIONSHIP

The underlying assumption of HFE is that there is a special relationship between technology and the human. It is impossible to conceive of technology without a human to design and exercise that technology. There is an inherent dynamic within the relationship. If one thinks of that relationship as an equation, behavioral inadequacies will reduce technological efficiency, and technological inadequacies produced by poor design will reduce the effectiveness of human performance. The human dominates the relationship, because technology occurs only when some human molds that technology into a form compatible with human requirements. Technological inadequacies are also produced by behavioral inadequacies. For example, a lack of technician awareness, a conceptual error, caused the Chernobyl nuclear catastrophe.

The proceeding is admittedly an extreme example, but human performance inadequacies occur constantly in situations that fortunately do not lead to catastrophes. Technology can also damage humans severely; a well-known example is muscular trauma which afflicts thousands.

If poor design can cause inadequate human and system performance, then it is logical to assume that more adequate design will produce more effective human and system performance. This is the rationale for HFE; if one did not believe this, it is unlikely that there would be an HFE at all.

4. STIMULUS–ORGANISM RESPONSE (S–O–R) PARADIGM

S-O-R, which was derived from psychology, is a fundamental paradigm. It describes the human as a reactive animal, who perceives stimuli, events and phenomena, assesses these in terms of actions required, and responds appropriately (presumably so, or the human animal would probably have disappeared in earlier ages).

From a more immediate standpoint, however, the significance of S-O-R is that to understand and deal with complex phenomena, it is necessary to *decompose* these into S-O-R elements. Phenomena may be too complex to deal with as they first appear, but decomposition into more molecular elements permits a more meaningful response. The decomposition of X into S-O-R is the first and most important expression of HFE analysis.

This is of course reductionism and one sees it most emphatically in HFE design and in the various behavioral analyses that aid that design. At the same time, this reductionism is balanced by a building-up (composition) process, which in design takes molecular elements like individual software instructions and builds them into a complex computer system. Composition also allows us to understand the relationship between these elements and the whole.

5. SYSTEM CONCEPT

The system concept, which derives from principles developed by Hegel, the 19th-century philosopher, and even more antique philosophers upon which he built (van Gigch 1974) serves as a principle underlying the development and understanding of human–machine systems. The essence of a system is *organization*; a system is created when formerly independent entities are organized in relation to each other. Inherent in the notion of

system is a *purpose* that in technological terms is transformed into the concept of *goal*; it is the goal that organizes the combination and interrelationships of functions and tasks to achieve that goal. The goal is, however, essentially an abstraction; it can exert its influence only by being itself transformed by the behavioral specialist into the *mission* that makes functions and tasks dynamic. It is obvious that we have here a series of conceptual transformations: from purpose to goal to mission to functions and tasks.

Inherent in the system concept is a hierarchy of *complexity*. The development of components into modules like a circuit board and from a combination of circuits to the individual equipment, and from the combination of individual equipments to subsystems and then full scale systems — all of these involve an increase of complexity. All systems, physical, chemical, biological, psychological, manifest varying degrees of complexity. In physical terms the number of components interacting with each other defines complexity. In behavioral terms complexity is conceptualized in information processing dimensions: as the number of changing system states that are displayed to the human as stimuli. Our knowledge of complexity is unfortunately slight, because the topic has not been studied to any extent, but we hypothesize that it is responsible both directly and indirectly for adverse performance effects such as difficulty, workload or stress. Since these conditions are deleterious to the human and to system performance, HFE works to reduce or compensate for complexity.

The system concept has a number of very practical consequences:

1. All behavioral analyses begin with an examination of the goal/mission and decomposition of the functional elements linked by the mission. The same is true of measurement; the performance of system elements and their interrelationships must always be related to the overall system goal.

2. The system concept has also meant the necessity of considering in system design all the elements that may affect goal accomplishment. If a possible system performance-shaping element can even be conceptualized, its possible effect on the goal/mission must be considered, even if, after that consideration, the element or factor is rejected as being unimportant.

3. The concept that human–machine systems are built up progressively from elemental to more complex structures means that different attributes must be considered at different levels of complexity. For example, with greater complexity the designer will be able to consider attributes like transparency (the display of internal system operations to assist in failure diagnosis) and flexibility (the availability of different modes of exercising the system under different conditions).

4. Because, in a Gestalt sense, the system is more than the sum of its parts, it is possible that at some level of complexity a degree of *uncertainty* enters the system, unknown to and uncontrolled by system developers. Because the design process involves transformations (to be discussed below), any transformation involves a certain degree of uncertainty, because the designer must cross domain boundaries. At sufficiently high levels of complexity this uncertainty may create what has been called "emergents," which are system properties that were not anticipated and which in

consequence may have unexpected and potentially negative effects on system functioning.

5. Complex systems and equipments exert control on their human operators because of the way in which the devices have been designed; the operator cannot make a device do what it was not designed to do. Extremely large ergonomic systems like an automobile factory require that its personnel perform in specific designated ways designed to permit accomplishment of the system goal. Human–machine systems are built into *organizations* which control these systems; the study of these organizations and how they affect subordinate system units is a speciality within HFE which is called macro-ergonomics (Hendrick 1997). In a nutshell, larger systems control smaller systems, into which these smaller systems are embedded. System developers may not think that what they are constructing have these organizational effects, but those who have studied this aspect of system functioning report that this control aspect does have significant effects on the performance of individual workers and their equipments.

6. TRANSFORMATIONS

HFE has a continuing problem that might be considered to be the essence of the discipline. This is the necessity to transcend domain boundaries when one attempts to transform principles and data (the behavioral part of the human–technology equation) into human–machine systems (the physical domain). HFE is not alone in being involved with such transformations: an example from biochemistry is the transformation of chemical changes at the synapse into electrical neural impulses. Transformations are a critical function of system design when a behavioral function (for example, to start an engine) is transformed by design into a switch for turning that engine on.

Would that all such transformations were that simple! As system complexity increases, the transformation process becomes more uncertain because the purpose of the transformation may be only partially achieved. Suppose that one has to monitor a complex process control system that is influenced by seven variables, each of which interacts with and may affect every other. The designer, in a more traditional mode, might provide the monitor with 7 discrete gauges; which would require the monitor to estimate the effect of variations in any single variable on each of the other six variables — a daunting cognitive problem. In a more advanced design the designer might provide a single multivariate display in which the computer controlling the process would present the interactive effects of the individual variables in terms of changes in a single display. Even so, the behavioral (cognitive) transformation process is significantly more complex than turning on an engine or monitoring a single display, because it requires the monitor to develop a mental concept based on what the display presents; this may be difficult and the way in which the information contained in the display has been packaged may determine how well the monitor can interpret that information. The display designer must not only conceptualize how to transform the information into a display, but also to imagine how the monitor will interpret that display.

What we see in the example above is the emergence not of a new element, but one which in the earlier days of behavioristic psychology was largely ignored — the mind (the modern term

for mind is something like *cognitive ergonomics*). Initial behavioristic concepts attempted to deal with the mind as a sort of "black box" which was covert and non-observable (observability was *a*, if not *the*, criterion of behaviorism). Objective stimuli and responses could be observed and measured, and since those stimuli and responses could only be received and transmitted through the agency of the black box, the box could be defined by those stimuli and responses.

This behavioristic orientation was perhaps acceptable until modern technology (i.e. the computer) forced HFE to attend to the black box directly, not by deduction. That is because the observable operations of the computer speak primarily to the mind. From the operator's standpoint the computer is little more than a display presenting information and the information displayed requires cognitive functions. Because proceduralized control of step by step operations have been taken over by the computer, only the mind is left as a controlling factor.

7. HFE FUNCTIONS

The primary functions of HFE are (1) to study, explain and predict human performance in the operational environment (OE); and (2) to assist in the development of SEP, which will be compatible with human needs and desires.

The ultimate referent for HFE is the OE, which is where workers perform their technological tasks under routine (i.e. non-laboratory) conditions. Because the OE presents difficulties in controlling phenomena, most HFE studies are conducted in the laboratory (Meister 1989, 1999). The laboratory is the antithesis of the OE, since it abstracts and simplifies that factors influencing the performance being studied (the only exception to this is if the experimenter has a sophisticated simulator that permits the researcher to reproduce OE conditions). Because in most cases laboratory conditions vary widely from those of the OE, it is necessary to validate experimental findings. Unfortunately, very little in HFE is validated in the OE.

Prediction is an inherent function of a science and so it is necessary for HFE to make predictions (preferably in quantitative form) of the kinds of performance one would expect to find in the OE. Such prediction would not only validate laboratory findings, if predictions from those findings were found to be correct, it would extend some degree of control (because of the knowledge gained) over operational system performance. Prediction of laboratory findings to OE situations has almost never been tried in HFE, although there is a speciality called "human reliability" which has been studied, although not under the aegis of HFE. Some HFE experimentalists would say that by developing a hypothesis prior to the research and then conducting the study to determine the accuracy of the hypothesis, this is a form of prediction. However, since the OE is the ultimate referent for HFE findings, prediction must describe OE performance or the prediction is essentially irrelevant.

If the prediction function is almost never performed in HFE, the second HFE function of assisting in the development of human–machine systems is flourishing. This function is the one that distinguishes HFE from all other behavioral disciplines, although, in working within an engineering framework, HFE can also be viewed as a quasi-engineering discipline.

Inadequacies in the study of HFE (the first function) hamper the effectiveness with which the second function can be accomplished. HFE professionals need to take an expanded view of HFE into the 21st century, if the discipline is to achieve its many potentialities. That expanded view must include a greater recognition of the role of concepts and theory in performing HFE functions.

It is of course possible to perform daily HFE tasks without tying them to fundamental concepts, but one cannot go much beyond these quotidian tasks without some awareness of the concepts involved in them.

REFERENCES

HENDRICK, H.W., 1997, Organizational design and macro-ergonomics. In G. Salvendy (ed.), *Handbook of Human Factors and Ergonomics* (New York: Wiley), 594–636.

MEISTER, D., 1989, *Conceptual Aspects of Human Factors* (Baltimore: Johns Hopkins University Press).

MEISTER, D., 1999, *The History of Human Factors Ergonomics* (Mahway: Erlbaum).

VAN GIGCH, J.F., 1974, *Applied General System Theory* (New York: Harper & Row).

Germany: Gesellschaft für Arbeitswissenschaft

A. Seeber

Universität Dortmund, Institut für Arbeitsphysiologie, Ardeystr. 67, D-44139 Dortmund, Germany

The Gesellschaft für Arbeitswissenschaft (GfA) is the Human Factors and Ergonomics Society for the German-speaking area. Predominantly it covers members from Germany, Austria and Switzerland, but single members come from other European countries. The GfA was founded in 1953. The roots of ergonomics in Germany can be seen in the Max-Planck-Institute of Occupational Physiology, founded in Berlin in 1912 and transferred later to Dortmund where now is the office of GfA.

The GfA represents a broad range of research and application of ergonomics in occupational health and safety, product design, work and organization design, as well as in organization development and labor policy. The target groups of GfA are all "ergonomically acting or interested people." On the one hand, it can be those taking on ergonomic tasks in practice, such as analysis, evaluation and design of work; on the other hand, it can also be those involved in the design of products and processes. Finally, all those who take responsibility for the realization (e.g. managers) have to be included.

Which principles and activities has the GfA? The specific task of the GfA is in the promotion of scientific and professional interests of Ergonomics/Human Factors. The GfA sees itself especially as an active forum for the dialogue between science and practice and all interested social groups.

The GfA develops and guarantees professional standards not only by participating in standardization or as a German partner for the certification as an Euro-Ergonomist, but also by assessment procedures for journal and congress contributions. As far as possible and as required, it represents the interests of its members in the professional field.

The fundamental principles in achieving this goals are:

- The (working) society is subject to continuous change, which requires an appropriate development of Ergonomics/Human Factors. In that case goal conflicts and tensions will occur and be accepted as normal.
- Ergonomic research and design is always aware of the uniqueness of a human being.
- Ergonomic-acting aims at a prevented design of work, technique and organization, which is oriented towards human and economic criteria.
- Ergonomic solutions are subject to a concept, which is useful to all stakeholders (e.g. employees, management, shareholders, society).
- The persons concerned by analysis, evaluation and design will be — wherever possible — involved in these processes.
- Ergonomics/Human Factors also deals with analysis, evaluation and design of non-gainful employments (f. e. housework, voluntary work, community work, etc.)

The GfA is responsible to its members and it feels obliged to the social goal and it makes contributions — wherever possible — to the maintenance, creation and (where appropriate) distribution of human, economically and environmentally compatible work.

The GfA wants to contribute to the dissemination of these contents not only among its members, but also for as many people as possible. The GfA supports the transfer of ergonomic knowledge by the organization of congresses, workshops and publications.

Twice a year, in the Spring (March) and Fall (September), the GfA organizes conferences for members and other participants. For example, the topic of the 44th Spring Conference 1998 was "Communication and Cooperation;" that for the 45th Spring Conference was "Systems to manage Work Protection — Chance or Risk?" Usually, ~300 participants visit the conferences where they can follow ~100 presentations. Additionally, workshops and tutorials are offered. The Fall conferences are more closed in their issues. The presentations characterize only one area of interests. For example, the Fall Conferences of the mentioned years dealt with "Man–Machine Interfaces" and "Future of Labour in Europe: Management of Company Changes." The Fall Conferences have differing numbers of visitors from ~100 to 400 participants.

The *Zeitschrift für Arbeitswissenschaft* is the scientific journal of the GfA. It has four issues per year. There is also a GfA Newsletter sent to the members of the GfA twice a year. Additionally, the GfA is represented with a web-site: www.gfa-online.de

1. HOW DOES THE GFA WORK?

Who can become a member of the GfA? As mentioned, those persons who are interested in Ergonomics can get the membership of the GfA. About 50% work at universities and similar educational institutions. About 25% work in enterprise. The rest covers retired members. In 1999 the GfA had 685 members.

Important decisions can be done only by the General Assembly of the members which takes part once a year during the annual Spring Conference. An Executive Board with eight members leads the society. Management of the daily tasks is done by the GfA-Office with a Secretary and the elected Secretary-General. The President, the two Vice-Presidents and the other members of the Executive Board are elected for 2 years. They can stay for 6 years on the Board, but the President has to change every 2 years.

Technical groups work in different fields. At the moment just two exist: "Management of Innovative Working Time Systems" and "Future of the Working Society." Besides these working groups, there is a substructure that includes the "Lecturers in Ergonomics," which covers the leading staff from the technical institutes and universities that discuss all the topics for training in ergonomics. There are no regional structures among the members of GfA.

Greece: Hellenic Ergonomics Society (HES)

N. Marmaras and T. Koukoulaki

Hellenic Ergonomics Society, Athens, Greece

1. INTRODUCTION

The Hellenic Ergonomics Society (HES) was founded in July 1988 and became a federated member of IEA in August 1994. It is the association of Greek ergonomists and is located in Athens. According to the society's bylaws, its goal is the development, promotion and propagation of the science of ergonomics through the cooperation, exchange of knowledge, methods and experience between ergonomists as well as other scientists whose fields of specialty include the research and understanding of the participation of the human factor to the design, production and use of machines, displays and production systems.

The membership policy of HES is strict regarding the qualifications of its members so that it maintains a scientific society with members that are both highly qualified and active in the field of ergonomics. According to the society's bylaws, full membership can be attained by any person of Greek nationality who holds a university degree in a field related to ergonomics or has at least 3 years experience in ergonomics.

The HES has 33 active full members. The majority (60%) is engineers; the rest are work physiologists, occupational physicians and other scientists. Among them three are academicians, four hold a PhD in Ergonomics, two hold a PhD in Work Physiology and the rest either hold a Masters degree or have taken a postgraduate course in Ergonomics. Four members are qualified European Ergonomists. The main domains of interest and work activities of the HES members are Human–Computer Interaction, Cognitive Ergonomics, Occupational Health and Safety, Ergonomics for the Disabled, Products and Workplaces Design, and Work Organization.

2. MILESTONES

Some important activities of HES throughout its 10-year existence are:

- A 2-year postgraduate training course in Ergonomics funded by the European Commission (1989–91). The course was jointly organized with the Greek Productivity Center. Two European Ergonomics Institutes collaborated in the course: the HUSAT Institute of Loughborough University, UK, and the IAO institute of Germany. Fifteen trainees holding a university degree in disciplines related to ergonomics (engineers, physicians, psychologists, etc.) graduated from this course.
- The society's participation in the European Year for Occupational Health and Safety (1992). A brochure explaining the role of ergonomics in the prevention of occupational accidents and diseases, as well as in the enhancement of productivity, has been published.
- Organization of two seminars on human–computer interaction and environmental factors (1993).
- Participation in the TE 59 "Health & Safety" Technical Committee of the Greek Standardization Organization providing consultation in ergonomics issues.
- Organization of the 1st National Ergonomics Conference entitled "Applications of Ergonomics in Greece" (Athens, November 1997). The conference was organized jointly with the Technical Chamber in Greece, the main body of engineers in the country (> 70 000 members). Twenty ergonomic studies and interventions carried out in Greece were presented, which covered four generic topics: ergonomic evaluation and prescriptions of workplaces, software ergonomics, ergonomic design, and ergonomics in occupational health and safety. The conference was attended by > 130 participants, which is a very satisfactory number for Greece, and enjoyed wide publicity.
- Participation in the national dialogue on the arrangement of working time and work schedules by publishing and distributing a related article (1998).

The HES publishes a tri-monthly newsletter, "Ergonomic Issues." It is distributed to > 300 professionals interested in Ergonomics.

History of Human Factors in United States

D. Meister
1111 Willow Avenue, San Diego, California 92109, USA

Historical writings can be chronological and/or cultural and/or intellectual. The history of a discipline has elements of all of these (Meister 1999), but this discussion of Human Factors (HF) and what has ultimately been termed Human Factors Ergonomics (HFE) will, because of space limitations, be confined largely to the chronological.

1. ANTECEDENTS TO HF

Technology, some of it quite sophisticated, goes back to ancient times (James and Thorpe 1994), but the antecedents of modern HFE can be traced to the pre-World War I studies in scientific management of Taylor (1919) and the work of the Gilbreths, which eventually led to the time and motion components of industrial engineering (see Meister and O'Brien 1996, Moroney 1995 for a more comprehensive description of the early work).

2. WORLD WAR I

The inception of World War I stimulated, as war always does, the development of much more sophisticated equipment. For example, the primitive flying machines of the Bleriot type were modernized into Spad, Neuport and Fokker fighters. The war also saw the development of the first tanks. Because of the inability of many men to operate these machines, particularly aircraft, much attention was paid to the development of selection tests, and this led in the USA to the recruitment of psychologists to develop and administer these tests. This in turn led to the establishments of aero-medical laboratories, which were continued after World War I and were used as a model for the utilization of experimental psychologists in World War II. Because of the need to expand the number of aviators, the Council of the American Psychological Association established a Committee on Psychological Problems of Aviation, which in November 1918 became a subcommittee of the National Research Council. Twenty-three mental and physiological tests were evaluated by trying them out on Army Aviation cadets.

3. BETWEEN THE WARS

The period between the end of World War I and the start of World War II was largely one of gestation, and with relatively few outstanding accomplishments. Many studies of driver behavior were conducted because of the increasing popularity of the automobile. The research interest in automotive themes has been maintained to the present, with the number of papers published by the Human Factors Society on this topic second only to the number of aviation psychology papers. By the end of World War I two aeronautical laboratories had been established, one at Brooks Air Force Base, Texas, the other at Wright Field outside Dayton, Ohio.

Toward the end of the war and following the armistice in 1918 many tests were given to aviators of the Army Expeditionary Force in Europe to determine the characteristics that differentiated successful from unsuccessful pilots. (A similar effort was made, with the same lack of success, following World War II.) Early aeronautical work explored human and machine performance thresholds at environmental extremes; for example, in 1935 an altitude record of 72 000 feet was set in a balloon. Initial work on anthropometry and its effect on aircraft design and crew performance was begun. Link, a pioneer inventor, developed the first simulator as an amusement device and in 1934 the Army Air Corps purchased its first flight simulator (Dempsey 1985).

During this period also noteworthy research was performed at the Hawthorne plant of the Western Electric Company from 1924 to 1933 to study the effects of illumination on worker productivity. The mere knowledge that they were experimental subjects induced all the workers to exert increased effort, a result now known as the "Hawthorne effect." This suggested that motivational factors could significantly influence human technological performance.

4. WORLD WAR II

In 1939, in anticipation of the coming war and following traditions developed in World War I, the Army established a Personnel Testing Section; the National Research Council created an Emergency Committee on Psychology, whose focus was on personnel testing and selection. This was followed in 1941 by the creation of the Army Air Force Aviation Psychology Program directed by John Flanagan (the creator of the "critical incident" method).

So much was very reminiscent of World War I and, if it had not been accompanied by other activities, would be of little interest to HFE history. The war saw an exponential leap in sonar, radar and high-performance aircraft; the list could go on.

Because this was total war, involving great masses of men and women, it was no longer possible to adopt the Tayloristic principle of selecting a few special individuals to match a pre-existent job. The physical characteristics of the equipment now had to be designed to take advantage of human capabilities and to avoid the negative effects of human limitations.

An outstanding example of the kind of work that was done is the now classic study of Fitts and Jones (1947), who studied the most effective configuration of control knobs for use in developing aircraft cockpits. Two points relative to this example: the system units studied were at the component level; and the researchers who performed the study were experimental psychologists who adapted their laboratory techniques to applied problems.

Early studies of signal discrimination were directed at auditory capabilities of sonar; similar research was performed to determine the visual capabilities needed to detect targets on radar. The aim was to make controls and displays "easier" for operators to perform more efficiently. In cataloging these areas of research it is necessary to point out that they required more than "pure" research; if equipment was developed and/or evaluated as a result of this research, it forced psychologists to work closely with design engineers to make practical use of the HF research. Slowly, but ineluctably, as a result of the enforced intimacy with engineers, applied experimental psychology (the title of the first text on the new discipline, Chapanis *et al.* 1949) was transitioning to HF.

Immediately after the war the military attempted to summarize what had been learned from research performed

during the war. The Army Air Force published 19 volumes that not only emphasized personnel selection and testing, but also contained studies of apparatus tests, psychological research on pilot training, and particularly one on psychology research on equipment design. A second significant publication on what became HF dealt with human factors in undersea warfare.

Why did the efflorescence of what later came to be termed HF and then HFE occur in World War II and not in the earlier war? It was a matter of "critical mass": US participation in World War I lasted only a year and a half; aircraft had been developing since 1903; the tank was put into use only in 1916. The "window of opportunity" for HF lasted only 2 or 3 years, after which the US fell back into its customary somnolence. However, the example of World War I use of psychologists and the research in aero-medical laboratories between the wars provided a pattern for World War II. Moreover, the rise of the "Cold War" between the West and the Soviet Union gave HF an added push. It is also possible that the exponential march of technology — the transistor, the jet engine, the computer, etc. — made a science dealing with human–technology relationships necessary.

5. POST–WORLD WAR II (THE MODERN ERA)

This covers ~20 years from 1945 to 1965. It includes the activities of those who worked in the military in the war and the first generation that followed them.

The beginning of the Cold War fueled a major expansion of Department of Defense (DOD) -supported research laboratories. The immediate post-war environment was particularly hospitable to government-supported research. Laboratories established during the war expanded, e.g. the University of California Division of War Research established a laboratory in San Diego, California. This became the US Navy Electronics Laboratory, which then evolved into the Naval Ocean Systems Center and subsequent incarnations. Each of the services either developed human performance research laboratories during the war or shortly thereafter.

Almost all HF research during and immediately following 1945 was military-sponsored. Universities were granted large sums to conduct basic and applied research (e.g. the Laboratory of Aviation Psychology at Ohio State University). "Think tanks," like the System Development Corporation in Los Angeles and the RAND Corporation that split off from it, were established by the military. Whereas during the war research had concentrated on smaller equipment components like individual controls and displays, the new studies performed by the laboratories embraced larger equipment units, such as an entire workstation or an entire system.

Some of the major psychologists in World War II continued their work. Paul Fitts (considered *the* Founding Father of HF) remained as chief of the Psychology Branch of the Aero Medical Laboratory until 1949. The Air Force Personnel and Training Research Center was built into a corporation-sized organization employing hundreds of specialists in Texas and Colorado.

At the same time opportunities opened up in "civilian" industry (the quotation marks mean that, although industry was in civilian hands, it was supported and dominated by the military). Large companies in aviation like North American, McDonnell-Douglas, Martin Marietta, Boeing and Grumman established HF groups as part of their engineering departments.

The introduction of HF to industry represented a major change in HF. HF was now no longer completely or primarily a research-oriented discipline. The interaction between HF researchers and designers that was fostered in World War II now expanded to HF groups which became integral elements of the system design team. What had formerly been the domain of those who performed "basic" research now had to incorporate the application of their work to the development of physical systems. Even when a formal laboratory was not established, the government, through agencies like the Human Engineering Division at Wright-Patterson AFB or the Army's Behavioral Sciences Research Laboratory, let contracts for human performance research that were awarded to departments of psychology and industrial engineering in universities up and down the country.

To bid on these contracts private companies were formed, like the American Institute for Research under the directorship of John Flanagan who had headed the Army Aviation program; these employed numbers of HF professionals.

All these activities expanded the number of HFE professionals from a very small cadre during the war to at least 5000 professionals (at present), almost all of whom have advanced degrees (MA/MS and PhD). The discipline drew into itself not only psychologists, but also those with training in industrial and other forms of engineering, those with a physiological or safety background, etc. The connection with engineering that had only been vaguely foreseen in the war was now firm and exerted its own pressures on the discipline.

6. POST–MODERNISM (1965 TO PRESENT)

The post-modern period has seen a maturation of the discipline. The Human Factors Society, which was established in Tulsa, Oklahoma, in 1957 with a membership of 60 now has ~5000 members. The number of universities offering graduate programs in what has become HFE has increased significantly. Technology has shunted the discipline in new directions. The development of the computer and the tremendous expansion of computer applications to technology have created a new specialty field for HFE. Empirical HF research in software did not become significant until after 1970, when the personal computer (PC) was developed. Use of the PC by the general public brought with it behavioral problems that stimulated a great deal of research in the development particularly of graphics. Because this was directed at information processing and cognitive capabilities, another new field has recently arisen, cognitive ergonomics, with considerable emphasis on highly sophisticated multivariate displays. The effect of increasing automatization on system performance also received attention, particularly in the field of aviation (the "glass cockpit") with its integrated computerized graphic displays. Organizational psychology has received a new breath of life in the form of macro-ergonomics. Along with this, a traditionally system-oriented HFE has now recognized to some extent the importance of industrialized ergonomics

7. OTHER COUNTRIES

Although much HFE activity developed under American auspices, there were corresponding developments in other countries. HF in the UK proceeded in a manner parallel to that of the USA but

on a much smaller scale. A great deal of cross-fertilization between US and UK ergonomics occurred.

In Western Europe emphasis was very largely on what is called in the US "industrial ergonomics (IE)," which has tended to concentrate more of its attention on worker satisfaction, muscular trauma, biomechanics, etc., topics that have a closer tie with macro-ergonomics that with the more traditional HFE interests in system development. The system development interest is gradually being melded with IE, as exemplified perhaps by the fact that the Human Factors Society (the primary society in US representing HF professionals) changed its name in 1993 to The Human Factors and Ergonomics Society.

The Western orientation to HFE (represented largely by US and UK publications) has influenced HFE activity in Canada and South America, Japan and Australia, as well as Western Europe. HFE in the former Soviet Union took a somewhat idiosyncratic course. The antecedents of modern Russian HFE go back to the late 19th century and to Pavlov, Anokhin and Bernshtein (who are unfortunately not well known outside their native country). The Russian revolution had a devastating effect on Soviet psychologists who were considered by the government to have "bourgeois" and anti-Marxian tendencies. As a result, in the late 1930s laboratories devoted to psychological work were closed and many of their practitioners were exiled or killed. Not until Stalin's death was there a gradual recrudescence, and presently there is a small cadre (perhaps 300?) of HFE professionals functioning. Unfortunately information on what they are doing is not readily available.

American and European publications were allowed to enter Russia, but, to protect Russians against ideological accusations, the Western terminology of the concepts was often changed, although the basic outlines of the concepts remained much the same. There are, however, discernible differences in the Russian orientation, which are described in Meister (1999).

8. SUMMARY

One cannot review the past without anticipating the future. Old problems still remain: precisely what is HFE? What is it supposed to do? Many great challenges face HFE in the 21st century, e.g. the need to develop methods of predicting technological human performance quantitatively; the need to develop acceptable HFE theory and usable human performance models. All of these give HFE professionals much opportunity for further growth.

REFERENCES

CHAPANIS, A., GARNER, W.R. and MORGAN, C.T., 1949, *Applied Experimental Psychology* (New York: Wiley).

DEMPSEY, C.A., 1985, *Fifty Years Research on Man in Flight* (Dayton: Wright-Patterson AFB).

FITTS, P.M. and JONES, R.E., 1947, Psychological aspects of instrument display. Analysis of 270 "pilot error" experiences in reading and interpreting aircraft instruments. Report TSEAA-694-1-12A (Dayton: Aero-medical Laboratory, Air Material Command).

JAMES, P. and THORPE, N., 1994, *Ancient Inventions* (New York: Ballentine).

MEISTER, D., 1999, *The History of Human Factors Ergonomics* (Mahwaw: Erlbaum).

MEISTER, D. and O'BRIEN, T.G., 1996, The history of human factors testing and evaluation. In J. Charlton and T.G. O'Brien (eds), *Handbook of Human Factors Testing and Evaluation* (Mahwah: Erlbaum), 3–11.

MORONEY, W.F., 1995, The evolution of human engineering: a selected review. In B. Weimer (ed.), *Research Techniques in Human Engineering* (Englewood Cliffs: Prentice-Hall), 1–19.

TAYLOR, R.W., 1919, *Principles of Scientific Management* (New York: Harper).

A History of Human Factors/Ergonomics in Power Systems

S. O. Parsons†, J. M. O'Hara‡ and J. L. Seminara§

†Parsons and Associates, 19740 Via Escuela Drive, Saratoga, CA 95070-4445, USA

‡Brookhaven National Laboratory, Department of Advanced Technology, Building 130, Upton, NY 11973, USA

§1235 Via Huerta, Los Altos, CA 94024, USA

1. INTRODUCTION

A modern large electrical power plant, powered by nuclear energy, fossil fuels hydro, etc., is a complex human–machine system that controls a thermodynamic process used to produce electrical power. The machine side of the system is a sophisticated arrangement of hardware and software components that are highly reliable, redundant and have a high degree of interconnectedness. The human side of the human–machine system is really a large socio-technological organization with management, engineering, maintenance, operations, and training personnel. Operations personnel interact with the plant through a wide variety of interfaces included in the main control room and various remote monitoring and control stations. Their actions are supported by plant operating procedures and a variety of human–system interfaces (HSI) such as alarm systems, information display systems, controls, diagnostic aids, and communication systems. The safety of such a complex system is supported by extensive analysis of failure modes and mechanisms, engineered features to handle process disturbances, and preplanned human responses to analyzed failures.

While a nuclear power plant (NPP) may be one of the most complex engineered systems, Human Factors/Ergonomics involvement in the design, operation, and maintenance of them was slow to develop and has, at times, been difficult to sustain. In this chapter we will review some of the major events that have shaped the development of human factors in the power industry and some of the individuals that are responsible for its accomplishments. While our main focus is on NPP in the USA, we also address developments in some other countries and in non-nuclear electrical power systems.

2. THE EARLY YEARS, 1950–80

2.1. United States of America

In the USA the field of Human Factors Engineering had become quite well established as a technology and a bridge between behavioral science and engineering by the US Department of Defense (DoD) and its large civilian contractors in the late 1950s and 1960s. However, very little attention was paid to Human Factors (HF) in power systems. There was little or no staff in HF at the US Atomic Energy Commission, predecessor of the US Nuclear Regulatory Commission (NRC), or at the major power plant design firms. Large companies such as General Electric and Westinghouse had human factors staff in many of their military and space divisions but no specialists in their commercial power groups. Nuclear plants that were designed before 1975 were not designed using HF analytical techniques or design standards. Rather, the industry used the same engineering methods that had been developed over the past 50 years in designing hydro and fossil fuel power plants. Interviews with 20 control board designers from a mix of nuclear steam suppliers and architect and engineering firms showed almost no knowledge of human factors/ergonomics and little use of human factors data or methodology in control room design (Parsons *et al.* 1978, Seminara and Parsons 1980).

In 1972 the US Atomic Energy Commission, which later became the US NRC, asserted that insufficient attention was being given to control room design and staffing and operator training and procedures. Alan Swain (1975), at Sandia Labs, identified human factors deficiencies in the design of the engineered safety panels at the Zion nuclear power plant. In 1976, the newly formed Electric Power Research Institute (EPRI) — the research arm of the electric power industry — contracted with Lockheed Missiles and Space Co. to review five typical control rooms. The study performed by Seminara *et al.* (1976) and Parsons *et al.* (1977) discussed the lack of HF principles and showed > 100 photographs of deviations from HF standards and potential areas which could lead to human errors and accidents. When an informal oral presentation of this study was presented to the NRC in Washington, DC, one of the comments was that the NRC considered Human Factors somewhat of a "black art" and was only interested in the back of the control room panels. Another review of 18 control rooms by Finlayson *et al.* (1977) at the Aerospace Corp. was equally negative.

As we have witnessed in numerous situations in various contexts over the years, it often takes a major accident to get the attention of management and the engineering community regarding the lack of good Human Factors Engineering. Such an event occurred on 28 March 1979 at the Three Mile Island Nuclear Plant #2 (TMI-2) in Pennsylvania. Accident investigations disclosed that this catastrophe was due to a variety of factors: inadequate training, a control room poorly designed for people, questionable emergency operating procedures, and inadequate provisions for the monitoring of the basic parameters of plant functioning. The event was a turning point for the nuclear power industry because it emphasized the central importance of human factors to safe plant operation. The President's Commission on the Accident at Three-Mile Island (Kemeny 1979) stated that: "There are many examples in our report that indicate the lack of attention to the human factor in nuclear safety. The control room, through which the operation of the TMI plant is carried out, is lacking in many ways. The control panel is huge, with hundreds of alarms, and there are some key indicators placed in locations where the operators cannot see them. There is little evidence of the impact of modern information technology within the control room … it is seriously deficient under accident conditions."

HF in nuclear safety was suddenly discovered and Joseph Seminara served as a member of the Rogovan Investigating Committee, Stuart Parsons was subpoenaed to testify before the White House Committee on Nuclear Safety, and the NRC contracted with John Snider at the University of Tennessee to assemble a cadre of noted Human Factors specialists (John Hungerford, Stuart Parsons, Earl Wiener, and Chris Wickens), to come to Bethesda, MD, and give all of the NRC managers, engineers and scientists a 1 week crash course in this technology.

In November 1979 the Institute of Electrical and Electronics Engineers (IEEE) devoted an entire issue of their journal, *Spectrum,* to Three Mile Island and provided a large section to human factors and safety using many of the EPRI/Lockheed pictures of control room design problems. The IEEE held their first meeting on Human Factors and Power Plants at Myrtle Beach, SC, in 1979. The industry was finally starting to appreciate that power systems were socio-technical and that technology, people, organizations and regulations interacted with one another in ways that had not previously been understood. Soon after TMI-2, the NRC contracted with the Human Factors Society, to conduct a study of the industry to determine the extent of the problem. The "blue ribbon committee" consisted of Charles Hopkins, Chairman, Richard Hornick, Robert Mackie, Harold (Smoke) Price, Harry Snyder, Robert Smillie and Robert Sugarman. The conclusions of the study were similar to the Lockheed and Aerospace investigations and reaffirmed the lack of human factors in the design of control room controls and displays. Hopkins famous quote was picked up by the press throughout the country, "This disregard for human factors in the control rooms was appalling. In some cases the distribution of displays and controls seemed almost haphazard. It was as if someone had taken a box of dials and switches, turned his back, thrown the whole thing at the board and attached things wherever they landed. For instance, sometimes 10 to 15 feet separated controls from the displays that had to be monitored while the controls were being operated. Also, sometimes no displays were provided to present critical information to the operators. There were many instances where

information was displayed in a manner that was not usable by the operators, or else was misleading to them. A textbook example of what can go wrong in a man–machine system when people have not been taken into account" (Machine Design 1981). Figure 1 is an example of operators adding arrows to the board to indicate the relationship between two related controls after repeated errors had occurred. Figure 2 is a typical control room of that era where no human engineering principles had been applied to the layout.

Some professional human factors specialists did join industrial firms designing nuclear power plants during the late 1970s. Lewis Hanes, later the president of the Human Factors Society, joined Westinghouse's R&D Center in 1973 and soon hired John O'Brien who later headed EPRI's human factors group. Len Pugh was at General Electric in San Jose and helped to lay out Nuclenet, an advanced highly computerized control room. Michael Danchak was at Combustion Engineering in Windsor, Connecticut and was performing human factors research studies on display colors and other variables.

2.2. United Kingdom

Towards the end of the 1950s the Central Electricity Generating Board (CEGB) became aware of ergonomics and they commissioned Dunlap and Associates in the USA to undertake a pilot survey of human factors aspects of control rooms. The resultant report convinced the CEGB that ergonomics should be taken seriously, and as they were then designing the control room for the Trawsfynydd Nuclear Power Plant, they immediately asked Ron Easterby to act as an ergonomics advisor on the design of
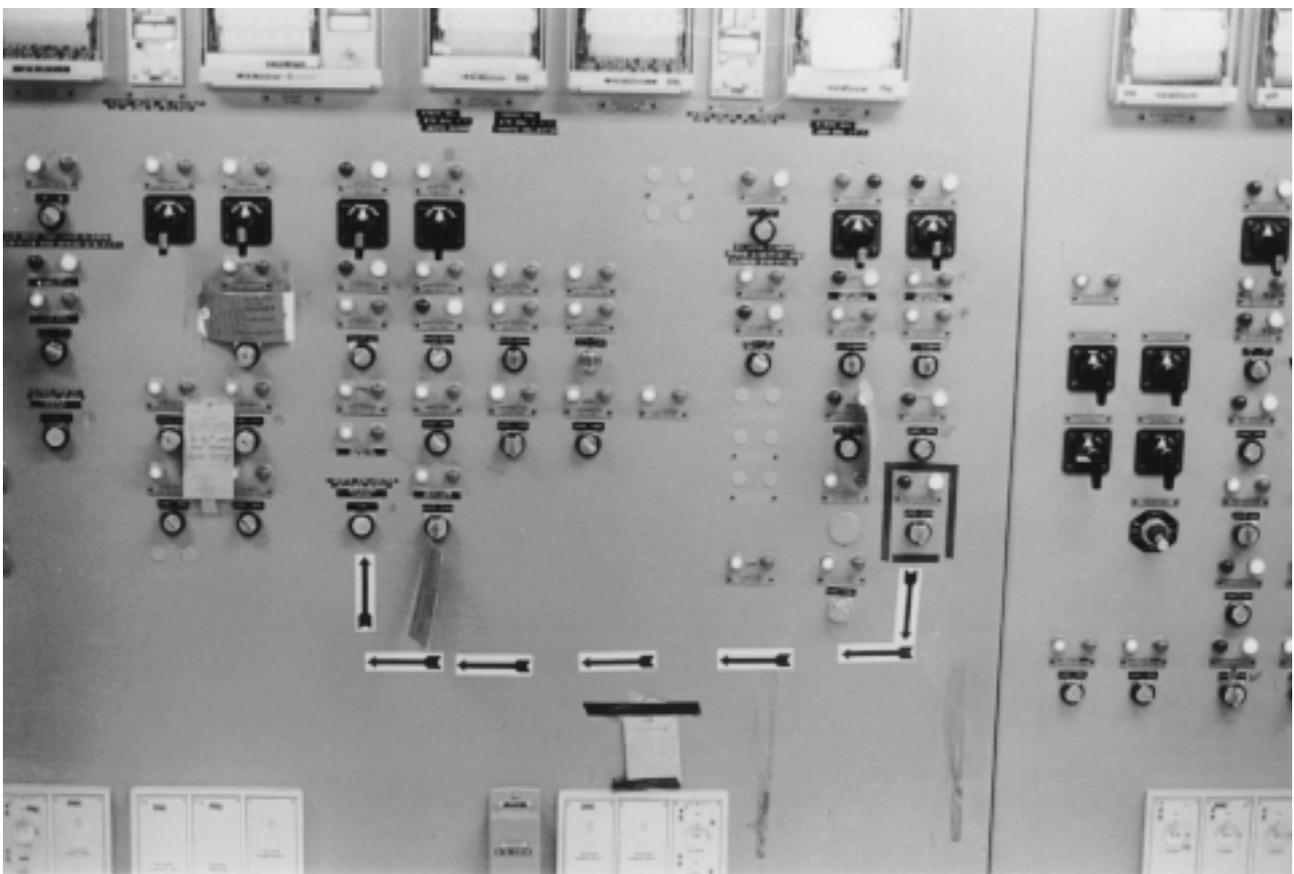
Figure 1.   Operators adding arrows to indicate the relationship of controls
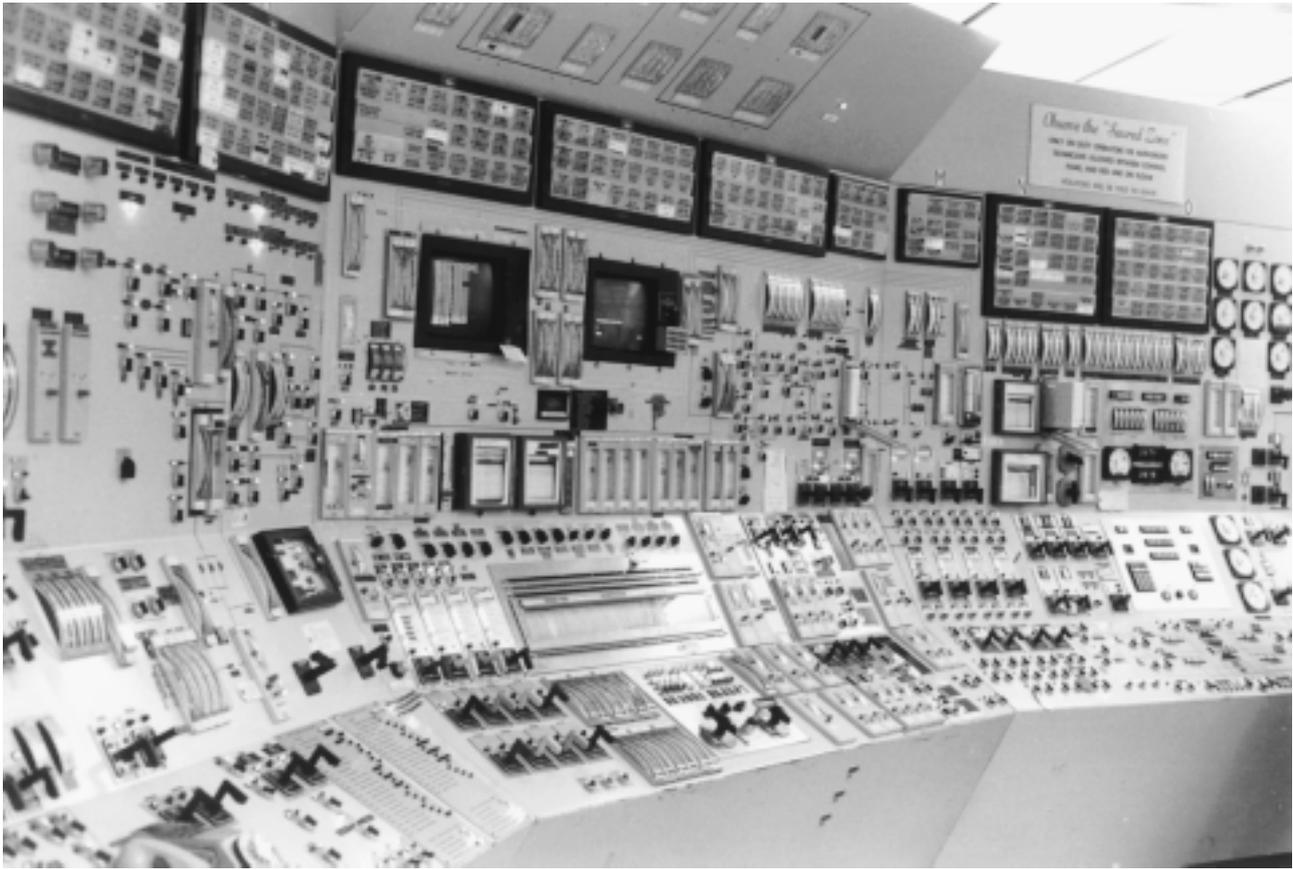
Figure 2.   Typical hardwired nuclear power plant control boards in 1980.

this control room. Easterby ensured that the control room complied with the limited ergonomics guidance that was then available and proposed some modifications to specific interfaces, including a novel mimic display of the control rods.

The resulting control room was acclaimed a success by the CEGB and Reg Sell was appointed as their in-house ergonomist to assist with the design of the control room for the Wylfa Nuclear Power Plant. However, shortly after this appointment, a significant blackout occurred in the south of England, which was attributed to a human error and Sell was diverted to improving the displays for grid control tasks. Thus, much of the design for Wylfa was undertaken by the manufacturer's own ergonomist, Peter Craft.

The design engineers at the CEGB learned much about ergonomics from their experiences on these programs, so that when Sell left the CEGB, it was decided that Reg Pope, an engineer with no formal training in ergonomics or psychology, should head up the ergonomics group. Despite his lack of an ergonomics background, Pope had a good understanding of the operators' requirements and became a powerful advocate of good basic ergonomics throughout the CEGB. This was sufficient while generically similar control rooms were being designed, but by 1979 the CEGB had decided to submit a proposal for a new pressurized water reactor (PWR) to built at Sizewell. This prompted renewed interest in ergonomics by the CEGB and the Nuclear Installations Inspectorate (NII), which was further reinforced as the impact of ergonomics deficiencies on the TMI-2 incident became appreciated. The NII responded to these changed circumstances by asking Lisanne Bainbridge and Neville

Moray to act as their advisors. Meanwhile the Ergonomics Development Unit at Aston University, under the stewardship of David Whitfield, began to build up expertise in this area, with a variety of projects on both nuclear and non-nuclear power plants, for the CEGB and the NII. Perhaps the most notable of these early studies was a comprehensive review of human reliability by David Embrey.

Interest was also being shown in ergonomics on the nuclear fuel reprocessing side. During the early 1980s some limited ergonomics had been undertaken, but by 1987 the tempo was increased significantly as the design work on the massive Thermal Oxide Reprocessing Plant (THORP) began in earnest. Barry Kirwan developed a comprehensive integrated program of task analysis and human reliability analysis to support the design of the computer-based control and display facilities for the plant. This program was to continue for over 5 years, with ergonomists such as Michael Carey, Johanne Penington, Julie Reid and Sue Whalley making major contributions.

2.3. The Soviet Union
The first implementations of human factors principles in the power industry in the Soviet Union were performed by K. M. Gurevich in the 1950s. He investigated Mosenergo's (Moscow Power Co.) operators' behavior during emergencies. The first serious attempts in ergonomic designing of control rooms were done in the Central Institute of Complex Automatics (Moscow) by E. P. Stephany and staff. Later this process was influenced by the Institute of Technical Esthetics (V. M. Munipov and

collaborators), and in the Institute of Psychology, USSR Academy of Sciences (V. F. Lomov and colleagues). Rare attempts to ergonomically influence nuclear power plant designs were made in the Institute of Nuclear Energy. Ergonomics existed mainly as a theoretical science. There were many interesting ideas and sophisticated investigations at the research labs, but they were rarely converted into practical systems. There were no technical standards for control room designs. Engineers applied general construction norms to the design of power plants. Western projects and common sense were the main guidelines. The first simulators were put into operations independently by Mosenergo and the various nuclear plants. There were numerous meetings on psychology and ergonomics on power plant safety. Strong recommendations were made by many outstanding scientists to Soviet Ministry officials, but these were almost always ignored.

3. THE 1980S

3.1. US Nuclear Regulatory Commission

Although the US AEC/NRC had funded research on human factors and human reliability as early as 1972, there was no long-range program and no human factors staff until 1980. In 1981 Dan Jones and other human factors professionals were hired in the Division of Human Factors Safety. In 1983 the NRC published its first long-range human factors research plan (NUREG-0961, 1983) aimed at:

- upgrading personnel qualifications and examinations;
- upgrading operating procedures;
- the utilization of computers;
- the impact on safety of organization and management;
- human contributions to risk and how to reduce them;
- human–machine technology changes that should be considered; and
- HF requirements for severe accident management.

In the 10 years from 1977 to 1987, the NRC funded 125 human factors research projects, studies and related efforts.

As the result of TMI-2, the NRC issued the requirement that operating reactor licensees and applicants for operating licenses perform a detailed human factors design review of their control rooms to identify and correct safety and design deficiencies. Toward this end, NUREG 0700, Guidelines for control Room Design Reviews, was issued in 1981. This 350+ page document was essentially a checklist for reviewing functions, tasks, workspace, communications, annunciator warning systems, controls, visual displays, labels and location aids, process computers, panel layout, and control-display integration. It was used to perform detailed analyses of currently licensed and pending licensed plants. The degree of compliance and amount of retrofitting was frequently dependent upon the utility's attitude toward human factors, and the dedication of various NRC review teams. Other countries, such as Taiwan and Korea, who had purchased nuclear power plants of US design requested that the program be extended to identifying and correcting human engineering deficiencies in their existing plants. Another NRC requirement involved all licensees to install a "safety parameter display system" to aid operators in the rapid determination of plant safety status, something they were unable to do effectively during the accident at TMI. The display system was designed to assist operators in detecting, interpreting and tracking process disturbances by providing a concise display of key parameters

and giving them the ability to observe trends in real time (NRC 1981b).

This Detailed Control Room Design Review (DCDR) program prompted research efforts to determine what might be done cost-effectively to enhance power plant control rooms that were expected to have a nominal 40-year life span. The Lockheed human factors group, under contract to EPRI, paved the way in establishing measures for enhancing existing control room panels with surface changes such as improved labeling, functionally demarcated groupings of related control-display elements, control coding, and marking meter scales to reveal normal and off-normal operating bands (EPRI November 1979). Compare before (Figure 3) and after (Figure 4) enhancement of a Steam Generator Feedwater System Control panel. In this effort to improve an existing panel, none of the control-display elements were moved or replaced. The DCRDR program transformed many deficient control rooms in this fashion.

In general, the NRC control room reviews of the 1980s identified a large number of HFE problems and resulted in significant improvements to the Human System Interfaces (HSI) (Eckenrode and West 1997, van Cott 1997). Despite the success of the reviews, there are limitations to the improvements that can reasonably be made to an already designed and operating plant. Many discrepancies are too difficult to correct without completely redoing entire panels. The design review experience
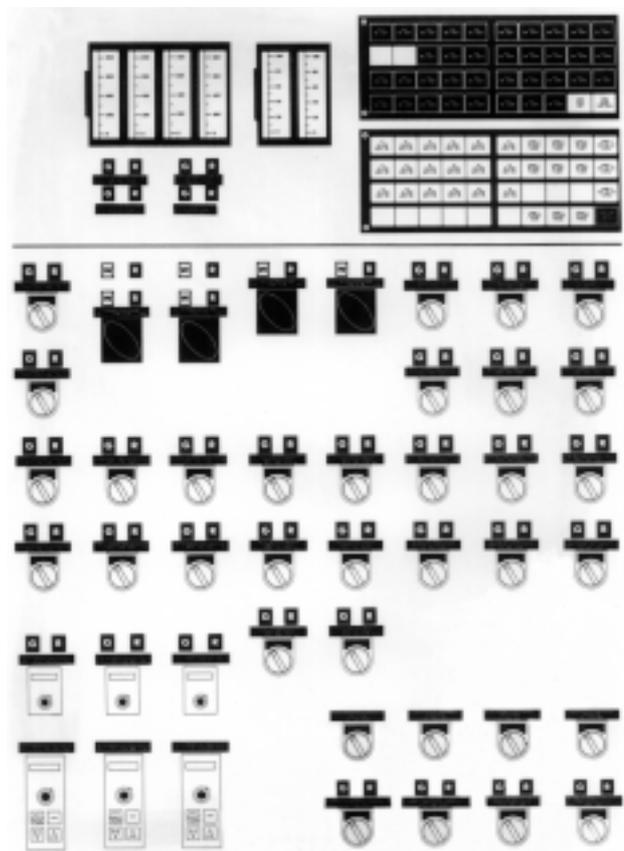
Figure 3.   Feedwater control panel from existing plant (non-human engineered).
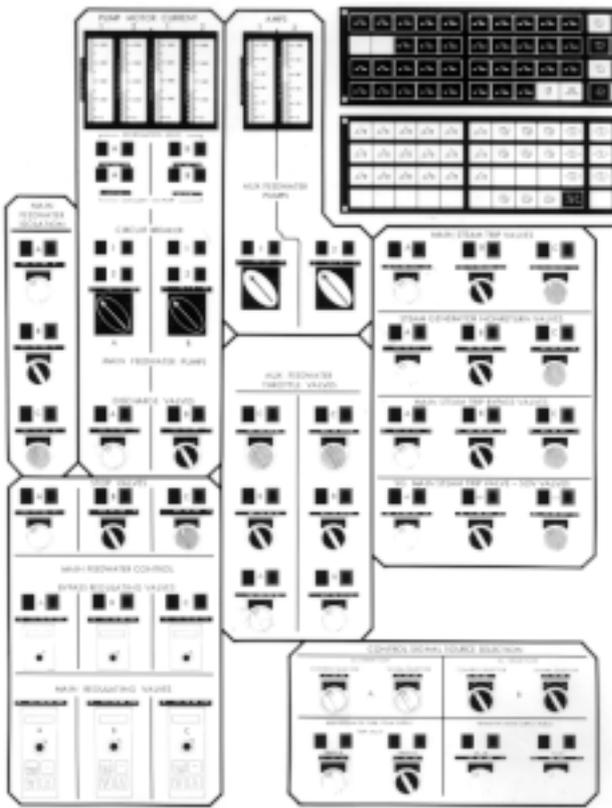
Figure 4.   Same feedwater control panel using labeling and demarcation techniques.

made clear that it is *certainly* preferable to incorporate HFE into the design process from the start.

3.2. Electric Power Research Institute

The long-standing human factors research program of EPRI has produced a series of high quality research reports and research products that are used widely throughout the industry. Because the scope and direction of its HF research program is defined by the industry, the results of the research are usually directly transferable. Other reasons for the success of this program is the overview and direction coming from a task force made up of utility human factors specialists and managers, the stability of EPRI human factors staff, and the fairly constant funding. Over the past 25 years three professional managers have directed the program — Randall Pack, Howard (Jack) Parris, and John O'Brien. The EPRI research program has covered such topics as control room design, alarm systems, computer-generated displays, operator alertness, lighting, qualifications and training, design for maintainability, industrial safety/radiation control, inspection and testing, preventive maintenance, plant enhancement techniques, simulation, labeling and coding, communications, shift length and scheduling, and organization and management. Some of the key technical reports are listed under EPRI in the list of references. It can be seen that these reports include fossil fuel plants and dispatch centers as well as nuclear plants.

3.3. Other Organizations

Soon after TMI-2, the power industry established the Institute of Nuclear Power Operation (INPO) for the transfer of technology from research to industry. INPO recognized that human performance problems were the largest contributors to system accidents and failures in power plants. Accordingly, INPO formulated a Human Performance Enhancement System Program (HPES) to encourage utilities to review significant events in the light of deficiencies in training, procedures, management, self-check techniques or user-unfriendly work environments. The US Department of Energy national laboratories at Brookhaven, Idaho, Livermore, Oak Ridge, Sandia, etc. hired and maintained a large number of human factors professionals with knowledge of power systems. These individuals conducted numerous research studies and provided support to the DOE and the NRC.

During the 1980s some firms including Anacapa Sciences, Essex, General Physics, Honeywell, Lockheed, MPR Associates and Westinghouse conducted numerous studies for the USNRC, EPRI and the utilities. Jens Rasmussen of Denmark created the skill-rule-knowledge paradigm that aided much of the theoretical and analytic work in the field. This work of Rasmussen has had an important impact on the thinking that went into the design of many products now incorporated in plants. Research by David Woods, Erik Hollnagel and Jens Rasmussen on plant safety information systems had a large impact on the development of cognitive engineering. Some of the better known human factors specialists who participated during this time and have not previously been mentioned in this history are: Richard Badalamente, Valerie Barnes, James Easter, Catherine Gaddy, James Geiwitz, Connie Goddard, Douglas Harris, Robert Kinkade, Michael Maddox, Randall Mumaw, Steven Pine, Emilie Roth, Sidney Seidenstein, Alan Spiker, Harold van Cott and Allan Williams,

In 1988, Neville Moray and Beverly Huey edited a report funded by the National Research Council and prepared by an international panel on Human Factors Research and Nuclear Safety. The report presents the panel's recommendations for an agenda of applied research on the human factors issues involved in the safety of nuclear reactors. Two other major publications during the 1980s are: a paper by Woods *et al.* in Salvendy's (1987) edition of the *Handbook of Human Factors*; and a paper by Hanes and O'Brien (1982) in the *IEEE Spectrum*.

3.4. United Kingdom

The 1980s was a time of great expansion for ergonomics in the British power industry, particularly in the nuclear area. The NII soon increased the number of its advisors to form an advisory group known as the Operations and Nuclear Safety Working Group (ONSWG), which consisted of an impressive array of ergonomists with experience of complex systems. By 1981 the UK Atomic Energy Authority (UKAEA) had also become involved in human factors, by appointing David Embrey and Jerry Williams, with a remit to focus upon human reliability issues. The Ergonomics Development Unit also remained at the forefront of practical assessments of UK control rooms, with Les Ainsworth undertaking a comprehensive assessment of different control room evaluation techniques for the NII.

It is interesting to note that by the early 1980s most of the ergonomics had moved from relatively global applications of established ergonomics, to become predominantly task-focused. This demonstrated that ergonomics within the British power industry was fast maturing and becoming accepted at all levels. Another important landmark along this road to increased maturity was passed when the UKAEA commissioned a study of the

ultrasonic inspection of welds, which was probably the first demonstration within the UK that ergonomics could make contributions outside of the control room.

In 1992 Reg Sell spearheaded a campaign for the Ergonomics Society to take part in the Public Enquiry into whether to build a PWR at Sizewell B. This was a momentous decision for British Ergonomics, as it was the first time that the Society had entered such a debate so publicly. The Society did not take a position on the question of whether the station should be built, but argued very forcibly that if it was to be built, ergonomics and human factors must be taken very seriously. The Ergonomics Society's team was chaired by David Whitfield and also included Lisanne Bainbridge, Fred Brigham, Keith Duncan, David Embrey, Andrew Hale, Andrew Shepherd and Ian Umbers. Almost immediately, the CEGB appointed Tom Singleton as their special advisor on ergonomics issues for Sizewell B and from then on, HF/E issues came to the forefront. Indeed the assessor at the Public Enquiry listed some specific ergonomics issues that had to be satisfactorily addressed before an operating license would be granted.

In early 1985 Jerry Williams took charge of the CEGB's ergonomics program for Sizewell B. Subsequently the in-house team was expanded and was later split into two groups. Geoff Pendlebury headed the group that specifically examined operational and procedural issues. This involved undertaking task analyses of critical tasks, then walk-throughs and finally simulator trials with most of the procedures. Pendlebury was supported in this work by Synergy, with Les Ainsworth undertaking most of the task analyses and walk-throughs, whilst Michael Herbert and Ed Marshall ran the simulator trials. Ned Hickling and Denise McCafferty were responsible for the hardware and software interfaces, with support from several different consultants throughout the project. The task analyses that were undertaken for Sizewell have been described in more detail in Ainsworth and Pendlebury (1995).

The Sizewell B project represented a major effort in British ergonomics and most of the input was specific to particular interfaces. However, there were two exceptions to this, which had more widespread practical application. These were a study of operator stress by Tom Cox and a detailed examination of defenses against cognitive or conceptual errors by Ian Umbers and Donald Ridley.

In 1986 and 1987 more ergonomists were appointed elsewhere within the CEGB. At the Central Electricity Research Laboratories in Leatherhead, Michael Herbert and Ed Marshall were involved in several projects, including generic studies of alarm handling and the development of some redesigned interfaces for a coal-fired station. At the Barnwood Offices Jon Berman and Ian Umbers were mainly involved in other Sizewell work, whilst Dave Collier and Jim Jenkinson were finishing the design work for the Heysham Nuclear Power Plant.

The UKAEA initiated an informal working group comprising human factors experts and other representatives from the nuclear industry. This group was started by Ian Watson as the Human Factors in Reliability Group (HFRG) and was soon split into a series of specific working groups. Lisanne Bainbridge presented a series of guidelines for human factors in process control during one of the early meetings, which was subsequently produced as The Short Guide to Reducing Human Error in Process Operations. The first major project of the HFRG was to produce an expanded

form of these guidelines under the chairmanship of Peter Ball, with major inputs from Les Ainsworth, Barry Kirwan, Andrew Shepherd, David Visick and Jerry Williams. Subsequently, the HFRG was extended beyond the nuclear industry, but its working groups are still active.

By 1987 the NII had decided that ergonomics was sufficiently important to justify an inspector specifically to deal with ergonomics issues. Subsequently, David Whitfield was appointed to this post. Soon ergonomics received a further boost from the NII, when it was listed as one of the prime areas for nuclear research. Since then, the NII has defined a series of general ergonomics issues where it feels that further ergonomics research is needed. A group comprising ergonomics representatives from the utilities and the NII then invites research proposals on specific issues that cover these areas and subsequently awards contracts that are funded by a levy of all the nuclear utilities. This has resulted in a wide range of ergonomics studies that are of general, rather than project-specific interest.

3.5. The Soviet Union

On 26 April 1986 the Chernobyl Nuclear Power Plant #4 in the Ukraine exploded and flooded the densely populated heartland of the Ukraine, Belarus and Western Russia with deadly radiation — > 300 times the amount unleashed by the atomic bomb of Hiroshima. There is still strong controversy about human error versus design error and roles of the operators, the designers and the scientists. However, there was no human factors engineering in the design of the control room, and inadequate training and procedures. Chernobyl has brought world attention to what can happen if managers, engineers and scientists disregard the necessity of including ergonomics in the process of designing and operating complex technical facilities (Munipov 1998)

4. THE 1990S

4.1. Nuclear Regulatory Commission

The HSI (Human System Interface) technology being developed for the new generation control rooms was changing rapidly and much of the design was not finalized when submitted to the NRC for design certification. This led to the development of an approach to the HF/E safety evaluation of advanced nuclear plant designs in which the review addresses the design process, as well as final detailed design. There are compelling technical reasons to support this approach. First, it is generally recognized that HFE issues and problems such as those identified above emerge throughout the design process. Second, when HF/E issues are identified before the design is complete, solutions can be considered in a more timely, cost-effective manner, i.e. before design details are locked in. Third, the evaluation of final designs is often based upon the quality of HF/E analyses conducted during the design process, i.e. task analyses to specify control and display requirements. It was further concluded that HF/E reviews should include a variety of assessment techniques.

John O'Hara and colleagues in the human factors group at Brookhaven National Laboratory (BNL) helped the NRC to develop a top-down approach to the conduct of HFE evaluation so that the significance of individual review topics could be seen in relationship to the high-level goal of plant safety. Top-down refers to a review approach starting at the "top" with high-level plant mission goals that are broken down into the functions

necessary to achieve the mission goals. Functions are allocated to human and system resources and are broken down into tasks. Operator tasks are analyzed for the purpose of specifying the alarms, information, and controls that will be required to allow the operator to accomplish assigned functions. Tasks are arranged into meaningful jobs assigned to individual operators and the HSI is designed to best support job task performance. The detailed design (controls and displays, graphical user interface software, maintainability, procedures and training) is the "bottom" of the top-down process. The HF/E safety evaluation should be broad based and include such HF/E aspects of normal and emergency operations, test, and maintenance.

The result of this effort was the HFE Program Review Model which is usually referred to as NUREG 0711 (O'Hara *et al.* 1994). The HFE Program Review Model (PRM) was developed as a basis for performing design certification HFE reviews that include design process evaluations as well as review of the final design. The PRM consists of ten component elements: HFE program planning, operating experience review, functional requirements analysis and allocation, task analysis, staffing analysis, human reliability analysis, HSI design, procedure development, training program development, and verification and validation. Each review element was divided into four sections: Background, Objective, Applicant Submittals, and Review Criteria.

The PRM was used to support several advanced reactor design certification reviews (e.g. GE ABWR, ABB-CE System 80+, and the Westinghouse AP600). The PRM was also used as the basic HFE criteria for the conduct of operational readiness safety reviews for DOE facilities (Higgins and O'Hara 1996).

The NRC research conducted at BNL also led to a significant revision to the guidance contained in NUREG-0700 that was used for the post-TMI control room reviews (NRC 1981a). Following the original control room guidance development in the 1980s, the NRC staff focused on human factors issues for which there were uncertainties in the scientific data needed to support the development of regulation. One such issue was the introduction into control rooms and local control stations of advanced, computer-based HSI technology, a technology that was not used in TMI-era NPP. Advanced HSI designs were emerging because of several factors including: (1) replacement of existing plant HSI with computer-based technologies when existing hardware is no longer supported by equipment vendors, (2) upgrading plants with new, computer-based monitoring and control systems, and (3) development of advanced control room concepts as part of new reactor designs. Each of these developments had the potential for significant implications for plant safety in that they affect the overall role (function) of personnel in the system, the amount of information available, the type of information available, the format of information presentation, the ways in which personnel interact with the system, and the requirements imparted upon personnel to understand and supervise an increasingly complex system. However, the guidance developed in the early 1980s, well before these technological advances, was limited in its applicability to new technology Accordingly, the human factors guidance needed updating.

BNL conducted extensive research into the effects of advanced technology on crew performance. The results were used to develop NUREG-0700 Revision 1 (O'Hara *et al.* 1996). The guidance consists of HFE guidelines, design review procedures,

a computer-based review aid called the "Design Review Guideline" (DRG). Guidance for specific HSI topics such as graphic displays, touch screens, expert systems and local control stations was developed through the application of a general guidance development process (O'Hara *et al.* 1995).

The NRC human factors research at BNL continues to address safety significant HSI technologies such as the alarm systems, advanced information systems, computer-based procedures, soft controls, interface management (e.g. navigation using menus), digital system maintenance, and the design and implementation process associated with control room modernization. The results of these ongoing programs will be used as the basis for the next revision to NUREG-0700, thereby maintaining the document as an up-to-date source of HFE guidance in design and development of electrical power systems.

Currently the NRC's HF Assessment Branch performs inspections of operating plants using professionals such as James Bongerra, David Desaulniers, Richard Eckenrode, Clare Goodman, and Garmon West while the HF Research Branch manages contracted research directed by key personnel such as J. J. Persensky and Jerry Wachtel.

4.2. Industry

Early in the 1990s, the DOE, EPRI, and the NRC provided funding to Combustion Engineering, General Electric and Westinghouse to develop safer and more efficient advanced nuclear power systems. Combustion Engineering, now a division of the Swiss-Swedish giant Asea Brown Baveri (ABB), developed their System 80+, General Electric developed the Advanced Boiling Water Reactor, and Westinghouse Nuclear, now part of British Nuclear Fuels, designed the AP600 System. All of these engineering groups had human factors staffs and utilized current human factors/ergonomics technology during design, development and NRC certification. However, due to the low cost of fossil fuels, gas turbine technology, NRC regulations, and public attitudes in the USA, there have been no orders from US utilities. Combustion Engineering, currently supported in human factors by Robert Fuld, Darryl Harman and Donna Smith, are designing and exporting two System 80+ plants at 1400 MWe each to Korea at the Yanwang and Ulchen plants. General Electric, using the human factors talents of Richard Gutierrez, and the Technatom of Spain in engineering analysis and design, and Stuart Parsons, Joseph Seminara and Linda Taylor as the independent review team, support the Advanced Boiling Water Reactor design project which is building the two Lungmen plants for Taiwan Power Co. The human factors independent review team for Stone and Webster, the Architect and Engineering firm for this project, is contracted to Synergy in the UK with Les Ainsworth, Ned Hickling and Ed Marshall providing technical support.

4.3. Electric Power Research Institute

At EPRI, one major program in the 1990s, managed by Joseph Yasutake, was a joint 5-year program with the Japanese for developing an array of intervention products for improving performance of nuclear power plant maintenance workers. The final summary report by Hanes, Parsons and Taylor describes these interventions (EPRI 1994). Madeleine Gross now manages the human factors long-range strategic program. The tactical program is directed toward better integrating human factors into

the total EPRI engineering effort. The key topics include: identifying key indicators, root causes, and selection of corrective actions; improving procedures, maintenance proficiency and training; and putting the key HF documents on the web for ease of use by utility personnel.

4.4. Professional Organizations

The Institute of Electrical and Electronic Engineers (IEEE) has continued to hold conferences every 3–4 years on Human Factors and Power Plants which have been attended by international experts from utilities, universities, regulation agencies, and researcher organizations. These conferences have played an important role in worldwide information transfer. In 1992, the International Ergonomics Association (IEA) established a Power Systems Technical Group and has been active in information exchange and promoting symposia at the IEA triennial congresses.

4.5. Human Reliability Analysis (HRA)

The topics of risk probabilities, human error, and human reliability analysis (HRA) have been studied in relation to nuclear power plants for the past 25 years by Allan Swain, Neville Moray, John Senders, Jens Rasmussen, Harold Blackman and others. However, this has been a controversial area focusing on: (1) the quantification of human reliability; (2) the lack of hard data for human reliability; and (3) the idea that human reliability debases humans to the level of mechanical system components. Although quantitative HRA data is currently used in plant probabilistic risk analyses, there is no universal agreement on the methods used to develop these data. Blackman and Byers (1998) has recently proposed a six-point program for solving this controversial problem.

4.6. Canada

In Canada the application of a structured methodology for HF/E to the design of CANDU (Canadian Deuterium Uranium) power reactor control rooms, whether for new designs or retrofit projects, has been increasing steadily over the past 10 years. At the present time, new designs and major retrofits have HFE applied in a structured fashion consistent with technical guidance provided in standards such as IEEE 1023 (Guidance for the Application of HFE to Systems, Equipment and Facilities of Nuclear Power Generating Stations). The Canadian planning document, the HFEPP (Human Factors Engineering Program Plan) as described by Beattie and Malcolm (1991) is the main mechanism for incorporating HFE activities in the overall systems design process, and has been used successfully on many projects, e.g. the Darlington Plant near Toronto and the new control room for the CANDU 600 plant under construction in China.

4.7. France

Research and engineering in energy systems including human factors technology has been very active in France. Electricite de France (EdF) is one of the worlds most active designers and marketers of nuclear plants. The French plants have been designed in an evolutionary fashion with one generation of plant design following its predecessor with specific design improvements. The most recent design, the N4, is one of the most advanced designs in operation. The N4 control room underwent extensive testing and evaluation. It is a computer-based control room with seated workstations for operating crewmembers and a large panel display providing high-level status information, important plant parameters, and alarms. With N4's operating in Civaux and Choose, EdF is designing the next generation of reactor.

While EdF has been conducting research and development activities related to plant design and operation, other organizations in the French nuclear industry have been conducting research and investigations into many aspects of nuclear safety and operations. For example, the Nuclear Protection and Safety Institute (NPSI) is a research and consultation organization for the Directorate of Safety and Nuclear Regulation. Their mission is to conduct research and provide evaluations in all aspects related to the control of nuclear risk. These activities include plant safety, protection of people and the environment, management of accidents and nuclear materials transportation. The evaluation activities involve incident investigations as well as reviewing and evaluating information submitted by plant operators for various stages of licensing. These evaluations serve as a technical basis for decisions made by the French regulatory authority.

4.8. Japan

The lessons of TMI and Chernobyl were not lost on the power industry in Japan. A consortium of government and industry initiated a program for human factors research and development. The government side was headed by the Ministry of International Trade and Industries (MITI) and utilized several organizations: the Nuclear Power Engineering Test Center (NUPEC) Institute for Human Factors (focus on basic research), The Japan Power Engineering and Inspection Corporation (focus on reliability improvements), and the Nuclear Power Operation Management Aid System Development Association (focus on man–machine systems). The industry side was made up of several organizations including the Central Research Institute of Electric Power Industry — Human Factors Research Center (focus on applied research), the Federation of Electric Power Companies (focus on human reliability, safety, and control room research), individual electric power companies such as Tokyo Electric Power Company and Japan Atomic Power Company (focus on operation support systems, abnormality diagnostic systems, and reliability), and manufacturers such Mitsubishi. The program is quite comprehensive and includes, for example:

- survey of the current status of HF research;
- human performance and cognitive modeling;
- team performance and performance measurement;
- human reliability research;
- function allocation and automation;
- advanced HSI technology development (including the operation support system — an advanced decision-aiding technology);
- maintenance support system; and
- human factors database development.

Control room development has been influenced by national efforts in Japan to develop advanced HSI technology and operator support systems. These efforts are the result of joint government and industry programs. Unlike control room developments in the USA, the evolution of HSI technology has been gradual.

The conventional (post-TMI) control room (CR) is referred

to as a first generation CR. It is basically a hardwired CR with individual indicators and switches arranged on large-sized panels (typical of many US CR). These control room included a number of post-TMI improvements designed to reduce human error. Control room modifications were assisted by a NUREG-0700-type analyses (many supported by human factors practitioners from US companies such as Essex Corp.). Improvements included color coding of alarm tiles, set point markers, shape coding, system organization of switches and indicators, mimic lines, and the use of CRT for trend monitoring. Work began on the second generation control room in the early 1980s and reflected design objectives to minimize the operators' working area (seated operations), enhance real-time CRT display functions so that plant status can be more easily understood, to reduce operator error during surveillance tests, and to reduce operator workload by utilizing computer controls and computerized operation guides. The "third generation" control rooms featured fully computerized operations, e.g. Toshiba's Advanced Boiling Water Reactor control room called the Advanced Plant Operation by Display Information and Automation (A-PODIA). Current research is being devoted to applying artificial intelligence information processing and operator support for the next generation control room.

4.9. Norway

Although Norway is a non-nuclear country, the premier human factors power systems laboratory in the world is located in Halden. This international joint research effort called the OECD Halden Reactor Project began back in 1958, and was in the beginning focusing on human–machine interaction (HMI). It was decided that all process information should be provided by means of CRT and displayed using process mimics, and where status information used mixed alphanumeric and graphics formats which included mimic diagrams, bar graphs, and trend diagrams. It was also realized that color was a necessary part of the information presentation.

This early work, led to the establishment of the simulator based Halden human–machine laboratory in 1983, which has been the focal point for this area of research and development in Halden. The nucleus in the laboratory has since 1983 been a full-scope pressurized water reactor based upon the Loviisa nuclear power plant in Finland. Today the laboratory is the object for a major upgrading which at the end of 1999 will include four simulators. The existing one, a new pressurized water reactor based on the Fessenheim nuclear power plant in France, a new boiling water reactor based on Forsmark unit 3 in Sweden, and finally a simulator based on the Oseberg petroleum production platform in the North Sea. The experimental facility consists of a cockpit control room for operators and supervisor using CRT, large screens, keyboards and trackballs for interaction. In addition, the laboratory includes an experimenter's gallery from where studies are run and controlled, and where all data collection equipment are supervised. Data collection includes equipment such as computer logging, video cameras, wireless microphones, audio and video mixing, eye movement tracking, etc.

The aim of the human–machine interaction research at the Halden Project is to provide knowledge which can be used by the funding organizations to enhance safety and efficiency in the operation of nuclear power plants by utilizing research about the capabilities and limitations of the human operator in a control room environment. A main premise of this research program element is that as systems evolve and new, more advanced technologies supplant older ones, greater automation of operator functions becomes possible, thereby changing the types of demands which the system will place upon the operator and the potential role which the operator serves in the control room. Understanding the impact of new technology on the role and performance of operating personnel is crucial in decision making for the safety of nuclear power plants.

Today the Halden Project is funded by twenty different countries world wide, and interacts with more than one hundred organizations coming from a variety of areas, including regulatory authorities, vendors, utilities and research organizations. Numerous research projects are being conducted in the area of human factors, for not only the nuclear industry, but also for all kinds of process industry applications. The human factors staff consists of eighteen professionals.

4.10. South Korea

In the late 1990s a small team was set up in South Korea to develop the human–machine interfaces for a new design of Korean reactors. This team under the direction of Yeong-Cheol Shin, developed a comprehensive plan for task analyses to support this design work. At the time of writing, Joongnam Kim and others have started this program.

4.11. UK

During 1992 and 1993 the Sizewell program was wound down (and the THORP project was also coming to a close). There were no immediate plans for further nuclear stations and so the power industry, which had by now been privatized into smaller companies cut down its ergonomics staff, so that only Nuclear Electric and Magnox retained any ergonomists, with small teams headed by Ian Umbers and Ray Hughes respectively. The focus also changed — from design to assessment — and stations that had been built with little or no ergonomics inputs, were subjected to rigorous ergonomics assessments as part of the process of extending the operational life of some of the earlier stations. Another concern lay in assessing the adequacy of new systems to provide a further source of emergency cooling at some of the older stations, and which involved many local-to-plant tasks. All of these assessments were based upon task analyses of selected tasks, and whilst they were mainly control room-based, they also included many manual tasks undertaken on plant. As many of these assessments were being made to support Safety Cases, many of them also involved some form of human reliability assessment. The stations themselves undertook some of these studies, but consultants from Synergy and Vectra also provided much support.

As the millennium drew to its close, Jerry Williams and Craig Reiersen, now both at the NII, started to broaden their interest further from the traditional control issues. Thus, the utilities were forced to provide evidence that human factors had also been adequately considered in a variety of tasks. Typical of these were studies of refueling by Les Ainsworth and of maintenance by Jane Mechan. The NII also sought assurances that personnel downsizing, which up to then had been undertaken with little coordinated planning, could be achieved without jeopardizing safety. To date, one station, Chapelcross has grasped this nettle and with support from Synergy, has attempted to use task analysis

to assess and plan the impact of downsizing proposals upon operational performance.

Away from the nuclear industry, ergonomics in other areas of the power industry was more limited. Les Ainsworth and Michael Herbert undertook two studies of grid control facilities, while Ed Marshall and Michael Herbert developed some guidelines for gas turbine plants and also assessed the proposed computer-based interface for some gas turbine plants.

4.12. International Atomic Energy Agency (IAEA)

The IAEA, with headquarters in Vienna, Austria, has been increasingly examining the role of human factors in plant safety. While human factors has long been identified as a key technology for nuclear safety, the IAEA has more recently focused on programs designed to provide human factors support to countries operating nuclear plants. From a macro-ergonomic perspective, the IAEA has provided support on the role of organizational factors to the safe operation of nuclear plants and the importance of a safety culture. From a micro-ergonomic perspective, the IAEA has been providing information on the human factors aspects of control room modernization. For a variety of reasons, many NPP have been upgrading their plants with digital technology. One reason for this trend is to achieve performance improvements. In other cases, old analog systems become difficult to maintain and are replaced with newer systems. The IAEA has held and is continuing to hold numerous workshops and meetings on digital upgrades initially focusing on software and hardware issues. One factor that frequently emerged as important to the successful implementation of a digital upgrade program is human factors. The human factors issues ranged from understanding the effects of advanced technology (such as computerized operator support systems) on human performance to issues related to operator acceptance and training. Some of the reports resulting from these efforts are IAEA (1991–95).

5. INTO THE NEW MILLENNIUM

As noted above, accidents, events and operating experience has demonstrated the importance of "defense in depth" and the role that human factors plays in it. Defense in depth includes the use of multiple barriers to prevent the release of radioactive materials and uses a variety of programs to ensure the integrity of barriers and related systems (IAEA 1988). These programs include, among others, conservative design, quality assurance, administrative controls, safety reviews, personnel qualification and training, test and maintenance, safety culture, and human factors. IAEA established "Human Factors" as an underlying technical principle that is essential to the successful application of safety technology for NPP. The principle states: "Personnel engaged in activities bearing on nuclear power plant safety are trained and qualified to perform their duties. The possibility of human error in nuclear power plant operation is taken into account by facilitating correct decisions by operators and inhibiting wrong decisions, and by providing means for detecting and correcting or compensating for error" (p. 19). Thus, human factors/ergonomics (HF/E) technology is now internationally recognized as essential to the safe design, operation and maintenance of not only NPP but also other large electrical power plants. Through the 20th century, HF/E has become integral to the design and operation of the control room. As we look toward the turn of the century, we see the extension of HF/E to even broader applications. Some of these applications have already begun, but should receive increased attention in the 21st century:

- Understanding and assessing the influence of organizational factors on safety.
- Focus on individual and crew performance assessment and on the management of unplanned and unanticipated events.
- Further application of advanced technology in the control room:
- staffing reductions;
- performance improvements.
- Extension of successful HSI technologies and programs to HSI out in the plant (outside the control room).
- Improvement of maintenance design and operations.
- Improving data collection and testing methods, including HRA, event reporting, and V&V.
- Understanding and assessing the influence of deregulation on safety.
- Understanding and assessing the influence of license extension on productivity and safety.
- Understanding and assessing the HFE aspects of decommissioning.
- Greater application of HFE to non-nuclear plants operations and maintenance.

While there is every expectation the HF/E considerations will strongly influence the power industry's future; we cannot afford to be complacent. Some serious problems to be faced and overcome are:

- There is a tendency to focus intensely on HF/E when a catastrophe or near-disaster occurs. We should find ways to sustain interest and effort in this discipline between newsworthy accidents.
- Some members of power industry management tend to regard the necessity for HF/E in terms of a temporary "magic bullet" to be applied primarily when a serious problem surfaces rather than developing systematic programs for addressing such concerns in terms of constant preventive measures.
- As regulatory pressures to consider HF/E subside usually due to budget cuts, utilities and government agencies tend to divert resources to other "squeaking wheels."

While many important inroads have been made in the nuclear power industry, there has been only a modest HF/E impact on the design and operation of fossil, hydro, solar, etc. electrical plants. There have been cases where the nuclear division of a given utility was spending enormous resources to enhance operational control rooms from a HF/E standpoint while the fossil division of the same utility was in the process of designing new control rooms with all the usual HF/E deficiencies.

There is a lack of standardization in the methodology and application of HF/E analytical procedures (i.e. functional requirements, function allocations, task analysis, and human reliability analysis) used during the design and development phase. A guidebook with specific examples needs to be developed.

Some people consider nuclear power, particularly in the USA, to be a dying industry making it difficult to recruit and train resources and top talent to be applied in this area of HF/E endeavor.

ACKNOWLEDGEMENTS

REFERENCES

AINSWORTH, L. and PENDLEBURY, G., 1995, Task-based contributions to the design and assessment of the man–machine interfaces for a pressurized water reactor. *Ergonomics*, 38, 462–74.

BEATTIE, J.D. and MALCOLM, J.S., 1991, Development of a human factors engineering program for the Canadian nuclear industry. In *Proceedings of the Human Factors Society 35th Annual Meeting*, vol. 2, Santa Monica, CA, 1248–52.

BLACKMAN, H.S. and BYERS, J.C., 1998, Unsolved problem: acceptance of human reliability analysis. In *Proceedings of the Human Factors and Ergonomics Society 42nd Annual Meeting*, vol. 1, Santa Monica, CA, 702–4.

ECKENRODE, R. and WEST, G., 1997,. Detailed control room design reviews — were they worth the effort? In *Global Perspectives of Human Factors in Power Generation: Proceedings of IEEE Sixth Conference on Human Factors and Power Plants*, Washington, DC, IEEE.

EDITORIAL, 1981, *Machine Design*, 55, 4.

EPRI, March 1976, *Human Factors Review of Nuclear Power Plant Design*. Project 501. NP-309 (Palo Alto: Electric Power Research Institute).

EPRI, February 1981, *Human Factors Review of Power Plant Maintainability*. Project 1126. NP-1567 (Palo Alto: Electric Power Research Institute).

EPRI, July 1981, *Human Factors Review of Electric Power Dispatch Control Centers*, vols 1 and 2. Project 1354-1, EL-1960 (Palo Alto: Electric Power Research Institute).

EPRI, September 1981, *Survey and Analysis of Communications Problems in Nuclear Power Plants*. Project 501-5, NP-2035 (Palo Alto: Electric Power Research Institute).

EPRI, April 1984. *A Procedure for Reviewing and Improving Power Plant Alarm Systems*, vols 1 and 2. Project 2011-2, NP 3448 (Palo Alto: Electric Power Research Institute).

EPRI, July 1984, *Maintainability Assessment Methods and Enhancement Strategies for Nuclear and Fossil Fuel Power Plants*. Project 2166-2, NP 3588 (Palo Alto: Electric Power Research Institute).

EPRI, August 1984, *Human Factors Guide for Nuclear Power Plant Control Room Development*. Project 1637-1, NP-3659 (Palo Alto: Electric Power Research Institute).

EPRI, October 1984, *Enhancing Fossil Power Plant Design, Operation and Maintenance: Human Factor Guidelines*, vols 1 and 2. Project 1752-1, CS 3745 (Palo Alto: Electric Power Research Institute).

EPRI, December 1988, *Effective Plant Labeling and Coding*. Project 1637-8, NP-6209 (Palo Alto: Electric Power Research Institute).

EPRI, December 1989, *Guidebook for Maintenance Proficiency Testing*. Project 2705-10, NP-6679 (Palo Alto: Electric Power Research Institute).

EPRI, December 1994, *EPRI-CRIEPI Joint Human Factors Program Summary Report*. Project 3111-10, TR-104324 (Palo Alto: Electric Power Research Institute).

FINLAYSON, F.C. *et al.*, 1977, *Human Engineering of Nuclear Power Plant Control Rooms and its Effect on Operator Performance*. Report ATR-77 (2815)-1 (Aerospace Corp.).

HANES and O'BRIEN, 1982, Control room design: lessons from TMI. IEEE Spectrum.

HIGGINS, J. and O'HARA, J., 1996, *Control Systems/Control Room Review*

Guidance for DOE Operational Readiness Review (ORR) (Upton: Brookhaven National Laboratory).

IAEA, 1988, *Basic Safety Principles for Nuclear Power Plants*. Safety Series No. 75-INSAG-3 (Vienna: International Atomic Energy Agency).

IAEA, 1991, *Safety Implications of Computerized Process Control in Nuclear Power Plants*. IAEA-TECDOC-581 (Vienna: International Atomic Energy Agency).

IAEA, 1992, *Expert Systems in the Nuclear Industry*. IAEA-TECDOC-660 (Vienna: International Atomic Energy Agency).

IAEA, 1993, *The Potential of Knowledge Based Systems in Nuclear Installations*. IAEA-TECDOC-700 (Vienna: International Atomic Energy Agency).

IAEA, 1994a, *Development and Implementation of Computerized Operator Support Systems in Nuclear Installations*. Technical Reports Series No. 372 (Vienna: International Atomic Energy Agency).

IAEA, 1994b, *Operator Support Systems in Nuclear Power Plants*. IAEA-TECDOC-762 (Vienna: International Atomic Energy Agency).

IAEA, 1994c, *Current Practices and Future Trends in Expert System Developments for Use in the Nuclear Industry*. IAEA-TECDOC-769 (Vienna: International Atomic Energy Agency).

IAEA, 1995a, *Computerization of Operation and Maintenance for Nuclear Power Plants*. IAEA-TECDOC-808 (Vienna: International Atomic Energy Agency).

IAEA, 1995b, *Control Room Systems Design for Nuclear Power Plants*. IAEA-TECDOC-812 (Vienna: International Atomic Energy Agency).

IEEE SPECTRUM, November 1979, *Three Mile Island and the Future of Nuclear Power* (Institute of Electrical and Electronics Engineers).

KEMENY, J., 1979, *Report of the President's Commission on the Accident at Three-Mile Island* (Washington, DC: US Nuclear Regulatory Commission).

MORAY, N.P. and HUEY, B.M. (eds), 1988, *Human Factors Research and Nuclear Safety* (Washington, DC: National Academy Press).

MUNIPOV, V.M., 1998, The problems of nuclear power are much too serious to leave to nuclear experts. In *Proceedings of the Human Factors and Ergonomics Society 42nd Annual Meeting*, Santa Monica, CA, 737–9.

NRC, 1984, *Standard Review Plan*. NUREG-0800, Revision 1 (Washington, DC: US Nuclear Regulatory Commission).

NRC, 1981a, *Guidelines for Control Room Design Reviews*. NUREG-0700 (Washington, DC: US Nuclear Regulatory Commission).

NRC, 1981b, *Human factors Acceptance Criteria for Safety Parameter Display Systems*. NUREG-0835 (Washington, DC: US Nuclear Regulatory Commission).

O'HARA, J., BROWN, W., STUBLER, W., WACHTEL. J. and PERSENSKY, J., 1996, *Human–System Interface Design Review Guideline*. NUREG-0700, Revision 1 (Washington, DC: US Nuclear Regulatory Commission).

O'HARA, J., BROWN, W., STUBLER, W., WACHTEL. J. and PERSENSKY, J., 1995, Comprehensive guidance for the evaluation of human–systems interfaces in complex systems. In *Proceedings of the Human Factors and Ergonomics Society 39th Annual Meeting*, Santa Monica, CA.

O'HARA, J., HIGGINS, J., STUBLER, W., GOODMAN, C., ECKINRODE, R., BONGARRA, J. and GALLETTI, G., 1994, *Human Factors Engineering Program Review Model*. NUREG-0711, Revision 0 (Washington, DC: US Nuclear Regulatory Commission).

PARSONS, S.O., ECKERT, S.K. and SEMINARA, J.L., 1978, Human factors design practices for nuclear power plant control rooms. In *Proceedings of the Human Factors Society 22nd Annual Meeting*, Santa Monica, CA, 133–9.

PARSONS, S. O., SEMINARA, J.L. and KATZ, G.H., 1977, Human factors

engineering in nuclear power plants. In *Transactions of the American Nuclear Society Conference on Reactor Operating Experience*, Chattanooga, TN, 7–10 August, 50–1.

SEMINARA, J.L. and PARSONS, S.O., 1980, Survey of control-room design practices with respect to human factors engineering. *Nuclear Safety*, 21, 603–17.

SWAIN, A.D., 1975, *Preliminary Human Factors Analysis of Zion Nuclear Power Plant*. Report 76-6503 (Sandia Corp.).

US ATOMIC ENERGY COMMISSION, 1972, *Evaluation of Incident of Primary Coolant Release from Boiling Water Reactors*. Technical Report WASH 1260 (Washington, DC: USAEC).

US NUCLEAR REGULATORY COMMISSION, 1981, *Critical Human Factors Issues in Nuclear Power Regulation and a Recommended Comprehensive Human Factors Long-range Plan*. NUREG/CR 2833, vols 1 and 2 (Washington, DC: USNRC).

US NUCLEAR REGULATORY COMMISSION, 1981, *Guidelines for Control Room Design Reviews*. NUREG-0700 (Washington, DC: USNRC).

US NUCLEAR REGULATORY COMMISSION, 1983, *Long Range Research Plan*. NUREG-0961 (Washington, DC: USNRC).

VAN COTT, H., 1997, Were the control room design reviews worth it? In *Global Perspectives of Human Factors in Power Generation: Proceedings of IEEE Sixth Conference on Human Factors and Power Plants* (Washington, DC: IEEE).

WOODS, D.D, O'BRIEN, J. and HANES, L.F., 1987, Human factors challenges in process control: the case of nuclear power plants. In G. Salvendy (ed.) *Handbook of Human Factors and Ergonomics* (New York: Wiley Interscience).

History of the Gesellschaft für Arbeitswissenschaft (GfA)

H. Schmidtke, A. Seeber, and K. J. Zink

GfA, Ardeystrasse 67, D-44139 Dortmund, Germany

1. HISTORY

The Gesellschaft für Arbeitswissenschaft (GfA), the ergonomics association of the German-speaking area was founded in 1953. It provided a cover organization for several outstanding research personalities and practical experts from Germany, Austria, and the German-speaking regions of Switzerland, who before were scattered over various scientific associations. The time of reconstruction after World War II in Germany and Austria and the structural change in industry in all Western countries focused the attention increasingly upon human capital as a productive factor.

However, more than 50 years earlier, numerous scientists in the fields of medicine, psychology, production engineering, pedagogy, and later sociology had considered the issue of workers and their performance in industry. Fechner deserves mention for providing the first serious experimental evidence of the connection between stimulus and sensation and the threshold term in his main work *Elemente der Psychophysik* in 1860; besides that he published several methodological studies. The first Institute of Experimental Psychology was founded by the physiologist Wundt in Leipzig in 1879. The reorientation from pure basic research towards application-oriented research was achieved by Stern, who coined the term "psychotechnics" in 1903, adopted by Münsterberg in his book *Psychotechnology*, published a decade later.

In addition, a series of associations and organizations such as the Verband für Arbeisstudien (REFA), an association on work studies, and the Rationalisierungskuratorium der Deutschen Wirtschaft (RKW), a curatory for rationalization in German industry, contributed considerably to the development of work science in the German-speaking area. REFA succeeded in distributing ergonomic knowledge over the entire range of industry with its books on time and motion study, industrial engineering, and later the six-volume works on methods of work study. Individual education and research units already existed at the universities before GfA was founded. However, these institutions were integrated into a wide range of different faculties and maintained relatively little contact among each other.

From today's point of view, the Max-Planck-Institute of Industrial Physiology, founded in Berlin in 1912 and transferred to Dortmund later on, has to be considered as the most important unit for ergonomic research in Germany. It undertook a great deal of application-oriented research on the issues of energy transmission in the body during muscular work, blood circulation control, sense physiology, environmental physiology, industrial psychology, much of it under the auspices of Lehmann (from 1927), Müller (from 1930), Graf (from 1929) and Schmidtke (from 1956). In addition, there existed institutions on work science in forestry with Hilf and work science in agriculture with

Preuschen. For practical work in industry REFA proved to be an important conveyor of ergonomic knowledge, since with Bramesfeld it was directed by an outstanding personality who was able to integrate the current knowledge of a large variety of research fields and apply it to the advantage of both workforce and industry.

Perhaps it is no surprise that all these activities in research and industry paved the way towards the foundation of an independent scientific association providing a roof for a variety of different fields. In 1953 the impetus was given by Kellner (Munich), who was then a leading official in the Bavarian Ministry of Labour and Social Affairs. The foundation assembly was held in Nuremberg on 16 October 1953. In contrast to already existing organizations, the foundation assembly put an emphasis on the promotion of research. However, particular strength and institutional power was to be given to the cooperation between industrial physiology, industrial psychology, work pedagogy, technology of work, and industrial sociology. The members agreed on the title Gesellschaft für Arbeitswissenschaftliche Forschung (Association of Ergonomic Research). The choice of the word ergonomic was intended to put special emphasis on the intention of a coordination between the individual research fields.

At the first general assembly in March 1954, some 44 specialists were able to participate. By 1998 the number of regular members had increased to 680. Theoretically there should be a much larger membership. However, several personalities in research and industry working in close connection with ergonomics are members of competing scientific organizations, such as the German Association of Industrial Medicine, the German Association of Psychology, the Professional Association of Psychologists, the German Association of Aerospace Medicine, the Association of German Engineers, and the German Association of Aerospace Engineering. It has to be considered a success that several outstanding personalities from these scientific organizations have been won over to GfA during the last 50 years. This can also be traced back to the fact that, following the tradition of the first scientific congress in 1954, a spring congress had been held every year. In 1958 the name Gesellschaft für arbeitswissenschaftliche Forschung was changed to Gesellschaft für Arbeitswissenschaft (Association of Ergonomics); this was to give a better reflection of its objectives, which cover basic research as well as the transfer of ideas into industry.

Scientific congresses aim at presenting new research results. But they seldom provide a suitable forum to discuss research concepts or methodological questions. In order to create a platform for these issues while giving younger members of research institutes a chance to discuss their unfinished work, an autumn conference has been held every year since 1954. This idea was initiated by Preuschen, who has hosted them in Bad Kreuznach for more than 15 years. As a rule, their scientific concept has been based on the current problems discussed in the fields of economic, labour market, and social policies. Here are some examples:

- higher age at work and office
- women at work
- problems linked to working hours
- young people in the work environment
- reduced performance and preretirement disablement
- definition and evaluation of work performance

- work performance and work environment
- contributions of ergonomics to automation
- application of ergonomic knowledge
- ergonomic aspects in office work
- work organization and motivation
- demand and strain
- systems ergonomics
- communication and cooperation
- management systems for occupational safety

In Germany the congresses have been held almost exclusively in various university cities, in Austria they are held in Vienna, and in Switzerland they are held in Zurich.

During its early years, GfA met with relatively little public interest in the German-speaking area. However, the industry associations and trade unions were interested in the research results in the fields of fatigue and recovery, impact of noise, climate or mechanical fluctuations on performance and health, and methodology of work and time studies. Research results from the wide field of equipment design, design of workplaces, and design of work environments did not attract much attention at that time. Nevertheless, there were ergonomic germ cells in the iron and steel industry, mining, and several producing enterprises. However, their tasks mainly involved work organization, wage definition, and staff selection.

By making it obligatory to follow established ergonomic knowledge in the concept and design of human work, a series of laws and state regulations have led to an increased public awareness of ergonomics from 1972 onwards. GfA profited from this development. On the one hand, the research results presented during the annual congresses have found a much wider audience; and on the other hand, they have stimulated discussion on the ideas and objectives of ergonomics.

In 1946, soon after the end of World War II, several leading personalities released the first ergonomic publication, *Zentralblatt für Arbeitswissenschaft und soziale Betriebspraxis*; their backgrounds were industrial physiology, industrial psychology, work pedagogy, industrial sociology, industrial technology, work organization, and worker protection. The quarterly *Zeitschrift für Arbeitswissenschaft* was founded in 1975 as an organ of GfA. It remains GfA's central publication and is produced alongside REFA's *Arbeitsgestaltung, Betriebsorganisation und Unternehmensentwicklung*. This editorial connection reflects the close relationship between GfA and REFA, the largest technical and scientific association in Germany.

In 1990 GfA created a medal for outstanding scientific achievement, and so far it has been awarded to the members Friedrich Fürstenberg, Walter Rohmert, Heinz Schmidtke, and Eberhard Ulich. Due to their special contributions to ergonomics, the members Hugo Hilf, Herbert Scholz, Gerhardt Preuschen, and Heinz Schmidtke were appointed as honorary members of the association. The Fritz-Giese-Preis donated in memory of Fritz Giese, one of the early pioneers of ergonomics, has so far been awarded to the members Heinrich Dupuis and Klaus Zink. And for several years now, awards have been given to young scientists who produce the best spring congress papers and posters.

The Gesellschaft für Arbeitswissenschaft is a founding member of the International Ergonomics Association.

2. MISSION

In 1998/99 GfA critically and intensively examined its mission and how it reacts to changes in basic conditions. Here is a summary of the discussions among its members. The Gesellschaft für Arbeitswissenschaft eV (registered society) wants to achieve a relevant contribution to the design of living and working conditions by promoting interdisciplinary and holistic research and design concepts. The analysis, evaluation, and design of human work as well as the design of products, services, systems, and environments suitable for human beings, all require a knowledge of different scientific and practice-oriented disciplines. Some of these disciplines are work organization and occupational psychology, occupational medicine, work physiology, industrial sociology, labor policy, vocational education and engineering sciences, economics, and jurisprudence.

GfA integrates this knowledge with a scientific basis. That means the uniqueness of ergonomics and human factors, hence the uniqueness of GfA itself, results from linking individual disciplines (work-related sciences) to achieve a holistic view. In this context holistic not only means the integration of individual disciplines, but also the compatibility of different goals, taking into account human and economic objectives of particular importance.

Work is considered human if it is suitable for people and can be done in a human way and is meant to have neither a negative effect on physical and psychological health nor a lasting and disturbing effect on human well-being. It should correspond to needs and qualifications and enable an individual and/or collective influence on work. Finally, it is necessary to pursue the development of personality by work according to the realization of potential and the promotion of competences. When assessing economic efficiency, one has to consider that solutions might be "efficient" at the level of an individual organization, but might turn out to be inefficient at an economic or social level. All these assessment levels should therefore be taken into account.

The specific task of GfA is the promotion of scientific and professional interests of ergonomics and human factors. For that reason, GfA fosters contact with those who are interested in ergonomics within the German-speaking area as well as at an international level. GfA sees itself especially as an active forum for the dialog between science and practice and all interested social groups. It develops and guarantees professional standards (e.g., by participating in standardization or as the German partner for certification as a Euro-Ergonomist, but also by assessment procedures for journal and congress contributions). As far as possible and as far as required, it represents the interests of its members in the professional field.

Here are its fundamental principles of realization:

- The (working) society is subject to continuous change, which requires an appropriate development of ergonomics and human factors. In that case goal conflicts and tensions will occur and be accepted as normal.
- Ergonomic research and design are always aware of the uniqueness of a human being.
- Ergonomic action aims at a design of work, technique, and organization which is oriented towards human and economic criteria.
- Ergonomic solutions are subject to a concept which is useful

to all stakeholders (e.g., employees, management, shareholders, society).

- Wherever possible, the persons affected by analysis, evaluation, and design will be involved in these processes.
- Ergonomics and human factors also deal with analysis, evaluation, and design of nongainful employment (housework, voluntary work, community work, etc.).
- GfA and its members feel obliged to the social goal and, wherever possible, they make contributions to the maintenance, creation, and (appropriate) distribution of human work that is economically and environmentally compatible.

GfA's target group is all "ergonomically acting or interested people." One group contains those who take on ergonomic tasks in practice, such as analysis, evaluation, and design of work; another group contains those involved in the design of products and processes. Finally, there are those who take responsibility for the realization (e.g., managers). Ergonomic knowledge will be acquired and disseminated by universities, high schools, and application-oriented research institutes; then completed and realized by the ergonomists within the organization.

Even associations, government institutions, and self-help organizations have employees working in the field of ergonomics and human factors. The target group of the GfA is accordingly broad. Here are the central points:

- occupational safety and health protection
- product design
- work and organization design
- organization development and labor policy

Starting from the fact that every person will be confronted with various forms of work in the course of his or her life and will have to deal with the results of work (according to products and services) every day, the broad spectrum of ergonomic topics will then become clear. GfA wants to help spread these ideas among its members and as many other people as possible.

This means GfA must integrate ergonomic knowledge into the education of all those who need these ideas either directly (e.g., production and product design, work organization and organization development, labor protection and work design, personnel development, and in the significant fields of labor policy) or indirectly (e.g., education and training, investment planning, purchase, human resources departments, and management functions). Apart from that, ergonomic knowledge and matters of concern should be imparted by schools of all-round education, but particularly by vocational schools

GfA supports the transfer of ergonomic knowledge by the organization of appropriate platforms (e.g., congresses, workshops, publications, and clearinghouses).

History of Work-related Musculoskeletal Disorders

M.J. Sanders

Department of Occupational Therapy, Quinnipiac College, Hamden, CT, USA

1. INTRODUCTION TO CUMULATIVE TRAUMA DISORDERS

Cumulative trauma disorders (CTD) are injuries that develop in the soft tissues structures of the body such as nerves, muscles, tendons and joints due to prolonged or repeated stresses to a particular part of the body. Such stresses may include awkward or static postures, forceful exertions, vibration and mechanical stresses and may be exacerbated by psychosocial or work organizational issues. While CTD have received notable attention in recent years due to the spiraling incidence in industry and its subsequent impact on industry profits and individuals' lives, CTD are not a new entity. CTD were first described in the 1700s and documented throughout the nineteenth and twentieth centuries as many countries endured an "epidemic" of CTD. These epidemics served as the catalyst to investigate the breadth and causative factors of CTD. Interestingly, the same risk factors proposed to influence the development of CTD in the eighteenth century are the same risk factors that are recognized today.

This chapter will provide an overview of the occurrence of CTD in each country that endured a significant rise of CTD. The discussion will include the focus of investigation in each country, the action taken and speculation as to the cause. This historical overview will illuminate the various theories that help to shape our perspectives of CTD today.

2. EARLY HISTORY OF CUMULATIVE TRAUMA DISORDERS

2.1. Eighteenth-Century Occurrences

Bernardo Ramazzini, known as the father of occupational medicine, was the first physician to chronicle the occurrence of CTD that arose in the workplace. He observed that many of his clients presented with symptoms apparently related to occupational exposures. To investigate these relationships, Ramazzini interviewed his patients about their work tasks, observed their workplaces, and related their physical manifestations to factors within their occupations. In doing so, Ramazzini clearly identified associations between his patients' use of their bodies and physical complaints.

In his *De Morbis Artificum Diatribe* (*The Diseases of Workers*; Wright 1940) Ramazzini (1717: 15) described the respiratory, dermatologic, musculoskeletal and emotional problems of his patients. He opened his treatise with the following observation:

> Various and manifold is the harvest of disease reaped by certain workers from the crafts and trades that they pursue: all the profit that they get is fatal injury to their health. That crop germinates mostly, I think from two causes. The first and most potent is the harmful character of the material that they handle for these emit noxious vapors and very fine particles inimical to human beings and induce particular disease; the second cause I ascribe to certain violent and irregular motions and unnatural postures of the body, by

reason of which the natural structure of the vital machine is so impaired that serious disease gradually develop therefrom.

He continued to describe the posture of scribes and notaries (1717: 421):

> The maladies that afflict the clerks afore said arise from three causes: First, constant sitting, secondly the incessant movement of hand and always in the same direction, and thirdly the strain on the mind from the effort not to disfigure the books by errors or cause loss to their employer when they add, subtract or do sums of arithmetic.

Finally, regarding bakers, he noted (1717: 229), "Now and again, I have noticed bakers with swelled hands, and painful, too; in fact, the hands of all such workers become thickened by the constant pressure of kneading the dough."

Ramazzini poignantly described each task or occupation with careful attention to the potential causes and effects of long-term exposure to toxins, stressful tasks or postures in the workplace. As early as 1717, Ramazzini identified that "constant sitting," "incessant movement of the hand," "strain on the mind" and "violent and irregular motions" were critical factors in predisposing patients to the neuromuscular problems seen today.

Later, in the mid-to-late 1800s, physicians described and designated upper extremity syndromes such as writer's cramp, seamstress' cramp, scrivener's palsy and telegraphists' cramp. These conditions were described as having both physical and psychosocial origins; many were recognized after the advent of technological advances (i.e. telegraphists' cramp coincided with the introduction of Morse code) (Melhorn 1998).

2. CUMULATIVE TRAUMA DISORDERS IN THE TWENTIETH CENTURY

2.1. The Industrial Revolution

The next well-documented occurrences of CTD appeared in the early 1900s during the Industrial Revolution as workers were enticed into higher wages, in exchange for repetitive tasks, assembly-line pacing, and overtime work. When worker's compensation laws upheld coverage for work-related injuries, such as tenosynovitis, insurance companies began to record and further investigate these injuries as related to their clients' occupations. Physicians became the purveyors of decisions regarding worker compensation awards and were compelled to compile data that equated neuromuscular symptoms with work-related activities (Conn 1931).

Evidence of cumulative trauma disorders was most apparent in hand intensive tasks. Early practitioners found that "high speed hand operations," "increased intensity of effort" and "being new to the job" predisposed workers to tenosynovitis (Conn 1931). Attempts to identify the parameters for tolerances of human tendons during repetitive tasks concluded that tenosynovitis would occur in human tendons if repetitions > 30–40 per min or 1500-2000 per h (Hammer 1934).

2.2. The Automated Office

As automation replaced manual labor, work became more efficient and productivity requirements became higher. Workers no longer partook in the "mini-breaks" inherent in operating traditional manual office equipment such as retrieving a typewriter carriage, changing paper or delivering a message. In the automated office, "high speed typewriters" and data-entry machines replaced such

tasks with a keystroke. Workers overloaded the small muscles of the arms and hands instead of using a variety of positions to accomplish the job. In essence, the automated office eliminated any extraneous movements not directly related to the task so that work became streamlined, routine and monotonous. Although each task demanded high levels of concentration, workers became disengaged from the job as a whole.

Hence, the "automated office" of the 1950s emerged as a problematic area for the development of neuromuscular problems. In 1960, at the Fifth Session of the International Labor Organization Advisory Committee on Salaried Employees and Professional Workers office personnel reported physical and "mental fatigue" problems such as eye strain, pain and stiffness in cervical regions, and numbness in the right hand. These complaints prompted an organizational review with ensuing recommendations (ILO Advisory Committee 1960). After the 1960s, many industrialized countries reported an "outbreak" of CTD at different times during their histories.

3. OCCUPATIONAL CERVICOBRACHIAL DISORDERS IN JAPAN

Japan reported a dramatic increase in occupational cervicobrachial disorders between 1960 and 1980. Workers such as typists, telephone operators and assembly-line workers reported pain in the hand and arm that interfered significantly with abilities to perform their jobs. These claims rose to such proportion that in 1964, the Japanese Ministry of Labor issued guidelines for keyboard operators, mandating that workers spend no more than 5 h per day on the keyboard, take a 10-min rest break every hour, and perform < 40 000 keystrokes per day. Companies that implemented these measures reported a decrease in the overall incidence of neuromuscular disorders from 10–20 to 2–5% per industry. However, the overall incidence of CTD continued to rise (Maeda 1977).

Japan subsequently formed the Japanese Committee on Cervico Brachial Syndrome to define the syndrome and fully identify contributing factors. The committee conducted a thorough screening of individuals with neuromuscular complaints (Maeda 1977: 200) and concluded that "how the workers use their muscular and nervous systems at work" and "how the task is organized into the work system as a whole" underlie most conditions. Specifically, researchers identified static loading of postural muscles, dynamic loading of the small arm and hand musculature and muscular fatigue as contributing factors to these disorders. The conditions were found to worsen with insufficient recovery time and excessive workload.

The Japanese committee proposed the name occupational cervicobrachial disorder (OCD) to reflect acknowledgement of problems with both proximal and distal structures. They defined the problem as a functional or organic disorder resulting from mental strain and/or neuromuscular fatigue due to performing tasks in a fixed position or with repetitive movements of the upper extremity (Maeda et al. 1982). Maeda et al. astutely identified a fundamental controversy that exists today; that is, whether OCD is caused solely by factors within the workplace or whether psychological factors such as personal anxiety or workplace stress are the core problems that become magnified by the physical aspect of the workplace.

4. REPETITIVE STRAIN INJURIES IN AUSTRALIA

In Australia the epidemic of repetitive strain injuries (RSI) started in the 1970s and peaked in 1985 with up to 35% of telecommunication workers reporting symptoms of arm pain and muscular fatigue. The "epidemic" spread to include data processors, process workers, bank tellers, musicians and workers in clerical and textile industries. While repetition was initially suggested as the causal factor medical studies could not establish a correlation. Investigators therefore proposed that "psychological rather than physical factors within both the work and social environments ... contributed to the outstanding prevalence" (Ferguson 1971: 284).

The Commonwealth Government of Australia organized a task force on RSI to investigate the factors involved. The task force concluded that both psychological and ergonomic factors contributed to RSI. However, many investigators struggled with whether RSI was a separate disease entity or a grouping of other conditions with similar occupational underpinnings.

The predominant belief holds that the RSI epidemic was based on psychosocial origins. Ireland (1998) chastised the Australian medical community for insinuating that RSI is a medical rather than psychological phenomenon and echoed Ferguson's (1971) and others' beliefs that repetitive strain injury is a sociopolitical phenomena rather than a physical entity. Ireland suggests that RSI is a "collection of inconsistent symptoms" with no objective clinical findings or lab findings to substantiate tissue damage. Ireland attributes the Australian epidemic to a prosperous economy with a waning work ethic, technological changes that increased efficiency but led to job deskilling and job dissatisfaction, and difficulty for older workers to adapt to these changes. Ireland submits that the legal, healthcare, and printed media contributed to the perception that this condition was a physically based injury.

5. OCCUPATIONAL DISORDERS IN SWEDEN, FINLAND AND NORWAY

Scandinavia is considered to be progressive in its laws that protect workers from environmental and ergonomic hazards. Since World War II legislation has encouraged the cooperation of labor and management in issues regarding the work environment (Jensen 1997). The Danish Working Environment Act 1975 emphasized that management held the responsibility for healthy and safe working conditions and that this tenet must be fulfilled in cooperation with employees. Norway and Sweden have corresponding legislation. Although this legislation is in place, these issues are not always fully supported, especially in smaller firms (Jensen 1997).

The Scandinavian countries changed their foci in the 1980s from problems associated with the low back to those associated with upper extremity discomfort in response to increasing complaints of neck and shoulder pain among blue-collar workers (Kvarnstrom 1983). Nordic researchers performed careful studies to identify the contributing factors to neck and arm pain and found significant associations with the type of work being performed (i.e. repetitive, monotonous) and demographic features such as family situation, participation in leisure activities, having a sick spouse, or marital status (Kvarnstrom 1983). The Nordic Council of Ministers recognized the difficulty of comparing studies from country to country and therefore devised the standardized Nordic questionnaire to enhance their ability to compare and compile information among countries.

6. CUMULATIVE TRAUMA DISORDERS IN NORTH AMERICA

Finally, the USA witnessed a gradual rise in CTD from 1980 to 1986 after which point the incidence of reported CTD skyrocketed from 50 000 in 1985 to 332 000 in 1994. This sharp increase was largely attributed to cases in the meatpacking and equipment manufacturing industries. However, a shift to service industry jobs coupled with media awareness has probably contributed to the increase in reporting (Melhourn 1998).

To address the rising incidence of CTD in absence of an ergonomic standard, the Occupational Safety and Health Administration issued Ergonomics Program Guidelines for Meatpacking Plants in 1990 to assist industry in developing CTD prevention programs. The incidence of CTD began a slow descent back to 281 100 in 1996 as businesses began to acknowledge and manage their problems with CTD. The term "cumulative trauma disorder" has been largely used in the USA. However, the National Institute of Occupational Safety and Health (NIOSH) recently proposed the term "work-related musculoskeletal disorders" (WRMSD) to describe these disorders.

Initial investigations in CTD began in the USA with carpal tunnel syndrome in the 1980s and have gradually expanded to examining the incidence of other CTD in high-risk industries such as industrial workers, dental hygienists, electricians and video-display workers.

Research has focussed on delineating the risk factors for CTD, both physical and psychosocial, developing methods to study ergonomics exposures and examining dose-response relationships. In their seminal study, Silverstein et al. (1987) examined the relationship between force and repetition in industrial job tasks. Researchers found that workers in high-force, high-repetition jobs were 15 times more likely to have carpal tunnel syndrome than workers in low-force, low-repetition jobs. This study formed a basis for examining case definitions, methods, and interpreting results relative to CTD exposures. To date, practitioners agree that workers' exposure to a higher number of risk factors will increase the probability of developing a CTD. However, no dose-response relationships have yet been established (NIOSH 1997). Video display work has emerged to be an area widely investigated relative to psychosocial, work design and work organizational factors.

7. CHARACTERISTICS OF CUMULATIVE TRAUMA DISORDERS

Although various names have been attributed to CTD, worldwide investigations of these disorders collectively yield common characteristics. Researchers agree that the following implications that incorporate the work task, the work environment, design of the job and psychosocial characteristics of the individual all contribute to the development of CTD:

- The causes of CTD are multifactorial; they involve physical, psychosocial and work organizational factors. The relative contribution of each factor has yet to be determined.
- CTD are related to the intensity, frequency, and duration of ergonomic exposures; the level of risk for CTD depends on exposure to conditions and the individual capacities.
- Jobs that present multiple risk factors will have a higher probability of causing neuromuscular problems.
- Symptoms of CTD may include distinct features such as objective signs or non-specific symptoms such as pain. CTD present with or without physical manifestations.
- CTD develop insidiously; they may appear after months or years on the job.
- CTD recuperate slowly; they may require extended periods of recovery.
- CTD may reduce worker productivity and cause worker dissatisfaction (Kvarnstrom 1983, Silverstein 1987, NIOSH 1997).

The incidence, causal factors, preventive strategies and legislative acts concerning CTD continue to be explored worldwide. The globally collaborative efforts of researchers are making strides in standardizing assessments and increasing the rigor applied to studies related to CTD. These efforts will increase our understanding and future prevention of CTD in the workplace.

REFERENCES

FERGUSON, D., 1971, An Australian study of telegraphists' cramp. British Journal of Industrial Medicine, 31, 447–53.

HAMMER, A., 1943, Tenosynovitis. Medical Record, 140, 353–5.

INTERNATIONAL LABOUR ORGANIZATION ADVISORY COMMITTEE, 1960, Effects of mechanisation and automation in officers, III. International Labour Review, 81, 350.

IRELAND, D.C.R., 1998, The Australian repetition strain injury phenomenon. Clinical Orthopedics and Related Research, 351, 63–73.

JENSEN, P.L., 1997, Can participatory ergonomics become "the way we do things in this firm" - The Scandinavian approach to participatory ergonomics. Ergonomics, 40, 1078–87.

KVARNSTROM, S., 1983, Occurrence of musculoskeletal disorder in a manufacturing industry with special attention to occupational shoulder disorders. Scandinavian Journal of Rehabilitation Medicine, 8, 1–61.

MAEDA, K., 1977, Occupational cervicobrachial disorders and its causative factors. Journal of Human Ergology, 6, 193–202.

MAEDA, K., HORIGUCHI, S. and HOSOKAWA, M., 1982, History of the studies on occupational cervico brachial disorder in Japan and remaining problems. Journal of Human Ergology, 11, 17–29.

MELHOURN, J.M., 1998, Cumulative trauma disorders and repetitive strain injuries: the future. Clinical Orthopedics and Related Research, 351, 107–20.

NATIONAL INSTITUTE FOR OCCUPATIONAL SAFETY AND HEALTH, 1997, Elements of Ergonomics Programs. DHHS (NIOSH) Publication no. 97–117 (Cincinnati: US Department of Health and Human Services).

RAMAZZINI, B., 1717, De Morbis Artificum Diatriba; repr. and trans. W. Wright, 1940, The Diseases of Workers (Chicago: University of Chicago Press).

SILVERSTEIN, B.A., FINE, L.J and ARMSTRONG, T.J., 1987, Occupational factors and carpal tunnel syndrome. American Journal of Industrial Medicine, 11, 343–58.

Human Factors, Politics and Power

R. J. Badham

Department of Management and BHP Institute for Steel Processing and Products, University of Wollongong, NSW 2522, Australia

1. WHAT IS HUMAN FACTORS?

Human factors as an engineering discipline that concerns the design of equipment in accordance with the mental and physical characteristics of operators. Since World War II responsibility for seeing that equipment designers take the characteristics of operators and maintenance people into account in their designs has been formalized in the position of human factors engineer, human engineer or ergonomist (Perrow 1983). In more recent years the rise of information or computer-based technologies has led to a rapid increase in human–computer interface or usability engineering. Human factors research may be interpreted more broadly in terms of those disciplines or subdisciplines concerned with the psychological and organizational aspects of the development, implementation and use of technology (Clegg 1993). Narrower concerns with physical or cognitive ergonomics and direct health and safety issues are then broadened to include the impact of technology on the jobs of direct and indirect users, the consequence of technology for changes in organizational structure and processes, and the processes by which technology should be designed and implemented. Despite arguments for adopting a broader view of human factors, survey and case study research carried out for the National Science Foundation's Manufacturing Processes and Equipment Program confirmed that health and safety and ergonomics continue to be the main human factors concerns among US firms (Lund *et al.* 1993).

2. MARGINALIZATION OF HUMAN FACTORS CONSIDERATIONS

Human factors professionals often point to a general failure adequately to consider human factors in technology projects, and commonly attribute this to a lack of awareness or interest in human factors issues by design engineers or the lack of sophisticated quantitative evidence for human factors judgements or easily usable human factors tools and methods. While some explanatory power must be given to such conditions, analysts of organizational power and politics do not regard them as the main reason for the frequent marginalization of human factors considerations. The neglect of human factors is, rather, attributed to organizational factors, in particular the ideas and interests of different affected groups and the distribution of power and resources between them. Perrow (1983) points to the lack of incentives and legitimization for human factors considerations provided by senior management; the culture, training and resources of design engineers; the lack of organizational authority of human factors professionals; and the relative weakness of operators or users of technology who are most immediately disadvantaged by inadequately designed technologies. Clegg (1993) observes that broader organizational, institutional and educational systems have evolved and operate to marginalize human and organizational issues without the deliberate intervention of any particular individual or group. He points in particular to the characteristics of technology development organizations, user firms, education and training cultures, regulatory institutions and public funding bodies. These institutions systematically foster: disempowering attitudes towards end users and a lack of end user skills, knowledge and organization; technology development processes that have goals, design criteria and control mechanisms that foster a narrow technical orientation; technology investment and commissioning practices that systematically undervalue human factors; an education and training system that is biased towards technical skills and creates two antagonistic scientific and humanistic cultures; and a research and development profile that under-resources and narrows the scope of human factors research. The analysis of such influences and how to address them leads the human factors professional into the broader area of organizational power and politics.

3. ORGANIZATIONAL POWER AND POLITICS

Power and politics are not straightforward terms and are highly contested. A key feature of power is, however, the capacity to produce intended effects in line with perceived interests. It is the ability to get things done your way, with the cooperation of others or in the face of their opposition. Political behavior can be seen as the practical domain of power in action, worked out through the use of techniques of persuasion, influence, coercion or manipulation. The roots of such political behavior lie in personal ambition, in organization structures that create roles and departments which compete with each other, and in major decisions which cannot be resolved by reason and logic alone but which rely also on the values and preferences of the key actors involved. Organizational politics is fostered by the existence of vertically and horizontally differentiated groupings, each of which develops its own culture, interests, procedures and practices. Typically, individuals within these subunits, their leaders, and the groups themselves, compete for scarce resources. They interpret organizational events and choices, and attempt to bring about changes, in line with their particular perceptions of their interests and those of the organization.

This view of power and politics directs attention towards the deliberate strivings of individuals and groups in situations of conflict over human factors issues. In line with commonly recognized prescriptions for political "stakeholder" analysis, it has often been noted that it is not in the career interests of highly mobile senior managers to invest heavily in enduring workable systems, that design engineers with different career aspirations control more of the resources and information relevant to human factors considerations than human factors professionals themselves, and that users are often too busy, uninterested or restrictively focused on potential negative consequences for pay and employment to participate effectively throughout the technology development process. Underlying these more obvious political influences are, however, the organizational and structural conditions that create an institutional "mobilization of bias" against the full consideration of human factors issues. Such a bias, consciously or unconsciously promoted within and between organizations, has been referred to as the "second face" or second and third dimensions of power (Buchanan and Badham 1999).

Examples of the mobilization of bias include such issues as

why senior managers are often moved onto other jobs by the time that the technologies they were responsible for acquiring are fully operational. It includes the causes and consequences of factors such as the fact that US managers rotate more frequently than Japanese managers, giving them less of a vested interest in the long term health and performance of their particular areas of responsibility. Perrow (1983) points to broader "error inducing" or "error avoiding" cycles built into the institutional contexts within which senior managers operate. The air transport industry, for example, has an "error avoiding" cycle as performance failures have a relatively immediate effect on profits and reputation, failures receive broad media coverage and are thoroughly and openly investigated by a number of parties, and the causes of failure has significant legal and insurance consequences. Moreover, users of the air transport system can go to other users quite easily, and technological fixes are often not that complex or expensive. In contrast, in the marine transport industry, the international system is made up of conflicting national interests, there are few effective regulations, economic losses are absorbed with little notice and passed onto the final consumer, and human losses are initially restricted to officers and crew, and production pressures are high and competition is a comparatively under-regulated, and the causes of accidents are difficult to determine and impeded by national interests, weak regulatory agencies, weak unions, and the lack of monitoring data.

The perceptions, interests and actions of design engineers are also affected by the broader cultural and institutional settings within which they operate. Engineers are educated in different subdisciplines and within different national contexts, and they work in different departmental and organizational locations with different objectives and constraints. Manufacturing and industrial engineers are often more sympathetic to technology implementation matters than design and development engineers, Scandinavian engineers operate within a training and legislative environment more conducive to considering human factors than their UK counterparts, Japanese manufacturing engineers are often noted for their greater attention to continuous improvement of operating practices than US engineers, influenced by their rotation through operational jobs, more frequent to line management, and less interfirm job mobility.

Human factors professionals may also operate within very different organizational conditions. In those organizations that have more or less strong human resource management representation at senior management level, initiatives such as human resource development or occupational health and safety strategies may alter the criteria against which the work of human factors professionals are judged. Many technology projects, for example, are used as initiatives for cultural change in organizations committed to human resource development, and ergonomic and health and safety considerations then become symbols of management's commitment to people.

Finally, the users or operators themselves differ considerably in character and degree of skill and power. Levels vary considerably in the degree of trust in employer/employee relations, unionism, dependence of employers on employee motivation, skill and expertise etc., all of which strongly affect the will and skill of operators towards participation in technology development. Airplane pilots are, for example, both highly skilled and unionized, and able to devote considerable resources to fighting

allegations of operator error as the source of aircraft malfunctioning and crashes. Representatives of user firms come into conflict with vendors or engineering development companies, line managers conflict with engineering departments, and maintenance people and direct operators each have their own interests and perspectives (Salzman and Rosenthal 1994).

4. POLITICS AND THE ART OF INFLUENCE

Much of the literature on organizational politics has been influenced by a negative view of politics as a black art, something to be reduced or avoided. This interpretation has been dominant among a number of theoretical traditions that have influenced human factors research, such as organizational design and development and socio-technical theory (Buchanan and Badham 1999). The predominance within human factors research of a harder scientific or engineering approach to research and education has tended to encourage such a view of politics as an unfortunate distraction or disruption of the real scientific work of the day. In contrast to such a view, however, has been a more positive view of organizational politics as the techniques and art of getting things done in organizations. Positive political skills are increasingly important when changes are being proposed that cut across organizational boundaries and involve people in changing their traditional patterns of behavior.

As the introduction of technology creates or requires behavioral change, it will inevitably trigger organizational politics. The human factors professional is inevitably part of this process. Badham (1993) argues that human factors professionals need to improve their understanding and skills in this area if they are to be effective actors in technology implementation. Pettigrew (1974) observes that "Specialists do not merely advise, they persuade, negotiate and exercise the power they can mobilize" (p. 27). In so doing they utilize five power sources: expertise; control over information; political access and sensitivity; assessed stature; and the amount and kind of groups support given to the specialist by his colleagues in his own and related specialist groups. The human factors professional, like other specialists, needs to establish credibility if he or she is to be effective. This inevitably involves *anticipating* the varying needs, expectations and reference groups of different groups of executives and specialists involved in or affected by a human factors project. Those specialists who work on their own tasks, become preoccupied with the intricacies of their own expertise, and only see clients when task issues are involved is unlikely to be able to anticipate such needs very well. Successful specialists develop multiplex relationships with other significant partners or clients in a project, and succeed in demonstrating competence in areas salient to the other actors. Buchanan and Badham (1999) argue that such "power skills" should be part of the skills of all professional innovators.

5. THE POLITICS OF TECHNOLOGY

It would be a mistake to restrict the political dimension of human factors research to a question of implementation. It also enters the fundamentals of technological design, what is considered valid knowledge, and criteria against which such goals as system productivity and success are to be measured. What counts as "manufacturing knowledge" is the product of the power relations in which it is constructed (Gillespie 1993) and technologies are shaped by the "technological frames" (Bijker 1995) of the

groups involved in their design and operation. As Perrow (1983: 534, 540) remarks, "The design of systems, and the equipment that is used, is not entirely determined by technical or engineering criteria; designers have significant choices available to them that will foster some types of social structures and operator behaviors rather than others. . . choices that are taken for granted because they are a part of a largely unquestioned social construction of reality — one that should be questioned." Despite such appeals, there have still been few detailed case studies about the actual political processes involved in introducing human factors considerations into technology design and implementation. Human factors professionals need to be aware of the nature of such processes, the political nature of the beliefs and interests of groups that shape technology and affect the role of human factors, as well as the bias inherent in their own assumptions and values.

REFERENCES

BADHAM, R., 1993, New Technology and the implementation process, *International Journal of Human Factors in Manufacturing*, 3, (1), 3–13.

BIJKER, W.E., 1995, *Of Bicycles, Bakelites, and Bulbs: Toward a Theory of Sociotechnical Change* (Boston: MIT Press).

BUCHANAN, D. and BADHAM, R., 1999, *Power, Politics and Organizational Change: Winning the Turf Game* (London: Sage).

CLEGG, C.W., 1993, Social systems that marginalize the psychological and organizational aspects of information technology, *Behavior and Information Technology*, 12, 261–6.

GILLESPIE, R., 1993, *Manufacturing Knowledge: A History of the Hawthorne Experiments* (Cambridge: Cambridge University Press).

LUND, R.T., BISHOP, A.B., NEWMAN, A.E. and SALZMAN, H., 1993, *Designed to Work: Production Systems and People* (Englewood Cliffs: Prentice Hall).

PERROW, C., 1983, The organizational context of human factors engineering. *Administrative Science Quarterly*, 28, 521–41.

PETTIGREW, A.M., 1974, The influence process between specialists and executives. *Personnel Review*, 3, 24–30.

SALZMAN, H. and ROSENTHAL, S., 1994, *Software by Design* (New York: Oxford University Press).

Human–Machine Systems: Written and Symbolic Communication

S.J.P. McDougall and D.J. Oborne
Department of Psychology, University of Wales Swansea UK

1. INTRODUCTION

The purpose of communication is to convey information. Communication is a process involving two (or more) information-processing individuals or machines. The need for effective communication between individuals and machines is at the heart of human—machine systems. Both words and symbols can form the basis of a visual language through which information is conveyed on interfaces.

The process underlying communication is best understood when one realizes that the linguistic origin of the word is the Latin "communis" meaning common. An essential prerequisite of communication is a common understanding by those involved in the communication process. Oborne (1995) points out that the role of interface designers is to arrange a situation which enhances the chances of the transmitter and the receiver having a maximum common understanding.

Influential psycholinguistic theories have suggested that the communication process is driven by the need to achieve the best possible communication for the least possible communication effort (Sperber and Wilson 1986). This constraint applies equally when communication is between individuals or between human and machine. Interfaces that use written or symbolic communication should therefore be designed in order to make the interpretation of information as easy as possible. Usability considerations therefore need to focus on how meaning can best be represented to allow effective communication via the interface.

2. REPRESENTATION OF MEANING

2.1. Similarities between Symbols and Words

Although icons appear very different from words, they share the same historical roots. Horton (1994) points out that the use of visual symbols is not new and, in fact, is one of the earliest forms of written language. Simple pictorial images have been used to convey information and tell stories in many cultures. The earliest true writing systems were "logographic" (from the Greek "logos" meaning word) and were based on a one-word-one-symbol principle. The alphabetic systems now commonly used came about as visual symbols began to represent sounds rather than words (see figure 1).

2.2. Differences between Symbols and Words

Symbols, by their very nature, are more ambiguous than words. When reading, words and letters act as a code which allows us to arrive at meaning. The code is based on sound and we use our knowledge of letter—sound rules to work out what words refer to (e.g. "cuh-aah-tuh" for cat when we are young). Providing we are skilled readers, the rules governing word "decoding" allow us to access meaning for all but the most difficult and rare words. Reading therefore provides unambiguous access to meaning in a way that is often not possible with symbols.

When symbols are used on interfaces there are no definite rules which allow us to decode their meaning. As a result, users often rely on visual associations to understand them. One of the keys to success of graphical user interfaces is the ease with which they enable use of visual associations by capitalizing on the user's existing world-knowledge. Typically collections of symbols or icons are created which represent real world objects. The user can then employ their existing knowledge about the nature of these objects and their interrelationships in order to understand the interface. One of the earliest and most influential systems using this visual metaphor was the Xerox "Star" office workstation (Smith *et al.* 1982). Symbols were designed to resemble papers, folders, filing cabinets, and mail boxes. Relationships between these icons and their meanings were much clearer because their functions related directly to operations performed on physical objects found in an office. This model underpins current interface design and provides a useful way round the problem of ambiguity.

Where the visual metaphor is used, there is usually a direct visual relationship between symbols and their meanings (e.g. when a picture of a house indicates returning to a home page within a series of displays, see figure 2). In other instances the relationship is implied rather than direct and inferences have to be made in order to arrive at meaning (e.g. figure 2(ii) uses our knowledge that rabbits run quickly to indicate that a system is running a process quickly). In other instances, the relationship is arbitrary (e.g. figure 2(iii) indicates no entry) and it is only our familiarity with the symbol that allows us to interpret it.

2.3. Symbols versus Words

Although symbols are inherently more ambiguous than words, the use of symbols on interfaces has mushroomed over the past two decades. This applies, not only to computer interfaces, but to a huge variety of other interfaces (e.g. cars, video players, and washing machines; chemical, oil, and nuclear process control; public information signs). At the very least, symbols are used as an adjunct to written instructions or function labels, but on many occasions completely replace them. Underlying this change is the fact they make communication possible with little effort, or prior knowledge, on the part of the user.

Ancient Egyptian hieroglyph	Sinai script	Moabite stone	Early Phoenician	Greek	Roman
				A	A
				M	M
					O

Figure 1: Evolution of the alphabet
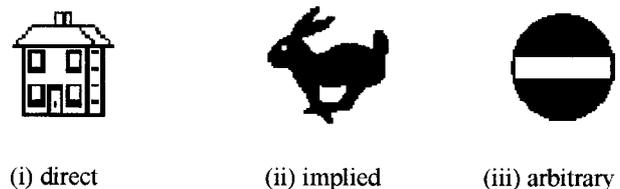
(i) direct (ii) implied (iii) arbitrary

Figure 2. Symbols with direct, implied and arbitrary relationships to their meaning

Graphical user interfaces have increasingly replaced command-driven text-based interfaces. This is because users do not have to rely on remembering and reproducing the correct command but simply need to recognize, or be able to guess, which symbol represents an appropriate function. Recall is much more effortful than recognition and this is reflected in users' performance. Users learning a graphical interface will make fewer errors, require less task-solving time and express more satisfaction than those using a text-based interface. Even where menus reduce the effort involved, novice users show a strong and consistent preference for symbol-based interfaces. Despite the popularity of graphical user interfaces, it is important to remember that written instructions still appear almost everywhere. In practice, text and symbols or graphics are often combined on interfaces. A number of general considerations determine the balance of their use on the interface. These are as follows:

- Symbols offer a universal and international mode of communication and will be used where users may not be skilled readers or share common language (e.g. public information signs).
- Symbols can sometimes convey information in a more spatially condensed way. This can be particularly important where clutter on the interface is likely to be a problem.
- Symbols and graphic displays are better when spatial information needs to be conveyed (e.g. in cockpits).

Written information is important where ambiguity needs to be resolved. Even where there are direct relationships between symbols and their meaning, 20% of users will not understand their meaning when they first encounter them so written labels are important for ambiguity resolution. Often users may initially scan interface symbols to find appropriate information and thereafter will use written labels to determine whether this is in fact the correct part of the interface.

3. DESIGN PRINCIPLES

The manner in which symbols, words, and text are displayed on an interface will either enhance or reduce the chances of effective communication taking place. The design principles employed when using words or symbols will now be considered.

3.1. Words and Text
3.1.1. Legibility
The distance at which the text will typically be read will largely determine type size. The font size should be large enough to be easily legible since type sizes which are too small increase reading times. Hard and fast rules are not appropriate since legibility depends upon context of use. Warnings, for example, need to be legible at a safe distance and placed where they will be noticed. Legibility is increased if simple type faces are used and text consists of a mixture of upper and lower case letters.

3.1.2. Sectioning prose
If text giving instructions uses prose, then dividing the text into meaningful sections allows the material to be structured both perceptually and conceptually since it informs the reader where one set of ideas ends and another begins. Sectioning of prose can be made more prominent by highlighting using increased display luminance, by typographical techniques such as underlining and emboldening, or by coloring different sections. Although the evidence to date suggests that highlighting may have beneficial effects on reading performance and comprehension, caution should taken about the overuse of such techniques since they will only create too many cues for readers and defeat the purpose for which they were originally intended.

3.1.3. Size of display area
While dividing text into distinct areas facilitates the reader's conceptual understanding of the text, the size of the text area relates to lower-level comprehension processes. Width of text should not be too narrow since it will make visual sampling of the text laborious and make comprehension more difficult. By the same token, lines of text which are over long will mean that the reader will take a long time to go back to the beginning of the next line and this will also affect comprehension.

3.1.4. Conciseness and clarity
It is important to be as concise as possible when providing instructions in text since users are less likely to read long sets of instructions. When text is used as a label for icons, long labels should be avoided because they are less likely to be meaningful and will create unnecessary clutter on the screen. Wherever possible use of abstract words should be avoided since their meaning is often "fuzzy" and difficult to define. Concrete words are clearer and more memorable.

Metrics have been used to estimate the "readability" or difficulty of texts. On average, difficult text has a higher percentage of longer words and sentences. This is because longer words tend to be less frequently used and are less familiar and longer sentences place greater processing demands on working memory which has direct consequences for comprehensibility.

3.2. Symbols
3.2.1. Visual complexity
Symbols should be kept as simple as possible since this reduces the overall complexity of the interface and helps make symbols and icons legible and easy to find. If symbol complexity cannot be avoided, increasing their size is important for legibility and this means that fewer can be displayed on the interface. As with text, size requirements to allow symbols to be easily seen will depend on the users' typical viewing distance and the quality of the interface.

3.2.2. Organization and layout
The display should be organized in a way that is compatible with the functionality of the system and allow the user to structure their understanding accordingly. For example, many software packages display clusters of symbols associated with particular functions. Color helps to direction our attention to appropriate parts of the displays and can be used partition displays or to indicate functional relationships between symbols.

As users learn an interface layout, location is likely to become an important cue for symbol search and, if symbols are organized into sensible "conceptual" clusters this will aid both perceptual and conceptual processing. The display should be organized in a way that is compatible with users' likely long-term understanding of the system, i.e. it should be cognitively compatible with the user to allow effective communication to take place.

3.2.3. Shape

Typically, shape is thought to help with visual perception of a display rather than with communicating meaning. When symbols in a set are created so that they differ in shape, this will make them easier to discriminate and reduce search times. Designers can make use of shapes with conventional meanings to represent functions.

3.2.4. Clarity

As we have already noted, the ease with which an interface might be understood is often determined by the pictorialness of the symbols used and the degree to which users might have access to visual metaphors with the real world. The importance of pictorialness decreases as users become more familiar with a system and rely on their familiarity with the system, rather than the visual metaphor, for understanding.

4. CONTEXT AND EXPECTATIONS

The context in which words or symbols are presented may lead to differing interpretations of a system interface. Figure 3 shows a simple example of how meaning can be altered by changing the context. In (i) symbol represents a geometric shape, whereas in (ii) it represents one level of shading. Similar effects can be observed with text as is shown in the sentences below:

> There was a tear in her jacket.
> A tear rolled down his cheek.

Although the words in each sentence are identical, the meaning is changed by the sentence context.

Context is important because it sets up expectations about what is likely to happen next, what procedures we need to follow, and what we can expect of a machine or system. For each contextual situation we have a series of expectations, or "slots", which we use to interpret and communicate in that context. This kind of mental structure is most often referred to as a schema (Schank and Abelson 1977). If users have access to an appropriate schema, it will be an effective tool in reducing the amount of processing effort we need to make in order to interpret information. Dixon (1987) measured how long people spent reading instructions that specified drawing simple geometric shapes which combined to make a picture. When the picture was mentioned at the beginning of the instructions (allowing access to a schema and a set of expectations and procedures), then instructions were read more quickly and drawings were more accurate. These expectancies have been effectively utilized when the visual metaphor is employed on

(i) geometric shape   (ii) shading/contrast
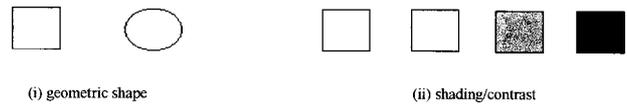
Figure 3. The effects of context on symbol meaning

graphical interfaces to reduce the processing required to arrive at an understanding of the interface.

The converse also needs to be considered in design. People will respond more slowly to events or signals which they are not expecting because they require more processing time. For example, system failure usually occurs rarely and users often respond to slowly or inappropriately. This is because users take time to process these events and do not have a set of expected procedures which they might draw upon in order to respond appropriately. Under these circumstances, designers need to ensure that rare events are made salient and obtrusive on the interface in order to allow users to respond effectively. Wherever possible the interface should indicate clearly what that response should be.

5. CONCLUSION

As can be seen from our consideration of the effects of expectancies, effective communication between human and machine is about producing an interface which creates appropriate expectancies in the user. If the interface is successful in accessing appropriate knowledge schemas, then communication will be facilitated and usability increased. Creating systems which are cognitively compatible with the user in this way is a key component of effective user-centered design.

REFERENCES

DIXON, P., 1987, The processing or organizational and component step information in written directions. *Journal of Memory and Language*, 26, 24–35.

HORTON, W., 1994, *The icon book: Visual symbols for computer systems and documentation* (New York: Wiley).

OBORNE, D.J., 1995, *Ergonomics at work: Human factors in design and development* (Chichester: Wiley).

SCHANK, R. AND ABELSON, R., 1977, *Scripts, plans, goals and understanding: An enquiry into human knowledge structures* (Hillsdale, NJ: Erlbaum).

SMITH, D.C., IRBY, C., KIMBALL, R., and VERPLANK, B., 1982, Designing the STAR user interface. *Byte*, 7, 242–82.

SPERBER, D. and WILSON, D., 1986, *Relevance: Communication and cognition* (Oxford: Blackwell).

Iceland: Icelandic Ergonomics Society

T. Svensdóttir

The Icelandic Ergonomics Society history is rather less ancient than the history of Iceland, which has been occupied for 1125 years. It was on 8 April 1997 that an ergonomics society was established in Iceland. The formal name of it is Vinnuvistfræðifélag Íslands or VINNÍS in short. The society has its address in Reykjavík. It might be of interest that in 1999 when the society was two years old the population in Iceland was 275 000.

The establishing of an ergonomics society in Iceland was at the initiative of physical therapists working in and interested in the field of ergonomics. A working committee was established in 1996 in cooperation with occupational therapists with the aim of establishing an ergonomics society . The concept of ergonomy was clarified with the assistance of the Nordic and the Swedish Ergonomics Societies statutes. The word "ergonomics" could not be accepted into the Icelandic language. It was therefore necessary to transform "ergonomics" into "vinnuvistfræði" (vinna = work, vistfræði = "ecology", i.e. ecology of work). The word "vinnuvistfræði" has a direct reference to the language and is synonymous with the international concept of ergonomics. We have defined ergonomics in the following manner:

"Vinnuvistfræði" deals with the interaction between human beings and the environment in which they live and work. The environment includes the workplace, equipment, organization of work, communication amongst other things with regard to human needs, well-being and security.

Various professional associations were contacted and encouraged to participate in the establishment of a multiprofessional society like VINNÍS. The result was the establishment of a society with 48 paying members with backgrounds in various professions such as medical, technical, and social science. Institutions, associations, and companies were also founding members. The members of the board are also of different professions.

The goal of VINNÍS is:
- to increase and disseminate knowledge on ergonomics in Iceland
- to facilitate the use of ergonomics in the design of the workplace, work organization, equipment and products.

Through the establishment of the society discourse on ergonomical problems from diverse points of view have been made possible. Members get in touch with other professionals which enables possibilities for direct communication, cooperation, and a flow of information.

The society is open to everyone interested in ergonomics and the working environment. Today there are 85 members from 11 different professional groups, institutions, companies, and associations. Private companies have status as supporting members. Less than 10% of the members have ergonomics as

Table 1. Membership composition by stated profession.

Profession	Total
Architects	4
Interior designers	6
Engineers	4
Ergonomist (MSc)	1
Occupational therapists	21
Industrial designers	2
Physicians	2
Psychologists	4
Physiotherapists	20
Sjuksköterskor	5
Sociologists	2
Other	3
Professional associations	2
Public institutions	4
Private companies	5
Total	85

their major occupation. The following table shows the diversity of members.

The first two years have been characterized by huge interest and enjoyable work. Members have been involved and we have been noticeable in society as a whole. Topic groups have been established with around 25 to 30% of members taking part. The groups themselves have chosen their topic of discussion, method of working and objectives. The following groups were established:
- School environment and school furniture
- VDU work
- Risk assessment in the work environment
- Indoor climate
- Study group on work environment
- PR group

The group on school environment and furniture has contributed considerably to the better ergonomical design of school furniture.

The board publishes two newsletters each year which are distributed to all members. Typical topics are information on VINNÍS activities, on the Nordic Ergonomics Society, latest courses and conferences, newly published books and periodicals on ergonomics. Two discussion meetings are held each year to discuss various issues in ergonomics. The annual meeting is hosted by one of the companies or institutions, to which we also make a site visit.

When VINNÍS celebrated its first year, the board and the PR group arranged an information week on radio and in the newspaper with the greatest circulation. We used the opportunity to draw attention to the society and its objectives. Open theme meetings with a lecturer have been organized twice a year in a café in the center of Reykjavík. Conferences are planned to be a part of the regular activities. In the autumn a half-day conference will be held on the theme "Work environment in a turbulent world".

VINNÍS has been a member of the Nordic Ergonomics Society for one year. It is very stimulating for us to be in contact with other Scandinavian countries with which we hope to work. Ergonomics is, as was stated in the beginning, a new concept for us in Iceland and there are not many who work directly in this field. Though the main emphasis in our Icelandic Ergonomics Society will be on ergonomical questions we will probably also include other work environmental matters. VINNÍS will strive to strengthen its economy in the near future and attract more

organizations and members with technological backgrounds. The future is exciting for us and for the development of ergonomics in Iceland, and we are committed to the future of our society. Those who are interested in further information, or want to make contact, can contact the chairman of the society, Torunn Sveinsdóttir, ph.: +354 567 2500 or e-mail: torunn@ver.is.

IEA Definitions of Ergonomics

1. INTRODUCTION

Ergonomics (or Human Factors) is the scientific discipline concerned with the fundamental understanding of interactions among humans and other elements of a system, and the application of appropriate methods, theory and data to improve human well-being and overall system performance. Derived from the Greek *ergon* (work) and *nomos* (laws) to denote the science of work, its scope extends to all human activity that involves the use of artifacts.

Ergonomists (ergonomics practitioners) contribute to the design of tasks, jobs, products and environments to make them compatible with the needs, abilities and limitations of people.

2. DOMAINS OF SPECIALIZATION

Ergonomics is a systems-oriented discipline and practising ergonomists must have a broad understanding of its full scope. That is, ergonomics espouses a holistic approach in which considerations of physical, cognitive, social, organizational, environmental and other relevant factors are taken into account.

Nonetheless, there exist domains of specialization within the discipline that represent deeper competencies in specific human attributes. These domains of specialization correspond to content knowledge about people rather than system attributes or economic sectors. Ergonomists often work in particular economic sectors or application domains, such as transportation and process control. However, application domains are not mutually exclusive and they change constantly (new ones are created and old ones take new directions), making it nearly impossible to define a useful and comprehensive set of application domains. Moreover, ergonomists can cross application domains and work effectively after an appropriate amount of familiarization. Hence, it is not useful to define application domains. It is far more difficult to cross-domains of specialization without extensive academic training.

The domains of specialization within the discipline of ergonomics include:

- Physical ergonomics is concerned with the compatibility between human anatomical, anthropometric, physiological and biomechanical characteristics, and the static and dynamic parameters of physical work. (Relevant issues include working postures, materials handling, repetitive movements, work-related musculoskeletal disorders, safety and health.)
- Cognitive ergonomics is concerned with mental processes, such as perception, human information processing and motor response, as it relates to human interactions with other elements of a system. (Relevant topics include perception, attention, workload, decision-making, motor response, skill, memory and learning as these may relate to human system design.)
- Social or organizational ergonomics is concerned with the optimization of work systems, including their organizational structures, policies and processes. (Relevant topics include human–system considerations in communication, crew resource management, work design and management, teamwork, participatory design, cooperative work, TQM.)

The International Ergonomics Association (IEA)

Secretary-General, IEA, Poste Restante, Human Factors and Ergonomics Society, PO Box 1369, Santa Monica, CA 90406-1369, USA

1. IEA MISSION STATEMENT

The IEA is the federation of ergonomics and human factors societies around the world. Working closely with its constituent societies and related international organizations, its mission is to elaborate and advance ergonomics science and practice, and to expand its scope of application and contribution to society to improve the quality of life. The following principal goals reflect the IEA mission:

- To develop more effective communication and collaboration with federated societies.
- To advance the science and practice of ergonomics at an international level.
- To enhance the contribution of the ergonomics discipline to global society.

2. ORGANIZATION AND GOVERNANCE

The IEA was organized pursuant to Article 60 *et seq.* of the Swiss Civil Code. The registered headquarters of the IEA is in Zurich with a business office in the USA.

The IEA has the non-governmental organization (NGO) status with the United Nations (UN), World Health Organisation (WHO), International Labor Office (ILO) and International Standards Organization (ISO).

The IEA is governed by a Council comprised of delegates from the member societies and by the Executive Committee of the Council. The IEA Executive Committee comprises of the elected Officers (President, Secretary-General, Treasurer), Chairs of the Standing Committees, Past President (non-voting), Newsletter Editor (non-voting), IEA Historian (non-voting), and the IEA Triennial Congress Chairperson (non-voting). The Standing Committees include: Policy and Development, Science and Technology, Professional Practice and Education, Industrially Developing Countries, Promotion and Publication, and Awards.

3. HISTORY

The origins of the IEA lie in a project initiated by the European Productivity Agency (EPA), a branch of the Organization for European Economics Cooperation. The EPA established a Human Factors Section in 1955, and in 1956 visited the USA to observe human factors research. The next phase of the EPA project was a technical seminar on "Fitting the Job to the Worker" at the University of Leiden in 1957. It was at this seminar that a set of proposals was presented for the formation of an international association of work scientists.

An informal (steering) Committee was appointed to develop the proposals and to organize an international association. The members of that committee were: Professor H. S. Belding, Professor G. C. E. Burger, Professor S. Forssman, Professor E. Grandjean, Professor G. Lehman, Professor B. Metz, Professor K.

U. Smith and Mr R. G. Stansfield. In various phases, members of the Ergonomics Research Society (UK) actively participated in the founding process.

Even though British scientists had founded an Ergonomics Research Society in 1949, "ergonomics" as the name of this new discipline became fully adopted only years later. The Committee decided to adopt the name "International Ergonomics Association" on a preliminary basis until a better name could be found. In 1957, the Council of Ergonomics Research Society (UK) agreed to support the establishment of the IEA.

A preliminary set of bylaws for the proposed international association was prepared and distributed to the informal committee. A special meeting was held in Paris in September 1958, and it was decided to proceed with organizing an association and to hold an international congress in 1961. The committee then designated itself the "Committee for the International Association of Ergonomic Scientists" and elected Professor Burger, President: Professor Smith, Treasurer: and Professor Grandjean, Secretary. Thus, even though the IEA was formally organized in Paris in 1958, the initial organization began with the naming of the informal (steering) Committee at the University of Leiden technical seminar in 1957.

The Committee for the International Association of Ergonomic Scientists met in Zurich during the last phase of the EPA project in March 1959 and decided to retain the name "International Ergonomics Association". The committee also met again in Oxford later in 1959 formally to agree upon the set of bylaws or statutes drawn up by the secretary with the help of committee members and the EPA staff. These statutes were later approved by the association membership at the First Congress of the IEA in Stockholm in 1961.

4. OFFICIAL JOURNALS

Ergonomics, published monthly by Taylor & Francis Ltd, London, is the official journal of the IEA. It published the proceedings of the first IEA Congress in 1961.

Beginning in 2000, the IEA is publishing the IEA *Electronic Journal of Ergonomics*.

5. IEA MEMBERSHIP

The IEA currently has 36 federated societies, two affiliated societies and two sustaining member organizations.

Federated Societies:
- All-Ukrainian Ergonomics Association
- Asociación Española de Ergonomia (Spain)
- Belgian Ergonomics Society
- Brazilian Ergonomics Association (ABERGO)
- Chinese Ergonomics Society
- Croatian Ergonomics Society
- Czech Ergonomics Society
- Ergonomics Society (UK)
- Ergonomics Society of Australia
- Ergonomics Society of the Federal Republic of Yugoslavia
- Ergonomics Society of Korea
- Ergonomics Society of South Africa
- Ergonomics Society of Taiwan
- Gesellschaft für Arbeitswissenschaft (Germany)
- Hellenic Ergonomics Society
- Association of Canadian Ergonomists/ACE

- Human Factors and Ergonomics Society (USA)
- Sociedad de Ergonomia y Factores Humanos de México (Mexico)
- Hungarian Ergonomics Society
- Indian Society of Ergonomics
- Inter-Regional Ergonomics Association (IREA: Russia)
- Irish Ergonomics Society
- Israeli Ergonomics Society
- Japan Ergonomics Society
- Nederlandse Vereniging voor Ergonomie (The Netherlands)
- New Zealand Ergonomics Society
- Nordic Ergonomics Society
- Österreichische Arbeitsgemeinschaft für Ergonomie (Austria)
- Polish Ergonomics Society
- Portuguese Association of Ergonomics (APERGO)
- Slovak Ergonomics Association
- Sociedad Colombiana De Ergonomía (Colombia)
- Società Italiana di Ergonomia (Italy)
- Societe D'Ergonomie de Langue Française (SELF)
- South East Asia Ergonomics Society (SEAES)
- Turkish Ergonomics Society

Affiliated Societies:
- European Society of Dental Ergonomics
- Human Ergology Society (Japan)

Sustaining Members:
- Bureau of the Hungarian Council of Industrial Design and Ergonomics
- Research Institute of Human Engineering for Quality Life (Japan)

6. OFFICERS OF THE IEA

The following individuals has served as elected Officers of the IEA:

1961–64:
- President: S. Forssman
- Secretary-Treasurer: E. Grandjean

1964–67:
- President: G. Lehman
- Secretary-Treasurer: E. Grandjean

1967–70:
- Professor: P. Ruffell-Smith
- Secretary-General: E. Grandjean
- Treasurer: A. Wisner

1970–73:
- Professor: B. Metz
- Secretary-General: F. Bonjer
- Treasurer: A. Wisner

1973–76:
- Professor: F. Bonjer
- Secretary-General: R. Sell
- Treasurer: J. de Jong

1976–79:
- Professor: A. Chapanis

- Secretary-General: R. Sell
- Treasurer: H. Scholz

1979–82:
- Professor: J. Rosner
- Secretary-General: H. Davis
- Treasurer: H. Scholz

1982–85:
- Professor: S. Sugiyama
- Secretary-General: H. Davis
- Treasurer: J. Rutenfranz/B. Shackel

1985–88:
- Professor: H. Davis
- Secretary-General: I. Kuorinka
- Treasurer: B. Shackel

1988–91:
- Professor: I. Kuorinka
- Secretary-General: H. Hendrick
- Treasurer: B. Shackel

1991–94:
- Professor: H. Hendrick
- Secretary-General: P. Rookmaaker
- Treasurer: I. Noy

1994–97:
- Professor: M. Helander
- Secretary-General: P. Rookmaaker
- Treasurer: I. Noy

1997–2000:
- Professor: I. Noy
- Secretary-General: W. Karwowski
- Treasurer: K. Kogi

7. TRIENNIAL CONGRESSES

The IEA has held Triennial Congresses as follows:
- 1961: Stockholm, Sweden
- 1964: Dortmund, Germany
- 1967: Birmingham, UK
- 1970: Strasbourg, France
- 1973: Amsterdam, The Netherlands
- 1976: College Park, USA
- 1979: Warsaw, Poland
- 1982: Tokyo, Japan
- 1985: Bournemouth, UK
- 1988: Sydney, Australia
- 1991: Paris, France
- 1994: Toronto, Canada
- 1997: Tampere, Finland
- 2000: San Diego, USA

The future IEA Congresses are scheduled as follows:
- 2003: Seoul, Republic of Korea
- 2006: Maastricht, The Netherlands

8. AWARDS OF THE IEA

Awards given by the IEA are presented at the Triennial Congress. They are:

- IEA Distinguished Service Award
- IEA Founders Award
- IEA Outstanding Educators Award
- IEA Ergonomics of Technology Transfer Award
- IEA Ergonomics Development Award
- IEA President's Award
- The Liberty Mutual Prize in Ergonomics and Occupational Safety
- IEA/K. U. Smith Student Award

The Awards Committee based upon nominations from the various federated societies selects recipients for the first five awards. Recipients for the IEA President's Award are nominated either by IEA Council or by the IEA Executive Committee: final approval for this award rests with the IEA President. The Student Awards Committee selects recipients for the K. U Smith Student Award

9. RECIPIENTS OF THE IEA AWARDS

IEA Distinguished Service Award:

- 1982: A. Chapanis, E. Grandjean
- 1985: M. Oshima, A. Wisner
- 1988: P. Davis, N. Lundgren, W. Singleton
- 1991: J. Rosner
- 1994: H. Davis
- 1997: H. Hendrick

IEA Founders Award:

- 1991: J. Scherrer
- 1994: K. U. Smith
- 1997: W. Floyd

IEA Outstanding Educators Award:

- 1991: E. N. Corlett
- 1994: W. Rohmert
- 1997: M. M. Ayoub

IEA Ergonomics of Technology Transfer Award:

- 1991: A. Wisner
- 1994: H. Shahnavaz
- 1997: R. Sen

IEA Ergonomics Development Award:

- 1991: K. Kogi
- 1994: J. Leplat
- 1997: D. Meister

IEA President's Award:

- 1997: T. Leamon

IEA/K. U. Smith Student Award:

- 1997: L. Ritmiller

REFERENCES

THE INTERNATIONAL ERGONOMICS ASSOCIATION, 1959, *Ergonomics*, 2, 400.

BONJER, F.H. *et al.*, 1958, *Fitting the Job to the Worker: Report on the Mission to the United States, 1956.* EPA Project 335, 1st phase (Paris: EPA/OEEC).

BURGER, G.C.E., 1958, *Fitting the Job to the Worker: Report on the Leiden Seminar, 1957.* EPA Project 335, 2nd phase (Paris: EPA/OEEC).

LeCOULTRE, D., 1985. Ergonomics and the European Productivity Agency: A historical note, *Ergonomics International*, 28, 1009–11.

METZ, B., 1960, *Fitting the Job to the Worker: International Conference of Zurich, 2nd–6th March 1959* (Paris: EPA/OEEC).

SMITH, K. U, 1988. Origins of ergonomics and the International Ergonomics Association, *Human Factors Society Bulletin*, 31, 2–5.

STANSFIELD, R.G., 1985, The founding of the International Ergonomics Association, *Ergonomics International*, 28, 1006–8.

International Ergonomics Standards: ISO/TC 159

W. Schultetus

Institut für angewandte Arbeitswissenschaft e.V., 50968 Köln, Germany

1. INTRODUCTION

Ergonomics is the study of human beings and their work. How can standardization help here? Do we want to standardize man? All ergonomists answer "no" to this question, as machines are supposed to adapt to human beings and not vice versa. However, to provide the means of achieving this objective, the human characteristics and abilities have to be researched and described. Only then can the builder of machines, the organizer of workplaces, and the designer of products for applications in industry, services, and the private sector take proper account of findings in the field of ergonomics. This, in turn, enables them to limit the workload borne by humans to a permissible level and to make a contribution to occupational health and safety.

Human engineering — designing technical products and services to take human characteristics into account — to improve the quality of life and work is becoming a central approach to making products and services more efficient, attractive, and competitive. Applied ergonomics is thus developing into a decisive market factor of the future.

The actual history of standards for ergonomics would have to begin with the first standard that was important for human activities, i.e. that took account of knowledge of human characteristics, abilities, and limits. This was done, for example, by setting the size or strength of the human body in relation to buildings, tools, the work environment, and work systems. Such standards certainly existed long before the concept of "ergonomics" found its way into standardization. It was gradually realized that standardization is an important and efficient way of allowing ergonomic findings to be applied. The International Ergonomics Association (IEA) was behind the establishment of ISO/TC 159, Ergonomics, back in 1975.

2. INSTITUTING STANDARDIZATION IN ERGONOMICS

The scope of ISO/TC 159 is standardization in the field of ergonomics including terminology, methodology, and human factors data. In the context of the scope, the committee, through standardization and coordination of related activities, promotes the adaptation of working and living conditions to the anatomical, physiological, and psychological characteristics of man in relation to the physical, sociological, and technological environment. Among the objectives are safety, health, well-being, and effectiveness.

However, many ISO standards involving aspects of ergonomics have not been developed by ISO/TC 159 for historical and organizational reasons. A survey carried out in Germany in 1996 concluded that more than 10 000 ISO, CEN and DIN standards contain aspects of ergonomics, even though the standard developers were unable to resort to fundamental ergonomic standards for the development of their own standards. This situation is now known and will be modified. The scope of

ISO/TC 159 promotes standardization in the adaptation of working and living conditions to the anatomical, physiological and psychological characteristics of man, with the objective of improving safety, health, effectiveness and well-being.

3. OBJECTIVES

Requirements for ISO/TC 159 to achieve its objectives are:

- to collect and critically review ergonomic data relevant to international standardization, pertinent to the design and manufacturing of machinery, the design and organization of work processes, and the layout of work equipment, as well as the control of the physical environment in the work premises;
- to identify those branches of industry, services, and trade where ergonomic needs will expand or arise with new technologies;
- to recognize the unavoidable time lags in producing as well as their enforcement;
- to set up and to implement comprehensive sub-programs for standardization activities in different fields of ergonomics;
- to make effective use of liaisons with other TCs and SCs through which standards with ergonomic specifications have been or are being produced, sometimes with insufficient consideration of granted ergonomic data and/or principles;
- to create within ISO/TC 159 a strategy planning function in charge of implementing and updating the present strategic policy statement.

Table 1. Organizational structure of ISO/TC 159, Ergonomics

(January 2000)

ISO/TC 159/SC 1, Ergonomic guiding principles

ISO/TC 159/SC 1/WG 1, Principles of the design of work systems

ISO/TC 159/SC 1/WG 2, Ergonomic principles related to mental work

ISO/TC 159/SC 1/WG 3, Terminology

ISO/TC 159/SC 3, Anthropometry and biomechanics

ISO/TC 159/SC 3/WG 1, Anthropometry

ISO/TC 159/SC 3, WG 2, Evaluation of working postures

ISO/TC 159/SC 3/WG 4, Human physical strength: manual handling and force limits

ISO/TC 159/SC 4, Ergonomics of human-system interaction

ISO/TC 159/SC 4/WG 1, Fundamentals of controls and signalling methods

ISO/TC 159/SC 4/WG 2, Visual display requirements

ISO/TC 159/SC 4/WG 3, Control, workplace and environmental requirements

ISO/TC 159/SC 4/WG 5, Software ergonomics and human-computer dialogues

ISO/TC 159/SC 4/WG 6, Human-centred design processes for interactive systems

ISO/TC 159/SC 4/WG 8, Ergonomic design of control centres

ISO/TC 159/SC 5, Ergonomics of the physical environment

ISO/TC 159/SC 5/WG 1, Thermal enviroments

ISO/TC 159/SC 5/WG 2, Lighting environments

ISO/TC 159/SC 5/WG 3, Danger signals and communication in noisy environments

4. WHAT STANDARDS ARE REQUIRED?

The first question asked at ISO was: What standards are required in the field of ergonomics? The scientists answered: "fundamental standards as guidelines for the design of work systems." But the product designers in industry replied: "standards with application-oriented, ergonomic details," — for example, making it possible to design the driver's compartment of a car or the lever on a machine. The organizational structure of ISO/TC 159, Ergonomics, tried to meet these different requirements (table 1).

Standards can be developed at various levels within the following structure:

- basic standards related to fundamental characteristics of man;
- functional standards related to human factors in the operation and use of equipment, processes, products of systems;
- environmental standards related to the effects of physical factors of the environment on human performance, in the range between comfort and health hazards;
- standards for test procedures and for processing ergonomic data, to be applied either in working out standards in the above three categories or in assessing conformity to already accepted standards.

Although these standards serve as aids for product standardizers, a considerable amount of ergonomic data is also required for machine safety. Thus close liaison with ISO/TC 199, Safety of machinery, became necessary.

Moreover, at the end of the 1980s and beginning of the 1990s, the European Single Market produced a demand for standards in the field of ergonomics. To this end, the European Committee for Standardization (CEN) set up Technical Committee 122, Ergonomics, which has liaised closely with ISO/TC 159 in many areas ever since.

5. RELEVANT AREAS OF STANDARDIZATION

The first ISO standard on fundamental ergonomics for the design of work systems, ISO 6385 (1981), Ergonomic principles in the design of work systems, was developed on the basis of the German standard, DIN 33400 (1975), Designing work systems according to ergonomic findings. In this fundamental standard, the committee attached central importance to the work system, to which it then assigned individual components, e.g. work equipment, workspace, work environment, and work organization. This top-down approach assumed that the framework first had to be established into which application-oriented standards could then be incorporated. It was clear from the outset that standards in ergonomics should not be drafted for just any application, but that fundamental standards were needed, providing the means for work instruments and products to be designed. The intention was, for example, to define the fundamental dimensions of the human body, instead of developing standards for anthropometric dimensions specifically applied to office chairs or car seats.

Standards for the design of work systems are therefore not only developed by SC 1, Ergonomic guiding principles, responsible for general principles, they are also included in the work of various subcommittees and working groups that deal with particular work systems, for example, office systems or control centers for process control.

6. INDIVIDUAL HUMAN WORKLOAD

One criterion for evaluating the quality of ergonomic design in each workplace component should be the individual human workload. To take an example, mental workload is becoming increasingly important as a result of mechanization and automation, which is why ISO/TC 159 developed a standard of its own (ISO 10075), dealing with ergonomic principles related to mental workload.

Anthropometry makes a significant contribution to ergonomic standardization. Nowadays, for instance, it is inconceivable that a car should be sold on the world market without appropriate account having been taken of the dimensions of the driver's body.

When designing workplaces, ever greater attention is being paid to the handling of heavy weights because the number of working hours lost due to spinal column disorders has markedly increased. One task of an ergonomics committee should therefore be to assist the workplace designer in this respect and to develop appropriate standards. This was actually done in SC 3, Anthropometry and biomechanics. The proposed standards will establish both maximum permissible and maximum acceptable weights of load to be handled manually (lifting and carrying) by adult males and females at work, occasionally and as a function of frequency of task. It provides recommended limits for manual lifting and carrying while taking into account the intensity, the frequency, and the duration of the manual handling task to enable an evaluation of risk to health in a working population.

Standards for ergonomics play an important role in making systems more usable, particularly through the user interface and the quality of interface components. The large manufacturers of visual display terminals operate internationally and therefore also ask for . Consequently, SC 4, Ergonomics of human-system interaction, has prepared a 17-part series of standards, ISO 9241. In addition, the European Committee for Standardization (CEN) has decided to adopt the ISO standards in this area for the Single Market. The European standards, in turn, replace the national standards in the EC and EFTA states. The importance of corresponding ISO standards has greatly increased as a result. Table 2 shows the results of work within SC 4. Special conditions apply in control centers, which explains why the series of standards was developed; it is shown in Table 3.

For many historical, political, and organizational reasons,

Table 2. ISO 9241, Ergonomic requirements for office work with visual display terminals (VDTs) (January 2000)

Part 1: General introduction
Part 2: Guidance on task requirements
Part 3: Visual display requirements
Part 4: Keyboard requirements
Part 5: Workstation layout and postural requirements
Part 6: Guidance on the work environment
Part 7: Display requirements with reflections
Part 8: Requirements for displayed colours
Part 9: Requirements for non-keyboard devices
Part 10: Dialogue principles
Part 11: Guidance on usability
Part 12: Presentation on information
Part 13: User guidance
Part 14: Menu dialogue
Part 15: Command dialogues
Part 16: Direct manipulation dialogues
Part 17: Form-filling dialogues

Table 3. ISO 11064, The ergonomic design of control centres (January 2000)

Part 1: Principles for the design of control centres
Part 2: Principles of control suite arrangements
Part 3: Control room layout
Part 4: Workstation and layout dimensions
Part 5: Displays and controls
Part 6: Environmental requirements for control rooms
Part 7: Principles for the evaluation of control centres
Part 8: Ergonomic requirements for specific applications

standards concerned with the physical environment of workplaces were drafted under headings other than "Ergonomics". These approaches, often specially oriented to technology and engineering, related more to physical aspects than to the importance of these influences on the human being — for example, standards for lighting, noise, and vibration. As a result, ISO/TC 159/SC 5, Ergonomics of the physical environment, took over standardization in this field and developed standards for the thermal environment, communication in noisy environments, visual and auditory danger signals, and skin contact with hot surfaces.

7. CONCLUSIONS

Standards will not determine workplace design but can provide a useful "starting point" for successful design. Therefore, the object of further ergonomic standardization at ISO — and also at CEN — must be a form of standardization that fosters human-centered, as opposed to technology-centered, development of products and work systems. The basis of this re-orientation is the insight that the gulf between man and technology will widen if products, work systems, and working conditions continue to be designed with nothing but technology in mind. Taking appropriate account of the user in the design process should put new technologies into more effective and more efficient use. ISO/TC 159, Ergonomics, will continue to make an important contribution to this process and ensure that it is not the human who must adapt to the machine, but the machine that is adapted to the human.

ACKNOWLEDGEMENT

Ireland: Irish Ergonomics Society

R. Saunders
Secretary of the Irish Ergonomics Society

An informal group was formed in October 1984 by Dr Tim Gallwey, and in March 1985 it was accepted as a Regional Group of the Ergonomics Society (UK). The Group ran seminars once or twice per year and was involved from 1985 onwards in the development of the ideas and processes for a "European Ergonomist". The aim was to provide the mechanism for establishing and controlling a recognized professional level for ergonomists throughout Europe, in order to provide for mobility between member states and to protect against the practice of those unqualified and/or not competent. The standards were agreed in 1992 and provided the mechanism for the Irish Ergonomics Group to become a fully Irish society.

A constitution was approved at an AGM in March 1993 and the Irish Ergonomics Society was formally launched on the 5 March 1994 when the first Annual Conference was held at the University of Limerick. The Society was admitted to the International Ergonomics Association as an Affiliated Organization at its 12th Triennial Congress in Toronto on 14 August 1994. This makes it the officially recognized society for ergonomics in Ireland. The IES became a Federated Member of the IEA with full voting rights at the IEA Council Meeting in August 1995. The Society also represents Ireland on the Council of the Centre for Registration of European Ergonomists (CREE) which was incorporated legally in the Netherlands on the 1 October 1995 and which administers the registration process for the European Ergonomist (Eur Erg). In June 1998, five applicants from Ireland were accepted as Eur Ergs.

Membership of the Society is available to people with a degree (or equivalent) in the ergonomics field, or with some other cognate or relevant discipline with either the equivalent of two years' full-time work in ergonomics or who have published two scientific papers in the subject. Those who do not meet these requirements but are interested in ergonomics can join as Associates. Appropriate post-graduate courses are on offer at the University of Limerick, National University of Ireland, Galway, and Queens University, Belfast.

THE IES TODAY: 1999

This year sees the Society at something of a landmark stage in its development. Those who have kept the flag flying for so long here (Dr Tim Gallwey, Eleanor McMahon, Steve Chan, Gerry Holmes, Enda Fallon, and others) have been joined over the past five years by new recruits and active members who have graduated from, or are involved in, the post-graduate courses mentioned. Today, the Society's membership stands at 60, almost a 100% increase since 1997.

November 1998 saw the Society organize and host its first ever National Conference, which was an overwhelming success due to the efforts of the conference organizing sub-committee. Over 120 delegates from industry, the public sector, and a number of NGOs attended the one-day event, which focused largely on the topics of immediate concern to the workplace: the law in relation to ergonomics; manual handling; industrial interventions and training schemes; and modern office environments.

The previous year, 1997, saw an international conference on Revisiting the Allocation of Functions: ALLFN '97. Organized by Enda Fallon of NUI Galway, those who attended will testify to its success and the quality of papers and debate.

The Society has organized its own internal annual conference each year since 1994; coinciding with the AGM, it provides a forum for current research to be presented to the membership. The Society also produces a newsletter, *The Irish Ergonomist*. With four editions a year, it keeps members up-to-date on national matters and events, etc., and is supplemented with information from the rest of the world, courtesy of the IEA. The Society also has a National Assessment Board for the processing of applications for European Ergonomist on behalf of CREE.

Finally, an ambitious Programme of Work was presented at this year's AGM, with plans for two industry-aimed workshops this year, the formation of two technical working groups and a Membership Review Committee, a major overhaul of our web site (http://www.ul.ie/~ies/), and the planning of the Second National Conference for early 2000. We have also seen the recent formation of the Occupational Health & Safety Institute of Ireland, a research-oriented institution composed of the universities, the social partners, the Health & Safety Authority, and a number of societies and professional bodies. The IES intends to play a leading role in its formative stage and will be a member of both its Board and Scientific Advisory Committee.

ISO and CEN Standardization Committee for Ergonomics (1999): Organizational Structure

H. J. Sälzer

ISO/TC 159 "ERGONOMICS"

Chairperson: W. Schultetus (interim), Germany
Secretary: H.J. Sälzer; DIN-Berlin
ISO Member responsible: DIN
Scope: Standardization in the field of ergonomics, including terminology, methodology and human factors data.

ISO/TC 159/SC 1 — Ergonomic guiding principles
Chairperson: W. Schultetus, Germany
Secretary: H. J. Sälzer, DIN-Berlin

ISO/TC 159/SC 3 — Anthropometry and biomechanics
Chairperson: Professor Dr T. Itani, Japan
Secretary: K. Suzuki; Japan

ISO/TC 159/SC 4 — Ergonomics of human–system interaction
Chairperson: T. F. M. Stewart, UK Secretary: Mrs C. Bassey, BSI-London

ISO/TC 159/SC 5 — Ergonomics of the physical environment
Chairperson: Professor Dr K. C. Parsons, UK
Secretary: Dr D. S. Fishman, BSI-London

CEN/TC 122 "ERGONOMICS"

Chairperson: Dr K.-P. Scheuermann, Germany
Secretary: H. J. Sälzer, DIN-Berlin
CEN Member responsible: DIN
Scope: Standardization in the field of ergonomics principles and requirements for the design of work systems and work environments, including machinery and personal protective equipment, to promote the health, safety and well-being of the human operator and the effectiveness of the work.

CEN/TC 122/WG 1 — Anthropometry
Convenor: Professor Dr Hans W. Jürgens, Germany
Secretary: N. Butz, DIN-Berlin

CEN/TC 122/WG 2 — Ergonomic design principles
Convenor: Dr K.-P. Scheuermann, Germany
Secretary: N. Butz, DIN-Berlin

CEN/TC 122/WG 3 — Surface temperatures
Convenor: Dr H. Siekmann, Germany
Secretary: R. Schmidt, DIN-Berlin

CEN/TC 122/WG 4 — Biomechanics
Convenor: Mrs Dr J. A. Ringelberg, The Netherlands
Secretary: Drs P. M. de Vlaming, NNI-Delft

CEN/TC 122/WG 5 — Ergonomics of human–computer interaction
Convenor: T. F. M. Stewart, UK
Secretary: Mrs C. Bassey, BSI-London

CEN/TC 122/WG 6 — Signals and controls
Convenor: M. I. Gray, UK
Secretary: Mrs C. Bassey, BSI-London

CEN/TC 122/WG 8 — Danger signals and speech communication in noisy environments
Convenor: Dr H. J. M. Steeneken, Netherlands
Secretary: Drs P.M. de Vlaming, NNI-Delft

CEN/TC 122/JWG 9 — Ergonomics of Personal Protective Equipment (PPE)
Convenor: Mrs Vibeke Andersen, Denmark
Secretary: Søren Nielsen, DS –Charlottenlund

CEN/TC 122/WG 10 — Ergonomic design principles for the operability of mobile machinery
Convenor: Drs K. A. Peereboom, The Netherlands
Secretary: Drs P.M. de Vlaming, NNI-Delft

CEN/TC 122/WG 11 — Ergonomics of the thermal environment'
Convenor: Professor Dr K. C. Parsons, UK
Secretary: Dr David Fishman, BSI-London

CEN/TC 122 ad hoc group — Integrating ergonomic principles for machinery design
Convenor: Dr J. A. Ringelberg, Netherlands
Secretary: S. M. van der Minne, NNI-Delft

STANDARDS DRAFT STANDARDS WORK ITEMS (UNDER PREPARATION)

(State: July 1999)

Ergonomic Guiding Principles

ISO 6385: 1981-06-00
ENV 26385: 1990-06-00
Ergonomic principles of the design of work systems
Abstract: Establishes these principles as basic guidelines. They apply for designing optimal working conditions with regard to human well-being, safety and health, taking into account technological and economic efficiency.

EN 614-1: 1995-02-00
Safety of machinery — Ergonomic design principles — Part 1: Terminology and general principles
Abstract: This document establishes the ergonomics principles to be followed during the process of design of work equipment, especially machinery. Although the principles in this document are orientated towards equipment for occupational use, they are applicable also to equipment for private use. It applies to the

interactions between the operator and the work equipment when installing, operating, adjusting, maintaining, cleaning, repairing or transporting equipment and outlines the principles to be followed in taking the health and safety of the operator fully into account.

prEN 614-2: 1997-12-00
Safety of machinery — Ergonomic design principles — Part 2: Interactions between the design of machinery and work tasks
Abstract: The document deals specifically with task design in the context of machinery design, but the principles and methods may also be applied to job design. It establishes the ergonomic principles and procedures to be followed during the design process of machinery and operator work tasks. It is directed to designers and manufacturers of machinery and other work equipment. It will also be helpful to those who are concerned with the use of machinery.

ISO 10075: 1991-10-00
pr EN ISO 10075-1: 1996-07-00
Ergonomic principles related to mental workload; Part 1: General terms and definitions
Abstract: This Standard represents an extension of ISO 6385, Subclauses 3.7 to 3.9, describing terms and definitions in more detail. Annex A forms an integral part of this standard.

ISO 10075-2: 1996-12-00
pr EN ISO 10075-2: 1998-07-00
Ergonomic principles related to mental workload — Part 2: Design principles
Abstract: Gives guidance on the design of work systems, including task and equipment and design of the workplace, as well as working conditions. Relates to the adequate design of work and use of human capacities.

Under preparation: Ergonomic principles related to mental workload — Part 3: Measurement and assessment of mental workload

Anthropometry

This International Standard is intended for use by ergonomic experts and non-experts, e.g. employers, employees and their representatives, system managers and designers, ergonomics and health experts and public authorities. Experts should be able to choose suitable methods based on the contents of this International Standard and/or to find information to be taken into account or to be followed when constructing methods for mental workload assessment. Non-experts will find useful information for their orientation in the field of assessment and measurement of mental workload, e.g. what kinds of methods are provided, who is able and qualified to use them.

EN 547-1: 1996-12-00
ISO/DIS 15534-1: 1998-04-00
Safety of machinery — Human body measurements — Part 1: Principles for determining the dimensions required for openings for whole body access into machinery
Abstract: The document specifies the dimensions of openings for whole body access as applied to machinery as defined in EN 292-1. It provides the dimensions to which the values given in EN 547-3 are applicable. Values for additional space requirements are given in annex A. This European Standard has been prepared primarily for non-mobile machinery, there may be additional specific requirements for mobile machinery.

EN 547-2: 1996-12-00
ISO/DIS 15534-2: 1998-04-00
Safety of machinery — Human body measurements — Part 2: Principles for determining the dimensions required for access openings
Abstract: The document specifies the dimensions of openings for whole body access as applied to machinery as defined in EN 292-1. It provides the dimensions to which the values given in EN 547-3 are applicable. Values for additional space requirements are given in annex A. This European Standard has been prepared primarily for non-mobile machinery, there may be additional specific requirements for mobile machinery.

EN 547-3: 1996-12-00
ISO/DIS 15534-3: 1998-04-00
Safety of machinery — Human body measurements — Part 3: Anthropometric data
Abstract: The document specifies current requirements for human body measurements (anthropometric data) that are required by EN 547-1 and EN 547-2 for the calculation of access opening dimensions as applied to machinery. The anthropometric data originate from static measurements of nude persons and do not take into account body movements, clothing, equipment, machinery operating conditions or environmental conditions. The data are based on information from anthropometric surveys representative of population groups within Europe comprising at least three million people. Both men and woman are taken into account. Measurements are given, as required by EN 547-1 and EN 547-2, for the 5th, 95th and 99th percentile of the relevant population group within Europe.

ISO 7250: 1996-07-00
EN ISO 7250: 1997-07-00
Basic human body measurements for technological design (ISO 7250: 1996)
Abstract: The document provides a description of anthropometric measurements that can be used as a basis for comparison of population groups. The basic list specified in this document is intended to serve as a guide for ergonomists who are required to define population groups and apply their knowledge to the geometric design of the places where people work and live.

ISO/DIS 14738: 1997-12-00
prEN ISO 14738: 1997-12-00
Safety of machinery — Anthropometric requirements for the design of workstations at machinery
Abstract: The document establishes principles for deriving dimensions from anthropometric measurements and applying them to the design of workstations at non-mobile machinery. It is based on current ergonomic knowledge and anthropometric measurements. It specifies the body's space requirements for equipment during normal operation in sitting and standing

positions. It does not specifically include space demands for maintenance, repairing and cleaning work. It does not give recommendations specifically for visual display terminal workstations at machinery. For these purposes EN ISO-9241-5 can be used in conjunction with this document.

Under preparation: Ergonomics — Computer manikins, body templates
Document scope: This European Standard establishes the general requirements for computer-aided manikins intended for the design and evaluation of workspaces at machinery (or: spaces, furniture and other equipment). These requirements concern the anthropometric and motional (and biomechanical) properties of the manikins, particularly the extent to which they represent the human body and the intended user population. The European Standard also specifies the requirements for the documentation of the manikins in respect to these properties and for the guidance of the user.
The European Standard is intended as a guide for the design and selection of manikins and for evaluation of their usability and accuracy for the specified use.
Under preparation: Selection of persons for testing of anthropometric aspects of industrial products and designs
Document scope: This European Standard determines the procedures to be applied to the composition of groups of persons whose anthropometric characteristics are to be representative of all prospective users of any specific object under test.
These procedures will be applicable to the anthropometric aspects of testing of industrial products and designs having direct or indirect contact with the human body and its functions, e.g. machinery, work equipment, personal protective equipment (PPE), consumer goods, working spaces, architectural details, types of transport.
Under preparation: Safeguarding crushing points by means of a limitation of the active forces
Document scope: Standardization of safe, low forces on crushing points (data concerning forces to which people can be exposed without risk of injury).
Target dates depend on financial support for pre-normative research.
Under preparation: Ergonomics — Reach envelopes
Document scope: This European Standard establishes ergonomic requirements for the metric design of workplaces. It specifies the operation's minimum and maximum reach (reach envelopes for hands and feet).
Under preparation: Anthropometric database
Document scope: The European Standard establishes an anthropometric database for all age groups to be used as the basis for the design of work equipment, workplaces and workstations at machinery.
Under preparation: Notation system of anthropometric measurements used in the European Standards EN 547 Part 1 to Part 3
Document scope: The Technical Report explains the complete notation system of anthropometric measurements used in the European Standards EN 547 Part 1 to Part 3.

Biomechanics

prEN 1005-1: 1998-12-00
Safety of machinery — Human physical performance — Part 1: Terms and Definitions
Abstract: The document provides terms and definitions on concepts and parameters used for EN 1005-2 to 4. Basic concepts and general ergonomic principles for the design of machinery are dealt with in EN 292-1 and 2 and EN 614-1. The study of human physical loading whilst handling/operating machinery relies on methods and techniques developed by different research disciplines (e.g. epidemiology, biomechanics, physiology, psychophysics). It aims to define relevant concepts to further cooperation between research disciplines in this field, and to improve the application of standards when designing machinery.

prEN 1005-2: 1998-12-00
Safety of machinery — Human physical performance — Part 2: Manual handling of machinery and component parts of machinery
Abstract: The document specifies ergonomic recommendations for the design of machinery involving manual handling in professional and domestic applications. It applies to the manual handling of objects of 3 kg or more. The standard provides data for ergonomic design and risk assessment concerning lifting, lowering and carrying in relation to the construction, transport and commissioning (assembly, installation, adjustment), use (operation, cleaning, fault finding, maintenance, setting, teaching or process changeover) and decommissioning, disposal and dismantling of machinery. This standard provides current data on the general population and certain subpopulations.

prEN 1005-3: 1998-12-00
Safety of machinery — Human physical performance — Part 3: Recommended force limits for machinery operation
Abstract: The document specifies force limits for pushing and pulling tasks, gripping, arm work and pedal work under the following conditions: pushing and pulling with whole body exertion and using a two-handed symmetrical grip whilst standing and walking; manual exertion (to/from, up/down, in/out), arm work whilst sitting and standing; leg exertion, one-foot or two-foot forces pushing or pressing of pedal(s) whilst sitting and standing.

prEN 1005-4: 1998-11-00
Safety of machinery — Human physical performance — Part 4: Evaluation of working postures in relation to machinery
Abstract: The document presents guidance to the designer of machinery or its components parts in assessing and controlling health risks due to machine-related postures and movements, i.e. during assembly, installation, operation, adjustment, maintenance, cleaning, repair, transport and dismantlement. The standard specifies recommendations for postures and movements with minimal external force exertion. The recommendations are

Under preparation: Safety of machinery — Human physical performance — Part 5: Risk assessment for repetitive handling at high frequency

Document scope: This European standard presents a risk assessment method for machinery related repetitive movements.
ISO/DIS 11226; 1999-02-00
Ergonomics — Evaluation of working postures
Abstracts: This International Standard establishes ergonomic recommendations for different work tasks. The standard will provide information for all those involved in design, or redesign, of work, jobs and products that are familiar with the basic concepts of ergonomics in general and working postures in particular.

The standard specifies recommended limits for working postures with minimal external force exertion, while taking into account body angles and in time aspects.

This standard is designed to provide guidance on the assessment of several task variables, allowing the health risks for the working population to be evaluated.

This standard applies to the adult working population. The recommendations will give reasonable protection for nearly all health adults. The recommendation concerning health risks and protection are mainly based on experimental studies regarding the musculosekeletal load, discomfort/pain, and endurance/ fatigue related to working postures.

ISO/DIS 11228-1: 1998-08-00
Ergonomics — Manual handling — Part 1: Lifting and carrying
Abstract: The part of this International Standard on manual handling establish ergonomic recommendations for different manual handling tasks. The standard apply to such vocational actions as manual handling. The standard specifies recommended limits for manual lifting and carrying while taking into account respectively the intensity, the frequency and the duration of the task. The standard applies to manual handling of objects with a mass of 3 kg or more.

Under preparation: Ergonomics-Manuel handling — Part 2: Pushing and pulling
Document scope: to be defined

Under preparation: Ergonomics-manual Handling — Part 3: Handling, at high repetition of low loads
Document scope: to be defined

Thermal Environments
ISO 7243: 1989-08-00
EN ISO 7243: 1993-12-00
Hot environments; estimation of the heat stress on working man, based on the WBGT-index (wet bulb globe temperature)
Abstract: Gives a method, which can easily be used in an industrial environment for evaluating the stresses on an individual. It applies to the evaluation of the mean effect of heat on man during a period representative of his activity but it does not apply to very short periods, nor to zones of comfort.

ISO 7726: 1998-11-00
EN ISO 7726: 1993-12-00
Ergonomics of the thermal environment — Instruments for measuring physical Quantities
Abstract: The document specifies the minimum characteristics of appliances for measuring physical quantities characterizing an environment as well as the methods for measuring the physical quantities of this environment.

ISO 7730: 1994-12-00
EN ISO 7730: 1995-07-00
Moderate thermal environments — Determination of the PMV and PPD indices and specification of the conditions for thermal comfort
Abstract: The purpose is to present a method for predicting the thermal sensation and the degree of discomfort (thermal dissatisfaction) of people exposed to moderate thermal environments and to specify acceptable environmental conditions for comfort. Applies to healthy men and women and was originally based on studies of North American and European subjects but agrees also well with recent studies of Japanese subjects and is expected to apply with good approximation in most parts of the world. Applies to people exposed to indoor environments where the aim is to attain thermal comfort, or indoor environments where moderate deviations from comfort occur.

ISO 8996: 1990-12-00
EN 28996: 1993-10-00
Ergonomics; determination of metabolic heat production
Abstract: Specifies methods for determining the metabolic rate, but can also be used for other applications, e.g. for the assessment of working practices, the cost of specific jobs or sport activities, the total cost of activity, etc. Annexes A to G contain: classification of metabolic rate for kinds of activities, classification of metabolic rate by occupation, data for standard person, metabolic rate for body posture, type of work and body motion related to work speed, metabolic rate for typical activities, example of calculation of the average metabolic rate for a work cycle, examples of calculation of the metabolic rate based on measured data.

ISO 7933: 1989-07-00
Hot environments; analytical determination and interpretation of thermal stress using calculation of required sweat rate
Abstract: Describes a method of calculating the heat balances as well as the sweat rate that the human body should produce to maintain this balance in equilibrium. The various terms used show the influence of the different physical parameters. It does not predict the physiological response of individual subjects, but only considers standard subjects in good health and fit for the work they perform.

EN 12515: 1997-06-00
Hot environments — Analytical determination and interpretation of thermal stress using calculation of required sweat rate (ISO 7933: 1989, modified)
Document scope: to be defined

ISO 9886: 1992-11-00
p EN ISO 9886: 1997-06-00
Evaluation of thermal strain by physiological measurements
Abstract: Describes methods for measuring and interpreting body core temperature, skin temperature, heart rate, and body mass loss. Annex A presents a comparison of the different methods concerning their field of application, their technical complexity,

the discomfort and the risks. The measurement techniques are described in annex B, limit values are proposed in annex C.

ISO 10551: 1995-05-00
Pr EN ISO 10551: 1997-06-00
Ergonomics of the thermal environment — Assessment of the influence of the thermal environment using subjective judgement scales
Abstract: Covers the construction and use of judgement scales for use in providing reliable and comparative data on the subjective aspects of thermal comfort or thermal stress.

ISO 11399: 1995-12-00
Pr EN ISO 11399: 1997-06-00
Ergonomics of the thermal environment — Principles and application of relevant international standards
Abstract: Purpose is to specify information which will allow the correct, effective and practical use of International Standards concerned with the ergonomics of the thermal environment. Describes the underlying principles concerning the ergonomics of the thermal environment.

ISO/DIS 12894: 1997-01-00
Pr EN ISO 12894: 1997-01-00
Ergonomics of the thermal environment — Medical supervision of individuals exposed to extreme hot or cold environments
Abstract: This International Standard provides guidance to those concerned with the safety of human exposures to extreme hot or cold thermal environments, about the medical fitness assessment and health monitoring which may be appropriate to and during such exposures. It is intended to assist those with responsibility for such exposures to reach decisions about the appropriate level of medical supervision in different situations. The standard presents guidance that should be read and used in the context of other guidance and legislation applying to each particular situation.

ISO/DIS 13731: 1997-04-00
Pr EN ISO 13731: 1997-04-000
Ergonomics of the thermal environment — Vocabulary and symbols

ISO 9920: 1995-03-00
Ergonomics of the thermal environment — Estimation of the thermal insulation and evaporative resistance of a clothing ensemble
Abstract: Gives methods for estimating the thermal characteristics (resistance to dry heat loss and evaporative heat loss) in steady-state conditions for a clothing ensemble based on values for known garments, ensembles and textiles. Does not deal with other effects of clothing, such as adsorption of water, buffering, tactile comfort. Does not take into account the influence of rain and snow on the thermal characteristics. Does not consider special protective clothing. Does not deal with the separate insulation on different parts of the body and discomfort due to the asymmetry of a clothing ensemble.

ISO/TR 11079: 1993-12-00
ENV ISO 11079: 1998-01-00
Evaluation of cold environments; determination of required clothing insulation (IREQ)
Abstract: Proposes methods and strategies to assess the thermal stress associated with exposure to cold environments. Cold stress is suggested to be evaluated in terms of both general cooling of the body and local cooling of particular parts of the body (e.g. extremities and face). The methods apply to continuous, intermittent and occasional exposure and in both indoor and outdoor work. Specific effects associated with certain meteorological phenomena (e.g. precipitation) are not covered and should be assessed by other methods.

Under preparation: Ergonomics of the thermal environment — Medical supervision of individuals exposed to extreme hot or cold environments

Under preparation: Ergonomics of the thermal environment — Comfortable contact surface temperature

Under preparation: Ergonomics of the thermal environment — Application of International Standards to the disabled; the aged and other handicapped persons

Under preparation: Ergonomics of the thermal environment — Assessment of the long-term thermal comfort performance of indoor environments

Under preparation: Ergonomics of the physical environment — Determination of the combined effect of thermal environment, air pollution, acoustics and illumination on humans

Under preparation: Ergonomics of the thermal environment — Working practices for cold indoor environments

Under preparation: Evaluation of the thermal environment in vehicles

Under preparation: Ergonomics of the thermal environment — Risk assessment of stress or discomfort in thermal working conditions *Under preparation*: Principles of the visual ergonomics — Lighting of indoor work systems

Fundamentals of Controls and Signaling Methods

EN 894-1: 1997-02-00
ISO/DIS 9355-1: 1997-09-00
Safety of machinery — Ergonomics requirements for the design of displays and control actuators — Part 1: General principles for human interactions with displays and control actuators
Abstract: This document applies to the design of displays and control actuators on work equipment, especially machines. It specifies the relationships that have to be maintained between the movements of control actuators, the response of any associated displays, and the human response to information given by the system, to minimize operator errors and to ensure an efficient interaction between the operator and the equipment.

EN 894-2: 1997-02-00
ISO/DIS 9355-2: 1997-09-00
Safety of machinery — Ergonomics requirements for the design of displays and control actuators — Part 2: Displays
Abstract: This document gives recommendations on the selection, design and location of information displays, so that they are adapted to the requirements of the operators, and take account of the circumstances of their use. It includes visual, audible and tactile displays. It applies to displays used in equipment (e.g. machines and installations, control panels, operating and monitoring consoles) for occupational and private use.

prEN 894-3: 1992-10-00
Safety of machinery; ergonomics requirements for the design of displays and control actuators; part 3: Control actuators
Abstract: This document gives recommendations on the selection, design and location of control actuators so that they are adapted to the requirements of the operators, and take account of the circumstances of their use. It applies to manual control actuators used in equipment for occupational and private use.

Under preparation: Safety of machinery — Ergonomics requirements for the design of displays and control actuators — Part 4: Location and arrangement of displays and control actuators
Document scope: This European Standard applies to the design of displays and control actuators on machinery. It establishes general principles for the location and arrangement of displays and control actuators, to minimize operator errors and to ensure an efficient interaction between the operator and the equipment. It is particularly important to observe these principles when an operator error may lead to injury or damage to health.

Ergonomics of Human–System Interaction
ISO 9241-1: 1997-06-00
EN ISO 9241-1: 1997-06-00
Ergonomic requirements for office work with visual display terminals (VDT) — Part 1: General introduction
Abstract: The document introduces multipart standard on ergonomic requirements for the use of visual display terminals for office tasks, provides guidelines for a user-performance approach, gives an overview of all parts of EN ISO 9241 currently published and of the anticipated content of those in preparation and provides some guidance on how to use EN ISO 9241. It describes also how conformance to EN ISO 9241 should be reported.

Under preparation: Amendment 1: Description and application of the software parts (Part 10 to 17)

ISO 9241-2: 1992-06-00
EN ISO 9241-2: 1992-06-00
Ergonomic requirements for office work with visual display terminals (VDT); Part 2: Guidance on task requirements
Abstract: Guidance is relevant to both the organization implementing the system and the people using the equipment and should be applied in accordance with local, regional or national agreements and regulations. The objective is to enhance the efficiency and well-being of the individual user by applying ergonomics knowledge in the light of practical experience, to the design of tasks. The ergonomics principles concerned are set out in ISO 6385.

ISO 9241-3: 1992-07-00
EN ISO 9241-3: 1992-07-00
Ergonomic requirements for office work with visual display terminals (VDT); Part 3: Visual display requirements
Abstract: Establishes image quality requirements (performance specifications) for the design and evaluation of single- and multi-color VDT. At present, the recommendations are based on Latin, Cyrillic, and Greek origin alphabetic characters, and Arabic numerals. Office tasks include such activities as data entry, text processing, and interactive inquiry. Annex A describes analytical techniques for predicting screen flicker, annex B an empirical method for assessing temporal and spatial instability (flicker and jitter) on screen, annex C a comparative user performance test method.

Under preparation: Amendment 1: User performance test
This annex is intended for testing the visual quality of VDUs where the entire set of physical requirements (defined in clause 5) cannot be applied completely. This alternative compliance route shall be used if some clause 5 requirements cannot be applied, for example to flat panel displays.

ISO 9241-4: 1998-08-00
EN ISO 9241-4: : 1998-08-00
Ergonomic requirements for office work with visual display terminals (VDT) — Part 4: Keyboard requirements
Abstract: The document applies to linear detachable keyboard designed for stationary use. It provides guidance on the design of keyboards used of typical office tasks so that the limitations and capabilities of users are considered

ISO 9241-5: 1998-10-00
EN ISO 9241-5: 1999-03-00
Ergonomic requirements for office work with visual display terminals (VDT) — Part 5: Workstation layout and postural requirements
Document scope: This part of ISO 9241 specifies ergonomic guiding principles that apply to the user requirements, design and procurement of workstation equipment for office tasks using VDT. In particular, the general principles and requirements specified in this part of ISO 9241 apply to the standards specifying technical design of furniture and equipment constituting the workplace.
ISO/DIS 9241-6: 1998-04-00
prEN ISO 9241-6: 1996-02-00
Ergonomic requirements for office work with visual display terminals (VDT) — Part 6: Guidance on the work environment
Abstract: The document gives guidance on basic principles for the ergonomic design of the work environment and the workstation, taking into account lighting effects of noise and mechanical vibrations, electrical and magnetic fields and static electricity, thermal environment, space organization and workplace layout. The document is applicable to the work environment and workstation in those work systems where a visual display terminal (VDT) is used for office work.

Under preparation: Ergonomics requirements for office work with visual display terminals (VDT) — Space organization and workplace layout — Consideration supporting the requirements presented in ISO/DIS 9241-6

ISO 9241-7: 1998-04-00
EN ISO 9241-7: 1998-04-00
Ergonomic requirements for office work with visual display terminals(VDT) — Part 7: Requirements for display with reflections
Abstract: This document establishes image quality requirements for VDT's used in luminous environments that may cause reflections from screen. It also applies to monochrome and multicolor displays.

ISO 9241-8: 1997-10-00
EN ISO 9241-8: 1997-10-00
Ergonomic requirements for office work with visual display terminals (VDT) — Part 8: Requirements for displayed colors
Abstract: The document describes minimum ergonomic requirements and recommendations to be applied to colors assigned to text and graphic applications and images in which colors are discretely assigned. The specifications thus exclude photorealistic images and graphics. The document applies to both hardware and software for visual display terminals, because both these sources control the presentation and appearance of color on the display screen

ISO/DIS 9241-9: 1998-05-00
prEN ISO 9241-9: 1998-10-00
Ergonomic requirements for office work with visual display terminals (VDT) — Part 9: Requirements for non-keyboard input devices
Abstract: The document applies to several types of non-keyboard devices designed for stationary use. It gives guidance based on ergonomic factors for the following input devices: mice, pucks, joysticks, trackballs, tablets and overlays, touch sensitive screens, stylist and light pens.

ISO 9241-10: 1996-05-00
EN ISO 9241-10: 1996-05-00
Ergonomic requirements for office work with visual display terminals (VDT) — Part 10: Dialogue principles
Abstract: Gives ergonomic principles formulated in general terms; they are presented without reference to situations of use, application, environment or technology. These principles are intended to be used in specifications, design and evaluation of dialogues for office work with visual display terminals (VDT).

ISO 9241-11: 1998-03-00
EN ISO 9241-11: 1998-03-00
Ergonomic requirements for office work with visual display terminals (VDT) — Part 11: Guidance on usability
Abstract: The document defines usability and explains how to identify the information which is necessary to take into account when specifying or evaluating usability in terms of measures of user performance and satisfaction. Guidance is given on how to describe the context of use of the product (hardware, software or service) and the required measures of usability in an explicit way.

The guidance is given in the form of general principles and techniques, rather than in the form of requirements to use specific methods. The guidance can be used in procurement, design, development, evaluation and communication of information about usability.

ISO 9241-12: 1998-12-00
EN ISO 9241-12: 1998-12-00
Ergonomic requirements for office work with visual display terminals (VDT) — Part 12: Presentation of information
Abstract: The document provides ergonomic recommendations for the presentation of information and specific properties of presented information on text-based and graphical user interfaces used for office tasks. It provides recommendations for the design and evaluation of visual presentation of information including coding techniques. These recommendations can be utilized throughout the design process. The coverage of color is limited to ergonomic recommendations for the use of color for highlighting and categorizing information(see ISO 9241-8 for additional recommendations for the use of color).
The document does not address auditory presentation of information.

ISO 9241-13: 1998-07-00
EN ISO 9241-13: 1998-07-00
Ergonomic requirements for office work with visual display terminals (VDT) — Part 13: User guidance
Abstract: The document provides recommendations for user guidance attributes of software user interfaces and their evaluation. User guidance are covered by this document includes recommendations specific to prompts, feedback and status, error management and online help as well as recommendations common to all these types of user guidance. The recommendations in this document are formulated to be independent of applications, environment, or technology. They correspond to typical situations involving special needs for information and actions.

ISO 9241-14: 1997-06-00
EN ISO 9241-14: 1999-04-00
Ergonomics requirements for office work with visual display terminals (VDT) — Part 14: Menu dialogues
Abstract: This part of ISO 9241 provides conditional recommendations for menus used in user-computer dialogues to accomplish typical office tasks. The recommendations cover menus presented by various techniques including windowing, panels, buttons, fields, etc. These recommendations can be utilized throughout the design process (e.g., as guidance for designers during design, as a basis for heuristic evaluation, as guidance for usability testing).

ISO 9241-15: 1997-12-00
EN ISO 9241-15: 1997-12-00
Ergonomic requirements for office work with visual display terminals (VDT) — Part 15: Command dialogues
Abstract: The document provides recommendations for command dialogues used to accomplish typical office tasks. Command dialogues are sequences of instructions provided by the user to the system which, when processed, result in associated system

actions. Users input (from recall, rather than selecting from a menu) complete or abbreviated command phrases (e.g. mnemonics, letters, function keys, hot keys) in the order required by the command language syntax and the computer performs the activities initiated by the command(s) and their associated parameters.

ISO/FDIS 9241-16: 1999-04-00
prEN ISO 9241-16: 1999-04-00
Ergonomic requirements for office work with visual display terminals (VDT) — Part 16: Direct-manipulation dialogues
Abstract: The document gives guidance on the design of direct manipulation dialogues the user directly acts on objects on the screen; e.g., by pointing at them, moving them and/or changing their physical characteristics (or values) via at the use of an input device.

ISO 9241-17: 1998-08-00
EN ISO 9241-17: 1998-08-00
Ergonomic requirements for office work with visual display terminals (VDT) — Part 17: Form-filling dialogues
Abstract: Ergonomic design of form-filling dialogues, including form structure and output considerations, input considerations and form navigation. Form filling dialogues are dialogues in which the user fills in, select entries for, or modifies labeled fields on a 'form' or a dialogue box presented by the system. This part of ISO 9241 pertains to form filling dialogues generated through both VDT character-based and bit-mapped screen displays and input through keyboards and optional pointing devices.

ISO 13407: 1999-06-00
EN ISO 13407: 1999-06-00
Human-centered design processes for interactive systems
Abstract: The document provides guidance on human-centered design activities throughout the life cycle of computer-based interactive systems. It is aimed at those managing design processes and provides guidance on sources of information and standards relevant to the human-centered approach. It is concerned with both hardware and software components of interactive systems. It addresses the planning and management of human-centered design. It does not address all aspects of project management.

Under preparation: Ergonomics of human system interaction — Guidance on accessibility for human–computer interfaces
Document scope: To be defined
Under preparation: Ergonomics of human–system interaction — Usability methods supporting human-centred design.
Document scope: To be defined
ISO/FDIS 13406-1: 1999-05-00
Ergonomic requirements for work with visual displays based on flat panels — Part 1: Introduction
This part of ISO 13406 establishes the rationale for ergonomic requirements and measurement methods for flat panel displays and defines a flat panel display. It is applicable to flat panel technology applied to displays for office work and other work.

ISO/DIS 13406-2: 1997-07-00
Ergonomic requirements for visual display units based on flat panels — Part 2: Requirements for flat panel displays

This International Standard establishes ergonomic image quality requirements for the design and evaluation of flat panel displays (ISO/CD 13406-1); defines terms needed to address image quality on flat panel displays; specifies methods of determining image quality on flat panel displays; establishes ergonomic principles for guiding these requirements.

This International Standard is applicable to flat panel display screens when used to perform office tasks; flat panel display screens that consist of a regular array of picture elements arranged in evenly spaced rows without built-in gaps; the presentation of alphanumeric characters on flat panel display screens when the characters are based on Latin-origin fonts; the presentation of Asian characters; flat panel display screens that are large enough to display at least 40 Latin-origin characters.

This International Standard is not applicable to: flat panel technology applied to a display that uses optics to form an image that is not the same size as the electro-optical transducer (projection application of flat panel displays); flat panel technology applied to a display limited to fixed-messages or segmented alphanumerics. See 2.13 IEC SC 47C (Central Office) 3: 1992.
Under preparation: Multimedia user interface design — Software ergonomics requirements — Part 1: Introduction and framework
Document scope: This International Standard establishes requirements, recommendations and principles for the design of interactive multimedia user interfaces that integrate and synchronize different media (static media such as text, graphs, images, and dynamic media such as audio, animation and video). This International Standard does not provide detailed design guidance for specific media but addresses issues related to applications integrating different media.
This International Standard applies to presentation and interaction techniques for computer-based multimedia applications in general, including stand-alone and network-delivered applications. It provides recommendations for design and evaluation and addresses issues related to the interaction of different media and to selection and combining media. It applies to the design of the software user interface and does not address hardware issues, such as input or output devices.
This International Standard mainly addresses ergonomic design issues. Due to the intended range of applications and user groups, areas like aesthetic quality of a system also are addressed. The recommendations provided in this International Standard may not apply to systems whose primary purpose is entertainment (such as games).

Under preparation: Multimedia user interface design — Software ergonomics requirements — Part 2: Multimedia control and navigation
Document scope: To be defined.

Under preparation: Multimedia user interface design — Software ergonomics requirements — Part 3: Selection of media and media combination
Document scope: This part of ISO 14915 establishes software ergonomics requirements and recommendations for interactive multimedia user interfaces that integrate and synchronize different

media (static media like text, graphs, images and dynamic media like audio, animation and video). The standard applies to presentation and interaction techniques for computer-based multimedia applications when the prime goal is to support the user's work task or provision of information. The standard applies to training and tutorial multimedia in so far as its recommendations bear on effective delivery of information. It does not deal with pedagogic design issues for tutorial applications and entertainment applications are also not within its *scope*. This standard does provide recommendations for design and evaluation of visual presentation of information but does not make detailed recommendations for dialogue design of multimedia interfaces which is addressed in ISO 14915-2. The focus is on design using appropriate media to convey different types of information, ensuring important parts of the message are comprehended by the user and that messages in multimedia presentations are delivered in a coherent, integrated manner.

Under preparation:
Multimedia user interface design — Software ergonomics requirements — Part 4: Domain specific multimedia interfaces
Document scope: to be defined.

Ergonomic Design of Control Centers

ISO/DIS 11064-1: 1999-02-00
pr EN ISO 11064-1: 1999-02-00
Ergonomic design of control centers — Part 1: Principles for the design of control centers
Document scope: This Part of ISO 11064 specifies requirements and establishes general principles for the ergonomic design of a new control center, as well as for control center expansion, refurbishment, and technology based upgrades. This Part of ISO 11064 covers all types of control centers, for example, those found in process plants, transport/-logistic control systems, and people deployment transportation systems.
Some of the principles regarding design procedures presented in this Part of ISO 11064 can be helpful in the design of mobile control centers, such as those found on trains, locomotives, ships and aircraft, though the standard does not set out to consider control centers of this type.

ISO/DIS 11064-2: 1999-03-00
pr EN ISO 11064-2: 1999-03-00
Ergonomic design of control centers — Part 2: Principles of control suite arrangement
Document scope: This Part of ISO 11064 covers ergonomic design principles of control centers, more specifically the various arrangements of rooms in a control suite. The principles are based on an analysis of tasks that have to be supported by the control room and functionally related rooms. They include identifying functional areas, estimating the space provisions for each functional area, determining operational links between functional areas, and developing preliminary control suite layouts to facilitate the smooth transition between all the activities conducted in the control suite. The control suite includes rooms or areas such as meeting rooms, equipment room, office(s), relaxation room and other rooms that support control room activities, as well as the control room itself. Associated ergonomic considerations include

not only architectural conditions, but social, cultural and environmental factors as well.

ISO/DIS 11064-3: 1997-02-00
prEN ISO 11064-3: 1997-02-00
Ergonomic design of control centers — Part 3: Control room layout
Document scope: ISO 11064-3 established ergonomic principles for the layout of control rooms. It include requirements, recommendations and guidelines on control room layouts, workstation arrangements, the use of off-workstation visual displays and control room maintenance.

ISO 11064-3 covers all types of control center including those for the process industry, transport and dispatching systems in the emergency services. Though this Part of ISO 11064 is primarily intended for non-mobile control centers many of the principles could be relevant/applicable to mobile centers such as those found on ships and aircraft.

Under preparation: Ergonomic design of control centers — Part 4: Workstation layout and dimensions
Document scope: ISO 11064-4 establishes general principles for the ergonomic design of workstations found in control centers. ISO 11064-4 covers all aspects of workstation design with particular emphasis on layout and dimensions, the type of workstations covered include seated and standing control room workstations, seated workstations with visual display terminals, and maintenance workstations. In addition considerations on the type and number of instruments and equipment will be given.

Under preparation: Ergonomic design of control centers — Part 5: Displays and controls
Document scope: ISO 11064-5 deals primarily with the display and input of data to monitor and/or operate processes from control centers. It establishes requirements that shall and recommendations that should be matched by displays and controls. Furthermore it gives permissions and reveals possibilities to support the designer.

It is impossible to cover in depth all relevant aspects of interaction design. Therefore this international Standard is limited to principles and selective in presenting requirements. To promote understanding and to ease access, the design procedure was applied for structuring the text.

The following topics are beyond the standard: The design of equipment like monitors, indicators, meters or input devices neither in hardware nor in software; the design of digital or analogue displays; the design of control actuators; location and arrangement of displays and controls; problems occurring with visual display terminals (VDT) applied to office tasks.

Under preparation: Ergonomic design of control centers — Part 6: Environmental requirements for control rooms
Document scope: ISO 11064-6 establishes environmental requirements for the ergonomic design, upgrading or refurbishment of control rooms, control suite and all other rooms such as offices, technical and auxiliary rooms. The following aspects

are: thermal environment; air distribution principles; lighting environment; acoustical environment; air composition; vibrations.
Under preparation: Ergonomic design of control centers — Part 7: Principles for the evaluation of control centers
Document scope: to be defined

Under preparation: Ergonomic design of control centers — Part 8: Ergonomic requirements for specific applications
Document scope: to be defined

Danger Signals and Communication in Noisy Environments

EN 457: 1998-11-12
Safety of machinery. Auditory danger signals. General requirements, design and testing
Abstract: This standard specifies the safety and ergonomic requirements and the corresponding test methods for auditory danger signals and gives guidelines for the design of the signal to be clearly perceived and differentiated as required in 5.3 of PrPN-EN 292-2. It does not apply to verbal danger warnings (e.g. shouts, loudspeaker announcements). Special regulations such as those for a public disaster and public transport are not effected by this standard.

ISO 7731: 1986-12-00
Danger signals for work places; Auditory danger signals
Abstract: Specifies the safety requirements and the corresponding test methods for work places in the signal reception area and gives guidelines for the design of the signals. May also be applied to other appropriate situations. Does not apply to verbal danger warnings. Special regulations such as those for a public disaster and public transport are not affected.

EN 842: 1996-06-00
Safety of machinery — Visual danger signals — General requirements, design and testing
Document scope: to be defined

ISO 11428: 1996-12-00
Ergonomics — Visual danger signals — General requirements, design and testing
Abstract: Describes criteria for the perception of visual danger signals in the area in which people are intended to perceive and to react to such a signal. Specifies the safety and ergonomic requirements and the corresponding physical measurements.

EN 981: 1996-12-00
Safety of machinery — System of auditory and visual danger and information signals
Document scope: to be defined

ISO 11429: 1996-12-00
Ergonomics — System of auditory and visual danger information signals
Abstract: Specifies a system of danger and information signals taking into account the different degrees of urgency. Applicable to all danger and information signals which have to be clearly

perceived and differentiated as specified in ISO/TR 12100-2. Does not apply to certain fields covered by specific standards.

ISO 9921-1: 1996-11-00
Ergonomic assessment of speech communication — Part 1: Speech interference level and communication distances for persons with normal hearing capacity in direct communication (SIL method)
Abstract: Provides a method for prediction of the effectiveness of speech communication in the presence of noise generated by machinery as well as in noisy environment. Parameters are the ambient noise at the speaker's position, ambient noise at the listener's position, distance between the communication partners and a great number of physical and personal conditions.

Under preparation: Ergonomic assessment of speech communication — Part 0: General description and criteria
Document scope: This part of the standard specifies the requirements for the speech communication quality in case of verbal alert and danger signals, information massages and speech communications in general. Methods to predict and to measure the performance in practical applications are described and examples are given.
Auditory warnings are also omnidirectional and may therefore be universal in many situations. It is required however that, in case of verbal massages, a sufficient intelligibility is offered. If this cannot be achieved synthetic warning signals may be considered.

Under preparation: Ergonomic assessment of speech communication — Part 2: Predictive assessment methods
Document scope: This part of the standard provides more universal methods for prediction of the performance of speech communications. The parameters covered by the prediction methods are: speaker related parameters as vocal effort, distance to listener or microphone; sound pressure level and frequency distribution of the ambient noise at the speaker and listener position: frequency transfer of the (electro acoustic) system; temporal distortion (reverberation, single echoes); hearing aspects of the listener (masking, reception threshold, hearing disorders). All predictive methods are based on the calculation of an effective signal-to-noise ratio in a number of frequency bands in which all types of relevant deterioration are included.

Under preparation: Ergonomic assessment of speech communication — Part 3: Subjective and subjective assessment methods
Document scope: This part of the standard specifies the use of a number of subjectives and objective intelligibility methods which are recommended for the assessment of the type of communications covered by this standard.
Subjective assessments require speakers to produce representative speech signals and listeners to determine the performance, selection of the speech utterances used for the test. For the objective assessment physical measuring methods are used. Depending on the method and application, selected speech or artificial signals replace the speaker. After transmission a physical

analysis is performed on the received signal which results in a performance score.

Touchable Surfaces

EN 563: 1994-06-00
Safety of machinery — Temperatures of touchable surfaces — Ergonomics data to establish temperature limit values for hot surfaces
Abstract: This document applies to hot surfaces of all products and equipment that must or can be touched during their normal use. That includes the area of safety of machinery as well as any other application. It provides data concerning circumstances under which contact with at hot surface may lead burns of the human skin. These data allow the assessment of risk of burning.

EN 563: 1994/A1: 1999-07-00
Safety of machinery — Temperatures of touchable surfaces — Ergonomics data to establish temperatures limit values for hot surfaces
Document scope: This document is an amendment to EN 563; 1994 adding contact periods below 1s.

prEN 13202: 1998-0400
Ergonomics of the thermal environment — Temperatures of touchable hot surfaces — Guidance for establishing surface temperature limit values in production standards with the aid of EN 563

Abstract: The document describes methods for the assessment of the risk of burning when a hot surface is touched by unprotected skin. It describes how surface temperature limit values can be established in product standards with the aid of EN 563. The document is applicable for all kinds of products where hot surfaces cause a risk of burning. It applies as well for electrically powered

Under preparation: Safety of machinery — Surface temperatures of touchable surfaces — Ergonomics data to establish temperature limit values for cold surfaces

Document scope: This standard applies to cold surfaces of all products and equipment that must or can be touched in accordance with their normal use. This standard provides data to assess the risk of injuries caused by contact of the skin with a cold surface. This standard also provides data to be used to establish temperature limit values for cold surfaces to protect against skin or body injuries. These data shall be used in the development of standards for specific equipment where temperature limits are required. (This standard does not provide data for the protection against pain.) This standard applies to solid surfaces. It does not apply to surfaces of liquids.

Ergonomics of Personal Protective Equipment

Under preparation: Ergonomic principles for the design of Personal Protective Equipment — Anthropometric measurements and principles for application to PPE design
Document scope: This European Standard defines basic anthropometric measurements for designing Personal Protective Equipment (PPE) to be worn by an individual against one or more health and safety hazards. This European Standard includes principles for research and application of the relevant anthropometric data to the design of PPE. It also includes principles for checking that a person wearing the PPE can perform the movements required in his activity. This European Standard may be used in conjunction with national or international documents and anthropometric data bases.

Under preparation: Ergonomic principles for the design of personal protective equipment (PPE) — Biomechanics
Document scope: This European Standard describes the biomechanics interaction between Personal Protective Equipment (PPE) on the static and dynamic biomechanics of the human body and its influence on task performance and workload. It describes test principles for acceptance of the PPE. (Care should be taken in applying the principles in work on the surface, under water and in hyperbaric environment.)
This standard includes the principles related to: the static distribution of weight and load on the human body when wearing different PPE and/or combinations of PPE; the dynamic or intertial forces on the human body when using different PPE and/or combinations of PPE; optimization of the influence of biomechanical aspects of PPE on work load and/or task performance.
This standard does not include aspects of biomechanics related to the mechanical deformation of the human skin. This item is included in the prEN xxx (the standard relating to sensory aspects of PPE).

Under preparation: Ergonomic principles for the design of personal protective equipment (PPE) — Thermal characteristics
Document scope: This European Standard is a guide for the evaluation of the thermal effects imposed by PPE/PPE-systems. It defines procedures, criteria and requirements to be applied for the assessment of thermal comfort and physiological strain resulting from the PPE/PPE-system under defined user conditions. The standard also gives guidance on the tests needed to evaluate the thermal characteristics of single PPE to establish the range of thermal conditions in which their use is safe.

Under preparation: Ergonomic principles for the design of Personal Protective Equipment (PPE) — Biological aspects
Document scope: This European Standard gives requirements and test methods for PPE materials to ensure that the materials will not generate substances during the use of the PPE which reversibly or irreversibly affect the user.
This standard provides lists of chemicals and references to chemical lists which are known to generate hazards if coming in contact with the human body and adversely affects health and safety of the user of PPE.
This standard identifies chemicals of which the manufacturers of PPE shall acquire sufficient knowledge in terms of health and safety.

Under preparation: Ergonomic principles for the design of Personal Protective Equipment (PPE) — Sensory aspects
Document scope: This European Standard provides details

regarding the ergonomic principles relating to the interaction between PPE and the human senses: sight, hearing, smell, taste, touch, vestibular orientation and proprioception. It specifies guidelines for evaluating and impairment of sensory perception arising from the use of PPE. It is concerned with the inward flow of sensory signals and the ways in which PPE may adversely affect that flow.

This European Standard does not cover the outward flow of signals (e.g. to give information), although this can also have an impact on the acceptability of any PPE.

Under preparation: Ergonomic principles for the design of Personal Protective Equipment PPE — Introduction to principles and application of European Standards on ergonomic principles for PPE

Document scope: Part 1 of this European Standards sets out the generic principles that shall be considered by those preparing product standards for personal protective equipment (PPE). It also provides background information and guidance on the application of these principles which it is intended should be helpful to those responsible for the drafting of product standards and the design of PPE. It also gives an introduction to subsequent parts of the standard which are a series of application standards that can be used in the preparation of product standards for personal protective equipment.

The aim of this generic standard is to provide: a basic reference for ergonomic requirements and test methods in product standards; a basic reference for ergonomic specifications where no relevant product standards are available; a source of information so that product standards can be drafted to specify PPE that is optimally suited to the people who will use it and the situations in which it is used; a basis for drafting application standards; an introduction to the use of application standards.

Under preparation: Ergonomic Principles for the design of Personal Protective Equipment (PPE) — Headforms and facial features
Document scope: This European Standard defines basic anthropometric measurements for comparing headforms in different population groups.

The standard gives recommendations for additional measurements (facial features) for designing headworn PPE which cover the face.

Ergonomics of Operability of Mobile Machines

Under preparation: Ergonomic design principles for the operability of mobile machinery
Document scope: This European Standard provides ergonomic principles, general design guidelines and basic dimensions for the operability of mobile machinery.

Integrating Ergonomic Principles for Machinery Design

Under preparation: Safety of machinery — Guidance for introducing ergonomic principles and for the drafting of the ergonomics' clauses
Document scope: The approach is based on EN 292-1, EN 292-2 + A1 and EN 414, using an amended version of the step by step method as set out in EN 1050. Each step has been adapted to include ergonomic principles, and practical guidance is given to apply horizontal standards. This European standard may be used for the drafting of C-type standards (product-machinery-standards) and by manufacturers where no relevant C-type standards are available.

Israel: Israel Ergonomics Society

I. Gilad

Faculty of Industrial Engineering and Management, Technion — Israel Institute of Technology, Haifa 32000, Israel

The Israel Ergonomics Society (IES) was founded in 1982 to enhance the state of the Ergonomics discipline in the State of Israel. To do this the founders outlined six goals:

1. To unite all Certified Ergonomists and qualified professionals who practice ergonomics in Israel under one umbrella by creating one national society.
2. To introduce the Ergonomics discipline and its benefits to increase safety and health measures for the working populations and individuals.
3. To spread out the knowledge of Applied Ergonomics in commercial enterprises, industrial and service organizations, for the improvement of productivity and economic status.
4. To aid governmental agencies when ergonomics issues are raised, help the legislator by providing solid ergonomics know how in the decision-making process.
5. To organize conferences, seminars, and professional forums where ergonomics issues are to be discussed, to support R&D activities, where ergonomics problems are to be studied.
6. To establish professional ties with Ergonomics societies around the globe, to build bridges of knowledge with centers of expertise and individual that masters all types of ergonomics.

Over time, the IES has become the representative of the Ergonomics discipline in Israel. The IES is a fairly well recognized organization, and its membership is being kept on a high level of professionalism. Since 1995, the core membership (Full Members of IES) is comprised of about 50 Certified Ergonomists who are practicing all aspects of Ergonomics. The IES is Israel's sole representative to the International Ergonomics Association. The IES's affairs are managed by its Governing Committee headed by a two-year elected Chairperson.

The IES achieves its goals by conducting a variety of activities. IES stands ready to serve the public, where professional activities require research and applied know-how and solving problems related to Ergonomics and Human Factors.

Ergonomics degrees in Israel can be earned on Graduate levels in one of the Universities: Ben-Gurion University and Technion — Israel Institute of Technology. Ergonomics and Human Factors courses are provided in the Industrial Engineering Departments, at MSc and PhD degree level.

1. IES MEMBERSHIP STATUS

The IES is comprised of four kinds of membership groups: Full Member, Associate Member, Representative Member and Student Member.

1.1. Full Member

One who holds a formal academic degree from a recognized university and has 5 years of proven experience in Ergonomics and Human Factors. He has to be recommended to the IES Membership Committee by two Senior IES Members. Alternatively, one who has finished formal Ergonomics studies in a recognized institution and has 10 years of proven experience in Ergonomics and Human Factors. He has to be recommended to IES Membership Committee by three Senior IES Members.

A Senior IES Member is a Professional Ergonomist who has been a Full Member, has been registered in IES for 5 consecutive years, and is serving the profession in a respectful manner.

A Full Member is eligible to vote in IES forums and receive IES publications. He can present himself to IES offices.

1.2. Associate Member

One who is not qualified as a Full Member, but is actively involved in the fields of Ergonomics and Human Factors, for at least 5 years. Alternatively, one who is a Certified Ergonomist and is a member of a foreign Ergonomics Society affiliated to the International Ergonomics Association (IEA). A recommendation from a Senior IES Member is required for this status.

An Associate Member may serve on IES committees and receive IES publications. An Associate Member may not vote in IES forums and may not present as a Full Member of the IES.

1.3. Representative Member

One who is interested in representing an organization, has formal knowledge in Ergonomics and is using as a daily activity.

A Representative Member may receive IES publications; they may not serve as member of IES committees. A Representative Member may not vote in IES forums and may not present as a Full Member of the IES.

1.4. Student Member

One who is a registered student in a university Ergonomics/Human Factors program. A student membership is limited to 3 years. To join the IES as a Student Member a recommendation from a Full IES Member is required.

A Student Member may receive IES publications; but may not serve on IES committees. A Student Member may not vote in IES forums and may not present as a Full Member of the IES.

Italian Society of Ergonomics (Società Italiana di Ergonomia, SIE)

G. Caterina

Università di Napoli Federico II/ Dip., Configurazione e Attuazione dell'Architettura, Via Tarsia 31, I-80135 Naples, Italy

Ergonomics began in Italy at the end of the 1950s, about a decade after K. F. H. Murrel founded the scientific group that gave birth in Oxford to the Ergonomics Research Society (July 1949), and a few months before SELF, a society devoted to the promotion of ergonomics in French-speaking countries. Several studies on the topic had already been developed at the turn of the 20th century, but the term 'ergonomics' was created by Murrel to stress the gathering of different disciplines and with their individual profiles playing a significant role in ergonomics.

In Italy ergonomics was developed in the context cultural background of physiology, hygiene and occupational medicine. Then psychology, sociology, engineering and architecture added their substantial contribution.

In 1961, the Italian Society of Ergonomics was founded in Rome, the same year in which in Stockholm the International Ergonomics Society (IEA) was founded to which the Italian Society of Ergonomics was among the first affiliated members. In 1964, the Italian Ergonomics Association was formed in Milan on the initiative of a multidisciplinary research team. In 1968, this new association merged with the Italian Society of Ergonomics to the present Società Italiana di Ergonomia (SIE). Among the scientists promoting the union of the two associations was an outstanding personage, Professor Caio Plinio Odescalchi, *ad hoc* delegate of the Italian Ergonomics Society.

In the 1960s there were attempts, mainly of an individual nature, at studies, contacts and bibliographic researches, almost entirely in the field of anthropometrics, work physiology, industrial hygiene, occupational medicine or plant design, even though, with special reference to the design context, theoretical intuition already contemplated objectives based on the contributions of a wider range of disciplines.

The first experiments on the workplace, starting from the late 1960s, were dictated by the need to develop analytical methods able to identify conditions of risk, rather than to draw up projects for a real change; this was perhaps due to the difficulty in achieving an organic approach that would take in account the requirements of both the company technicians and the workforce in a period marked by a strong social tension.

The aims of these studies consisted of those aspects that occupational medicine had so far addressed only partially or even neglected, i.e. work postures related to plants and instruments of production, evaluation of energy expenditure for each operation and of parameters for the quantification of heat stress, distribution of information in space and time. Nevertheless, attempts were also made to develop methods to acquire overall knowledge about conditions of production able to cause 'chronic fatigue'.

In this phase, there was a declared awareness of the importance, for ergonomics purposes, of organization analysis but, under the impetus of cultural development and trade union conflicts, participation and subjective aspects were more influential than structural aspects in the assessment of risk conditions.

In the context of workspace experience, the development of ergonomics in the steel industry is of special importance. In 1965, the European Coal and Steel Community (ECSC) initiated research programs and studies on ergonomics within coal and steel industries. In this field the Italian steel industry has developed a significant experience, thus qualifying, with that contribution, as one of the most important reference points for ergonomics in Italy.

In the second half of the 1970s with the beginning of the technological changes, especially in micro-electronics, taking place in a large number of enterprises, a planned and analytical view began to gain ground, taking into more careful consideration the interaction between man and technical systems and thus offering new space to ergonomics. The field of ergonomics activity widened covering all choices concerning the organized work with contributions of knowledge originating mainly from the organization theories.

While in the 1970s the main subject of ergonomics was the man at work, in the 1980s its interest was focused mostly on the consumer with the consequence that the usability context became hard to identify. Therefore, ergonomics is asked especially to design the interface between man and product in which the man is the consumer. There is in this way the passage from the trend man–product to those man–technical system and man–man, the last one mediated by VDT.

In 1996 an important step in the development of ergonomics culture in Italy has been the promulgation of Decree no. 626, in which for the first time ergonomics is officially recommended in a national rule devoted to the prevention of workers risk.

In this new scenario SIE is aiming today to improve education in ergonomics inside Italian universities to define the professional profile and core competences for ergonomists certifying in agreement with the Center for Registration of European Ergonomists (CREE).

Congresses of the Società Italiana di Ergonomia:
- 1° Congresso Nazionale Rimini, 2 October 1974, c/o Centro Pio Manzù
- 2° Congresso Nazionale Milano, 15–16 June 1979, c/o Museo della Scienza e Della Tecnica
- 3° Congresso Nazionale Roma, 22–24 March 1984, c/o National Research Council (Cnr)
- 4° Congresso Nazionale Ischia, Naples, 12–14 May 1988, c/o Grand Hotel Terme Regina Isabella
- 5° Congresso Nazionale Palermo, 4–7 October 1993, c/o Grand Hotel Villa Igiea
- 6° Congresso Nazionale Bologna, 15–17 September 1997, c/o Faculty of Sociology

Macroergonomics

B. M. Kleiner

1. INTRODUCTION

Many environmental and organizational factors have created the need to develop a macroergonomic perspective, cultivate macroergonomic research, and apply macroergonomic methods and tools. These forces include increased competition, restructuring, changes in technology, changing workforce demographics, increased litigation, shorter product shelf-life, increased global competition, and increased information and communication requirements for products, processes, and people. For example, in industry, increased automation has in many cases inadvertently led to technology dominant designs, creating "leftover" functions for operators to perform. As a result, many jobs are monotonous, boring, fatiguing, unhealthy, unrewarding, and/or unsafe. Other design and redesign efforts have been more subtle in their inattention to the sociotechnical characteristics of work systems, but the results have been equally disappointing. While there have been some who have attempted to improve organizations, systems, and jobs, most often, these components are treated as independent entities — having no relationship to one another. Macroergonomics takes the perspective that these work-system components are interdependent. As interdependent components, work-system components need to be analyzed systematically and jointly in order to maximize performance and wellbeing of workers, groups, and organizations.

2. DEFINITION

As presented by Hendrick and Kleiner (in press), macroergonomics is a sub-discipline of the human factors engineering/ergonomics profession with the stated focus of work system analysis and design. In fact, macroergonomics and work system analysis and design are synonymous. A work system is comprised of personnel (e.g. operators) interacting with hardware and/or software (e.g. computers), an internal physical environment (e.g. illumination, humidity), external sub-environments (e.g. legal, political, technological, cultural) and organizational design, including structures, processes, and management systems. Macroergonomics has also been referred to as the ergonomics sub-discipline concerned with human–organization interface technology. Other sub-disciplines are concerned with other technologies such as human–machine interface technology, human–environment interface technology, human–software interface technology, and human–job interface technology.

3. MACROERGONOMICS AS A SUB–DISCIPLINE

As a formal sub-discipline of ergonomics, there are macroergonomics technical groups in the Human Factors and Ergonomics Society (HFES) and the International Ergonomics Association (IEA), as well as in other national professional ergonomics bodies. For the ergonomics generalist or microergonomics specialist, macroergonomics offers a systems perspective which creates a useful context for microergonomics intervention. In addition, this potentially increases the relevance and potential impact of ergonomics because work system factors often hinder system performance or user acceptance downstream. By considering these factors upstream, success is more likely. For the macroergonomist, the sub-discipline offers more than a perspective. Like other scientific sub-disciplines, macroergonomics has its own research stream of knowledge, research methods, and practical methods and tools.

4. MACROERGONOMICS PERSPECTIVE

For the ergonomics generalist or microergonomics specialist, macroergonomics represents a perspective that provides the ergonomist with an appreciation for the larger system — a perspective that will increase the likelihood that the microergonomic interventions will flourish. Too often, a product, process or interface has been designed without attention to the larger system. In systems terms, the design was optimized, but without attending to the larger system first, the total system can be sub-optimized. That is, ultimately implementation success does not depend only on the merits of a micro-level design but whether and to what extent micro design is aligned with macro-level factors and forces from the environment, including organizational, political, economic, cultural, and legal factors.

5. MACROERGONOMICS RESEARCH

Macroergonomics basic research is concerned with the relationships among technological, personnel, internal and external environmental, and organizational factors and their interactions. These factors and their individual and combined effects on performance have been studied in both the laboratory and in the field that is in actual work systems. The sociotechnical systems research body of knowledge and the ergonomics literature offer a foundation for this research stream. Laboratory research offers the advantage of control, while field studies are often more relevant, but lack the precision afforded by controlled laboratory studies. In both cases, experiments or studies are designed and conducted, typically using humans as research subjects. Various objective performance, usability and subjective data are collected at the individual, group, and organizational levels. These data are statistically analyzed with inferences about causal relationships drawn that contribute to the ongoing development of a research body of knowledge.

6. MACROERGONOMICS METHODS AND TOOLS

Other research methods, often used in combination with experimentation, include survey research, simulation, systems dynamics modeling and ethnography, each with its own set of tradeoffs.

Complementing the empirical research, systematic macroergonomic methodologies for analysis and design of work systems have emerged as well.

The design of a work system's structure and related processes involves consideration of three major sociotechnical system components that interact and affect optimal work system design. These are the technological subsystem, personnel subsystem, and relevant external environment. Each of these components has been studied in relation to its effects on the organizational design dimensions of complexity, formalization, and centralization. Respective empirical models exist to analyze work system structure. Complexity refers to the extent to which the work

system is differentiated or segmented (e.g. vertical, horizontal, spatial) and the extent to which there is integration in order to link the various segments. Formalization refers to the extent to which there is standardization present in the work system (e.g. procedures). Centralization refers to the degree of decision-making performed by relatively few persons, usually at the helm of the organization.

In addition to analyzing and designing the structure of the work system, it is desirable to analyze and design the processes of the work system. One approach is first to scan the external environment of the work system in order to understand the forces that will necessarily affect performance. Next, a technological subsystem analysis will define the type of production system and will detail the work process steps and variances. Then, control of these variances is addressed by identifying the roles needed to control or prevent variances. Following this, joint design of ergonomic solutions is carried out. This involves function allocation and the design of technological and training support for human functions. It is also helpful to evaluate the operators' perceptions of their roles and to redesign the reward sub-system and any other sub-system to support human performance. Finally, a cycle of continuous improvement will periodically review and seek additional changes.

As methods such as these are implemented, a philosophy of "participatory ergonomics" is adopted. Participatory ergonomics implies workers participate in both ergonomic analysis and design, especially as these procedures relate to their jobs. This requires some degree of decentralization or sharing of decision-making. The process of participatory ergonomics can be as simple as involving users in their own usability testing or involving workers in formative or summative evaluation of their workplaces or processes, but may be as complex as training non-professional workers in the concepts and methods of ergonomics for more sophisticated participation. The basic principle is that ownership and quality of results and involvement in the processes of creating results go hand-and-hand.

7. MACROERGONOMICS RESULTS

When a work system has been designed effectively from a macroergonomics perspective, and that effort has been carried through to the microergonomic design of jobs and human–

machine and human–software interfaces, then the work system design is harmonized. As a result, synergistic functioning becomes possible, and system productivity, safety, employee satisfaction, commitment, and perceived quality of work life will be much greater than the simple sum of the parts would suggest or predict.

Macroergonomics has been demonstrated to improve work system performance throughout sub-system areas, including productivity, quality, safety and health, quality of work life, user satisfaction, etc. In terms of the magnitude of improvement, Hendrick (1986) suggested that improvements of 60–90%, or more, are to be expected by macroergonomics approaches, while microergonomic approaches can at best achieve 10–20% improvement.

8. THE FUTURE

As macroergonomics continues to be developed as a sub-discipline, the need increases. Global competition and restructuring continue, creating changes in the demographics of the workforce. As many organizations struggle with their turbulent external environments, they have turned to such approaches as total quality management, reengineering and restructuring. Macroergonomics can be the organizing framework for such approaches to organizational change and process improvement. In this context, the macroergonomist has played the role of change agent.

In addition, systems, products, and processes continue to become more information-dependent, creating heightened needs for the knowledge worker. Optimizing work systems within the context of this unprecedented level of information content is a daunting task for the ergonomist. Macroergonomics offers the perspective, research-based knowledge, methods, and results needed for such a challenge.

REFERENCES

HENDRICK, H.W., 1986, Macroergonomics: a conceptual model for integrating human factors with organizational design. In O. Brown, Jr., and H.W. Hendrick (eds) *Human Factors in Organizational Design and Management, II* (Amsterdam: North Holland), pp. 467–78.

HENDRICK, H.W. and KLEINER, B.M., in press, *Macroergonomics: An Introduction to Work System Design* (Santa Monica, CA: Human Factors and Ergonomics Society).

Ontology

V. Raskin

Department of English and Interdepartmental Program in Linguistics, 324 Heavilon Hall, Purdue University, West Lafayette, IN 47907-1356, USA

1. INTRODUCTION

"Ontology" is an ambiguous term. It may be used synonymously with the philosophical discipline of metaphysics, an ancient and sometimes maligned theory of most basic categories. It may refer to a much newer discipline of formal ontology, based on mereology and topology, dealing with the formal, mathematical methods of representing elements of reality. And it may refer to engineering artifacts, conceptual structures, used in various applications.

Ergonomics should be particularly and obviously interested in the last, engineering aspect of ontology because of the ergonomic focus on specific applications for specific tasks in specific domains. It is important, however, to lay out the philosophical and formal foundations of ontology because their understanding affects the engineering enterprise directly: a specific ontology can be designed better or worse, more or less costly, more or less effective depending on the developers' understanding of the nature of the phenomena they are dealing with in the process.

Much of ergonomics depends on natural language, especially on the meaning expressed by and in it and on the comprehension of that meaning both by the speakers and hearers. It is in the area of natural language semantics and its application, computational semantics, that the ontological approach has been particularly successful.

2. FACETS OF ONTOLOGY
2.1. Basic Notion and Example

An ontology is understood here as a tangled hierarchy of concepts, such that each concept is a frame with a set of slots and fillers. The specific examples come from one of the largest and richest ontologies in existence, the one developed initially for the MikroKosmos knowledge-based system of machine translation (MT) and available for browsing on the Web at crl.nmsu.edu. Its top level consists of the concept ALL, which has three "children," PROCESS, OBJECT and PROPERTY — the names of the concepts are, however, mere mnemonic labels for human user convenience and can be easily replaced without any changes in the ontology as long as the frames remain constant and consistent. All the other concepts in the ontology are descendants of at least one of those general concepts and inherit their slots and fillers at least monotonously, i.e. vertically down. Thus, the process PERMIT may be represented as:

```
PERMIT
    is-a          directive-act
    agent         human
    beneficiary   human
    theme         event
    agent         permit.beneficiary
    precond       have-authority
    agent         permit.agent
```

theme   OR   permit.beneficiary
permit.theme

The concept is characterized as a DIRECTIVE ACT, which is a descendant of MENTAL ACT, which is, of course, ultimately a descendant of PROCESS (and more closely, a descendant of ACT or ACTION, one of the two children of PROCESS, the other one being STATE). The act has a human agent and a human beneficiary and it concerns an event involving that beneficiary as an agent: provided that the agent of PERMIT, permit.agent has the authority to allow the beneficiary to be the agent of that event (i.e. to perform a certain act) and/or the authority over that kind of act, the beneficiary of PERMIT, permit.beneficiary, is licensed to bring about the event by committing an appropriate act.

Each of the slots and fillers in the concept entry is itself a concept in the ontology (and was not put in the customary small caps for easier reading). In the real ontology, PERMIT will inherit several slots and fillers (certainly and at the very least, the agent/human and beneficiary/human slot/fillers) from its parent, DIRECTIVE-ACT, which, in turn, will inherit (at least some of) them from its parent, COMMUNICATIVE-ACT.

2.2. Foundational Issues in Ontology
2.2.1. Ontology as metaphysics

The term "ontology" is used by some philosophers interchangeably with metaphysics. Metaphysics, the most ancient part of philosophy, dating back at least to Aristotle, deals with the most basic categories of reality: for Aristotle and many others, primarily objects and processes. In the empirical 19th century, metaphysics was perceived as experimentally unverifiable and, therefore, irresponsible, and parts of it may have been so. There has been no going away, however, from the necessity of understanding the basic categories underlying both human thinking and the spectacular successes of empirical sciences. In the 20th century, the concern about foundational issues of physics and other hard sciences has become common and acute. The term "ontology" was used for a while as a "polite" substitute for metaphysics, but towards the end of the century, the later term regained its respectability. Accordingly, "ontology" started being used more correctly as the actual result of metaphysical research, a specific system of basic concepts.

The major issue of metaphysics is the status of properties. The realists recognize the existence of both individuals, such as houses or cars, which are spatio-temporal entities, and of their shared properties, such as costliness or redness, which are not spatio-temporal but rather abstract. Their opponents, the naturalists, recognize only individuals. Both camps, existing in a large number of variations, have a difficult philosophical issue to deal with: for the realists, the different status of properties; for the nominalists, the existence of those properties. For a specific ontology, the matter is only important in one sense, namely, whether the ontology is to contain properties or not, and the nature of any reasonable application usually imposes the realist solution.

2.2.2. Formal ontology

Formal ontology was developed late in the 20th century as a cross of mathematical disciplines of mereology, which studies the relations between parts and wholes, theory of dependence, and topology. But it was originally Edmund Husserl earlier in

the century, who saw a need to establish formal ontology as the "theory of things" parallel to logic, the "theory of truths."

The agenda of formal ontology typically includes explorations into notions of time and space, modality, especially the deontic (imperative, must-do) phenomena, taxonomy of artifacts, and types of inheritance. A specific ontology, in its implementation, has to take a position on each of those and other issues of formal ontology, so the awareness of these issues is essential for the developers.

2.2.3. Ontologies as engineering artifacts

The MikroKosmos ontology, containing over 7000 actual and many more virtual concepts (i.e. concepts derivable from the actual concepts with the help of well-defined rules), was created by a team of computer scientists and linguists over almost a decade at a considerable cost. After the original time- and effort-consuming basic research, which established the top-level concepts and the format of each type of entry, a major effort was devoted to the development of semi-automatic tools of acquisition enabling the participation in the work, at its simplest level, of rank-and-file enterers with a minimal training. This was achieved by an inventive use of easy-to-understand templates for several basic types of concepts. The introduction of the tools made the acquisition of numerous concepts very speedy and inexpensive.

The resulting ontology meets and exceeds the industry's standards of clarity, coherence, and extendibility, and it stands to reason to conclude than any ontology of the future, comparable to this particular one in size and depth, will have to be developed more or less along the same lines: a combination of the tiered approach to acquisition, the use of semi-automatic tools, and the emphasis on cheap labor whenever and wherever practical.

2.3. Ontology and Natural Language

It is in the development of MT and other natural language processing (NLP) systems that the ontological approach proved to be particularly successful. Early MT was of the transfer type: a system was designed on the basis of a minimally required rules of transfer from words of the source language into the words of the target language. While the simplest transfer, word for word, does not work in translation, inventive rules were created to overcome this difficulty in many cases. The resulting translation was, however, pretty rough and not very accurate, making it hard to use without expensive human post-editing.

The alternative to transfer MT was the interlingua approach. Interlingua was a formal representation of text. The input text was analyzed and represented as an interlingua text and the latter was used to generate an equivalent text in the target language. An ontology was introduced as a rich and sophisticated form of interlingua in the late 1980s and won a dominant position in the interlingua approach in the 1990s.

The MikroKosmos ontology is the foundation of the whole knowledge- and meaning-based approach to language processing. As a universal language-independent resource, the ontology is the metalanguage for the lexicon: each lexical entry in a natural language is defined in ontological terms. In the simplest case, the meaning of a word corresponds exactly to an ontological concept; in this case, the entry contains a simple pointer. More often, the match is not exact and an additional constraint, formulated in terms of ontological properties, must be added to

the entry. Yet in other cases, the entry must contain a reference not to a concept but rather to a certain slot and/or filler in its frame.

Even more importantly, the ontology provides a conceptual foundation and world grounding for whole texts. Working with the lexical entries making up a sentence and using the pertinent syntactic information for the natural language being processed, the analyzer produces ontology-based formula, the text meaning representation, for the sentence.

The approach has been successfully tested in non-toy systems of MT, information extraction and summarization, and other sophisticated NLP applications. It is particularly powerful in limited professional domains, which are precisely what ergonomics usually serves in a specific application.

3. STANDARD OBJECTIONS AND REBUTTALS ABOUT ONTOLOGY

While ontology has been gaining ground in natural language and other research, there has been considerable resistance to it from some quarters on a variety of grounds.

3.1. Natural Language Fallacy

Some computational semanticists insist that ontological concepts are simply words of a natural language. For some of them, this is perfectly acceptable and they proceed with their own methodologies, explaining meanings of natural language in terms of the same natural language. Theoretically, this is how the later Wittgenstein's notion of meaning is usually seen. Practically, this is how traditional lexicographers wrote their monolingual dictionary entries. For others, a charge of circularity is raised and it is indicated that the connection between natural language and extra linguistic reality (world grounding mentioned above) is not then addressed.

Whether accepted or rejected as a method, ontological terms cannot be confused with words of natural language. Developed for use by computer, the former are totally devoid of ambiguity, which removes them from natural language. What confuses these critics of ontology is that when a human user reads these terms he or she may know the other, unintended meaning(s) of the word in a natural language that is used as a concept label. The computer does not possess this ability and is, therefore, unaffected by the ambiguity, which renders the ontological labels non-natural.

Besides resulting in disconnecting natural language from extralinguistic reality that it expresses, this natural language fallacy leads to the confusion between object language and metalanguage. All approaches use metalinguistic terms for the categories that they use in language description or representation. Only the ontological approach introduces these categories explicitly and completely, usually in the PROPERTY branch.

3.2. Irreproducibility Charge

Other objections to ontology question the scientific method of the approach. A sound scientific research should be reproducible, they say, but it is clear that two different people, let alone groups, will come up with different ontologies for the same conceptual domain. The confusion here is twofold. First, the creation of ontology is not equal to a scientific experiment, and extending the burden of reproducibility to it requires a considerable

metaphorical extension — of the kind that Chomsky accused B. F. Skinner of in 1959, effectively removing him and eliminating his influence from cognitive psychology. Second, given the same parameters and conditions, as in a scientific experiments, which includes the semi-automatic acquisition tools, different people and groups do indeed come up with identical or compatible results, and this has been tested in practice — unlike the speculative argument of ontology foes.

3.3. Engineering Objections

The objections here are serious because they are cost-related. Ontology acquisition is costly, especially at its initial stage, when methodology and tools have to be developed and deployed prior to and along with the massive acquisition stage. There is, however, a major catch: once done, ontologies can be expanded, borrowed, interchanged, etc. at a nominal cost. A special effort by formal ontologists has led to the development of a standard Knowledge Interchange Format, and both the MikroKosmos ontology and other ready-made ontologies have been successfully borrowed, adjusted, and deployed by other groups in new domains. In other words, the high costs do not need to be incurred more than once, and this was already done.

4. ONTOLOGY AND ERGONOMICS

Ontology is essential for any ergonomic application which deals with conceptual representations. It is especially in limited domains of technology that ergonomics so often operates in that ontology provides a complete theory of the world of the domain, i.e. all the possible objects, processes, properties in minute detail, thus defining formally — mathematically, that is — every possible and impossible state of affairs. Any form of communication, be it in natural language, art, or other visual or audial media, is incomparably improved if underlain by the total clarity guaranteed by the ontology. Ergonomics does not have to expend any resources on developing ontologies. Ready-made ontologies should be used — and have been used — to improve ergonomic methods and results.

REFERENCES

GROSSMANN, R., 1992, *The Existence of the World: An Introduction to Ontology* (London: Routledge).

GRUBER, T.R., 1995, Toward principles for the design of ontologies used for knowledge sharing. *International Journal of Human and Computer Studies* [special issue on "The Role of Formal Ontology in the Information Technology", eds N. Guarino and R. Poli], 43, 907–28.

GUARINO, N. (ed.), 1998, Formal ontology in information systems. In *Proceedings of the First International Conference (FOIS'98)*, 6–8 June, Trento, Italy (Amsterdam: IOS Press).

HUSSERL, E., 1900–01, *Logische Untersuchungen* (Halle: Niemeyer; 2nd ed., 1913–21); English trans. of 2nd ed., *Logical Investigations* (London: Routledge, 1970).

NIRENBURG, S. and RASKIN, V., 2000, *Principles of Ontological Semantics* (Cambridge, MA: MIT Press).

An Outline of Ergonomics, or the Science of Work Based upon the Truths Drawn from the Science of Nature

W. B. Jastrzebowski

PART I (NO. 29, PP. 227—31)

A word is as good as a wink to the wise man

***

INTRODUCTION

Hail, Thou great unbounded idea of work! God, Who, as the Bible teaches us, cursed mankind and subjected him to work, cursed him with a father's heart; for the punishment was also a consolation. He who complains against his work knoweth not life; work is an uplifting force by which all things may be moved. Repose is death, and work is life[1])!

If there were as many people preaching the idea of work in other words the idea of the deed as there are those who preach the idea of the word, the thought, the feeling, form, of Mysticism, Idealism, Materialism, Pantheism, Egoism and many other such things, what a great deal of good there would be happening in the world!

What would be the fate of two societies, of which one would be engaged in useful work and praiseworthy deeds, such as the tilling of the soil and the improvement of the people, while the other consisted of persons busy with words, thoughts, feelings, and other things of those just enumerated?

Affecta nihil aliud sunt, nisi accessoria ad facta.

Affections are nothing else, but accessories to deeds.

By their deeds shall ye know them.

Not every one that saith unto Me, Lord, Lord, shall enter into the kingdom of heaven, but he that doeth the will of My Father (Matt. 7:21).

Work enriches or *divitiates* us, making us more like unto the *Divine* . . .

Work is the mother of all good.

He who relies on others shall be forsaken by God.

God cannot save us without us.

Were God to come down from heaven and appoint the Archangels His Ministers and the Angels his servants, there shall still be no happiness in the world until mankind comes to love work and until all people apply all of the forces given them by the Creator to work.

One idle man, in other words one squanderer of the forces and faculties with which Man is endowed by God, may do more evil in the world than a thousand industrious people can do in amends for him.

One bad man can do more damage than ten good men can help.

As the body of a particular man suffers and incurs loss from the ailment, idleness or malfunction of even one of his limbs, so too the body of society suffers and needs must incur detriment if even one part of it is in a similar condition.

What a thousand good men gain in a thousand years of work, one wicked man given to laziness may destroy in one day[2]).

Just as there may be no good wall without good bricks, and no good edifice without good walls, so without good people there may be no good peoples, and without good peoples mankind shall not be good.

Which people, and which of their labors is best — is it not that whereby the most: the most things, the greatest number of people, and they themselves — have been bettered?

The above-cited sentences, whatsoever source they may come from (which is of least import to the truth that through them is revealed, and to the good that may come to us and to all men from the acknowledgement of this truth as indisputable) have been placed at the head of this treatise on work, in other words on the good use of the forces and faculties given Man by his Maker, so that, knowing the need for such work on the strength not merely of our own but of the general conviction, and acknowledging it as an indispensable condition of our happiness and well-being in this world, we many endeavor to bring first our own will unto work, and then the wills of all those of our fellow-brothers whom we have the power to influence by word or example. By this inclining of ourselves and our fellow men unto work we may contribute to the fulfillment of their and our own hopes for the improvement of their and our existence. Which improvement and increase both in the material as in the moral respect haveth its surest warranty only in the said work, *source of all good.*

That this be truly so, that it is only in work that the surest hope lies for the improvement and increase of our own and the general well-being — in work which is useful and commendable, such as the tilling of the soil or the betterment of people and things — no man appears to doubt in the least, since it is only through such work that all things may be improved and rendered fit to serve the common good.

And apart from this our *forces* and abilities, whereby and through the guidance of which we perform our work, develop through our exercise of work, perfecting in the proper respect, and thereby contributing to the advancement and perfection of our entire being, which is the condition for our felicity and without which our existence is meager and ever under the threat of doom. For it is well known that our vital *forces* grow weak and impoverished as much by the *lack* of their exercise as by their *abuse*; and that they are maintained in their proper condition, growing and increasing by their proper and moderated exercise, which we call work, and whereby we improve things, people,

and ourselves, making them and us more conducive to the service of the common good.

But lest that *exercise* of our *forces*, which is the principle and essence of our lives, be too burdensome upon us; that it may bring us to the source of felicity and contentment — we ought ever to arrange it in such a manner that it may not hamper those vital forces in us, but that it may extend to include all those forces we have in us and by which we may serve the said common good.

For it is only through such application of all our forces united that they may be mutual supports one unto another, not only making our work lighter but also bringing us greater profit, as we have seen elsewhere, for instance in the *application* of those forces to the enhancement of the earth's fertility. For if the earth's fertility be improved merely by the least of our vital forces, that is by the physical or *motory* force, the force of movement whereby we subsist only in sleep, in sleepwalking and in the mere unconscious carrying out of our ordinary chores, she shall prove but so little enhanced in her action and capacity to act that for each grain we entrust unto her she shall bring forth but *two* in harvest[3]).

But if we endeavor to raise the earth's fertility by applying both our physical and our aesthetic forces, that is our motory and our *sensory* or *emotional* forces, which will ensue when we cultivate the land industriously and with a sense of taste, the earth's fertility will be doubled by the engagement of our effort so as to bring forth a *fourfold* crop at harvest for every grain sown.

This gives us grounds to suppose that if we were to apply three of our vital forces for the enhancement of the earth's fertility, our physical force, our aesthetic force, and our intellectual force, that is our forces of *motion*, of *feeling*, and of *reason*, then the earth's fertility would become even more efficient, such that for every grain sown the crop would be *eightfold* at harvest-time. This is generally confirmed by all our present-day well-managed farms, that is those that are managed industriously, and according to the precepts of taste and reason, with the use of the three above-mentioned forces. This leads us to the premise that if we could enlist yet a fourth of our forces, adding it to the other three, the earth's fertility could be raised to a still higher power.

And since the relation between the three former forces with respect to the increase of the earth's fertility is, as we believe, a *doubling effect of the previous outcome*, for as we have seen the ration between them is 2: 4: 8, we may thus expect that this fourth force, the moral or *spiritual* force which induces us to work not only for our own and the common good (which entails the glory of God, the welfare of our neighbors, of our fellow creatures and of ourselves) could advance the earth's fertility so much as to yield twice eight, or *sixteen* grains at harvest from every grain sown.

This is not far from the truth, for we have observed through our own experiment, carried out in the strictest conditions, that given our appropriate application of all the four vital forces just mentioned to the earth's fertility, it may be brought to the state where for every grain sown, e.g. of wheat, the harvest from even modestly fertile soil not particularly suited for wheat will be three-hundredfold. But such an occurrence is still not the highest on record, for elsewhere there have been experiments carried out which have shown that with the engagement of but three of man's vital forces, that is his *motory* force, his *emotional* or *sensory* force, and his *intellectual* force, the harvest obtained was over three times the just quoted value of a three-hundredfold crop, that is about one thousand grains of crop from each grain sown — one hundred and twenty-five times eight, which marks the richest ordinary yield to be expected from the application of the three just mentioned forces, and 62 1/2 times 16, that is the crop to be expected from the engagement of all four of our forces, and without their very great exertion.

What has been said here of the earth's fertility, that is of *its increase by the application of our vital forces to it*, holds for all the other forces of living and inanimate Nature. For even those forces, as for example steam power and electrical power, which previously in their natural condition were of so little effect under the exercise of merely our *sensory* forces upon them, have now in our own times progressed so far in their marvelous and beneficial effects to a much higher importance by the application to them of the first three of the human forces. We shall now turn our attention to the forces of living Nature only, of which we say there are only as many as we have observed in our own being, that is four (viz. 1. the physical, kinetic or *motory* force of movement, such as is given to all living creatures, even the lowest of them, the Plants; 2. the aesthetic or *sensory* force which is possessed by all creatures capable of feeling, even the lowest of them, the invertebrates or *Primitive Animals*; 3. the intellectual forces of *reason* of which all comprehending and thinking creatures are endowed, even the lowliest of them such as the fishes, the reptiles, the birds, and the quadrupeds, in other words all *Animals*; and finally 4. the moral or *spiritual* force which has been ascribed only to human beings, that is to *Mankind* proper). If we now turn our attention to these forces of living Nature we shall see that they are even more conducive to give bigger and bigger yields as we exert our vital forces on them than are the forces of inanimate Nature, which make the things on which they act and in which they are lodged produce only inanimate returns, in other words only the utilities, for the common good. But the forces of living Nature can generate, through their development and perfecting brought about by our own forces acting on them, gains which are animate, that is the services, favors or goods, as demonstrated by properly conducted bee-keeping, by well-managed animal husbandry, and by a similarly well-ordered education and management of human resources. As a result of the appropriate education and suitable government of human beings, in other words through the furtherance and management of man's vital forces by himself, those human vital forces turn out to be of incomparably higher value and manifest a greater ability to serve the afore-mentioned common good than the corresponding value and ability in other people who are no so organized.

Such other people, who have not had their vital forces developed and organized by our own forces, are often in the same or even in a worse position than land that is uncultivated as regards their faculty to apply their forces. For uncultivated land, which as is well-known does not yield a harvest of cereals, nevertheless brings forth grasses and herbs, bushes, trees, and other such like beautiful and useful produce; whereas people who are badly educated and not organized, in other words whose vital forces have been neglected or are badly managed, very often not only bring no profit or gain, but are even harmful as regards the common good.

This is clearly the outcome of the failure of our vital forces to

exert an effect on their forces; or it may be the result of an inappropriate application of those forces. For it is well-known that the earth's yield is meager or none at all not only when we have neglected its fertility, but also if we have applied our four vital forces in an inappropriate manner, contrary to the nature of the land and the crops which are to grow on it.

If it is indeed so — and hardly any who considers it well shall have any doubts — as has been shown on the example of the earth, the animals, and mankind, that by the *application of our forces on the external forces of Nature, living and inanimate, these latter forces, and hence also the things and the creatures in which they are lodged and through which they manifest their visible effects, acquire a greater value or ability to serve the common good the more efficiently and the more appropriately we exert our own vital forces upon them*; if it is indeed so, then the exertion of our vital forces for the common good, which is called work, deserves our special and scrupulous attention. All the more so as by such exercise we gain still one further advantage profitable and useful for the common good, for, *like the magnet*, the magnetic force of which, as is well known increases through habitual use, we shall *promote and augment the vital forces within us by their frequent employment, and in consequence we shall achieve an ever greater and more effective increase in the value or virtue of other things and other people, determining their ability to serve the common good*, to which everything is devoted, following the goodness of God, and to which thanks to Man the most perfect creature made in the likeness of God more and more may and should be devoted[4]).

The importance of the proper exercise of our vital *forces* here described, the importance of our work, by which we are to bring and encourage other things, and other creatures like unto us or not, to work, that is to exercise their own vital *forces* for their own, our and the common good, supplies the reason and strong incentive for us to undertake this work of scholarship on work, and even to establish a new discipline regarding a subject which is of no lesser significance than other subjects of scholarly inquiry, not to mention its superiority over other questions of lower standing which merely entertain our curiosity but do not show us what we are to do and how we are to act in this age in which we are now living and which we call the Age of Action; how we should live to make our lives as bountiful as possible at the least expenditure of toil and with the greatest inner contentment, bringing forth fruit for our own and for the common good; ultimately with all due equity to be judged by others and by our own consciences.

Lest this Science of Work, understood as Work in the comprehensive and integral sense, not merely its part that is physical labor or *toil*, but physical, aesthetic, rational, and moral work, that is *Labor*, *Entertainment*, *Reasoning*, and *Dedication*, that is Work to be performed by all the *forces* assigned to us by our Maker and to relate to all the purposes of our existence as intimated by pure religion and an untarnished sense of personal dignity — lest this Science of Work deter us both by its novelty and by its expansiveness, we shall first give its brief outline or structure, and then, if it prove necessary, we shall expound and develop the subject more fully.

The Science of Work, or more accurately works, which we shall venture to call Ergonomics, after the manner of the other sciences, from the Greek *ergon* — work, and *nomos* — principle or law, is similar to the Ananconomics, or the Science of the

Needs which we presented a decade ago, because together they form one Science of Life, that is Bionomics, and are sustained in a quaternal composition because life is determined by the operation of the four already described vital forces to which all that is in nature corresponds, and by which we and our fellow creatures attain to all that is good for ourselves and for the common welfare.

PART II (NO. 30, PP. 236—8)

Work is the mother of all good.
***

(1) By the term Ergonomics, derived from Greek *ergon* (ἔργου) — *work*, and *nomos* (νόμοζ) — *principle* or *law*, we mean the Science of Work, that is the use of Man's forces and faculties with which he has been endowed by his Maker.

(2) The Science of Work, understood in the widest possible sense of the term "work", may be divided into two main disciplines, the science of useful work, which brings improvement or is commendable, by which we mean the good use of Man's forces and faculties with which he has been endowed by his Maker, or their use for the common good; and the science of harmful work, work that brings deterioration, and discreditable work, by which is meant the contrary use and intention to use the said forces and faculties.

(3) Useful work, which is the only kind of work we shall discuss here as the good from which all other goods come, and which rests chiefly in the improvement or perfecting of things, people, and ourselves, is divided, in accordance with the nature of the forces we devote to it and under whose guidance we carry it out, namely into:
Physical, Aesthetic, Rational, and Moral Work;
that is, Work which is
Kinetic, Emotional, Intellectual, and Spiritual,
or *Motory*, or *Sensory*, or *Rational*,
and which may otherwise be known as:
Labor, Entertainment, Thinking, and Devotion
or Toil, or Pastime, or Reasoning, or Dedication

(4) In each of these kinds of work, examples of which are as follow:

the breaking of stones	the playing with stones	the investigation of their natural properties	the removal of stones from the road[5]

there are four chief considerations:
(i) with what creatures do we share this work?
(ii) in which periods of our lives are we particularly suited for this work?
(iii) in what manner may we proceed in this work?
(iv) what are the benefits to be drawn from this work for ourselves and the common good?

Chapter 1: The shared or common nature of human work with the work of other living creatures

(5) The four kinds of work just mentioned above, that is *Labor, Entertainment, Reasoning,* and *Dedication,* are the categories of our own and the general employment; such that the first

category is performed chiefly by our *motory* forces, which we share with all the living creatures, that is plants and vegetables, primitive animals, animals, and humans; the second category is accomplished chiefly by the *sensory* forces, with which we and all the creatures capable of feeling have been endowed, that is humans and the animals and primitive animals; the third is done chiefly by the forces of *reason*, which has been given to all the thinking and reasoning creatures, that is to humans and to the animals; and the fourth is carried out chiefly by the *spiritual* force, which has been apportioned only to the creatures capable of self-dedication to the common good, and such are only human beings. Therefore, as has already been said, these four categories of work are divided as follows:

Labor	Entertainment	Thinking	Dedication

may be and are performed by:

Plants	Primitive Animals	Animals	Humans
	Primitive Animals	Animals	Humans
		Animals	Humans
			Humans

(6) The proof of this obvious truth is to be had in the following creatures:

esparto grass	wrynecks	wise serpents,	humans who devote themselves to work for the common good
earthworms	squirrels	birds, animals	
moles	persons given to entertainment	and similar humans	
common laborers			

The esparto grass which has been mentioned here, like many other plants which send down their roots very deep into the soil, is thus regarded to be performing physical work or Labor. Esparto grass pierces and shakes the soil with its roots just as the earthworm does, thereby facilitating the penetration into it of rainwater and all the substances washed out of the atmosphere, and contributing very significantly to the soil's fertility. Not only esparto grass and other herbal plants, but all vegetal creatures whatsoever perform this kind of work. The most conspicuous proof of this is offered by the trees, which, when they grow in the ground, push aside and move the soil so mightily with their roots, that even were it as dense as can be, even if it contained boulders, it could never withstand the force of pressure exerted by the roots of trees. And when they grow on rocks trees give rise to the breaking up and splitting asunder of those rocks. Even the herbal plants seem to make no meager contribution to this process, as evidenced by the name itself of a species of plant which involves many varieties, the Saxifrage species, the greatest number of specimens of which occurs in the rocky parts of high mountain areas, and which appears to be as good at breaking up the rock surface which is softened by moisture as esparto grass and other plants growing in the soil are at riddling the earth with their roots. And then upon the decay of those roots not only do they enrich the soil with their remains, but they also leave a network of orifices through which rainwater seeps in more easily. The rainwater in turn makes the soil moist and also more fertile

through the draining into it of all the atmospheric gases and emanations in the atmosphere from the burning, fermenting, decomposing, evaporating, and respiring creatures, living or dead.

As we have proved with unquestionable examples, and in accordance with our profoundest conviction, the *Plants*, along with all other living creatures, are engaged on motory work, otherwise called *Labor*. Likewise we could prove that the Lower or *Primitive Animals*, along with all creatures capable of feeling, busy themselves with the work called *Entertainment*; that the *Animals*, along with all the creatures capable of learning, are involved in the intellectual work called *Thinking*; while only *Man*, that is human beings, busies himself with spiritual work otherwise known as *Dedication*. However, since this matter calls for a rather extensive discourse supported by numerous examples to demonstrate its truth, it cannot be entailed in the outline of the Science of Work, but rather in its exposition, which does not concern us here, and the subject of which, relating to the nature of the work carried out by the four kinds of living creatures, may be found by the reader seeking enlightenment on the matter in our *General Natural History* in Jaworski's Yearbook for 1857.

Chapter 2:
On the aptitude of humans to undertake various kinds of work in particular periods of their lives

(7) Since the physical or motory force manifests itself in humans in the active condition in all periods of their lives, that is both in infancy, youth, maturity, and old age, the aesthetic or sensory force manifests itself only in the last three periods; the intellectual or mental force manifests itself chiefly in the last two periods, that is in maturity and old age; and the spiritual force manifests itself chiefly (at the present stage of Mankind's history) only in old age[6]), hence the four above-listed types of work, viz.

Labor	Entertainment	Thinking	Dedication

which are carried out chiefly by the corresponding force, or under its leadership are carried out, or at any rate are most frequently carried out by Man in his

infancy	*youth*	*maturity*	*old age*
	youth	*maturity*	*old age*
		maturity	old age
			old age

A proof for the four circumstances in which Man undertakes work is supplied by the following example:

The taking and eating of useful things[7]	passing one's time on things which are pleasant and beautiful	The examination of unknown and interesting things	The bringing of all things to serve the common good

Chapter 3:
On the methods in which the four kinds of work may be performed

(8) There are four chief methods whereby Humans may undertake, and indeed do undertake work, viz.,

The first method

When they carry out the kinds of work already defined as

Labor   Entertainment   Thinking   Dedication

solely by means of the corresponding main force, that is by the

Motory force   Sensory force   Intellectual force   Spiritual force

In such cases only a minimum effect, or even no effect at all is achieved, for at least the three latter forces, unless assisted by the motory or *executive* force, cannot bear any fruit, they cannot perform any useful deed for the common good[8]). The forces corresponding to these three kinds of work, namely the sensory, intellectual, and spiritual forces, have no power to effect; they are merely such powers the first of which *encourages* or *incites*, the second of which *manages* or *guides*, and the third of which *sanctifies* or *devotes*.

The second method

(9) The second method of undertaking the four different kinds of work is the one undertaken, as in the previous instance, through the corresponding or chief force assisted through the support of the lesser forces, viz. when the four kinds of work, that is

Labor   Entertainment   Thinking   Dedication

may be regarded as manifestations of the following human forces:

| *motory* | *motory* *sensory* | *motory* *sensory* intellectual | *motory* *sensory* intellectual *spiritual* |

or in other words of the following forces:

| *executive* | *executive* *inciting* | *executive* *inciting* guiding | *executive* *inciting* guiding *sanctifying* |

PART III (NO. 31, PP. 244–7)

This will ensue when the different kinds of work are combined; namely, the first kind with toil; the second with toil and pleasure; the third with toil, pleasure, and thought; and the fourth with toil, pleasure, thought, and the love of the common good, that is such good which only man can know and love, and which is unique in that it contains all other goods, chiefly the following four: the glory of the Supreme Perfection, the good of our neighbors, the good of our fellow creatures, and our own good.

In this manner the four mentioned kinds of work are undertaken. All of them are activated by the *executive* force; while the latter three are also supported by the other superior forces; and they shall be bound to render their fruit unto us, all the more bounteously the more forces there are contributing to its achievement.

The third method

(10) The third method of performing work is when we carry out the different kinds of work in the manner described above

but also avail ourselves of the assistance of the forces of other living creatures, that is:

Plants   Primitive Animals   Animals   Humans

endowed, as we know (5), with the following forces:

| *motory* | *motory* *sensory* | *motory* *sensory* intellectual | *motory* *sensory* intellectual *spiritual* |

and hence able to undertake, like ourselves, the kinds of work corresponding to these forces. Other creatures may thus be helpful to us, or at least they can encourage us in our work by their example and induce us to lead a more active life which will yield a richer harvest than when we are relying solely on our own forces in our work and when we receive no external encouragement.

The fourth method

(11) The fourth method of undertaking the four above-mentioned kinds of work ensues when we perform work in the manner described for the two previous instances, but in addition enlist the forces of inanimate Nature, for example the force of steam, wind, water, gravity, or cohesion; or crystalline, organic, electrical, magnetic force, etc. Even if we do not use these forces directly, we endeavor to copy the works produced by these forces, in the production of which the following forces of Nature are particularly manifest:

| *attractive* | *cohesive* | *crystalline* | *organic* |

or in other words,

| the drawing force | the joining force | the ordering force | the vitalizing force |

for it is precisely these four forces or powers of Nature which give rise to and sustain in the true sense of the word such things as, for example,

air	ice	snow crystals	insects trapped in amber
water	chalk	crystalline salt	mammoths
the soil	marble		ammonites
clay	flint		
sand	sandstone		

which things are commonly known as:

| the *Elements* | *Stones* | *Crystals* | *Fossils*, |

or the inanimate entities the forms of which are:

| variable | stable | Regular or ordered | derived from biological forms |

or, in yet another classification, they are:

| originals | rocks | orders | images |

which, even if they were all to consist exactly of the same type of material, e.g. silica, in

sand	sandstone	rock crystal	silica fossils
sandy soil	flint	amethyst	

although they are all composed of silica, yet each conveys an entirely different property. The cause of this is the fact that each of the four arise through the action upon matter of the four different forces just enumerated, not acting in isolation but in their consecutive order. Thus the Elements are the product of the *attractive force*; the second entity, named the Stones, is made as a result of the action of the *attractive and cohesive forces*; the third, the Crystals, come about as a result of the action of the *attractive, cohesive, and ordering forces*; and the fourth entity, the Fossils, is produced when matter is acted upon by all four forces, the *attractive, cohesive, ordering, and vitalizing force*.

(12) These four different things, created in the described manner by the four above-described forces, are characterized chiefly by the circumstance that their natural properties, otherwise known as their lesser or internal properties, consist of an increasing number of qualities proper to inanimate objects and things which have devolved to the inanimate condition, and which may be described as follows:

variable	*variable*	*variable*	*variable*
stable	*stable*	*stable*	
ordered	ordered		
vitalized			

(13) As regards their usefulness or purpose, or their external and superior properties, these four entities are distinct from each other most of all in that if we make active use of them, in order to improve or copy them, or to obtain other objects like unto them as regards the qualities of *variability*; *variability and stability*; *variability, stability and regularity*; and *variability, stability, regularity and vitalization*, as for example:

cultivated land	artificial rocks	ornaments and utensils	mummified and embalmed bodies
fertilizers	man-made caves	tools and instruments	pictures and statues
food and drink	ruins and debris	edifices	machines

We shall of necessity be applying our vital forces: our first force in the *first* case; our *first* and *second* forces in the second case; our *first*, *second*, and *third* forces in the third case; and *all four* of our vital forces in the last case, that is:

motory	*motory*	*motory*	*motory*
	sensory	*sensory*	*sensory*
		intellectual	intellectual
			spiritual

or, in other words, the following forces:

fixative	fixative	fixative	fixative
	pleasure-giving	*pleasure-giving*	*pleasure-giving*
		activating	activating
			virtue-endowing

(14) Thus these four things, or more precisely the four forces to which they owe their creation and properties, may serve us as means to develop our own forces, and hence for their engagement in the carrying out of the four above-described kinds of work (3), and for the yielding of more and more abundant and useful fruit for the common good, the further we will manage to develop that application, and the more we will manage to engage other forces appertaining both to living and inanimate Nature, in the process.

Chapter 4:

On the advantages accruing from the undertaking of the four described kinds of work

(15) There are four chief advantages or benefits which may and indeed do proceed for us from the undertaking of the four described kinds of work, in other words from the good use of our corresponding four vital forces, assisted by the forces of the external world, and these advantages are our Property, Ability, Perfection, and Felicity, which are the same as our *external good*, our *first inner good*, our *second inner good*, and *our third* and *supreme good*.

1. Property

(16) *Property*, otherwise known as the *external good* or *asset*, is acquired by each of the four described kinds of work, viz.

Labor	Entertainment	Thinking	Dedication;

but it is not achieved in equal proportions by each of them. For if we apply them in the second manner of work described above in (9) of the preceding chapter, that is not in isolation but in an increasing combination of the four kinds of our vital forces, then the outcome of such work, even though it be concentrated on one object, shall be all the greater, the greater the participation in it of the four vital forces, according to the relation which has already been shown on the particular examples and unquestioned facts cited in the essay *On Man's vital forces and their significance in his productive or creative livelihood*. These facts and particulars show, for instance, that land which is cultivated by the application of the first of the vital forces, that is *Labor, the work which uses only one vital force*, usually produces (in outcome of its natural fertility being thereby enhanced) two seeds of crop for every seed sown; whereas land cultivated also with the application of Entertainment, Reasoning, and Dedication, in other words work engaging two, three, and four vital forces, for every seed sown yields harvests in a ratio of 4, 8, and 16. This shows that each successive vital force applied to the working of the land to enhance its fertility doubles the preceding outcome. Other forces may be utilized in this manner on all the other things of this world, both animate and inanimate, that is on the one hand, on the

originals	rocks	orders	images,

and on the other hand on the

Plants	Primitive Animals	Animals	Humans

outside of which there are no other categories of things in this

earthly world, and which therefore constitute the full scope of our Property. Therefore we have grounds to believe that each of these eight earthly things and creatures may have its value augmented in the following ratio:

$$2 : \quad 4 : \quad 8 : \quad 16;$$

or in some other similar ratio when we act on it by means of work engaging

| one force | two forces | three forces | four forces |

that is,

| Labor | Entertainment | Thinking | Dedication, |

that is, endeavoring to enhance the value of each of the eight earthly things through the following of our forces:

motory	motory	motory	motory
	sensory	sensory	sensory
		intellectual	intellectual
			spiritual

or, in other words, through the following forces:

performing	performing	performing	performing
	activating	activating	activating
		directing	directing
			virtue-endowing

or, to use foreign loan-words, the forces that are

physical	physical	physical	physical
	aesthetic	aesthetic	aesthetic
		intellectual	intellectual
			moral

2.   Ability

(17) The second chief advantage which we draw from work is that through it we acquire the skill to perform work itself more and more easily and with an ever-growing satisfaction, accuracy, and liking for it. In other words, that we can and are able to undertake work at the expense of a lesser and lesser amount of toil and drudgery, but to the ever-increasing gain of ourselves and the common good. Such an aptitude for the undertaking of work with an ever-increasing facility, satisfaction, accuracy and liking, and with the ever greater saving on effort, time, and material, is called *Ability*, and is our *first inner good* (15), which is acquired, just like *Property* or the *external good*, by each of the four described kinds of work, that is by work which engages

| one force | two forces | three forces | four forces |

that is,

| Labor | Entertainment | Thinking | Dedication. |

But *Ability* will not always be of the same import to us. For if it is the result of the first of our kinds of work, it will be called a *Craft*. If we apply the second, third, and fourth kinds of work respectively, the applicable names for the corresponding abilities will be decorative *Art*, precise *Skill*, and exemplary *Conduct*.

(18) The above shows that just as we have four different kinds of work, so too there are four different aptitudes or abilities for their performance, that is the above-mentioned

useful	decorative	precise	exemplary
Craft	Art	Skill	Conduct

These abilities will be all the more important to us the more we endeavor to devote of our forces for their acquisition, or at least the more we strive to make them the outcome of our undertaking of the four kinds of work; not by the *first manner* for the performance of work (8), which in Chapter 3 we found was the least effective of all, or even totally fruitless; but by the *second*, *third*, and *fourth means* (9, 10, and 11), requiring the application not only of the *physical*, *aesthetic*, *intellectual*, and *moral* forces, but also of other, inner or external forces; or by the combination of the latter with the former if they may profitably be used in combination. This applies particularly to the latter, that is the inner forces, which only then attain their true qualities in the sense of the work referred to as

| Labor | Entertainment | Thinking | Dedication, |

and may serve as means for the acquisition of our first internal good, known as *Ability*, and in particular

useful	decorative	precise	exemplary
Craft	Art	Skill	Conduct

when they are combined in the way described above in Chapter 3, with the use of the second manner of undertaking the four kinds of work just mentioned. For it is only when these forces are combined in this manner that we may perform any of our life's jobs, tasks, professions or offices with the application of

toil	toil	toil	toil
	satisfaction	satisfaction	satisfaction
		cogitation	cogitation
			love

In other words, we shall then be applying a greater and greater number of the forces given us by our Maker, by which these forces will be developing more and more within us, which in turn will become the source of an ever-increasing store of *Ability*. The result of this for us will be not only that the profession or occupation to which we devote our entire lives will be not merely mechanical or in other words motory and pertaining to the *crafts*; but it will also be motory and sensory, in other words *artistic*; or it will be motory, sensory, and intellectual, in other works *skilled*; or finally motory, sensory, intellectual and spiritual, for which no other appropriate name exists at present in this or any other language we know. For this profession, which will only emerge from the state of slumber in which now it persists, at some future time, when the spiritual or moral force attains its rightful place, which is now usurped by the intellectual force of reasoning; and

when the moral force becomes as active as the motory, sensory, and reasoning forces, in other words as the mechanical, aesthetic, and intellectual forces. It is by these three last-mentioned forces that in this the third era of man's history[9] that the useful *Crafts*, the decorative *Arts*, and the precise *Skills* flourish and thrive. But when the fourth era comes[10] it will no doubt come to the blossoming of the exemplary *Sciences*, *Conduct*, and *Deeds*. This may first be expected to come about within our virtuous agrarian Slavonic people and its noble and most innocent occupation of agriculture; an occupation which, since it may be conducted not only industriously as a useful *Craft*, not only industriously and with taste as a decorative *Art*, and not only industriously, tastefully, and skillfully as a precise *Skill*, but also honestly as an exemplary *Science* — will in time acquire that ultimate, and most noble quality, thereby providing an example for other occupations to follow. To this Mankind will be drawn not only by the awakening in those people who practice these abilities of a sense of Man's dignity in his position as the most perfect creature and therefore also the creature expected of the best conduct; but Mankind also will become aware of that important circumstance, especially as regards its material condition, that the outcome of human work in any profession or occupation whatsoever is all the grander the more people come to activate their forces for the acquisition thereof, that is the more they perfect their active, productive, and perfecting forces, the aims and purposes of which are to better things, people, and their own persons and hence also their own and other people's aptitude to serve the common good. Thus they will strive to raise themselves from the state of vegetal or mechanical activity to the state of mechanical, sensory, intellectual, and moral activity, which is the truly human business and the emblem of a truly Christian life founded upon the principle of serving with all of one's forces for the common good, which is the work of God, Guardian of this good, Who has made Man in His own image and likeness, wishing to have him as His helper to care for and attend to this business.

3.   Perfection

(19) The third chief advantage, known as *Perfection* (15), ensuing from the undertaking of work, or in other words from the good exercise of our vital forces, somewhat resembles the second advantage, *Ability*, but differs from it in that while Ability is often regarded as an external as it were property of ours (as the names of its four main kinds[11] indicate), thus separate from us and often finding external expression, either through writing or other means; *Perfection* on the other hand, the advantage now under discussion, is always seen as one of our inner properties, a thing strictly connected with us and a direct consequence of *Ability* in the entire range of its senses as evidenced by the Crafts, Arts, Skills, and Sciences, and an indirect outcome of *Work* or the good exercise of our four vital *forces*, known as the physical, aesthetic, intellectual, and moral forces.

(20) Thus Perfection, which is now under discussion, is the fruit of our *work*, in other words the outcome of the above-mentioned *forces* which always constitutes an integral part of our being, considered in that state which it may and should attain as our forces grow, and therefore known as the potential for growth and development. This potential for development, we observe, is the condition for *Perfection*, which is the chief ornament of our being and the principal condition for our *Felicity*, which we shall discuss in due course.

(21) *Perfection* is acquired like *Ability*, through the exercise of each of the four different kinds of *Work*, viz. by work that is

motory	sensory	intellectual	spiritual

or in other words through:

Labor	Entertainment	Thinking	Dedication (3)

But it will not be of equal import to us, for if it is the fruit of the first kind of work, or the effect of the exercise of the first kind of our *Ability*, that is a Craft, it will be known as health, wholesomeness, perseverance in toil, fitness, firmness, soundness, and most aptly Efficiency. But if we may regard it as the outcome of our second, third, or fourth kind of work, or as the effect of our application to the second, third, or fourth kind of *Ability*, that is to the Arts, Skills, or Sciences, that Perfection will be known in the first case as taste, decorum, courtesy, politeness or Affability; in the second case as wisdom, prudence, proficiency, faculty, or Capability; and in the fourth case [*sic*] goodness, mercy, humanity, equity, honesty, or Virtue.

(22) Hence it may be observed that just as there are four chief kinds of our *Work* and four corresponding *Abilities*, so there are derived from them four kinds of our *Perfecting*, the above-cited

Health	*Taste*	*Wisdom*	*Goodness*

or,

Efficiency	Affability	Capability	Virtue or Worthiness

Apart from their absolute value, by which our being is endowed with a similar value, these Perfections also have a relative value, which concerns the objectives of our active, improving and productive life, the chief purpose of which is to *bring people and things away from the harmful or neutral state, and from the state of little utility, to the state of their being most useful and best suited for the common good, which contains our own good too as a constituent part of its entirety*.

(23) The latter of these values, that is the relative value of the four *perfections*, may be learned from the very names used for them. The first of the *Perfections* is called Efficiency, since it makes us fit to carry out or effect the most difficult tasks without strain, fit to accomplish without fatigue even the most arduous tasks or useful work the aim of which is to make people and things useful, or to make people and things serve the common good. The second perfection is called Affability because if we are endowed of it, it will assure us of acceptance and favor by other people and beings capable of feelings, thereby making them assist us in the performance of the above-mentioned useful work more readily than if they were indifferent to us or did not like us. The third of our four perfections, Capability, is so named because it helps to make us more capable to render service to other people for the common good than merely as simple performers of useful work; so that apart from being endowed of the first and second perfection, that is of Efficiency and Affability, we should also be the capable managers of our own and all other

forces by which that work is carried out. Finally the fourth kind of our universal perfections is called Virtue or Worthiness, goodness or humanity, because if we are endowed of it, it makes us behave in such a manner in the referring of our own and general affairs to the common good that we come to be invested with the supreme virtue or worth, at least in the eyes of other people who also act in this way. Virtue or Worthiness is the only good which deserves the name of ''good'' of all people, since to this alone all peoples have given the appellation of ''goodness''. And since it is also called humanity, which term may not be applied to the other three perfections, Efficiency, Affability, and Capability, as these we share with other living creatures, therefore Virtue or Worthiness is a distinctly human quality, one which marks humanity more distinctly among the other creatures than reason or speech, which hitherto have been recognized as the chief factors distinguishing Man from the animals. But the first three Perfections alone, that is Efficiency, Affability, and Capability, can never give Man the right to call himself a humane, good, or virtuous creature, be these three perfections ever so patently manifested in him.

4.   Felicity

(24) *Felicity*, which is the fourth chief advantage, the fourth chief good which ensues from *work*, and by which we mean the ultimate and supreme contentment we receive from what is external and what is in us as the fruit of our industry that is of our good and active will to use the forces and faculties given us by our Maker; that *Felicity*, like the three preceding goods, that is *Property*, *Ability*, and *Perfection*, may be acquired by each of the four above-described kinds of *work*, that is by work which is

> *motory   sensory   intellectual   spiritual*

or in other words through:

> Labor   Entertainment   Thinking   Dedication

but that *Felicity* will not be of equal value, that is it will not mark the same state of our contentment and beatitude, but will depend on the nature of the kinds of *work* or of the corresponding *forces* which we apply to achieve that Felicity.

(25) If *Felicity* is the fruit of the first kind of work, that is *motory* work or Labor; or in other words, if that Felicity is the direct outcome of the *Perfection* acquired by that work which is called health, wholesomeness, fitness, firmness, soundness, or Efficiency, then that felicity will be of that same import to us as it is to every healthy creature which enjoys its health and the attributes governed by health achieved through the undertaking of motory work. Since this kind of work may be undertaken even by the plants, and also by humans who abide in a somnolent state, merely moving about but neither feeling, thinking, nor loving aught, when they merely apply their physical force only without the three superior forces; since therefore, as I say, such work may be undertaken by the plants and humans abiding in a vegetative condition (the proof of which, in the former case, is the already-mentioned perforation of the soil effected by the roots of plants, or the

motion of the blossoming sunflower as all day its head follows the sun, or the rising up off the ground of any young shoots that might perchance have been pushed down or dropped); hence the *Felicity* which is the resultant of this kind of work and which is called Merriment, is such a form of general felicity in mankind and other beings that it may be manifested by creatures not employing the sensory, intellectual, and spiritual forces, that is the aesthetic, rational, and moral forces, but merely exercising their motory or mechanical force.

(26) Since even the plants are also endowed of this force (the proof of which is supplied, apart from the three above-quoted instances, by 1. the movement of their roots during sprouting downwards into and towards the fertile soil, and of their stems upwards and towards the sunlight; 2. the movement of their young stems, shoots and leaves upwards towards the sunlight and away from the darkness; 3. the opening up of their blossoms in the daytime and during fine weather, and their closing up for the night and during inclement weather; 4. and finally their surfacing in the summer (for the water-growing species) and submerging for the winter after the flowering season is over etc.); since, as I say, even the vegetable creatures are endowed of this force, therefore even they can manifest the felicity which comes of the happy condition of this force, and after the manner of a healthy, sleeping infant, they too may be said to be merry.

(27) This, too, is commonly observed in them when they are in their healthy condition; and furthermore it is attested by the nature of our speech, which allows us to call trees and other plants which enjoy the described *Perfection* or Health merry creatures, thereby granting us the right to acknowledge their potential for the exercise of the first kind of general *Felicity*, which we have called Merriment and which is regarded both in these and in all other creatures and in Man himself as the indirect outcome of the undertaking of the *work* known as Labor, exercised by the *Ability* known as Craft, and therefore the direct consequence of the *Perfection* thereby accruing and named Health or Efficiency.

(28) If our lowest *Felicity*, which has just been described, accordingly may be said to be a felicity bereft of feeling and thus a vegetal kind of felicity, as the ultimate outcome of the application of our lowest force known as the motory force or Power which we share with the plants, and the direct outcome of the physical work known as *Labor*, or of the livelihood pursued through Labor, that is Craft, and of the resulting *Perfection*, which we have called health, wholesomeness, fitness, firmness, soundness, or most aptly *Efficiency*; then presumably it must be no otherwise with our second *Felicity*, the sensory Felicity which we share with the primitive animals, and which by analogy with the name for the first felicity, we may call sensual felicity or Delight.

(29) Indeed, this second felicity may not be assigned any another source but our second vital force, the sensory force or Sensibility, which we share with all the animals capable of feeling, and hence even with the least of them, that is with the Primitive Animals. Since the sensory force is the essence of the *Work* known as Entertainment, which is also performed by the Primitive Animals, as evidenced by the animated movements these creatures carry out reminiscent of our dances, such as the somersaults in the air done by insects

like Hilara and Chironomus, or the leaps over water by the insect Gyrinus; or by the pleasant sounds similar to our music uttered of a summer's evening by the crickets, Cicada orni and Acridium viridissimum; therefore Entertainment should be regarded as the source of the *felicity* described, and known as Delight. And since of this all are thoroughly convinced, it wants no further proof. Furthermore, since Delight is closely connected with the fair Arts, as confirmed by the just quoted dancing and music, which at the appropriate level of perfection are accounted even as belonging to the sublime arts; therefore the similar yet even more perfected or profuse Arts show they are derived from Entertainment as the source of their *Felicity*.

(30) *And since* in addition from Entertainment and the Arts a third good issues, namely the sensory Perfection, or taste, decorum, courtesy, politeness or Affability, which wins for us the acceptance and favor of other people and other creatures capable of feeling, therefore also this third good may and should be acknowledged as the source of the *Felicity* under discussion, that is Delight. Of this all are thoroughly convinced, for whosoever possesses this Perfection, or has ever in any way contributed to the pleasure of other people or creatures endowed with feeling knows how much innocent pleasure or Delight he himself has thereby received. According to what has been said of it here, Delight is a likewise a fruit or outcome of the good exercise of our second vital force which we share with all creatures endowed of feeling; just as Merriment is the fruit or outcome of the exercise of our first vital force, which we share in common with all living creatures. If this is how things stand with the first two kinds of our general *Felicity*, that is Merriment and Delight, then it can be no otherwise with the two superior Felicities, that is intellectual felicity and spiritual felicity, or respectively Consolation and Joy, which we shall now briefly discuss.

(31) The first of these latter felicities, and successively the third felicity, is called intellectual felicity because by analogy with our first and second felicities, Merriment and Delight, which correspond to their respective vital forces and which we share with all living and feeling creatures respectively (that is the plants and the primitive animals) — it is an outcome of our third vital force, the intellectual or rational force, which we share with all the creatures capable of thinking, that is with the Animals. And since this third vital force, otherwise known as capacity or docility, is the essence of the corresponding *work*, that is Thinking, which is practiced by the Animals as well, especially when they lay snares for animals that are more powerful than they; therefore this kind of *work* too may be the source of a *Felicity*, the one now being discussed and called Consolation. For we are made happy or pleased by our rational force when, having developed that force appropriately for the undertaking of the given work, we make plans to benefit or hurt other creatures, and discover the appropriate means to carry out those designs.

(32) This also applies to the Animals, since (as may be observed in trained dogs) they too experience the highest *Felicity*, superior to the ordinary felicity of the senses, that is superior to Delight, when they can do something witty for themselves or for us. And the fact that, along with human beings, animals too achieve such things, is the result of their *ability* or Skill.

This Ability is demonstrated, for instance, by skilled hunters and the dogs and hawks trained by them for the hunt. Therefore this kind of *Ability* may be and indeed is for us humans and for the animals as well a source of the *Felicity* known as Consolation.

(33) As the described *Ability* or Skill denotes the same as the *felicity* of the intellect, generally known as proficiency, wisdom, learning, or most aptly Capability, which makes ourselves and the Animals capable of carrying out more than just the simple, repetitive and mechanical tasks (viz. such that may be carried out by a machine), therefore this *Perfection*, attained by the exercise of the kind of *Ability* corresponding to what we call the exact Skills, should be regarded as the means — more direct than the *Work* known as Thinking which marks this category of *Ability* — to achieve that general Felicity of the intellect known as Consolation.

(34) This Consolation, we observe, is a far higher kind of our general beatitude than Delight, the felicity of the senses; or than Merriment, the felicity of movement, which we may exhibit even when asleep, through a pleasant disposition of the features of our countenance, and which may be considered desirable merely because it is a sign of health and because through it we may stimulate the higher felicities in other people, such as the felicity of the senses, which may be evoked even by the merry plants that enjoy this Merriment thanks to their health, although they presumably do not feel this, yet manifest it to other creatures that are capable of feeling.

(35) Thus if our first three vital *forces*, that is Physical Motion, Sensibility, and Intellectuality, which are the essences of the first three categories of *work*, that is Labor, Entertainment, and Thinking, and also of the first three kinds of *Ability*, that is the useful Crafts, the decorative Arts, and the exact Skills, are the sources of our first three *Perfections*, known as Efficiency, Affability, and Capability, and hence also of the directly resulting first three *Felicities*, that is carefree Merriment, innocent Delight, and lofty Consolation; and if, now, the fourth of these forces which vitalize us, Spirituality or the moral force, is the agent within us of the *work* known as Dedication, thereby evoking that general *Ability* of ours known as the exemplary Sciences, and the fourth kind of our *perfection*, which we have denoted by the terms goodness, humanity, or Virtue; it may not be doubted that this force is correspondingly the fount of a separate Felicity.
(to be continued)

PART IV (NO. 32, PP. 253–5.)

(36) In our own tongue, Polish, apart from the terms Merriment, Delight, and Consolation, we have another homonymous word, that is Joy [Radoos´c´ — Lat. gaudium, translator's note], which nevertheless denotes something superior and occurs especially in our ancient books, such as the translations of the Bible. This is the word most frequently used to mean the moral or spiritual felicity[12] — that felicity which is accessible only to the moral creatures, those able best of all to develop their moral force. Only humans, or the humane, most truly human and humanitarian among the humans, may and indeed do experience this, through the accomplishing of the most perfect and sweetest deeds pertaining to the supreme Perfection.

(37) If now we consider those deeds by which we perfect ourselves, other people and creatures, and the things pertaining to this world which are in an inferior condition due to human error or negligence or through misadventure, as the highest and sweetest deeds of the supreme Perfection — then that *perfecting of people and things*, that is bringing them from a harmful or indifferent condition, or from a condition of little utility, to the state in which they are most useful and best suited to serve the common good; if, thus, we bring them to that condition in which they carry the traits of the supreme Felicity and its most perfect creature, which we consider (and should out of our religious duty consider) to be ourselves — *we will regard this as the loftiest kind of all our deeds*, the supreme mark of our moral *perfection*, known as humanity, goodness or Virtue, which is the condition for the corresponding *Felicity*, which as has already been said, is called Joy.

(38) That Joy — Man's true *Felicity*, which indeed no other being can experience save Man leading a human, humane and humanitarian life, at least at the present stage of Man's development — that Joy is the fruit of its corresponding *Perfection*, the reward of perfect deeds, that is such that are intended to bring man and things to perfection, of this none can be convinced by reasoning, that is through the agency of the third vital force, the force of the intellect or of reason, since the latter force is restricted in its powers of judgement only to matters concerning itself or lesser matters. As regards deeds effected by the force superior to reasoning and by its attendant consequences, such matters may be judged only by that self-same higher force. Hence true judgements in such matters may not be expected of Man the Reasoner but of the Dedicated or Devoted Man, that is of him who accomplishes the said perfect deeds, that is such deeds which are intended to bring men and things to perfection.

(39) If Man the Thinker, that is those who manifest the working of the third of the vital forces as the supreme emblem of their lives, were to claim the right to judge those who work by Dedication or Devotion, they would thereby invest all those inferior to themselves, that is those whose highest activity is feeling, in other words those who are manifest only through the second vital power, the right in turn to judge the thinkers. But the thinkers would not allow this, and indeed could never allow it, as superior things cannot be known by things that are inferior. Likewise, without insult to common sense, by which no doubt they are ruled, the thinkers should not pretend to make themselves, and the generality of Mankind abiding in the third vital force, true and equitable judges of those dwelling within the realm of the fourth vital power and displaying all the vital forces with which Man has been endowed by his Maker.

(40) Now if Man the Thinker, that is those who live by the exercise of the third vital force, were to endeavor truly and equitably to judge the deeds of those who work by Dedication, that is of those whose aim is the perfection of things, other people, and themselves, thereby making them more apt to serve the common good; if, as I say, the former wished to judge the latter, and to enjoy the Felicity enjoyed by the latter, that is Joy; they would only be able to do this if they themselves endeavored to discover and practice this work, or at least instead of leading their hitherto *partial* life, consisting chiefly of thinking, in other words of carrying out intellectual work, if they endeavored to emulate at least for a brief time the *full* life of Man, consisting in movement, feeling, thinking, and dedication, that is in the performance of all four kinds of work which humans may pursue, motory work, sensory work, intellectual work, and spiritual work, in other words Labor, Entertainment, Thinking, and Dedication. They would be qualified to judge in this respect if they themselves enjoyed all the fruits which may be attained through all four kinds of work, that is through the good exercise of all four corresponding forces, assisted by the external forces of nature, and making all four forces their undisputed assets.

(41) Of the fruits obtained from Work and otherwise known as the benefits or goods (15), the first is called Property or the external good (16), which accrues (as we Poles have observed) in geometrical proportion, that is in relation to the square of the value for the previous vital force, viz. a 2 : 4 : 8 : 16 ratio. This good is the least bound to our nature, since it may be easily lost and easily regained. Nevertheless this benefit called Property is an indispensable condition not only for our human happiness, but even of human existence. For no person is able to enjoy a happy life, nor even to live without the external fruits of his own or others' work.

(42) As regards the second fruit of our four kinds of Work, which in (11) [! — cf. 18] we gave the general name of Ability, and the particular appellations of the useful *Crafts*, the decorative *Arts*, the precise *Skills*, and exemplary *Sciences*; as regards this second fruit, which we have called the first inner good, we know it is much more closely connected with our nature that the former good. For those people who have no permanent property at all must have at least a relative mastery of one of the above-described abilities, that is of the useful Crafts, the decorative Arts, the precise Skills, or of the exemplary Sciences. And whatever profession they have chosen, if they practice it diligently, or both diligently and with taste; or diligently, with taste, and rationally; or diligently, with taste, rationally and honestly — they shall in time acquire such proficiency in their profession that they will soon develop such a liking for their occupation as to consider themselves proponents of one of the four categories of general human Ability. We have already explained this for the case of agriculture, which, depending on the four different means of application of our four vital forces, may assume the status of a useful Craft as practiced by all peoples whatsoever, a decorative Art as delineated by Delille in his French Georgics, a precise Skill in the understanding of Thaer and his disciples, or finally as an exemplary Science as presented by Krasicki and Koźmian in their works, Pan Podstoli [The Steward] and Ziemiaństwo Polskie [The Polish Landed Gentlemen] respectively.

(43) As regards the third chief benefit which is the outcome of work, that is Perfection (19), which just like work is of a fourfold nature and consists of *Health*, *Taste*, *Wisdom*, and *Goodness*; or *Efficiency*, *Affability*, *Capability*, and *Virtue* (22); as regards this, the second inner benefit, we have no hesitation whatsoever that it is essentially and integrally connected with us, for not only can it not be removed from us once we have acquired it, but without it we cannot exist at all, or at least

without it we cannot make any significant contribution to human society nor enjoy a position of respect within it.

(44) Since this significant contribution to human society and the related social position are clearly the results of the above-described fourfold Perfection known as *Efficiency*, *Affability*, *Capability*, and *Virtue*, without which we cannot effect any good for the well-being of all or win respect, especially of those people who are also in the service of that universal well-being and who therefore have the right to expect us to help them and work with them to achieve that aim; therefore it is clear that albeit this Perfection, which we have called the second inner good (15), may seem to be our very own and exclusive good, not shared with any other persons, yet in fact it also belongs to other people, and even partially to our fellow-creatures, too. For they all have a certain right to demand this Perfection of us, in return for their own Perfection and the favors granted us thereby.

(45) The fourth benefit which we draw indirectly from the forces given us by our Maker and from the work they accomplish, and hence directly from the Property, Abilities, and Perfections (15) thereby acquired; particularly this fourth and last benefit, known as Felicity (24), the third inner good, is the most closely related with our nature, apparently pertaining to ourselves alone. For only we ourselves, who have acquired this benefit thanks to our work and perfection, appear to be enjoying our own Felicity, while other creatures seem incapable of sharing in our Felicity.

But this is true only of creatures which have no morality, in other words those which have not been endowed with the fourth vital force, or in which this force lies dormant. For as regards all those who partake of this fourth force, manifesting it actively (and this includes above all the humane and humanitarian, truly human people, that is such as, in outcome of an accomplished state of perfection or good education, have activated all their vital forces, including the moral, truly human force, which according to the religious and general conviction they share only with the Heavenly Beings); as regards such moral creatures, who are the most perfect and closest to the Supreme Perfection Itself by virtue of their own perfection, they therefore desire our attaining unto felicity, since without it they can never fully achieve their own perfect felicity, and thus they clearly share in our own state of felicity.

(46) This state of affairs should provide us with the best possible incentive encouraging us, first out of regard for ourselves, secondly out of regard for those creatures, and thirdly out of regard for the Supreme Perfection or Divine Goodness Itself, which desires our Felicity and the Felicity of all of Its creatures according to their faculties for the enjoyment of felicity — to strive to attain to our own felicity, the complete felicity for the achievement of which we have been endowed by our Maker of our four vital forces, and which we may accomplish through the performance of the respective kinds of work and hence also through the exercise of the thus accruing Perfections.

(47) Since we possess four each of the Forces, kinds of Work, and Perfections, which are the means to our achievement of felicity (and would have less if we were not humans, but animals, primitive animals, or plants; or if by keeping dormant our forces proper to the superior creatures we dwelled in the condition of the lesser creatures); hence also our felicity must be fourfold in nature, that is

| *physical felicity* | *aesthetic felicity* | *intellectual felicity* | *moral felicity* |

or in other words,

| *motory felicity* | *sensory felicity* | *rational felicity* | *spiritual felicity* |

The Felicity which we enjoy in common with the four categories of our most perfect fellow-creatures, that is the

| *Plants* | *Primitive Animals* | *Animals* | *Humans* |

may be named, in order of increasing power to bring an increasing degree of happiness, to the point of ultimate beatitude, for ourselves and other living creatures:

| carefree *Merriment* | innocent *Delight* | lofty *Consolation* | heavenly *Joy* |

Above this Felicity none other can be experienced or desired, save perhaps only the assurance that , along with all our neighbors we shall be able to enjoy for all eternity the ultimate of the four felicities, which may be secured by our and their continual dedication to the common and eternal good. For only through dedication, not by fruitless thinking nor ratiocination and investigation, can we and they be convinced that we hold sufficient power to achieve this, and therefore that we are eternal creatures, and that we are capable of enjoying the Eternal Felicity, which we have just defined as Heavenly Joy.

Notes

1 The friends of work will not hold it against the author, he trusts, that he is here repeating some of his main tenets already published elsewhere.

2 As evidenced by the histories of Herostratus and Cleopatra.

3 See our treatise on the *vital forces* and their importance in Man's productive life.

4 To this property of advancement in the human vital forces through their working on the forces of other creatures and things, which resembles the properties of the magnet, we particularly draw the attention not only of the philosophers and the economists, but also of the natural scientists.

5 So that they should not give rise to untidiness or of suffering for people or other creatures of God.

6 A proof of this is supplied in this age by the disposition developed chiefly in aged persons, to undertake even small tasks of dedication for the common good, such as the removal of stones from the road into a stream, that is the elimination of harmful conditions in favor of generally useful conditions.

7 The taking and eating of useful things, and the similar performance of all mechanical activities by means of the motory force alone is manifested not only by all *humans*, but also all the living creatures not excepting the *plants*, as evidenced for example by the flycatcher (Dionaea muscipula) and all the varieties of the sundew (Drosera).

8 Without the exertion of the motory force it is impossible even to speak or write.

9 That is, the Industrial Age, which has followed on the Barbarian Age, which in turn came after the Savage Age.

10 Which shall be the Age of Accomplished Christianity or the Age of Universal Love.

11 That is the Crafts, Arts, Skills, and Sciences.

12 "The angels shall rejoice at the conversion of one sinner".

Person-centered Ergonomics

D.J. Oborne

Department of Psychology, University of Wales Swansea, Singleton Park, Swansea SA2 8PP, UK

1. INTRODUCTION

To emphasize the centrality of the person within the ambit of ergonomics would suggest a tautology. Ergonomics is, and has been since its inception, centered about the person within the system. Stress has always been placed on the need for the environment to be designed to fit the person rather than adopting the Procrustean view of making people fit the environment.

Even a cursory overview of current ergonomic practices, however, demonstrates that much of it is not person-centered in its approach. A systems view, for example, emphasizes the importance of "fitting" environmental outputs to the operator's input capacities, and vice versa. In doing so, it places equal emphasis on the "needs" of people and their environment; people are not central to the approach. Similarly, a cybernetic view perceives people merely as error-inducing and error-correcting mechanisms; not as being central to the effectiveness of the operation as a whole.

Like many of the previous views of ergonomics, the person-centred view argues that the individual and the working system must operate in close harmony for effective operation of the system. It diverges from previous views, however, by arguing that it is the operator who controls the system, who operates it, steers its course and monitors its outcomes. It is individual characteristics such as these that need to be considered in more detail and which determine the system's eventual potential.

Branton (1983) emphasized that human operators within a system inevitably turn it from being a closed loop system in which information flows from one component to the other with (theoretically) maximum efficiency for correcting deviations, to an open-loop system. Instead of a "designed" error correcting system; therefore, the deviation corrections are effected by the operator on the basis of his/her "mental model" of the system and its operation. In this way, rather than emphasizing a mechanistic approach to improving information flow between components within the system, person-centered ergonomics takes as its central point the need to accommodate the human attributes that the person brings to the system.

Thus, the person-centered perception is one in which the operator and his/her abilities define and control the working system. In the course of these interactions, individuals bring to the system a collection of inherent strengths and weaknesses (from such characteristics as experiences, expectations, motivations, etc.) which themselves will work with the system to change it. Such changes are usually to the good, from the viewpoint of such criteria as efficiency and safety. Sometimes, though, the features that the individual brings to the system will include variability, fallibility and maybe even perversity — any of which are more likely to lead to errors and inefficiency. As Wisner (1989) argues, operators will use their "own story" when being asked to control and manipulate events.

2. PERSON-CENTERED COMPONENTS

Arguing that ergonomics needs to take a more central view of the wishes, abilities and needs of the operator begs a major question regarding the nature of the individual's characteristics that need to be considered and which are so important. person-centered ergonomics stresses individual characteristics that help to direct the person's behavior within the system and develop his/her shared identity with the system. Thus, characteristics such as having a sense of purpose, motivation, and awareness are considered, as, too, are more nebulous concepts such as the individual's sense of self and consciousness of his/her actions.

3. PURPOSIVITY

Purpose provides meaning to an action; without a sense of purpose the action will have no meaning as far as the system is concerned. Actions without purpose and meaning become random and lack direction.

The person-centered approach argues that one of the primary features brought to the system by the individual is a sense of purpose, of meaning and action. People are not just "doers"; they are also thinkers. They generate mental models about the world and impose into these structures the potential consequences of their actions. By understanding the individual's reasons for interacting with the system ergonomists will be better placed to design systems that facilitate the actions and are fit for the purpose.

4. ANTICIPATION AND PREDICTION

The existence of purposivity implies some concept of prediction and anticipation. By implication, therefore, the concept of purposivity also suggests that the user will have some notion of the desired outcome and thus some "internal" model of the course of future events. This is an important linkage as far as person-centered ergonomics is concerned. Prediction, and the anticipation that arises from it, concerns the operation of the system within the individual's perceived control and will have major implications for the nature and the quality of the information required for the task. Thus, such features are not merely precursors to efficient operation, they also *determine* the efficiency of the operation.

5. INTEREST, BOREDOM AND ADAPTATION

Allied to a consideration of the purposeful nature of the operator are questions about how far the operator has an interest in the operation under control. Clearly, interest is related to purpose insofar as reduced purposivity can lead to reduced interest; reduced interest can lead to reduced awareness of the environment.

Further analysis of the construct, however, leads to a consideration that it is not so much interest *per se* but the reason for the interest in the job being undertaken that are important. In this respect, interest can be expressed in terms of stimulus-seeking behavior; the nervous system adapts and habituates quite rapidly to any prolonged and unvaried stimulation, leading to cortical inhibition of stimuli and extinction of responses. So boredom, a result of reduced interest, can soon lead to a reduced ability to perceive and respond to important stimuli. Adaptation to the stimuli can have similar effects.

In addition to adaptation, a second effect of reduced interest and stimulation from the external environment is often expressed in terms of a lack of attention to the task and to mind wandering and daydreaming. Such behaviors could be taken to be a means by which people attempt to inject some (internal) stimulation of their own.

6. UNCERTAINTY REDUCTION

A further aspect of the person-centered equation that must be considered, and one that may even be thought of as having an overarching effect, is the reduction of uncertainty. From a psychological viewpoint uncertainty reduction plays a major role in everyday life; we attempt to reduce it if for no other reason than to reduce the stressful effects of unpredictability. Of course, like stimulation and arousal, individual attempts to reduce uncertainty are likely to be tempered by some concept of an optimum level. Indeed, in many respects life would be very boring if all events were perfectly predictable.

The concept of uncertainty is thus important when considering the nature of stressful events at work. Stress is likely to result in a person who is uncertain about what has happened, particularly during a period of "mind wandering." This is particularly the case if one also has a position of responsibility when carrying out the task (a concept considered below).

7. CONTROL AND AUTONOMY

Since increased control over a situation improves the likelihood that the events will occur in the desired and expected manner, it is clear that uncertainty reduction is also related to control. The evidence for "mini-panics" resulting from "mind wanderings" demonstrates that uncertainty occurs when control is relaxed. Or, put in another way: "Only what I do is certain for me."

Discussing the design of computer screens, for example, Branton and Shipley (1986) highlighted the importance of control as an important factor in stress causation and management. They proposed that such display technology has propagated a new "breed" of individuals — "Houston Man" — in which mismatches may occur between the interpretation of events (on the part of the operator) and the intention (on the part of the system). This mismatch creates feelings of uncertainty and mistrust in the system — which is exacerbated by the remoteness of the operator from the action.

The impact of uncertainty on working practices is likely to increase the extent of information-seeking behavior, and also the application of imagination to the job in the absence of information. As a result of a poor quality of external information; therefore, the operator may resort to having to retrieve material "internally."

The ergonomics and organizational psychology literature is full of examples of the need to provide autonomy in working situations. Its absence has been shown to be a major factor in job satisfaction, employee turnover, stress at work, etc. A person-centered view of the situation clarifies the reasons for its importance: increased autonomy and personal control implies increased certainty within the working situation; we are more in charge of the consequences of our actions. This leads both to lower stress levels and to less likelihood of mind wanderings.

8. RESPONSIBILITY

Any discussion of autonomy is incomplete without a consideration of the concept of responsibility. Autonomy and responsibility are clearly interrelated concepts; people who have responsibility ideally have conscious control over their own actions. Thus, an operator who has a sense of responsibility is one who is also self-determined. People at work generally act out their duties in sane and rational ways; in that respect they behave in a reliable way.

Having a sense of responsibility, however, can also have negative effects — particularly in the likelihood of increased stress resulting from it. This becomes particularly important in situations when mind wanderings may occur — the resultant mini panics can be considered to be a direct result of the responsibility (and fear of loss of responsibility) that the operator feels over the situation.

9. THE SELF AT THE CENTER

The six areas of person-centered ergonomics elucidated above illustrate some of the fundamental processes that a person can bring to a situation and which helps to shape behavior when interacting with it. People do not simply interact passively with the environment, merely responding to ergonomically designed displays, using ergonomically designed controls, within an ergonomically designed environment. They are active and creative doers. They develop mental models of the system: how it operates, how it should be operated, how to interact with it, how they should interact with it.

A further feature of person-centered ergonomics is to impress on the six features a meta-view of the roles of various "unconscious" processes in determining and developing these mental models. To disregard such processes is to dismiss crucial features of the operator and the ways in which s/he interacts with the system. As Branton (1987) argues, large sectors of real life will be ignored, while the problems of explaining such basic functions as selective attention of directed awareness still remain with us. 'The largest part of our knowledge is still "tacit knowledge' (Polyani 1958), a concept which has an important role in the general case for the purposive view" (p. 6).

Knowledge could be said to be derived from two sources:
- External knowledge appears from sources that are external to the individual, from others, from the system with which s/he is interacting, from past experience, etc.
- Internal knowledge derives from the "me" part of the self. This includes both conscious and unconscious information about our physiological states, knowledge of our actions, purposes, etc.

As individuals we must be able to differentiate between these two, often competing, sets of information. Unconscious (unselfconscious) behavior helps to shape our conscious behavior through the development of preconceptual thought. It thus helps to mould how we approach an activity and how we eventually carry it through:

To argue for the importance of unconscious and "preconceptual" behavior to eventual conscious activity, however, we must also consider the bridge between this kind of thinking and eventual "doing." This bridge is created by the individual's propositions about the information being assimilated and

manipulated. Such quasi-propositions contain immediate and certain knowledge. One knows the position of one's foot without having to say to oneself "There is my foot," or even without having to think about the foot at all. Indeed, perhaps more importantly, the way in which such material is dealt with forms the structure of the bridge — the individual's own internal model of the system's structure.

By helping to build the bridge between thought and action, therefore, from the ergonomist's perspective such quasi-propositions play an important role in determining how the individual will act within his/her environment.

10. A PERSON-CENTERED APPROACH TO MEASURING BEHAVIOR

The person-centered view of ergonomics extends beyond simply arguing that the operator's individuality is placed centrally within the system. Such an argument also begs the need to study the science of ergonomics and gather ergonomic data within a person-centered framework.

10.1. Empathy

An empathatic approach lies at the center of the person-centered approach to measuring behavior. As Branton (1984) has argued:

> To find ways out of this conceptual maze, one guiding thread may be offered to the researcher. It is to adopt an "epistemic" strategy: to ask himself in the first place at each stage, "How do I know whether this or that statement by the operator is true?" "What is the source of my knowledge?" "How direct or indirect is my perception of the measurement?" "How far is my conjecture based on an analogy and how far does it penetrate to the 'real thing?'" Having thus become conscious of their own inevitable bias, observers (and their readers) are better able to speculate on their subjects' knowledge, values and actual purposes. (p. 505)

Thus, unless the ergonomist can empathize with the individual's situation s/he cannot fully enquire into, or understand, its effects on the individual. It is rather like trying to imagine oneself in the other person's shoes. Empathy thus has important implications for decisions to be made about the ways in which studies are carried out, which variables are to be considered, and how interpretations are to be made of the data. Without such insight ergonomists would simply be attempting to understand the external forces that make the individual act in a particular way — rather than the internal knowledge that the person brings to the situation.

10.2. Multiple Methods

Since the bases on which the behavior is to be analyzed are multifactorial, using both intrinsic and extrinsic knowledge, the person-centered approach argues that data collection is similarly multifaceted. All kinds of approach need to be employed, including cognitive, physiological and psychophysiological, and social, probing into both conscious and unconscious behaviors of the operator. Often, no single measure will provide the insight needed to understand both the user needs and those of the system.

11. SUMMARY

The person-centered view of ergonomics exhibits significant departures from the traditional view of a system. The traditional view sees the system as a closed-loop process in which the two components — "man" and "machine" — operate as equal partners. Rather, it is argued that it is the human actor within the system that plays the crucial and dominant role in the safe and efficient operation of that system. Thus, it is the intrinsic features of "being human" that are the most important.

In this conceptualization people have a sense of purpose in their work; they approach the task with this purpose "in mind." So the system and the task must be designed to accommodate this need. Indeed, the individual's purpose extends beyond just wishing to get the job done. Individuals exhibit activity over and above that required to accomplish the task successfully. This sense of purpose often also involves anticipation and prediction (sometimes even at a quasi-mathematical level) of the consequences, the effects of the action.

It is in the nature of people that they will approach their tasks with a sense of responsibility. This means performing their tasks in the most efficient way and having some kind of control over their actions and that of the system. When these requirements are contravened, when instances occur over which individuals lack control or sense that they have failed, even momentarily, in their responsibility, stress and other negative effects typically occur.

The strength of the person-centered view, therefore, lies in the need to consider not only just the person's actions at work, but also the purposes that underlie the actions. Thus, the person is central to the ergonomic system and brings to the situation needs, wishes and desires, as well as the physical and cognitive abilities that are necessary to carry out the task — along with some limitations that need further understanding.

REFERENCES

BRANTON, P., 1983, Process control operators as responsible persons. Symposium paper on Human Reliability in the Process Control Center, Institution of Chemical Engineers, Manchester.

BRANTON, P., 1984, Transport operators as responsible persons in stressful situations. In H. Wegmann (ed.), *Breakdown in Human Adaptation to Stress: Towards a Disciplinary Approach*, vol. 1 (Boston: Martinus Nijoff).

BRANTON, P., 1987, In praise of ergonomics — a personal perspective. *International Reviews of Ergonomics*, 1, 1–20.

BRANTON, P. and SHIPLEY, P., 1986, VDU stress: is Houston man addicted, bored or mystic? In B Knave and P Wildbaeck (eds), *Proceedings of International Scientific Conference on Work with Display Units* (Stockholm).

OBORNE, D.J., BRANTON, R., LEAL, F., SHIPLEY, P. and STEWART, T., 1993, *person-centered Ergonomics: A Brantonian View of Human Factors* (London: Taylor & Francis).

POLYANI, M., 1958, *Personal Knowledge* (Chicago: Chicago University Press).

WISNER, A., 1989, Fatigue and human reliability revisited in the light of ergonomics and work psychopathology. *Ergonomics*, 32, 891–8.

Poland: Polish Ergonomics Society

L. M. Pacholski and J. S. Marcinkowski

The biggest professional ergonomic organization in Poland is the Polish Ergonomics Society (PES), established in 1977. PES provides educational services in the field of ergonomics, and helps in development of the ergonomic culture in Polish society. PES is a member of the International Ergonomics Association (IEA), and since 1980, the president or representative of PES has been a member of the IEA Council.

The society has 13 local branches in the main cities of Poland, including Warsaw, Poznan, Kraków, Wroclaw, Szczecin, Gdansk, Lodz, Lublin, Zielona Góra, Bydgoszcz-Torun, Katowice, Czestochowa and Kielce. The former Presidents of PES are: Professor Zbigniew Jethon (1977–80), Dr Halina Cwirko (1980–86), Professor Leszek Pacholski (1986–92), Dr Halina Cwirko (1992–98). Since 1999, the President has been Dr Jerzy S. Marcinkowski.

PES has organized the 7th IEA Congress in Warsaw, September 1979, with more than 250 participants from many countries. An important outcome of this meeting was participation in the IEA Congresses for the first time by many people from Eastern and Middle Europe. During the 7th IEA Congress, Professor Jan Rosner was elected the President of IEA and served in that capacity from 1979 until 1982.

The ergonomists associated in PES are actively participating in the IEA Congresses of IEA. During the 13th Congress IEA held in Tampere, Finland, ~40 members of PES presented papers and posters in many of the technical sessions. The society has also coorganized the International Ergonomics Association Conference on Ergonomics of Materials Handling and Information Processing at Work, held in Warsaw, June 1993.

The society cooperates in organizing many ergonomics conferences and international and national seminars for both scientists and practitioners. For the past 50 years, PES has been organizing an annual international seminar of ergonomics teachers concerning the problems of training and education in ergonomics, labor protection and work safety. These seminars are held in beautiful places of the Wielkopolska region of Poland in the last week of June. Professor Leszek Pacholski and Dr Jerzy S. Marcinkowski have headed the Organizational Committee.

The national ergonomic conference of PSE has been held for the past 26 years every October in Kapacz (near Wroclaw). Dr Stefan Frejtak organizes these conferences. Another annual meetings of PES concerning the problems of introducing ergonomic solutions into practice, and accompanied by exhibitions of technical solutions, have been held near Zielona Góra for the past 14 years. Professor Witold Rybarczyk heads the team of organizers for these events. In addition, an annual meeting of PES concerning the problems of ergonomics for people with disabilities has been held in Lodz for the past 5 years. Professor Jerzy Lewandowski heads the team of organizers for these events.

Since 1978, the Polish Ergonomics Society in cooperation with the Committee of Ergonomics of the Polish Academy of Science has published an ergonomics journal (*Ergonomia*; in Polish). In addition, since 1991 the PES in cooperation with the *Centre of Application of Ergonomics* at Zielona Góra has edited a journal on the applications of ergonomics (*Zastosowania Ergonomii*).

Professional Certification in Ergonomics in the USA

Board of Certification in Professional Ergonomics (BCPE)
PO Box 2811, Bellingham, WA 98227-2811, USA

1. INTRODUCTION

As a professional, an ergonomist uses the skills and knowledge from the various human sciences and engineering sciences. The ergonomist matches jobs/actions, systems/products, and environments to the capabilities and limitations of people. According to the Human Factors and Ergonomics Society, '"ergonomics professionals apply human–system interface technology to the design, analysis, test and evaluation, standardization, and control of systems for such purposes as improving human and system performance, health, safety, comfort, and quality of life" (HFES 1998).

Ergonomics is a body of knowledge about human abilities, human limitations and other human characteristics that are relevant to design (Chapanis 1988). *Ergonomic Design or Engineering* is the application of human factors information to the design of tools, machines, systems, tasks, jobs, and environments for productive, safe, comfortable and effective human functioning (Chapanis 1988). Ergonomists apply their skills in business, industry, government, and academia to:

- increase human productivity, comfort, health, and safety; and
- to reduce injury, illness, and the likelihood of errors.

1.1. Operating Philosophy and Code of Conduct

The operating philosophy of ergonomists is to adapt technology to the scientifically established characteristics of people. Applications of technology should not be based on unwarranted assumptions about human capabilities and their adaptability. In support of this philosophy, ergonomists shall:

- hold paramount the safety, health and welfare of the public in the performance of their professional duties;
- perform services only in the areas of their competence;
- shall issue public statements only in an objective and truthful manner;
- shall act in professional matters for each employer or client as faithful agents or trustees, and shall avoid conflicts of interest;
- shall build their professional reputation on the merit of their services and shall not compete unfairly with others;
- shall act in such a manner as to uphold and enhance the honor, integrity, and dignity of the profession;
- shall continue their professional development throughout their careers and shall provide opportunities for the career development of those ergonomists and support staff under their supervision.

1.2. Body of Knowledge

The body of ergonomics knowledge resides in the open literature of textbooks, handbooks, guidelines, standards, journals and electronic databases and from the experience of practitioners. A practitioner acquires ergonomic knowledge through formal course work (lectures, labs, workshops, seminars); self-study; supervised fieldwork and practical experience. Ergonomists integrate knowledge about human function, structure and behavior for practical uses in the design process. Formal courses and degree programs in ergonomics/human factors are available from universities and colleges throughout the industrialized world, and professional societies have existed since 1957 for educational and information exchange purposes. So, while self-study is possible, formal, advanced education specifically in ergonomics/human factors is strongly recommended for people wanting a professional career in ergonomics.

Representative sciences that *contribute* to the body of ergonomics knowledge needed for the design process include:

- Anthropometry and anatomy — human size and shape data for work space design and arrangement, tool/equipment design and job design.
- Biomechanics and kinesiology — kinematic and kinetic methods for the assessment of human physical characteristics and performance relevant to the design of musculoskeletal work.
- Engineering sciences and/or physical sciences — basis for engineering/design opportunity and/or constraints in adapting physical and material nature to human needs and welfare.
- Physiology — energy demands of different activities and responses to heat, cold, vibration, toxins and other environmental and situational conditions that need to be considered in design for safe, healthy and productive human performance.
- Psychology — human sensation, perception, information processing, decision-making, mental workload, learning, memory, adaptation, motivation, stress, individual behavioral differences, maturation, and psychomotor-skill acquisition and performance abilities relevant to the design of systems.
- Management disciplines — planning, organizing, staffing, directing and controlling the activities of multidisciplinary design teams as governed by socio-technical and cultural conventions: standards, codes, regulations, laws, ethics and value to society. This area entails what has become known as *macro-ergonomics*.
- Mathematics and statistics — methodological and numerical basis by which relevant data can be summarized, analyzed, interpreted and synthesized for use in the design process.

The *unique knowledge base of ergonomics* derives from the methods and techniques its researchers and practitioners have developed to distinguish ergonomic design technology from the contributing disciplines enumerated above. This knowledge focuses on the systems approach to human integrated design. Systems design involves the complementary application of natural and human sciences. System design is viewed as an organized approach to decision-making (in any design context) with a proper emphasis on the human factors to achieve system utility without detriment to the humans who manage, control, use or maintain the system. In practice, ergonomic design becomes as much an art as any other engineering design profession. But it is an art based on science, not conjecture on how people do (or should) function.

1.2.1. Ergonomist Formation Model (EFM)

In coordination with the Center for Registration of European Ergonomists (CREE) and in cooperation with the Education and Training Committee of the International Ergonomics Association (IEA), the BCPE, in 1995, adopted the EFM as the fundamental architecture for professional competence in ergonomics.

The subject categories and topics in the EFM (table 1) are taken directly from Rookmaaker *et al.* (1992). The EFM closely matches the curriculum criteria used by the Human Factors and Ergonomics Society (HFES) in the accreditation process for graduate programs in ergonomics/human factors (HFES 1990). Additional information on the education, training, and skills of ergonomists can be found in Bernotat and Hunt (1977), Jahns (1991), van Cott and Huey (1992) and Rentzsch (1994), as well as, in various meeting proceedings of the IEA and its Federated Ergonomics/Human Factors Societies. Table 1 provides the categories and topics of core-ergonomics course work for which documented evidence must be provided. The hours (clock hours for contact and/or studies) shown differentiate between two levels of practice in ergonomics, as well as, the different certification designators awarded by BCPE (Certified Professional Ergonomist — CPE and Certified Ergonomics Associate — CEA).

Table 1. The Ergonomist Formation Model

CATEGORY and TOPICS	CPE (MS/PhD)	CEA (BS)
A. Ergonomic Principles	20 hours	15 hours
1. Ergonomics Approach		Required
2. Systems Theory		Elective
B. Human Characteristics	80 hours	50 hours
1. Anatomy, Demographics and Physiology		Required
2. Human Psychology		Required
3. Social and Organizational Aspects		Elective
4. Physical Environments		Required
C. Work Analysis and Measurement	100 hours	65 hours
1. Statistics and Experimental Design		Elective
2. Computation and Information Technology		Elective
3. Instrumentation		Elective
4. Methods of Measurement and Investigation		Required
5. Work Analysis		Required
D. People and Technology	100 hours	65 hours
1. Technology		Required
2. Human Reliability		Elective
3. Health, Safety and Well-Being		Elective
4. Training and Instruction		Required
5. Occupational Hygiene		Elective
6. Workplace Design***		Elective
7. Information Design***		Elective
8. Work Organization Design***		Elective
E. Applications (projects pursued by the individual during education/training)	6 weeks	Part of on the job Training
F. Professional Issues (ethics, practice standards, marketing, business practice, legal liabilities)	20 hours	5 hours

*** CPE must be competent in at least one design domain

The EFM, however, is incomplete without also consideration of on-the-job experience. For the CPE, at least 4 years are necessary; for the CEA, 2 years are required.

1.2.2. Objectives and points of reference

In addition to the Ergonomist Formation Model, Rookmaaker *et al.* (1992) developed objectives and points of reference for each category topic of the EFM. These are:

(A) Ergonomics Principles — *to understand the guiding principles and theory of ergonomics*
1. Ergonomics approach
   Objective: To recognize the integrated (systems) nature of ergonomics, the centrality of human beings, to use its breadth of coverage and the available knowledge base to adapt the environment to people.
   Points of reference: History of work; current developments; paradigms (designing for individuals versus populations; working in normal versus extreme circumstances); interaction between society and work.
2. Systems theory
   Objective: To recognize the principles of systems theory and how they apply to ergonomics situations.
   Points of reference: Structure and dynamics of systems; human as a system component; system analysis and design (e.g. allocation of functions).

(B) Human characteristics — *to describe and recognize the contributions and effects of the factors contributive to the ergonomics knowledge base on people's physical and psychological well-being, and on their performance*
1. Anatomy, demographics and physiology
   Objective: To recognize and measure the physical characteristics of people and their responses to their activities and their environments with particular reference to health and performance.
   Points of reference: Anatomy; biomechanics and posture; anthropometry; energy and force production; adjustments (stress and strain); individual, gender-related, developmental, racial and cultural variability; chronobiology (e.g. circadian rhythm).
2. Human psychology
   Objective: To recognize behavioral characteristics and responses and to understand how these affect human behavior (including health and performance) and attitudes.
   Points of reference: Psychophysiological and cognitive aspects of information intake, information handling and decision making; individual motivation; human development.
3. Social and organization aspects
   Objective: To recognize the social dimensions of ergonomics and organizations and to specify systems structures suitable to achieve a good quality of working life or performance.
   Points of reference: Motivation and attitudes related to needs of individuals and to working in groups; individual and group functioning; socio-technical systems.
4. Physical environment
   Objective: To understand the human senses and to be able to recognize, measure and specify the appropriate levels of and the characteristics of the physical environment to be suitable for human activities.

Points of reference: Climatic environment; visual environment; acoustic environment; vibration; human senses.

(C) Work analysis and measurement — *to understand, select and use the appropriate methods for investigating ergonomics problems, and for presenting data for evaluating future design solutions to these problems*

1. Statistics and experimental design

Objective: To collect, aggregate, manipulate and evaluate data in a reliable and valid manner.

Points of reference: Descriptive and inferential statistics; probability theory; correlational techniques; estimation and sampling; experimental design; non-parametric statistics.

2. Computation and information technology

Objective: To use (digital) computers, particularly utilize standard packages, for the effective prosecution of ergonomics investigations.

Points of reference: Computation for data collection; computation for calculation; computation for storage, computation for database searches, computer-aided design.

3. Instrumentation

Objective: To use the major measuring instruments, sensors, etc., required by the ergonomist to gather data for investigations, design or evaluation of workplaces, work procedures or work equipment.

Points of reference: Simple and complex equipment; their potential and their limitations.

4. Methods of measurement and investigation

Objective: To understand the major methods and procedures of measurement used in ergonomics investigations, and to know when to use them and how to interpret the results.

Points of reference: Simulations (dynamic and static); methods for observing activity and performance; interviews and questionnaires; epidemiological approach; sampling procedures; checklists.

5. Work analysis

Objective: To describe and understand the determinants and organization of workers' activities in the field or system.

Points of reference: Activity analysis; task analysis; function analysis; task interdependency, communication and cooperation; the importance of strategies in task execution.

(D) People and technology — *to understand an area for application of ergonomics expertise, some models and concepts related to applying ergonomics, and* at least *one special form "ergonomic design" may take on in any area*. *Three "optional" topics must be chosen, at least one being a "design" topic.

1. Technology

Objective: To understand the factors in the chosen area of application that are relevant to the creation of ergonomic situations; in particular to recognize those aspects of the technology that are flexible/changeable. One of the following application areas may be chosen: consumer products; manufacturing; office work; transport; process industry; healthcare; automation; architecture; recreation, arts and leisure activities (see also E, Applications).

Points of reference: Functionality, operation and construction of the technology.

2. Human Reliability (optional)*

Objective: To design and evaluate work situations using 'best practice' in working towards error-free performance.

Points of reference: Accident models; attention, effort and vigilance; error taxonomies.

3. Health, safety and well-being (optional)*

Objective: To design and evaluate work situations to achieve healthy and safe work, as well as contribute to quality of work life.

Points of reference: Safety management; occupational injuries and work-related disorders; safety technology; legislation; characteristics of good quality of work life.

4. Training and instruction (optional)*

Objective: To understand the fundamentals of learning, of training programs and of instruction, and to specify requirements of those programs to achieve successful performance to new or changed work activities.

Points of reference: Learning skills; learning knowledge; assessing job requirements and worker capabilities; designing training programs to bring workers to the level of requirements; designing manuals.

5. Occupational hygiene (optional)*

Objective: To recognize, measure and cope with the presence of adverse physical and chemical conditions and other major pollutants.

Points of reference: National and international recommendations and requirements; their variations and limitations; measurement, protection, control and monitoring.

6. Workplace design (optional)*

Objective: To investigate and design workplaces to match the physical and psychological dimensions of their users and to measure their effect on ergonomically relevant dimensions.

Points of reference: Measurement of activities and performance; workspace layout; use of mockups/simulations to improve designs; evaluation; compatibility between workplace requirements and human capabilities.

7. Information design (optional)*

Objective: To investigate and design the major modes of information transfer to the human for effective and efficient performance of the system.

Points of reference: Signal detection; information processing and attention; display characteristics; information overload; stimulus–response compatibility.

8. Work organization design (optional)*

Objective: To investigate, design and implement work organizations for effective and efficient performance and good quality of work life.

Points of reference: Cooperative analysis and design of new work systems; basics and applications of work–rest schedules; introduction of change.

(E) Application — *to apply ergonomics skills and knowledge to human-oriented systems and products*

Objective: To understand the integrative nature of applying ergonomics, the need for and structure of a specification, and the interactive and iterative nature of work in an applied research or design group, recognizing the practicalities and limitations of applying ergonomics, including the introduction of change.

Points of reference: The applied research/design process is applied in a chosen area such as: consumer products; manufacturing; office work; transport; process industry;

healthcare; automation; architecture; recreation, arts and leisure activities, etc. (see also D1, 'Technology'); intervention techniques.

(F) Professional issues — *to understand how ergonomics affects an individual's way of life and to know how legislative actions and current economic situations affect the application of ergonomics*
Objective: To recognize the impact of ergonomics on people's lives, the costs and benefits accruing from ergonomics activities, the social and psychological impact of ergonomics investigations, and the professional responsibilities and requirements for the ergonomics practitioner.

Points of reference: Legislation; economics; the ergonomist in the organization; ergonomics and society; role of ergonomist in social settings with different interest groups; ethics; development and marketing of the ergonomics profession.

1.2.3. Behavioral objectives for the professional ergonomics practitioner

In addition to the EFM, there are specific behavioral objectives that are applicable to the ergonomics professional. These objectives are the results of Hendrick (1981), who was a member of the US Air Force task force to recommend goals and changes for the human factors/ergonomics profession. The objectives are:

1. Sufficient background in the behavioral sciences to respond to ergonomics questions and issues having psychological or other behavioral implications. Implies the equivalent of a strong undergraduate behavioral science minor.

2. Sufficient background in the physical and biological sciences to appreciate the interface of these disciplines with ergonomics. Implies the equivalent of an undergraduate minor.

3. Sufficient background in hardware engineering to (a) understand design drawings, electrical schematics, test reports and similar design tools; (b) appreciate hardware design problems and the general engineering process; and (c) communicate effectively with hardware design engineers. Implies formal knowledge of basic engineering concepts at the familiarization level.

4. To be able to (a) evaluate and, (b) assist in performing classic man–machine integration, including workspace arrangement, controls, displays, and instrumentation. Implies formal knowledge of ergonomic human–equipment integration technology.

5. To be able to apply knowledge of human performance capabilities and limitations under varying environmental conditions in (a) evaluating design and (b) assisting in the development of design requirements for new or modified systems. Requires formal knowledge at the familiarization level of human performance in the various physical environments (e.g. noise, vibration, thermal, visual).

6. To be able to (a) evaluate and (b) conduct the various kinds of traditional ergonomic analyses (e.g. functional task, time-line, link). Requires formal training in these techniques at the familiarization level.

7. Have sufficient knowledge of computer modeling, simulation, and design methodology to appreciate their utility in systems development, including ergonomic utilization in function allocation, task time-line analysis, workload analysis, crew station layout evaluation, and human performance simulation (this does not include the ability to actually design the models and simulations). Requires math through calculus and introductory computer science, and knowledge at the familiarization level of measurement, modeling, and simulation as these are applied in ergonomics.

8. To be able to apply knowledge of learning and training methodology to the evaluation of training programs and to instructional systems development (ISD). Requires knowledge of that portion of learning theory and research applicable to training and training methodology at the familiarization level. Also requires familiarization with the ISD process.

9. To be able to apply the organizational behavior and motivational principles of work group dynamics, job enrichment and redesign, and related quality of work life considerations in (a) developing ergonomic system design requirements and (b) evaluating the design of complex systems Requires formal knowledge of organizational behavior at the introductory level.

10. To be able to assist in the development and evaluation of job aids and related hardware. Requires knowledge of the state-of-the-art at the familiarization level.

11. To be able to evaluate the adequacy of applied ergonomics research and the generalization of the conclusions to operational settings. Requires formal knowledge of the basic statistical methods and the principles of experimental design at the introductory level.

12. Have at least one area of specialized expertise that goes beyond the introductory graduate level of understanding and application. Requires additional coursework or thesis project in a specialized area as covered in the ergonomics topics descriptions.

1.3. Scope of Practice

Because ergonomics is principally a design discipline, the skills needed for effective practice are similar to those needed for engineering. The focus, however, is on solving human performance problems rather than problems of technology application which engineers emphasize to create hardware and/or software "tools" for enhanced living and working. The stage for ergonomists is set by "tasks" (i.e. "work" or, in Greek, *ergon*) while the stage for engineers is set by "things" (planned-for, created, objects and artifacts) that serve a functional objective. The stage for other career groups such as, healthcare providers, psychologists, sociologists, educators, etc. is the "changing human condition" by providing education, training, therapies, counseling, and practical advice on how to change the human-related factors as affected by the natural and technological worlds.

Differentiating amongst these three "operating stages" becomes important because the scope-of-practice for ergonomists overlaps, to some extent, those of other career fields. This overlap has resulted in two levels of ergonomics practice:

- CPE — a career problem solver who applies and develops methodologies for analyzing, designing, testing, and evaluating systems. A CPE addresses complex problems and advances ergonomics technologies and methods.
- CEA — an interventionist who applies a general breadth of knowledge to analysis and evaluation. A CEA reacts to performance, safety, health and/or quality issues in currently operating work systems.

While the scope of practice for the CPE covers the entire breadth and depth of ergonomics knowledge, the scope of practice for the CEA is limited to the use of commonly accepted tools and techniques for the analysis and enhancement of human performance in existing systems. (Commonly accepted tools and techniques include those which have extensive use, are widely reported in the literature, have established protocols, and which are broadly accepted by practicing CPE. Human performance is defined as including physical, cognitive, psychosocial, organizational, environmental and system factors. Existing systems include current operating systems and their components, but excludes conceptual systems, extensive redesign or design of new systems.) In general, a CEA:

- Analyzes and assesses human interactions in the system.
- Intervenes by making recommendations for improvement of identified mismatches.
- Evaluates effectiveness of implemented recommendations.

Furthermore, the CEA has to recognize those projects which exceed this scope and for which the skills of a CPE must be used. Table 2 compares the fundamental differences in scope of practice of the CPE and CEA. Tables 3–5 provide more detailed comparative examples of the scope of practice of a CEA compared with a CPE.

1.3.1. Analysis and assessment
In table 3, the left column has scope of practice analysis and assessment examples for the CEA while the right column lists the additional functions that a CPE performs.

1.3.2. Intervention
In table 4, the left column has scope of practice intervention examples for the CEA while the right column lists the additional functions that a CPE performs.

1.3.3. Evaluation
In table 5, the left column has scope of practice evaluation examples for the CEA while the right column lists the additional functions that a CPE performs.

1.4. Practitioner Experience
Practitioner knowledge and skills are developed in a large variety of settings and situations leading to specialization in the ergonomist's job functions and systems design applications. Specialization can occur along design criteria lines (e.g. health and safety, marketability) or job function (ergonomics analyst, designer, evaluator or manager). Even so, experience will always involve one or more of the following: analysis, design, testing, evaluation, research, consultation and/or management of human

performance in an operational or developmental engineered-system context. Experience on one system often will be applicable in another system. Specialization among system lines can include:

- Mobility/transportation systems (ground, water, air and space).
- Industrial systems (processing, manufacturing, distribution, etc.).
- Business systems (offices, services, etc.).
- Communication systems.
- Information management systems/computers.
- Educational systems (instructional systems design).
- Regulatory/legal systems.
- Consumer systems.
- Architectural systems.
- Healthcare systems.

Surveys have shown that about 4 years of practical experience are needed before the ergonomist is comfortable with all the principles and practices of his or her profession related to a system design team. For CEA ~2 years of practice are needed to get a thorough familiarity with the methods and protocols used in an intervention strategy. The profession is not static, and continuing education by means of active participation in professional society meetings, workshops, seminars and short courses is strongly recommended to stay abreast of rapid advances in ergonomic technology and information.

1.5. Summary
Three key concepts characterize ergonomics: human performance (work) + ergonomic design (engineering) + systems integration (technology management). Subtract any component concept, and the essence of ergonomics is compromised. The ergonomist strives to balance technological factors with personnel selection and training options to achieve state-of-the-art system performance. The profession's goals are to "humanize technology" by creating jobs and work systems which provide:

- Reasonable human performance.
- Reasonable human workload.
- Reasonable health maintenance.
- Reasonable hazard control and injury-risk management.

2. CERTIFICATION
Certification is a way to assure that something is what it is supposed to be. Applied to professionals, certification ensures that individuals have met a set of standard requirements. The certification process is a structured approach to evaluate individuals in their discipline by assessing compliance to technical standards. Certification is generally a voluntary process by a non-

Table 2. Comparative chart of the scope of practice of a CEA and CPE

CEA	CPE
• Conducts basic workstations analyses	• In addition to conventional problems, addresses more complex and non-conventional problems.
• Applies widely-established techniques to address conventional problems.	• Develops and applies advanced methodologies, mathematical models and/or simulations.
• Works within the "intervention model," i.e., ergonomics-related problems identified at an existing workstation.	• Works within the broader 'system design model' in which intervention may be **one** particular strategy among others.

Table 3. Comparison of the scopes of practice of a CEA and a CPE — Analysis and Assessment

CEA	CPE

SYSTEM

- Conducts basic analysis of a facility to identify problem areas using widely recognized methods such as:
  — reviewing the workers compensation records to identify the three jobs with the highest incidents of musculoskeletal disorders in a plant.
  — conducting a walkthrough survey using a checklist.

- Recognizes musculoskeletal and performance problems from organizational and management factors such as:
  — increased injuries related to downsizing, change of supervision, department restructuring and changes that incorporate policies.

- Conducts basic to complex analyses (multiple facilities, national and international, diverse products, unconventional organizational structures) using existing to novel methods, such as:
  — developing the survey and analysis method to identify multiple factors to prioritize ergonomics-related problems.
  — developing an unique survey which integrates production bottlenecks and turnover, with injuries to prioritize ergonomic projects within a facility.

- Analyzes the organizational and management structure and processes to identify the root causes of injury and performance decrements.

- Analyzes work system structure and processes to determine appropriateness in light of key sociotechnical systems, such as:
  — reviewing organizational structure to develop and implement rule based procedures to lower decision making.
  — altering organizational culture to achieve quality management implementation.

- Uses computer modeling, simulation, and design methodology in systems development, such as:
  — allocating human-machine functions for an advanced manufacturing assembly line.

Task

- Conducts basic task analyses using tools such as checklists, questionnaires, videotaping and force measurement, for example:
  — investigating 50 person office workplace using standard VDT checklist.

- Uses commonly accepted tools and techniques for analysis, for example:
  — assessing manual transfer of chemical bags into a mixing bin using the NIOSH equation.

- Recognizes basic mismatches between the job requirements and human capabilities including physical, cognitive, psychosocial and enviromental, for example:
  — identifying worker injuries that result from workload and deadline pressures rather than from workstation layout.
  — recognizing that a quality problem results from excessive noise levels.

- Identifies obvious sources of error such as viloations of occupational stereotypes, inconsistent displays, controls, control-display arrangements and poor coding design.

- Bases analysis decision on widely recognized design criteria and ergonomics principles.

- Uses simple statistical techniques such as descriptive statistics and performs simple statistical analyses such as t-tests and correlations.

- Recognizes when a sophisticated level of analysis is required.

- Develops and applies advanced analytical methods, mathematical models and simulation, for example:
  — using a simulation model to review stairs in new stadium design
  — using a CAD system to determine adequacy of maintenance accesses.

- Designs and conducts analytical studies, for example:
  — recognizing that multiple workplace changes within a manufacturing cell warrant pilot studies and mock-ups to verify changes.
  — determining a safe workload based on workload analysis by body part, force requirements, pace and work duration.

- Develops and applies theoretical constructs, for example:
  — developing a diagnostic tool for incorporating a knowledge based system for scientists and rule-based system for technicians.

- Conducts error analysis, for example:
  — reviewing five years of operational data to determine human error likelihood for a refinery.

- Interprets study results to determine design requirements, for example:
  — reviewing literature to specify design for a new road grader utilizing either an integrated control stalk or seperate controls.

- Uses advanced statistical techniques such as multi-variate analysis and regression analysis, for example:
  — using multi-variate analysis to predict service call rates based on interaction of overtime requirements and shift assignment.

- Recognizes when the analysis requires a level of expertise in a specialty area at a level that exceeds ones own.

Table 4. Comparison of the scopes of practice of a CEA and a CPE -Intervention

CEA	CPE
• Applies ergonomics principles at the workstation level by: — modifying a packaging station by replacing a lift motion with a slide motion resulting in fewer risk factors and enhanced productivity. — rearranging displays on a control panel based on ergonomics principles of importance and frequency of use.	• Applies ergonomics principles at the work system level by: specifying methods and tools for different types of terrain for tree harvesting to optimize efficiency and safety. specifying resources for enhanced situation awareness in transportation systems.
• Makes basic recommendations for improvement of existing systems, encompassing engineering, administrative and work practice modifications, for example: — rearranging a workplace fixture to eliminate an awkward posture. — rearranging a sequence of work tasks to reduce cycle times. — using anthropometric data to determine the height of an overhead control valve for 5th percentile female worker. — recommending off-the-shelf hardware such as lift tables, hoists, conveyors. — recommending job rotation on an assembly line to control overuse injuries.	• Makes recommendations for the improvement of existing sophisticated systems by applying human-equipment integration technology, for example: — replacing a small motor assembly line with a manufacturing cell to control labor costs, enrich jobs and provide just-in-time output. — restructuring departmental jobs by recombining work modules to better distribute work loads and enhance output. — redesigning a shift schedule to reduce vigilance decrement of radar surveillance operators. — adding test functions to electronic sub-assembly in order to provide knowledge of results, thus enhancing motivation and reducing rework. — designing cognitive aids, such as a procedural storyboard for packing medical instruments.
• Trains employees in ergonomics principles and task techniques, for example: — training office workers on how to adjust workplaces. — training employees to include ergonomics in Job Safety Analyses.	• Develops training programs and materials, for example: — training people in ergonomics analysis, interventions, and evaluation, such as training ergonomics teams or design engineers. — using the training system development process to determine training requirements for computer maintenance personnel. — specifying use of whole versus part task simulators for training pilots.
	• Tailoring conventional intervention methods for new or unique opportunities, for example: — using aviation crew resource management procedures to optimize decision making in a chemical control room.
	• Designing and developing new systems or systems requiring major revision by determining design criteria and applying those specifications to the design, for example: — designing a new software package. — designing a nuclear power plant control room. — revising the manufacturing line to shift from conventional CRT monitors to flat panel displays.
	• Applies knowledge of engineering concepts to communicate with design engineers, for example: — understanding design drawings and appreciating hardware design problems of a new packing line to determine potential design options for reducing the physical job demands.
• Recognizes when the scale or complexity of intervention requires a CPE.	• Recognizes when the intervention requires a level of expertise in a specialty area at a level that exceeds ones own.

governmental agency that recognizes individuals for advanced knowledge and skill (NOCA 1996). According to the Council of Engineering and Scientific Specialty Boards (CESB 1998), "Certification is the recognition of special capability conferred on individuals who meet the education, experience, competency and other requirements for certification in a specified area of practice." Certification is used by many professions to ensure a defined level of competence.

The certification of individuals claiming to be ergonomists was initiated by the BCPE in 1992. The basis for the development of certification criteria, procedures and the content topics of the written examination were several reviews of job/task analyses performed by committees of the Human Factors and Ergonomics Society, the International Ergonomics Association, the Department of Defense, NATO, and the National Academy of Science/National Research Council. These efforts spanned the period between 1970 and 1992. Knowledge, skills and experience domains and topics were then categorized by the *Ergonomics Abstracts* (London: Taylor & Francis, 1993) schema and evaluated for validity and reliability against the most widely used textbooks and handbooks in ergonomics/human factors academic degree programs.

Table 5. Comparison of the scopes of practice of a CEA and a CPE — Evaluation

CEA	CPE
• Uses simple structured evaluation tools and methods to measure effectiveness of ergonomic interventions, for example: ? — administering before and after discomfort surveys to ensure a raised packing station height reduced back discomfort. ? — administering work environment survey to evaluate employee satisfaction with the new work station. ? — conducting a survey to elicit employee preferences for laboratory stools. — using structured interviews to gather feedback on the employee acceptance of a semi-automated ticketing system at an apparel assembly workstation.	• Designs and develops evaluation tools and methods to measure effectiveness of ergonomic interventions, for example: — evaluating effects of introducing robotic packing of television sets, by developing a program which includes determining evaluation requirements and designing or selecting evaluation tools. — specifying human performance measures suitable for meausuring vehicle guidance with cell-phone task sharing.
• Uses simple statistical evaluation such as means and standard deviations, for example: — calculating average area roofed per shift by construction workers and standard deviation for manual nailing versus using nailing guns.	• Conducts sophisticated evaluations including multivariate statistical analysis, for example: – evaluating the interaction effects of increasing lighting and reducing noise on error rate of pill dispensing in pharmacies. — evaluating effectiveness of 12 hour versus 8 hour work shifts in a rotating shift chemical plant.
• Performs cost justification of simple and relatively inexpensive projects, for example: — as part of cost justification, calculating changes in annual lost time accidents and injuries following ergonomic redesign of shoe assembly work stations and determining savings in workers' compensation costs.	• Performs cost benefit analyses, especially of complex and expensive projects, for example: — justifying the cost of modifying a harvesting machine cab to improve visibility in terms of savings from increased harvesting rate and reduced equipment damage from accidents.
• Recognizes when a sophisticated level of evaluation is required.	• Recognizes when the evaluation requires a level of expertise in a speciality area at a level that exceeds ones own.

2.1. Criteria for Certification

Based upon the EFM, the BCPE has defined two career paths for certification. To differentiate between the two levels of practice in ergonomics, different certification designators are awarded by BCPE. The Certified Professional Ergonomist (CPE) is the principal designation for career professionals. Individuals who desire to use the phrase "Human Factors" in lieu of the word "Ergonomics" can chose the Certified Human Factors Professional (CHFP) designation. The designation, Associate Ergonomics Professional (AEP)/Associate Human Factors Professional (AHFP), is granted to individuals who have met the initial CPE criteria but have not taken the complete CPE examination and have not met the 4-year experience requirement. AEP/AHFP is an interim designation. AEP/AHFP individuals have 5 years to complete the requirements for the CPE/CHFP designation. The Certified Ergonomics Associate (CEA) designation are given to individuals who are limited to an interventionist role for system analysis and evaluation.

2.1.1. CPE/CHFP

The BCPE job/task analyses led to the following minimum criteria for certification at the CPE/CHFP level:

- A master's degree in ergonomics or human factors, or an equivalent educational background in the life sciences, engineering sciences *and* behavioral sciences to comprise a professional level of ergonomics education.
- Four years of full-time professional practice as an ergonomist practitioner with emphasis on design involvement (derived from ergonomic analysis and/or ergonomic testing/evaluation).
- Documentation of education, employment history and ergonomic project involvement by means of the BCPE "Application for Certification."

- A passing score on the BCPE written examination.
- Payment of all fees levied by the BCPE for processing and maintenance of certification, in accordance with the following schedule (all fees are non-refundable):
- Application fee: US$10.00.
- Processing and examination fee: US$290.00.
- Annual certification maintenance fee: US$100.00.

2.1.2. AEP/AHFP

On 26 March 1995 the BCPE created an interim or transient Associate category of certification. An individual with the AEP/AHFP designation is considered to be an "ergonomist in training" who has 5 years to transition to the higher designation of CPE/CHFP after completing 4 full-time years of experience. A person can be certified as an AEP or AHFP if he or she:

- Meets the education requirements for BCPE certification (MS in human factors/ergonomics or related field).
- Has passed Part I (on "Basic Knowledge" of human factors/ergonomics) of the BCPE certification examination.
- Currently is working toward fulfilling the BCPE requirement of 4 years practical experience as a human factors and ergonomics professional.
- Note: a person takes the Basic Knowledge portion (Part I) of the BCPE exam immediately after fulfilling the education requirement. Parts II and III of the exam are taken after fulfilling the other BCPE requirements. A person who has graduated from a human factors/ergonomics degree program accredited by an IEA Federated Society (e.g. HFES), will not have to take Part I of the exam. BCPE established the "Associate" level to create a path by which individuals could achieve professional certification in progressive steps. BCPE also wanted to link this path with accredited educational

degree programs by giving some preference (by waiving Part I of the exam) to applicants who have graduated from programs accredited by IEA Federated Societies. BCPE believes this linkage is an important step in strengthening the human factors and ergonomics profession.

- Fees associated with the AEP/AHFP designations are (all fees are non-refundable):
- Application fee: US$10.00.
- Processing and examination: $100.00.
- Annual AEP/AHFP maintenance fee: $60.00.

2.1.3. CEA

On 17 April 1998, the BCPE established a technician level of certification to meet the growing need for certified ergonomists who use commonly accepted tools and techniques for analysis and enhancement of human performance in existing systems, but are not required to solve complex and unique problems, develop advanced analytic and measurement technologies, provide a broad systems perspective, or define design criteria and specifications. The minimum criteria for certification at the CEA level are:

- A bachelor's degree from an accredited university.
- At least 200 contact hours of ergonomics training.
- Two years of full-time practice in ergonomics.
- Documentation of education and employment history by means of the BCPE "Application for CEA Certification."
- A passing score on the BCPE CEA examination.
- Payment of all fees levied by the BCPE for processing and maintenance of certification, in accordance with the following schedule (all fees are non-refundable):
- Application fee: US$10.00.
- Processing and examination fee: US$190.00.
- Annual CEA maintenance fee: US$75.00.

2.2. Procedures for Certification

Certification candidates request application materials by sending a US$10.00 check to: BCPE, PO Box 2811, Bellingham, WA 98227-2811, USA.

The application materials consist of three or four pages of instructions and four or seven pages of forms to be filled out by the applicant. These include:

- Section A — Personal data.
- Section B — Academic qualifications.
- Section C — Employment history (CPE); ergonomics training hours (CEA).
- Section D — Work experience in ergonomic analysis, design, and testing/evaluation (CPE); employment history (CEA).
- Section E — Work product description (CPE).
- Section F — Signature and payment record.

The candidate completes the application and submits it with (1) the appropriate processing and examination fee, (2) an official academic degree transcript and (3) (CPE only) a work product (article/technical report/project description/patent/application, etc.).

The processing and examination fees are non-refundable.

A review panel evaluates all submitted materials and makes recommendations to the Board whether or not the applicant qualifies to take the written examination. The qualifying candidate takes the written examination and is certified after receiving a passing score. This candidate has up to 2 years after the application is received to take the exam. The non-qualifying applicant has *2 years* to correct any deficiencies or missing elements in the application as pointed out by the review panel for completing the application without additional payment.

3. EXAMINATIONS

As the knowledge, skills and experience domains of a maturing career field evolve, it is advisable to develop self-assessment and self-regulatory instruments. These are needed to gauge the status of the profession in relation to societal demands for services and to gauge the preparedness of individuals to meet those demands successfully. For ergonomists, academic standards have been evolving since the 1960s and resulted in the academic accreditation procedures adopted by the Human Factors and Ergonomics Society (HFES) in 1987. Since BCPE established the certification process in 1992, a multi-part examination was defined to assess how well individuals meet the objectives of the EFM.

3.1. Self-screening for Eligibility

People can enter into the practice of ergonomics from other academic disciplines and occupational experiences, not solely ergonomics. Knowledge and skills range from "novice" (those who have attended seminars, short courses, and/or technical/trade meeting presentations) to "expert" (those who have received advanced academic degrees in ergonomics and then practiced ergonomics in work system environments). Certification eligibility criteria reflect the topics of knowledge and skills that together differentiate ergonomics from other career fields and academic disciplines.

As an aide to decide whether an individual should take the CPE or the CEA examination after meeting the academic and experience requirements, a "Self-Screening Score Sheet" (see Appendix C) was developed for guidance.

3.2. Examination Administration

Applicants who have demonstrated eligibility for the examination will be notified regarding the date and location of the next examination ~2 months before the testing date. The examination, requiring a full day for CPE candidates and 4 hours for CEA candidates, will be scheduled generally for the spring and fall, and will probably be offered as an adjunct to the meetings of ergonomics related professional societies and associations. Qualified applicants needing accommodations in compliance with the Americans with Disabilities Act (ADA) are asked to specify their accommodation needs to the BCPE before signing up for taking the examination.

3.3. Scoring Methods

A panel of BCPE certificants with expertise in psychometrics determines the method of establishing passing scores to be used for the examination. Passing scores are established to ensure the applicant's mastery of the knowledge and skills required for the related level of ergonomics practice. The BCPE will periodically review, evaluate, and, as necessary, revise the examinations and scoring to assure that valid and reliable measures of requisite performance capability for ergonomics practice are maintained.

3.4. Retaking the Examination

An applicant who does not pass the examination may retake the examination at the next regularly scheduled examination date and place. A reduced fee will be applied to retakes, and the retake must take place within 2 years of the original application date.

3.5. Examination Development

An examination-development working group is functioning under Board direction to assist in constructing the examination. The examinations were developed from contributions from ergonomics academics, researchers and practitioners (both BCPE certified and non-certified). The Board developed the subject outline and assigned weighting factors to each subject area in accordance with the Ergonomist Formation Model.

3.5.1. CPE/CHFP subject areas

The examination is designed and constructed to examine:

- Fundamental knowledge in ergonomics by means of multiple-choice items. (Part I: 2.0 hours).
- Application skills by means of multiple-choice items and essays to questions spanning the practices and principles appropriately applied in a setting of ergonomists' involvement with systems analysis, ergonomic design, and systems/human performance evaluation. (Part II: 3.0 hours; Part III: 2.0 hours).

The answers to questions in Part II of the test reflect the complexity of problems typically encountered in professional practice. To answer a given question or problem, for example, may require the synthesis and integration of knowledge and techniques from several sources or subject areas. Although questions may involve the application of standard techniques to novel situations, the situations posed will be fundamental to all system designs from an ergonomic perspective as reflected in the cited books.

Part III of the test uses a scenario-based approach to allow the applicant to demonstrate ergonomic knowledge and skills in a particular setting (e.g. industrial, office, transportation, communication, computer systems). The examinee will be expected to select and respond to a subset of presented scenarios requiring ergonomic solutions. The selectivity is designed to enable truly qualified persons to respond adequately to the realistic work situations posed and to allow them to demonstrate professional expertise and practice.

3.5.2. CEA subject areas

The CEA examination is designed and constructed to examine:

- Fundamental knowledge in ergonomics by means of multiple-choice items. (Part I: 2.0 hours).
- Application skills by means of multiple-choice items to questions spanning the practices and principles appropriately applied in a setting of ergonomists' involvement with systems analysis, ergonomic design, and systems/human performance evaluation (Part II: 2.0 hours).

3.5.3. CPE/CHFP and CEA approximate weighting of subject areas

Table 6 summarizes the weighting factors and reflects the Board's best estimate of the relative importance of topics covered based on interpretation of the job/task analyses. By invitation of the

Table 6. CPE/CHFP and CEA Examination Approximate Weighting of Subject Areas

1. Methods and Techniques (M&T)	30%
2. Design of Human-Machine Interface (DHMI)	25%
3. Humans as Systems Components (Capabilities/Limitations) (HSC)	25%
4. Systems Design and Organization (SDO)	15%
5. Professional Practice (PP)	5%

BCPE, ergonomists are invited to draft test questions for the Test Item Data Bank. After evaluation and peer review, the Board selects questions to be included in a scheduled examination. Statistical data are collected and analyzed by the BCPE on the demographics of applicants and their test performance for test-validity and reliability assessment purposes.

3.6. Application Scenarios

Ergonomists work in a variety of settings and situations. Moreover, a variety of professions (engineering, hazard/risk control, project management, labor representatives, system operators/maintainers, health-and-safety specialists, etc.) work with ergonomists to develop technological systems. Different systems may also have different emphases on the type of human factors involved. For example, manual systems may be more concerned with musculoskeletal issues, while automated systems may place a greater emphasis on cognitive factors and/or supervisory-decision-aiding. However, the CEA and CPE should be able to transfer their fundamental knowledge and skills in ergonomics to a comprehensive set of work domains.

The CPE and CEA examinations sample from the following list of "system application scenarios":

1. Mobility/transportation systems (ground, water, air and space).
2. Industrial systems (processing, manufacturing, distribution, etc.).
3. Business systems (offices, services, etc.).
4. Communication systems.
5. Information management systems/computers.
6. Educational systems (instructional systems design).
7. Regulatory/legal systems.
8. Consumer systems.
9. Military systems.
10. Architectural systems.
11. Healthcare systems.

The application of ergonomics will be similar in all these systems to achieve:

1. Reasonable performance (human and system).
2. Reasonable workload.
3. Reasonable health maintenance.
4. Reasonable hazard control and risk management.
5. Reasonable cost/benefit ratios.

3.7. List of Books for Core-Topics Review

The purpose of providing a list of references is to assist applicants in their study preparation for the BCPE examinations. *The suggested references cited herein are not intended as an all-inclusive and complete list; rather, they are provided to enable applicants to direct their study efforts to the subject areas covered in the examination.* The books are general references that address specific system applications at various levels of depth. Consequently, there is

overlap among the topics they cover. Many of the books are available in public, university and/or employer libraries.

The Test Development Working Group of BCPE uses the book list as some authoritative resources for determining appropriate answers to questions and scenarios contained in the BCPE written examinations. That does not mean that test items are directly derived from those books. Test items are a representative sample of core knowledge and skills which ergonomists/human factors personnel working at the various certification levels should be able to demonstrate in writing. The books only provide the current, scientific basis for the needed knowledge and skills. Later editions of a given book generally incorporate advances in the science and/or technology of the subject matter covered.

3.7.1. Preparation for the CEA written examination

The following annotated bibliography is provided to facilitate review of the knowledge-base-of-ergonomics data, methods and techniques germane to the CEA level of practice:

- BAILEY, R.W. 1989, *Human Performance Engineering* (2nd ed.) (Englewood Cliffs: Prentice-Hall; 3rd ed. in press).

This book can be categorized as covering cognitive ergonomics. The system focus is on human–computer–environment performance and the analytical, design and test/evaluation methodology available to create user-centered work systems. In the words of the author: "The primary purpose of this book is to provide designers, particularly those with limited background in psychology, with some knowledge of how people sense, process information, and respond; as well as to introduce data, principles, and methods that are useful in eliciting an acceptable level of human performance in systems."

- CHAPANIS, A. 1996, *Human Factors in Systems Engineering* (New York: Wiley).

As the title implies, this book shows how to integrate human factors into the design of tools, machines, and systems so that they match human abilities and limitations. Unlike virtually all other books on human factors, which leave the implementation of general guidelines to engineers and designers with little or no human factors expertise, this book shows the reader how to prepare project specific system requirements that engineers can use easily and effectively. Additionally, it fully explains the various work products — the standards and specifications — that engineers must produce during development, and shows what human factors inputs are required in each of them.

- EASTMAN KODAK CO. 1983, *Ergonomic Design for People at Work*, vol. 1 (New York: van Nostrand Reinhold).
- EASTMAN KODAK CO. 1986, *Ergonomic Design for People at Work*, vol. 2 (New York: van Nostrand Reinhold).

These are definitive works on industrial human factors/ergonomics. Volume 1 is directed to a practical discussion of workplace, equipment, environmental design and of the transfer of information in the workplace. The second volume includes guidelines for job design, manual materials handling, and shift work.

- KROEMER, K.H.E. and GRANDJEAN, E. 1997, *Fitting the Task to the Man*: *A Textbook of Occupational Ergonomics* (5th ed.) (Philadelphia: Taylor & Francis).

This book is similar to the Kodak books, but has a greater emphasis on work physiology and how to use technology to overcome the stresses and strains of work systems. This popular

book provides a comprehensive introduction to the field of ergonomics. Human physical and mental characteristics, anthropometry, workplace design, heavy physical work, control/display design, work hours and shifts, noise and vibration, and many additional topics are addressed.

- KONZ, S. 1995, *Work Design*: *Industrial Ergonomics* (4th ed.) (Scottsdale: Gorsuch).

This book bridges industrial engineering concepts/methods and ergonomics by providing specific guidelines for human-centered design of jobs. In 34 chapters the analysis, design and evaluation issues for enhancing human performance, workload, safety/health and economic value of production systems are covered with examples drawn from typical manufacturing domains. The 4th edition has extensive coverage on cumulative trauma and manual material handling (including the NIOSH lifting guideline). Chapter 12 gives guidelines for reduction of human error.

- SANDERS, M.S. and MCCORMICK, E.J. 1992, *Human Factors in Engineering and Design* (7th ed.) (New York: McGraw-Hill).

This is the "classic" textbook in the USA for upper-division and graduate level ergonomics/human factors courses. Its previous editions span a period going back to 1957, and it has also become an important resource for practicing professionals in ergonomics/human factors. The book stresses basic concepts, with an empirical research basis, for ergonomic systems development, and provides numerous references to the scientific literature for those who want to delve deeper into a particular topic area.

- WICKENS, C.D, GORDON, S.E. and LIU, Y. 1998, *An Introduction to Human Factors Engineering* (New York: Addison Wesley Longman).

Although it uses the term "human factors," this book comprehensively spans all the topics covered in the BCPE ergonomist formation model. The authors also cite most of the other books listed by BCPE for review by CEA candidates, although those may have somewhat different emphases on various topics covered. In 19 chapters, the general aspects of work systems design methodology and principles are discussed. The writing style is excellent for comprehending the balance needed in using technologies with a human factors focus in different aspects of system and human performance development. From research methods to application domains, performance, safety and health criteria are introduced based on the most current consensus among researchers and practitioners in ergonomics.

In conclusion, these books can serve as a "refresher" to ergonomics for those who have had formal education and training in ergonomics. Trying to study them all two months before the written examination is likely to create high stress with little payoff in knowledge gained.

3.7.2. Preparation for the CPE/CHFP written examination

While the list of references suggested to review for the CEA examination discussed above provides fundamental data, guidelines and methodologies, the following book list spans ergonomics topics to a greater depth and/or more complex relationships among tasks, tools, talents and technology found in modern work science and systems. These books have been,

and continue to be, the most often cited for graduate course work in ergonomics/human factors.

- BOFF, K.R. and LINCOLN, J.E. (eds) 1988, *Engineering Data Compendium — Human Perception and Performance* (Wright Patterson Air Force Base, OH: Harry G. Armstrong Aerospace Medical Research Lab).

In three volumes, the *Engineering Data Compendium* (a) identifies and distills information of value to system design from the existing research literature on human perception and performance, and (b) presents this technical information in a way which is accessible to engineers working in harmony with ergonomists who can interpret the information and judge its applicability in system analysis, design and/or evaluation processes. Topics covered are:

- Visual acquisition of information.
- Auditory acquisition of information.
- Acquisition of information by other senses.
- Information storage and retrieval.
- Spatial awareness.
- Perceptual organization.
- Attention and allocation of resources.
- Human language processing.
- Operator motor control.
- Effects of environmental stressors.
- Display interfaces.
- Control interfaces.

- BOOHER, H.R. (ed.) 1990, *Manprint: An Approach to Systems Integration* (New York: van Nostrand Reinhold).

This book outlines a comprehensive approach to work science and systems development that enables organizations to convert "equipment-dominated" mindsets to those that are more 'people-oriented.' The concepts and theories of economics, ergonomics, engineering and business management are integrated to show a relationship to quality performance and human well-being needed for a rapidly changing global marketplace.

- CHAFFIN, D.B. and ANDERSSON, G.B. 1991, *Occupational Biomechanics* (2nd ed.) (New York: Wiley).

This book reveals how to work and how to design work tools to prevent musculoskeletal disorders and improve manual (physical effort) working conditions for optimal productivity and safety. Based on solidly scientific research findings, the topical structure of this new edition examines:

- Structure and function of the musculoskeletal system.
- Anthropometry in occupational biomechanics.
- Mechanical work-capacity evaluation.
- Bio-instrumentation in occupational biomechanics.
- Occupational biomechanical models.
- Methods of classifying and evaluation manual work.
- Manual materials handling limits.
- Guidelines for seated work.
- Biomechanical considerations in machine control and workplace design.
- Hand tool design guidelines.
- Worker selection and training criteria.

- HAYS, W.L. 1988, *Statistics* (4th ed.) (New York: Holt, Rinehart & Winston).

The behavior of "physical" systems can be described generally by the deterministic mathematics of geometry, trigonometry and calculus.

Because human performance tends to be less predictable and more variable than "machine performance," the mathematical tools of descriptive and inferential statistics play generally a greater role than calculus in ergonomics. This book provides information on elementary probability theory, frequency and probability distributions, measures of central tendency and variability, sampling distributions and point estimation, hypothesis testing (*t*-test, *F*-test, c^2, non-parametric tests, ANOVA, MANOVA, regression analyses, ANCOVA, etc.) which should be part of every professional ergonomist's tool kit. Without knowledge of how and when to use statistics, all ergonomics work becomes only "guess work."

- KARWOWSKI, W. and MARRAS, W.S. (eds) 1999, *The Occupational Ergonomics Handbook* (Boca Raton: CRC Press).

This handbook significantly complements the other books in the list by specifically focusing on the needs of the industrial practitioner faced with issues of physical effort, safety and health, cost/benefit ratio, and regulatory criteria. Nearly 150 international experts in ergonomics have provided their knowledge and skills throughout 111 chapters spanning seven categorical sections:

1. Background.
2. Foundations of Ergonomics Knowledge.
3. Musculoskeletal Disorders — Engineering Factors.
4. Administrative Controls.
5. Organizational Design.
6. Ergonomics and the Working Environment.
7. Applications of Ergonomics in Occupational Settings.

- RASMUSSEN, J. 1986, *Information Processing and Human Machine Interaction — An Approach to Cognitive Engineering* (Amsterdam: Elsevier/North-Holland).

This book provides the concepts and model for an integrated approach to the design of human–machine systems which use advanced information technology for support of human decision making during supervisory control tasks and emergency management. The systems of interest are highly automated and therefore rely more on the cognitive functioning of human operators than on physical functioning and/or effort. The book does not provide specific design methods, but rather a "means-goal" structure which ergonomists can use to create "human-centered" automation.

- SALVENDY, G. (ed.) 1997, *Handbook of Human Factors and Ergonomics* (2nd ed.) (New York: Wiley Interscience).

This "handbook" reads more like a comprehensive textbook with both theoretical and practical knowledge being covered on most aspects of physical, cognitive and socio-technical issues in ergonomics. Sixty chapters, written by more than 100 of the foremost authorities around the world, provide case studies, examples, figures, tables and synopses of the empirical foundations and broad applications of ergonomics/human factors in work systems.

- SIEGEL, S. and CASTELLAN, N.J. Jr 1988, *Nonparametric Statistics* (New York: McGraw-Hill).

The non-parametric techniques of hypothesis testing are uniquely suited to many data of ergonomics/human factors. The tests involved are generally "distribution-free," are usable with simple ranking scales as measures and are suitable for use with

small samples from the total population of interest. The book does a very good job of presenting suitable statistical tests in accordance with various research/evaluation designs involving human performance parameters.

- WICKENS, C.D. 1992, *Engineering Psychology and Human Performance* (2nd ed.) (New York: HarperCollins).

The author states: "I wrote this book because I saw a need to bridge the gap between the problems of system design and much of the excellent theoretical research in cognitive experimental psychology and human performance. … This is not a handbook of human factors or engineering psychology. … The chapters correspond to the flow of information as it is processed by a human being and are not generally organized from the perspective of different system components or engineering concerns." The book is written for: "the student in psychology, … the engineering student, and … the actual practitioner in engineering psychology, human performance, and human factors engineering."

- WOODSON, W.E., TILLMAN, B. and TILLMAN, P.L. 1992, *Human Factors Design Handbook* (2nd ed.) (New York: McGraw-Hill).

Since its publication in 1981, this handbook has become a definitive guide to all aspects of human factors design. This revised and expanded second edition is even more useful for integrating human factors into the design of systems, facilities, equipment, and products to ensure safe and efficient use, and to prevent user error, injury, inconvenience, or dissatisfaction.

The *Handbook* is fully applications-oriented rather than theoretical or tutorial, and recommendations are based on many years of tested usage and effectiveness in the marketplace. It offers guidelines, illustrated with nearly 500 diagrams, checklists, charts, and tables to help the user interpret and apply key concepts and data.

APPENDIX C: SELF–SCREENING FOR BCPE CERTIFICATION ELIGIBILITY

People enter into the practice of ergonomics from a variety of academic backgrounds and occupational experiences. The range of knowledge and skills is from "novice" (people who have attended seminars, short courses and/or technical/trade meeting presentations) to "expert" (people who have received advanced academic degrees in ergonomics and then practiced the career in the real world of work systems). *The Ergonomist Formation Model (EFM)* adopted by the BCPE spans the middle of this broad-range distribution of ergonomics talent. Certification eligibility criteria reflect the topics of knowledge and skills, which together differentiate ergonomics from other career fields and academic disciplines.

As an aide to deciding where any particular individual may fit into the overall ergonomics arena, this "self-screening" score sheet is offered for completion and guidance. More specific information is available in BCPE's *'Candidate Handbook: Certification Policies, Practices and Procedures.'*

A.1.1.Instructions

1. Read over the major categories (lettered) and associate detailed topics (numbered) in the score sheet.
2. Place your score on the line by each topic in accordance with the following "key for point assignments."

Point Assignments:

0 = Never exposed to this topic.
1 = Learned from self-study or "on the job" experience.
2 = Learned from seminars/short-courses/workshops.
3 = Part of my undergraduate program.
4 = Part of my *graduate* program in another field.
5 = Part of my *graduate* program in HF/ergonomics.

3. Compute the subtotal for each category and enter it in the space provided for each category.
4. Compute the grand total and enter it in the space provided.

A.1.2.Interpretation of Scores

The thing to keep in mind for "self-scoring" is that a person qualified in ergonomics will be able to honestly assign some points to each category. The point spread will be balanced across categories and *not* very high in one and very low in another. The latter situation is more indicative of the contributing disciplines to ergonomics than to ergonomics *per se*.

Table 7 provides the score distributions most likely to lead to successful certification for CEA and CPE candidates.

If the grand total falls between 43 and 73 the person should consider the CEA route; if between 79 and 150, the CPE route; those scoring below 43 may be familiar with some aspects of ergonomics, but they should not provide ergonomics services to others. If your score is between the two, 74–78, use your academic background to help you decide — Masters in Ergonomics = CPE; Bachelors = CEA. If more than two categories' scores crossover from CEA to CPE (or vice versa), an indication for lack of depth and/or breadth in ergonomics may be present.

Table 7: Score Distributions CEA and CPE

CATEGORY	CEA	CPE
A	4–6	7–10
B	9–15	16–25
C	9–15	16–30
D	9–15	16–30
E	11–20	21–50
F	1–2	3–5
Grand Total	43–73	79–150

Score Sheet

A.   Ergonomics
1. Ergonomics Approach to Systems Development _____
2. Ergonomics and Society _____
     (10) Subtotal _____

B.   Ergonomic approaches to people at work
1. Work Evaluation and Investigation _____
2. Work Activity/Analysis _____
3. Introduction to Ergonomic Design _____
4. Design Requirements Analysis and Report _____
5. Instrumentation _____
     (25) Subtotal _____

C.   Human characteristics and humans at work
1. Anatomy _____
2. Physiology _____
3. Biomechanics and Anthropometry _____
4. Human Psychology _____
5. Organizational Design and Management _____
6. Physical Environment of Work _____
     (30) Subtotal _____

D.   Supporting courses
1. Quantitative and Qualitative Design and Analysis _____

E.   Systems Theory _____
3. Technology/Engineering _____
4. Physics _____
5. Business/Economics _____
6. Ethics and Regulation _____
     (30) Subtotal _____

F.   Application areas for ergonomics
1. Workplace Design _____
2. Information Design _____
3. Work Organization Design _____
4. Health, Safety and Well-being _____
5. Training and Instruction _____
6. Occupational Hygiene _____
7. Architecture _____
8. Participatory Design Processes _____
9. Work-related Musculoskeletal Disorders _____
10. Human–Computer Interaction _____
     (50) Subtotal _____

G.   Field work
1. Applications (5) Subtotal _____

GRAND TOTAL _____

NOTE: 'work' is defined as any purposive human activity requiring effort and skill

Scenario-based Design

J. M. Carroll

VPI & State University, McBryde 660, Blacksburg, VA 24061, USA

Scenario-based design elaborates a traditional theme in human factors and ergonomics: namely, the principle that human characteristics and needs should be pivotal considerations in the design of tools and artifacts. In scenario-based design, descriptions of usage situations become more than orienting examples and background data, they become first-class design objects. Scenario-based design takes literally the adage that a tool *is* what people can do with it, and the consequences for them and for their activities of using it.

1. WHAT SCENARIO-BASED DESIGN IS

In scenario-based design, key situations of use in established work practice are enumerated and described in schematic narratives. Each narrative depicts one or more actors, a defining goal and possibly several ancillary goals, a context of tools and other artifacts, and a sequence of actions and events through which the defining goal is achieved, transformed, obstructed, and/or abandoned. Figure 1 describes the use of a multimedia education system. At an early phase of design, this might be an *envisioned scenario*, vividly conveying a new system concept, or, if this type of system is already in use and being refined, this might be an *observed scenario*, summarizing typical use and current design

Harry is interested in bridge failures; as a child, he saw a small bridge collapse when its footings were undermined after a heavy rainfall. He opens the case study of the Tacoma Narrows Bridge and requests to see the film of its collapse. He is stunned to see the bridge first sway, then ripple, and ultimately lurch apart. He quickly replays the film, and then selects the associated course module on harmonic motion. He browses the material (without doing the exercises), saves the film clip in his workbook with a speech annotation, and then enters a natural language query to find pointers to other physical manifestations of harmonic motion. He moves on to a case study involving flutes and piccolos.

Figure 1. A scenario of use for a multimedia educational system.
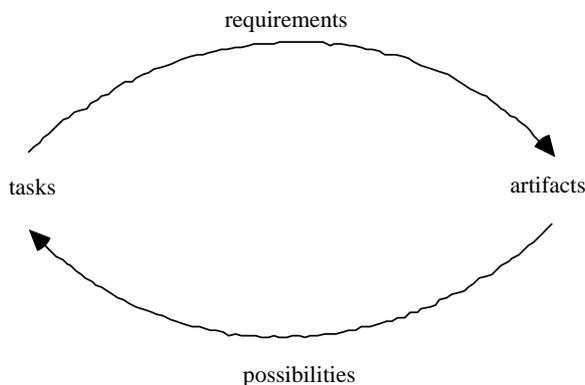
Figure 2. The task-artifact cycle. Tasks help articulate requirements for new technology that can support them; designed artifacts create possibilities (and limitations) that redefine tasks.

challenges (for example, how to get users to carry out practice exercises). In either case, the scenario guides the development of requirements and prototypes, and the specification of system functions. Subsequently, it could guide the development of training and documentation, and the design of evaluation tasks.

The ability of scenarios to depict vividly both current and future situations has encouraged their use as working representations in a wide variety of disciplines, from strategic management and operations research to requirements engineering, object-oriented design, and human-computer interaction. In the main, work and technology co-evolve: people engage in tasks and activities. They make discoveries and encounter difficulties; they experience insight and satisfaction, frustration and failure. At length, their tasks and experiences help to define requirements for future technology. Subsequent technological tools, in turn, open up new possibilities for tasks and activities, new ways to do familiar things, and entirely new things to do. They entrain new skills to learn and perform, new opportunities for insight and satisfaction, and new pitfalls of frustration and failure. Ultimately, the new tasks and new technology become a new baseline, helping to evoke requirements for further technology development. Scenarios have been proposed as an integrative representation of work and technology for managing this pattern.

2. WHY SCENARIO-BASED DESIGN IS IMPORTANT

Scenario-based design addresses five key technical challenges in the design of technology: Scenarios evoke reflection in the content of design work, helping developers coordinate design action and reflection. Scenarios are at once concrete and flexible, helping developers manage the fluidity of design situations. Scenarios promote work-oriented communication among stakeholders, helping to make design activities more accessible to the great variety of expertise that contributes to design, and addressing the challenge that external constraints designers and clients often distract attention from the needs and concerns of the people who will use the technology. Scenarios afford multiple views of an interaction, diverse kinds and amounts of detailing, helping developers manage the many consequences entailed by any given design move. Finally, scenarios can be abstracted and categorized, helping designers to recognize, capture, and reuse generalizations, and to address the challenge that technical knowledge often lags the needs of technical design.

2.1. Scenarios evoke reflection in design

Constructing scenarios of use as focal design objects inescapably evokes reflection in the context of doing design. The scenario in figure 1 succinctly and concretely conveys a vision of student-directed, multimedia instruction. It is a coherent and concrete vision, not an abstract goal, not a list of features and functions. Elements of the envisioned system appear in the scenario embedded in the interactions that will make them meaningful to people using the system to achieve real goals — perhaps revelatory, perhaps cryptic, but definitely more than just technological capabilities. For example, the role of natural language query is exemplified as a means of locating further case studies that illustrate the principles of harmonic motion.

The scenario emphasizes and explores goals that a person might adopt and pursue, such as watching the film clips twice,

or skipping the exercises. Some of these goals are opportunistic, such as investigating the Tacoma Narrows collapse because of experience with a totally unrelated bridge collapse, or deciding to branch from bridge failures to flutes. The scenario implicitly articulates the usage situation from multiple perspectives: the student stores and annotates a video clip with speech, raising specific requirements for user interface tools and presentation as well as for particular data structures and memory. The scenario impels the designer to integrate the consideration of such system requirements with consideration of the motivational and cognitive issues in education that underlie the person's actions and experiences.

2.2. Scenarios are at once concrete and flexible

Scenarios reconcile concreteness and flexibility. They are concrete in the sense that they simultaneously fix an interpretation of the design situation and offer a specific solution: the scenario in figure 1 specifies a particular usage experience that could be prototyped and tested. At the same time, scenarios are flexible in the sense that they are deliberately incomplete and easily revised or elaborated: in a few minutes, a piece of the scenario could be re-written (for example, stipulating a system-initiated prompt for the associated course module on harmonic motion) or extended (for example, the objects in the scenario could be described as anchors for a variety of link types).

Scenarios embody concrete design actions and evoke concrete move testing on the part of designers. They allow designers to try things out and get directive feedback. But their flexibility facilitates innovative and open-ended exploration of design requirements and possibilities, helping designers to avoid premature commitment.

The power of scenarios to convey concreteness through tentative and sketchy descriptions derives from the way people create and understand stories. Like the strokes of an expressionist painting, scenarios evoke much more than they literally present. The human mind seems especially adept at overloading meaning in narrative structures, both in generation and interpretation, as illustrated by the remarkable examples of dreams and myths. Indeed, dreams and myths are universal tools for coping with uncertainty. Part of their power stems from being recalled and discussed throughout a community. In a more modest scale, design teams do this too: sharing and developing scenarios helps to control the uncertainties inherent in the work, while strengthening the team itself.

2.3. Scenarios promote work orientation

Scenarios are work-oriented design objects. They describe systems in terms of the work that people will try to do as they make use of those systems. A design process in which scenarios are employed as a focal representation will *ipso facto* remain focused on the needs and concerns of prospective users. Thus, designers and clients are less likely to be captured by inappropriate gadgets and gizmos, or to settle for routine technological solutions, when their discussions and visions are couched in the language of scenarios, that is, less likely than they might be if their design work is couched in the language of functional specifications.

Scenarios help to integrate the variety of skills and experience required in a design project by making it easier for different kinds of experts to communicate and collaborate. They support the direct participation of users and other client stakeholders, helping to anticipate organizational impacts of new technology. Scenarios also ease problems associated with handoffs and coordination in the design process by providing all groups with a guiding vision of the project goal to unify and contextualize the documents and partial results that are passed back and forth.

Scenarios directly support the development of summative evaluation tasks. But perhaps more importantly, maintaining work-oriented scenarios throughout the development process allows formative evaluation walkthroughs to be carried out continuously.

2.4. Scenarios have many views

Scenarios are multifarious design objects; they describe designs at multiple levels of detail and with respect to multiple perspectives. A scenario briefly sketches tasks without committing to details of precisely *how* the tasks will be carried out or *how* the system will enable the functionality for those tasks. The multimedia education scenario in Figure 1 is at an intermediate level, with some detail regarding task flow; it could be elaborated with respect to Harry's moment-to-moment thoughts and experiences in order to provide a more elaborated cognitive view, or with respect to individual interactions to provide a more detailed functional view. Alternatively, the scenario could be presented from the point of view of someone watching Harry, perhaps in order to learn how to operate the system. It could be elaborated in terms of hardware and software components that could implement the envisioned functionality in order to provide a system view. Each of these variations in resolution and perspective is a permutation of a single underlying use scenario.

Scenarios can leave implicit the underlying causal relationships among the entities in a situation of use. For example, in Figure 1 the envisioned speech annotation capability allows adding a personal comment without the overheads of opening an editor and typing text. However, the annotation is noncoded, and thus cannot be edited symbolically. These relationships are important to the scenario, but often it is enough to imply them. This is an aspect of the property of being concrete but rough that we discussed above.

Sometimes it is useful to make these relationships explicit. For example, in another scenario Harry may wish to collect and revisit the set of film clips he viewed and annotated as "reakthroughs in forensic engineering". Unfortunately, his noncoded voice annotations cannot be searched by string. Thus, this new scenario would end in failure. To understand and address the variety of desirable and undesirable consequences of the original annotation design move, the designer might want to make explicit the relevant causal relationships in these two scenarios. Doing so provides yet another view of the envisioned situations, as shown in Figure 3.

Scenarios help designers manage trade-offs. For example, the data structures for the workbook might differentiate between annotated and non-annotated items, allowing annotated items to be retrieved and browsed as a subset. This would not allow Harry to directly retrieve the set of items with a particular annotation, but it would still simplify the search. Alternatively, the search problem might be addressed directly by speech recognition or audio matching, or by including the option of

A video clip of the Tacoma Narrows Bridge collapse
    provides an engaging introduction to a course module on harmonic
        motion
    evokes curiosity and self-initiated learner exploration
    but may not sufficiently motivate a learners to attempt the course
        module
Speech annotation of a saved video clip
    allows easy attachment of personal metadata
    but does not support indexing or search of metadata

Figure 3. Consequences associated with the orientational video clip and the speech annotation capability in the multimedia education scenario.

text annotation. Each of these alternatives would entrain different elaborations for both the annotation scenario in Figure 1 and the search scenario discussed immediately above. These elaborations could then be explored for further consequences and interdependencies.

2.5. Scenarios can be abstracted and categorized

Scenarios exemplify particular themes and concerns in work and activity situations. Earlier we discussed two scenarios for a multimedia education system. In one (Figure 1), a person is at first opportunistically exploring an information structure, but eventually adopts a particular interest that guides his exploration. In the other, a person wishes to search and organize information that has previously been browsed. Described at this level of generality, these are not scenarios unique to multimedia education systems, or even to computers. They are general patterns for how people work with information. Therefore, it is likely that some of the lessons learned in managing the "opportunistic exploration" pattern or the "searching under a description" pattern in the design of any given situation might be applicable in the subsequent design of other situations. Such a taxonomy of scenarios provides a framework for developing technical design knowledge.

Scenarios can also be classified in terms of the causal relations they comprise. In Figure 1, for example, providing speech annotation simplifies the actions needed to personalize a piece of information. In this causal relation, the consequence is the simplification of organizing and categorizing — a general desideratum in designing interactive systems. Generalizing the relation in this way allows the feature associated with the consequence (in this example, speech annotation) to be understood as a potential means for that consequence, and employed to that end in other design contexts. There is of course no guarantee that the generalization is correct; that can only be settled by trying to use it and succeeding or failing. The point is that such candidate generalizations can be developed from scenario descriptions.

The generalization of the causal relations comprising scenarios can also be carried out across features: speech annotation of data helps people to create a personalized view of their information. But this relation holds independent of whether the data is annotated by speech, by text, or by handwriting. Understanding the relation more generally allows designers to consider any medium for annotation as a potential means of facilitating a personalized data view.

Scenarios can also be taken as exemplars of model scenarios; for example, Figure 1 illustrates a model of opportunistic control.

Harry pursues the link from bridges to piccolos, because that is the aspect of the information that interests him. The system was designed to support this style of use; to that extent it embodies a model of opportunistic control. Other models are possible of course; many instructional systems would require a student to complete the current module before allowing a branch to related material in some other module. Seeing the scenario in Figure 1 as an opportunistic control scenario allows the designer to benefit from prior knowledge pertaining to this model and to contribute further design knowledge of the model based on the current project.

3. WHERE SCENARIO-BASED DESIGN IS GOING

Scenario-based design provides a framework for managing the flow of design activity and information in the task–artifact cycle. Scenarios evoke task-oriented reflection in design work; they make human activity the starting point and the standard for design work. Scenarios help designers identify and develop correct problem requirements by being at once concrete and flexible. They help designers to see their work as artifacts-in-use, and through this focus to manage external constraints in the design process. Scenarios help designers analyze the varied possibilities afforded by their designs through many alternative views of usage situations. And scenarios help designers cumulate their knowledge and experience in a rubric of task-oriented abstractions. This is depicted in Figure 4.

In current practice, scenario-based design techniques are widely used in human–computer interaction. Many new methods for capturing, generating, and classifying scenarios are emerging. For example, scenarios can be directly negotiated with end-users (participatory design), generated by systematically instantiating the signature phenomena of theories of human activity (theory-based design), or by various heuristic brainstorming techniques, like question generation and analogy. Scenarios can be applied in a wide variety of ways in system development. Besides their direct roles as descriptions of current or envisioned system use, they can be employed as bounding contexts for developing design rationales, for theories of human–computer interaction, and for software development (e.g. as use cases).
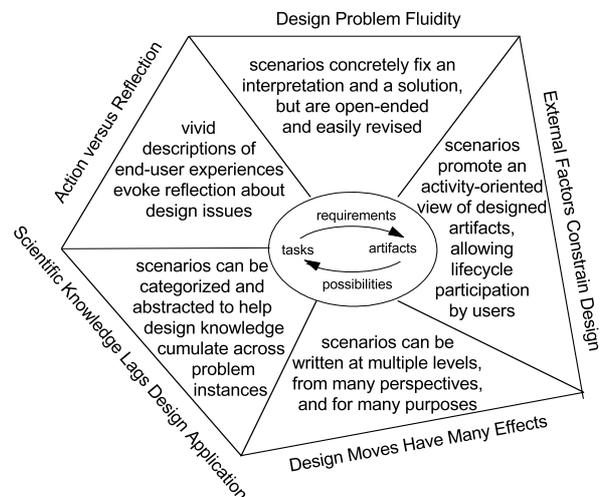
Figure 4. Challenges and approaches in scenario-based design.

A key challenge is to leverage this wide variety of scenario-based activity more systematically within the development process. The most promising recent developments involve closer integration of practice in human–computer interaction with the development of standard schemata, and support for libraries and reuse frameworks developed in requirements engineering.

REFERENCES

CARROLL, J.M. (ed.), 1995, *Scenario-based Design: Envisioning Work and Technology in System Development.* New York: Wiley.

CARROLL, J.M., 1999, *Making use: Scenario-based Design of Human-Computer Interactions.* Cambridge, MA: MIT Press.

CARROLL, J.M. and ROSSON, M.B., 1992, Getting around the task-artifact framework: how to make claims and design by scenario. *ACM Transactions on Information Systems,* 10(2)*,* 181–212.

JARKE, M., BUI, X.T. and CARROLL, J.M., 1998, Scenario management: an interdisciplinary approach. *Requirements Engineering,* 3(3–4), 155–73.

SUTCLIFFE, A.G. and CARROLL, J.M., 1999, Designing claims for reuse in interactive systems. *International Journal of Human-Computer Studies* 50(3), 213–41.

South Africa: Ergonomics Society of South Africa (ESSA)

P.A. Scott and J. Charteris
Department of Human Kinentics and Ergonomics, Rhodes University, PO Box 94, Grahamstown 6140, South Africa

The informal beginnings of ergonomics in South Africa, as in many countries, included both academic and professional endeavors of great relevance. The contributions of early luminaries cannot be documented without risking omission of major contributors, many of whom never used, or may not even have known, the term "ergonomics." Among these, Wyndham and Strydom, applied physiologists working in the human performance laboratories of the Chamber of Mines, stand out as international leaders in the field of environmental adaptations whose work was dedicated to ameliorating working conditions in the world's deepest mines.

Not surprisingly, then, it was at the Human Resources Research Laboratory of the Chamber of Mines that preliminary meetings were held in mid-March 1983 to discuss the formalization of Ergonomics in South Africa. A small group of interested individuals took tentative steps towards establishment of an organization for putative ergonomists. In February of the following year a group met at the Design Institute of the South African Bureau of Standards in Pretoria. Thirty participants from academic institutions, governmental bodies and industry offered insights. A study group comprising mainly of academics and businessmen was set up with Brian Hill acting as secretary and Mike Cooke as draftsman of a Constitution.

In February 1985 the Ergonomics Society of Southern Africa was formally inaugurated at the Council for Scientific and Industrial Research Conference Centre in Pretoria on the occasion of the Society's First Conference "Ergonomics '85." The Keynote Speaker, Stuart Kirk from the UK, leant stature to this early attempt to obtain international recognition in troubled socio-political times. Over 100 delegates attended and about 40 papers were presented. The first AGM was held and the society's first Council was instituted with Tony Golding being the first Chairperson of the society.

For the Second Conference, "esa '86" in April 1986, Cape Town was the venue and although the number of attending delegates drop, about 40 papers were presented covering diverse aspects of ergonomics.

From 1987 to 1994 the fluctuating fortunes of the Society mirrored the political turmoil in the country. Without fanfare, however, the society dug-in and survived. In 1987 "Ergonomics SA '87" was staged in Pretoria, again at the CSIR, its theme, "Consumer Ergonomics." No conference was held in 1988 but a small symposium attracted 30 delegates to the theme "Aspects of Work Stress" in Johannesburg, where 10 papers were read. In May 1989 "ergos '89" was the Society Conference, the venue being the University of the Witwatersrand, Johannesburg.

The next year it was decided to establish the society's own journal, and the first of *ergonomics: SA* was published in 1989. Although the articles have been primarily from local contributors, there have been international authors and the Editorial Board includes international referees. The present emphasis of the journal is to encourage articles on applied ergonomics, specifically based on research conducted in industrially developing countries. From 1989 to 1994 membership dropped to a low of 20. In 1994 under revived administration, taking advantage of a revitalized national milieu, the Fifth Conference "esa '94" was held in Grahamstown. Hal Hendrick, distinguished President of the IEA, was keynote speaker at a poorly attended conference where about 20 papers were read. Jack Charteris was elected Chair and a new look, leaner society with a smaller Council was elected. At this meeting it was agreed to focus on establishing a more solid society within the country and to modify the name of the society to the Ergonomics Society of South Africa (ESSA). The Society's flagging fortunes began to turn around, due particularly to the eventual acceptance by the IEA of ESSA as a Federated Member Society. Pat Scott was elected to represent ESSA at the joint IEA/South American Conference in Rio de Janeiro in 1995. The years 1996–98 saw a steady increase in ESSA membership. Houshang Shahnavaz and Pat Scott ran five successful Roving Seminars to introduce the new ILO/IEA "Ergonomics Checkpoints" and in 1996 the Sixth Conference, "esa '96" was staged in conjunction with the National Occupational Safety and Health Association's Convention NOSHCON '96 at the World Trade Centre in Johannesburg. African ILO representative Joshua Nkurlu delivered the Keynote Address.

In 1998 the society's crowning achievement to date, an international conference under the auspices of the IEA, was held in Cape Town. Some 300 delegates from 41 countries attended this conference, the largest and most prestigious gathering of international ergonomists in the history of Africa. The collated papers delivered at this conference, numbering over 100, were published in monograph form in the volume *Global Ergonomics* (Elsevier, 1998). The society has established criteria for, and given effect to, recognition of professional expertise in ergonomics. To date, nine of its members have been accorded the status of "Registered Ergonomist" in recognition of expertise in, service to, and professional consultative contributions in respect of, South African ergonomics.

Although most of the interest in Ergonomics in South Africa has been in the realms of academe with substantial research being conducted and published internationally, over the last decade of the century there has been a significant increase in an awareness of the benefits Ergonomics within the working environment. This has resulted in the establishment of a number of ergonomics consultancies throughout the country. South Africa, as with all industrially developing countries, needs Ergonomic input and the potential for growth in Ergonomic research and application is enormous, making it a challenging and rewarding field to work in.

Spain: Spanish Ergonomics Association

P. R. Mondelo

Dept. Proyectos de Ingeniería, Institut d'Organització i Control de Sistemes Industrials (UPC), Av. Diagonal, 647, Planta 10, E-08028 Barcelona, Spain

The Asociación Española de Ergonomía (AEE)/Spanish Ergonomics Association is a non-profit organization created in 1988 and is a federated member of the International Ergonomics Association (IEA). The aims of the AEE are:

- to link together all Spanish institutions and individuals working, teaching and researching in the fields of ergonomics;
- to promote ergonomics in Spanish companies and various governmental bodies; and
- to assess, according to the IEA standards, existing or projected ergonomics-related activities within Spain.

To reach these goals, the AEE strives:

- to promote and strengthen the role of Spanish ergonomists at a socio-professional level;
- to develop training curricula in ergonomics for university students (undergraduate, graduate, postgraduate levels), and for the working professionals in the fields relevant to ergonomics (i.e. occupational safety and health);
- to publish and disseminate relevant educational materials and research results; and
- to organize ergonomics-related events, gatherings and presenting awards (like the Annual Preemie National de Ergonomía/National Ergonomics Award).

As an example of current activities, the AEE is today working on the organization of a Federation of Spanish-speaking ergonomics societies that will strengthen ties between Latin America and Europe.

Symvatology: The Science of an Artifact–Human Compatibility

W. Karwowski

Center for Industrial Ergonomics, Lutz Hall, University of Louisville, Louisville, KY 40292, USA

> Theories can only be disproved, they can never be proved.
>
> Sir Karl Popper (1902–95)

1. INTRODUCTION

Ergonomics (*ergon* + *nomos*) or the study of work, as originally proposed by W. Jastrzebowski (1857), defines a discipline with very broad scope, i.e. the wide subject of interests and applications, encompassing all aspects of human activity, including labor, entertainment, reasoning, and dedication. Jastrzebowski divides work into two main categories: the useful work, which brings improvement for the common good, and the harmful work that brings deterioration (discreditable work). Useful work, which aims to improve things and people, is classified into physical, aesthetic, rational, and moral work. According to Jastrzebowski, such work requires utilization of the motor forces, sensory forces, forces of reason (thinking and reasoning), and the spiritual force. The four main benefits of the useful work are exemplified through the property, ability, perfection, and felicity.

Contemporary ergonomics, as proposed by Murrell in 1949 (Edholm and Murrell 1973), is considered an applied science, the technology, or both. Furthermore, ergonomics in the recent years has often been politicized and, as a consequence, is trivialized, at least by some, by projecting its solutions onto a common sense, and by pushing towards proving its immediate usefulness in practice. It should be noted here that practicality is a noble goal (Kotarbinski 1965), and is the subject of praxiology, or the "science of efficient actions," developed by Polish philosopher Tadeusz Kotarbinski. However, given the weak methodological basis of ergonomics, its practical applications often cannot be justified due to lack of objective measures of results.

Owing to its broad scope, contemporary ergonomics also lacks identity, and is often perceived by outsiders as a "discipline of everything that is related to the working environment and which in any way affect the human well-being." That is why the physical therapists, occupational physicians, experimental or cognitive psychologists, industrial hygienists, safety professionals or industrial engineers, have as much to say about ergonomics as the "true" ergonomists, i.e. those who have gained their professional status based on relevant education and training. The former category can be called the "ergonomically oriented professionals," while the latter category is the ergonomists.

2. COMPATIBILITY AND ERGONOMICS

According to the Human Factors and Ergonomics Society (HFES 1999), ergonomics advocates systematic use of the knowledge concerning relevant human characteristics in order to achieve compatibility in the design of interactive systems of people, machines, environments, and devices of all kinds to ensure specific goals. Typically such goals are: improved (system) effectiveness, safety, ease of performance, and the contribution to overall human well-being.

Although a key to the above definition of the HFES is the word "compatibility," in the human factors circles the term compatibility is mainly used in a narrow sense only, and is typically related to the problem of information processing. Specifically, the term compatibility is used in relation to design of displays and controls, such as the classical spatial (location) compatibility, or the intention–response–stimulus compatibility related to movement of controls (Wickens and Carswell 1987).

The use of compatibility in a greater context of ergonomics systems has been advocated by Karwowski and co-workers (Karwowski *et al.* 1985, 1988, Karwowski 1991). Recently, Karwowski introduced the term "human-compatible systems" (Karwowski 1997) to focus on the need for comprehensive treatment of compatibility in the human factors discipline.

The *American Heritage Dictionary Of English Language* (1978) defines "compatible" as: (1) capable of living or performing in harmonious, agreeable, or congenial combination with another or others; and (2) capable of orderly, efficient integration and operation with other elements in a system. Probing further, the word "congenial" is defined as "suited to one's needs, agreeable," while the "harmony" (form Greek *harmos*, joint) is defined as: (1) agreement in feeling, approach, action disposition, or the like, sympathy, accord; and (2) the pleasing interaction or appropriate combination of elements in a whole.

From the beginning of contemporary ergonomics, the measurements of the actual or intended compatibility between the system and the human, and the results of ergonomics interventions were based on those measures that best suited different ergonomically oriented professionals. The measures that have typically been used (both negative and positive) include the specific psychophysiological responses of the human body (for example heart rate, EMG, perceived human exertion, satisfaction, comfort or discomfort), as well as a number of indirect measures, such as the incidence of injury, economic losses or gains, system acceptance, or operational effectiveness, quality or productivity. The ergonomists have accepted that status quo for measurements, as they have not developed a unique measure of the system–human compatibility themselves. This lack of the universal way to quantify and measure compatibility is the very essence of ergonomics identity crisis (Karwowski 1998). That is perhaps, or at least in part, the reason why ergonomics is still perceived today by some (Howell 1986) as highly unpredictable area of human scientific endeavor, with great uncertainty about its future as an independent and unique discipline, if such a status is attainable at all.

3. AN ARTIFACT SYSTEM AND THE ARTIFACT–HUMAN COMPATIBILITY

To introduce the concepts of human-compatible systems, instead of traditional use of the name: "human–machine system," it is proposed here to use the term of an "artifact–human system," resulting from combining two of its integral parts, i.e. the artifact system and the human system, as follows:

artifact system + human system = "artifact–human system."

An "artifact system" is defined as a set of all artifacts (meaning objects made by human work), as well as natural elements of the environment, and their interactions occurring in time and space afforded to us by nature (or God). Owing to its interactions, the artifact system is often a dynamic system with high level of complexity, and it exhibits a nonlinear behavior. A human system is defined as the human (or humans) with all the characteristics (physical, perceptual, cognitive, emotional, etc.), which are relevant to an interaction with the artifact system.

Consequently, instead of the: "human–machine compatibility," it is proposed here to use the "artifact system-to-human system compatibility," or, in short, the artifact–human system compatibility. This notation expresses the premise that it is the artifact system that should be compatible with the human system, and that such compatibility should primarily be assured by design of the artifact system.

It should also be noted that an artifact system and human system are subject to mutual adaptation. In order to survive, biological systems must be adaptable, that is capable of functioning in an uncertain environment (Conrad 1983). The same applies to the artifact system. Adaptation means here an alteration or adjustment by which an individual or an artifact improves its condition in relationship to its environment. Adaptation implies such changes that allow for a system to become suitable to a new or special use or situation. This notion admits that optimal compatibility may be achieved by adaptation (primarily design or learning) of both systems.

4. SYMVATOLOGY

It is clear from the above discussion that we need to develop a new science that will aim to discover laws of the artifact–human compatibility, propose theories of the artifact–human compatibility, and develop quantitative matrix for measurement of such compatibility. Only then, when such measure(s) are available and used to objectively demonstrate the value of ergonomics discipline to the outside world, there can be hope for ergonomics to gain its claim for the independence and uniqueness among other sciences. Shortly, ergonomics needs help from another science that is yet to be developed. This new science is called the symvatology.

Symvatology, is coined by using two Greek words: *sumbatotis* = *symvatotis* (compatibility) and *logos* = *logos* (logic, or reasoning about). Specifically, the new word is the result of joining the "symvato-" (from symvatosis) and "logy" (a combining form denoting the science of).

Symvatology is proposed here as the science of the artifact–human (system) compatibility. As such, symvatology will be the corroborative discipline to ergonomics, in that it will help to build solid foundations of the science of ergonomics.

To optimize both the human and system well-being and their performance, the system–human compatibility should be considered at all levels, including the physical, perceptual, cognitive, emotional, social, organizational, environmental, etc. This requires a way to measure the inputs and outputs that characterize the set of system–human interactions (Karwowski 1991). And yet, we do not know today any more than Jastrzebowski knew when he proposed the science of work, how

to measure the system–human compatibility. Without the ability to identify, evaluate and measure the artifact–human compatibility one cannot claim to improve the system–human well-being and performance. Ironically, today we do not have a matrix for measurement of such compatibility, which is critical to credibility of ergonomics as a discipline.

Furthermore, the goal to measure the artifact–human compatibility can only be realized if we understand its nature. We need to discover and understand the laws governing the compatibility relations between the people and the natural and artificial systems that surrounds them. We need to develop a discipline that will aim to observe, identify, describe, perform empirical investigations, and produce theoretical explanations of the natural phenomena of the artifact–human compatibility. In short, as the above are the elements of any science, we need to develop the new science of artifact–human compatibility.

Given below is the discussion of some critical concepts of this new science that must be further developed and used in order to drive a universal measure of compatibility for evaluation of the artifact–human systems. These concepts include system complexity, incompatibility, entropy, and system regulation.

5. SYSTEM COMPLEXITY

The *American Heritage Dictionary of English Language* (1978) defines "complex" as consisting of interconnected or interwoven parts. Based on our earlier work (Karwowski *et al.* 1988, 1995), the artifact–human system (S) can be represented as a construct which contains the human subsystem (H), an artifact subsystem (A), an environmental subsystem (E), and a set of interactions (I) occurring between different elements of these subsystems over time (t). Further discussion on this subject can be found in (Karwowski 1995).

The set I is viewed here as a set of all possible interactions between people, artifacts, and various environments that are present in a given state of the system. These interactions reflect the existence (or non-existence) of the relationships between the subset of all relevant human characteristics (H), such as anatomical, physiological, biomechanical, or psychological, the subset of characteristics of the artifact subsystem A, and the elements of N, representing the subset of environmental conditions (physical environment, social support, organizational structure, etc.). These interactions also impact the system complexity.

6. MEASUREMENT OF THE ARTIFACT–HUMAN SYSTEM COMPATIBILITY

The compatibility is a dynamic, natural phenomenon affected by the artifact–human system structure, its inherent complexity, and its entropy or the irreducible level of incompatibility between the system's elements (Karwowski 1995).

The structure of system interactions (I) determines the complexity and related compatibility relationships in a given system. Therefore, compatibility should always be considered in relation to the system's complexity. The complexity–compatibility paradigm representation for the artifact–human system is shown in Figure 1.

The system space, denoted here as an ordered set [(complexity, compatibility)] are defined by the four pairs as follows [(high, high), (high, low), (low, high), (low, low)]. Under

COMPATIBILITY	LOW, HIGH ——————— LOW, LOW	HIGH, HIGH ——————— HIGH, LOW

COMPLEXITY

Figure 1. The compatibility-complexity paradigm representation for the artifact system.

the best scenario, i.e. under the most optimal state of system design, the artifact–human system exhibits high compatibility and low complexity levels. It should be noted that the transition from high to low level of system complexity does not necessarily lead to improved (higher) level of system compatibility. Also, it is often the case in most of the artifact–human systems, that an improved (higher) system's compatibility can only be achieved at the expense of increasing system's complexity.

7. COMPLEXITY–INCOMPATIBILITY PRINCIPLE

As discussed by Karwowski *et al.* (1985), the lack of compatibility or ergonomic incompatibility (EI), defined as degradation (disintegration) of the artifact–human system, is reflected in the system's measurable inefficiency and associated human losses. In order to express the innate relationship between the system's complexity and compatibility, Karwowski *et al.* (1988, 1994) proposed the Complexity–Incompatibility Principle, which can be stated as follows: "As the (artifact–human) system complexity increases, the incompatibility between the system elements, as expressed through their ergonomic interactions at all system levels, also increases, leading to greater ergonomic (non-reducible) entropy of the system, and decreasing the potential for effective ergonomic intervention." The principle was illustrated by Karwowski (1995), using as an example of design of the chair and the design of a computer display, two common problems in the area of human–computer interaction. In addition, Karwowski (1996) discussed the complexity–compatibility paradigm in the context of organizational design

It should be noted that the above principle reflects the natural phenomena that others in the field have described in terms of difficulties encountered in human interacting with consumer products and technology in general. For example, according to Norman (1989), the paradox of technology is that added functionality to an artifact typically comes with the trade-off of increased complexity. These added complexities often lead to increase human difficulty and frustration when interacting with these artifacts. One of the reasons for the above is that technology which has more features have also less feedback. Moreover, Norman argues that the added complexity cannot be avoided when functions are added, and can only be minimized with good design that follows natural mapping between the system elements (i.e. the control–display compatibility).

8. ENTROPY

Entropy, related to the probability of an order/disorder, is s useful

concept for analysis of the artifact–human systems, as changes in entropy (disorder) can be measured. The entropy of the universe, which is an isolated system, is continually increasing, that is, the disorder of the universe is increasing. Natural processes spontaneously move from the more ordered to the less ordered (more probable) state. However, the low entropy (or well-ordered systems) can be found in an open or closed system.

As discussed by Silver (1998), life maintains low entropy against the universe tendency to higher level entropy. Furthermore, so called dissipative structures are capable of maintaining their own entropy constant, but at the expense of increased entropy of the surrounding environment (or its part), leading to an increase in the total entropy of the whole system.

Of great importance to the artifact–human systems is that the dynamic equilibrium is more ordered and, therefore, less probable than the normal state. This is because the system in the sate of dynamic equilibrium has lower entropy. Depending on how widely open or isolated the system is, the maintenance of the lower entropy level over time may be more or less difficult.

The earth is an open system. It allows for the exchange of energy and matter. The ordered state survives as long as either energy or material is continuously fed to them. To maintain their equilibrium, living organisms (human system) need an input of material and energy.

Isolated systems are closed to the transfer of matter or heat. Such systems move toward states of higher entropy (lower order) and stop changing when they reach an equilibrium state, characterized by the maximum entropy. This does not preclude local islands of decreasing entropy, but the overall entropy of the closed system (like the universe) always increases (Silver 1998).

9. ENTROPY OF THE ARTIFACT–HUMAN SYSTEMS

The artifact–human systems can be an open, isolated or closed, depending on their purpose, structure, and dynamic state of interactions with the outside environment, which is not always intentional, or where the intentions change. As such, these systems will show changing levels of entropy (or order) over time.

It should be noted here that a change (increase) in incompatibility, which we will denote here as (change in) an ergonomic disorder, that is exhibited by many artifact–human systems, is natural. Such a change follows nature by tending to increase the disorder or increase the system entropy as much as possible. Incompatibility is considered an "attractor," that means a state or phenomenon to which the system will gravitate whatever the initial conditions, or a steady state point at which the system starts (if poorly design to begin with) and where it remains.

As discussed by Karwowski (1995), the entropy of the artifact–human system (S) (system entropy: E(S)) will be modeled using the entropies (E) due to the human-related interactions (H-subsystem), artifact-related interactions (A-subsystem), environment-related interactions (N-subsystem), and time (T). In view of the above, the system entropy E(S) can be defined as a function of entropy of the set of ergonomic interactions E(I) and time (T) as:

$$E(S) = f[E(I), T],$$

where E(S) is an expression of the total system incompatibility with the human.

Since the set of ergonomic interactions (I-subsystem) consists of the possible relations between the subsystem elements of {H}, {A}, and {N}, the E(I), can be expressed as:

E(I) = {E(H), E(A), E(N)},

where E(H) is the contributing entropy due to human subsystem, E(A) is the contributing entropy due to artifact subsystem, E(N) is the contributing entropy due to environmental subsystem.

It should be noted that E(H), or the contributing entropy due to human subsystem, is interpreted here as "the entropy due to natural human limitations, which adversely affect the compatibility of an artifact system with the human characteristics defined by a subset (H)."

The entropy of an artifact–human system (S) depends on the variety and complexity of all relevant interactions for the subsystems H, A and N. Any manipulation of elements of the artifact–human system may result in subsequent changes in the number or structure (or both) of other interactions, and, therefore changes in system entropy.

10. COMPATIBILITY REQUIREMENTS AND SYSTEM REGULATION

In design of human-compatible systems, one must consider the structure of system interactions (I) at the given system level, which induces complexity of that system. It is the complexity of the {A} and {N} subsystem interactions that defines the ergonomic compatibility requirements (CR) of the system. In general, the greater the system compatibility requirements (CR), the greater the need for the ergonomic intervention efforts, called here the system regulation. Therefore, it follows that an entropy of the artifact–environment {A, N} system, or the system regulator, E(R), is also defined by the system's structure complexity.

It should be noted that the optimal (ergonomic) system has minimal ergonomic compatibility requirements, i.e. minimal system incompatibility. In other words, the optimally designed artifact–human system satisfies most (if not all) of the compatibility requirements.

The artifact (A) and environmental (N) subsystems can be regulated (at least to some degree), and are jointly called the system regulator (R). Consequently, the sum of their respective entropies defines the entropy of the system regulator E(R):

E(R) = E(A) + E(N), or equivalently

E(R) = E{(A) + (N)}, denoted as E(R) = E(A,N).

It follows from the above discussion that an entropy of the regulator E(R), also defines the system's compatibility requirements (CR). This is because at any given level of human entropy E(H), the E(R) is directly related to the artifact–human system entropy E(S), which is the outcome of such regulation.

The above can be illustrated as the process system regulation, with E(H) as the system input, E(R) as the system regulator, and the artifact–human system entropy E(S) as the output, with the regulation structure as:

$$E(R)$$
$$E(H) \quad E(S)$$

According to Ashby 's (1964) law of requisite variety, for the

system shown above, the outcome, i.e. the entropy E(S) can be defined as:

E(S) >= E(H) + E(H(R)) – E(R),

where E(H(R)) is the entropy of the regulator (R) when the state of human subsystem (H) is known.

When E(H(R)) = 0, i.e. when the regulator subsystem R={A, N} is a determinate function of the human subsystem (H) functioning under R, the above equation transforms as:

E(S) ≥ E(H) – E(R),

showing that system entropy can be defined as a difference between the human entropy and entropy of the regulator.

The above equation also indicates that system entropy E(S), if minimal at a given design stage (and given the specific value of the human entropy E(H)), can only be further reduced by increasing entropy of the regulator (R). In the context of ergonomics design, this means increasing the current system design complexity. On the other hand, this also implies a need for increasing the system's compatibility requirements (CR).

Following the Ashby's (1964) law of requisite variety, Karwowski (1995) proposed the corresponding law, called the "law of requisite complexity," which states that only design complexity can reduce system complexity. The above means that only added complexity of the regulator (R = re/design), expressed by the system compatibility requirements (CR), can be used to reduce the ergonomics system entropy (S), i.e. reduce the overall artifact–human system incompatibility.

It should be noted that the minimum value of the artifact–human system's entropy E(S)MIN is equal to the human entropy E(H), and occurs when the value of E(R) = 0, indicating the system state where no further system regulation is possible. This is called the ergonomic entropy of the system:

E(S)MIN ≥ E(H), when E(R)= 0.

As discussed by Karwowski *et al.* (1995), the ergonomic entropy E(S)MIN is the "non-reducible" level of system entropy with respect to ergonomic intervention (regulation) efforts. It should be noted, however, that this entropy can be modified by the non-ergonomic means, that is through the human, rather than system, adaptation efforts. This can be done by reducing the value of human entropy E(H), through improvements in human performance and reliability by training, education, motivation, etc.

11. CONCLUSIONS

An artifact system was defined as a set all artifacts and natural elements, and their interactions occurring in time and space afforded to us by nature, and subject to adaptation. The focus of the proposed new science: "symvatology" is the investigation of natural phenomena that characterize the artifact–human system compatibility. The purpose of such investigations is to provide the necessary design knowledge for development of the human-compatible systems (Karwowski 1997). The benefits of such systems have been advocated by the ergonomics/human factors discipline for years.

Symvatology should be a useful science in that it should help to advance the progress of ergonomics discipline by providing methodology for design for compatibility, as well as design of

compatibility between the artificial systems (technology) and the humans. In the above perspective, the goal of ergonomics should be to optimize both the human and system well-being and their mutually dependent performance. It is not enough to assure the well-being of the human, as one must also optimize the well-being of a system (i.e. the artifacts-based technology and nature) to make the proper uses of life (Hancock 1999).

ACKNOWLEDGEMENTS

REFERENCES

ASHBY, W.R., 1964. *An Introduction to Cybernetics* (London: Methuen).

CONRAD, M., 1983, *Adaptability* (New York: Plenum).

EDHOLM, O.G. and MURRELL, K.F.H., 1973, *The Ergonomics Society: A History 1949–1970* (London: Ergonomics Research Society).

HANCOCK, P., 1997, *Essays on the Future of Human–Machine Systems* (Eden Prairie: BANTA Information Services Group).

JAMALDIN, B. and KARWOWSKI, W., 1997, Quantification of human–system compatibility (HUSYC): an application to analysis of the Bhopal accident. In P. Seppala, T. Luopajarvi, C.-H. Nygard and M. Mattila (eds), *From Experience to Innovation: Proceedings of the 13th Triennial Congress of the International Ergonomics Association*, Tampere, Finland, vol. 3, 46–8.

KARWOWSKI, W., 1991, Complexity, fuzziness and ergonomic incompatibility issues in the control of dynamic work environments, *Ergonomics*, 34, 671–86.

KARWOWSKI, W., 1992, The human world of fuzziness, human entropy, and the need for general fuzzy systems theory, *Journal of Japan Society for Fuzzy Theory and Systems*, 4, 591–609.

KARWOWSKI, W., 1995, A general modeling framework for the human–computer interaction based on the principles of ergonomic compatibility requirements and human entropy. In A. Grieco, G. Molteni, E. Occhipinti and B. Piccoli (eds), *Work with Display Units 94* (Amsterdam: North-Holland), 473–8.

KARWOWSKI, W., 1997, Ancient wisdom and future technology: the old tradition and the new science of human factors/ergonomics. In *Proceedings of the Human Factors and Ergonomics Society 4th Annual Meeting* (Albuquerque, Human Factors and Ergonomics Society), 875–7.

KARWOWSKI, W. and JAMALDIN, B., 1995, The science of ergonomics: system interactions, entropy, and ergonomic compatibility measures. In A.C. Bittner, Jr and P.C. Champney (eds), *Advances in Industrial Ergonomics and Safety VII* (London: Taylor & Francis), 121–6.

KARWOWSKI, W., MAREK, T. and NOWOROL, C., 1988, Theoretical basis of the science of ergonomics. In *Proceedings of the 10th Congress of the International Ergonomics Association*, Sydney (London: Taylor & Francis), 756–8.

KARWOWSKI, W., MAREK, T. and NOWOROL, C., 1994, The complexity–incompatibility principle in the science of ergonomics. In F. Aghazadeh (ed.), *Advances in Industrial Ergonomics & Safety VI* (London: Taylor & Francis), 37–40.

KOTARBINSKI, T., 1965, *Praxiology: An Introduction to the Sciences of Efficient Action* (Oxford: Pergamon).

MORRIS, W. (ed.), 1978, *The American Heritage Dictionary of English Language* (Boston: Houghton Mifflin).

NORMAN, D., 1988. *The Design of Everyday Things* (New York: Doubleday).

SILVER, B.L. 1998, *The Ascent of Science* (Oxford: Oxford University Press).

WICKENS, C.D. and CARSWELL, M., 1997, Information Processing. In G. Salvendy (ed.), *Handbook of Human Factors & Ergonomics* (New York: Wiley), 89–1129.

Taiwan: Ergonomics Society of Taiwan

T.–Z. Lee

Ergonomics Society of Taiwan, Department of Industrial Management Science, National Cheng Kung University, Tainan 701, Taiwan

The Ergonomics Society of Taiwan (EST) was founded in February 1993 by more than 150 founding members. Compared with the Human Factor/Ergonomics Society of the USA or the Ergonomics Society of the UK, EST is relatively small and young. Yet, EST is eager to be part of the international ergonomic society.

The purposes of EST are:

- To promote the basic research and practical application of ergonomic knowledge and rules.
- To provide qualified ergonomic services.
- To enhance interaction with international ergonomic society.
- To serve its members by necessary means such as publishing a newsletter and a journal, by holding conferences.

For the past few years, EST has worked hard to serve these purposes. An annual meeting was held regularly with a heavy academic touch. Conferences and workshops were given to academia as well as practitioners. For those activities, EST has been working closely with the National Science Council, Institute of Occupational Safety and Health, and Ministry of Transportation of Taiwan, and of course, with various companies.

To serve the domestic ergonomic society better, the 'EST Bulletin' newsletter is published quarterly. The academic journal of EST *Journal of Ergonomic Study* was inaugurated in March 1999.

The interaction between EST and the international ergonomic society began with Swedish professors Shahnavaz and Winkel in 1993, and continued with scholars from China, France, Japan, Korea, Sweden and the USA. This has been especially because of the Sino-France Joint Conference, which has been held every 2 years in alternating countries.

In October 1995, EST was approved to be a Federated Member of International Ergonomic Association (IEA), and it has maintained its active role ever since. In November 1996, EST sponsored the 4th Pan-Pacific Conference on Occupational Ergonomics, which was an IEA-approved conference.

To keep up with the development of the Taiwanese economy and international trends, the EST will keep on growing and booming.

Task Analysis

L. Ainsworth

Synergy Consultants, Royal Oak Barn, Cumbria, LA11 6QB, UK

1. INTRODUCTION

Task analysis is the study of the tasks that have to be undertaken by individuals or groups in order to achieve particular system goals. Thus, it could be argued that task analysis is almost as old as tasks themselves and certainly the systematic investigations of manual tasks by Gilbreth and others in the early 1900s could be considered as task analyses. However, the term "task analysis" was first popularized within the psychological and ergonomics communities by R. B. Miller in the early 1950s (Miller 1953). Miller considered that all task investigations should start by breaking down a complex task into a series of sub-tasks or task elements to produce task descriptions, consisting of an action verb and a noun, that described each of these task elements. To Miller, task analysis referred to the subsequent methodical examination of these task descriptions. In order to ensure that all the important task features were considered during this process, Miller specifically obtained the following data for each task element:

- Cues initiating the action
- Controls used
- Decisions involved
- Typical errors
- Response required
- Criterion of acceptable performance
- Feedback of task completion

Subsequently, other workers in the field have modified Miller's original approach to suit their own particular requirements. This has meant that the split between task description and task analysis has become blurred, with most workers in the field taking the pragmatic view that task analysis encompasses all aspects of the investigation of task-based performance from data collection through to the analysis of those data.

2. THE PURPOSE OF TASK ANALYSIS

In its simplest form, task analysis merely provides a set of task descriptions that can be used to ascertain whether persons have been provided with the basic control and display facilities to enable the tasks to be undertaken, or to check the accuracy of any operating instructions. However, the real value from task analysis comes from more in-depth investigations that probe into the relationship between various aspects of the task and the psychological/physiological requirements of the person undertaking that task. For new systems this should ensure that all aspects of a task will be designed in a manner that assists the person to operate safely and efficiently, whilst for existing systems, task analyses should identify any task features where there is a potential for human errors or other problems.

There are several task analysis methods that can be used to obtain deeper insights into different aspects of task design and by selecting these carefully, it is possible to address a range of issues, which include:

- Functional allocation
- Design/assessment of interfaces or procedures
- Development/assessment of training
- Human resource planning
- Workload measurement
- Human reliability assessment
- Incident investigation

3. THE TASK ANALYSIS PROCESS

It is appropriate to consider task analysis as a three-stage process, which is preceded by careful planning and preparation, as depicted in figure 1. However, it must also be stressed that task analysis is a highly iterative process in which it is often necessary to amend the approach that is being taken, dependent upon the findings that are obtained.

The main task analysis methods that can be used in these three stages are mentioned briefly below and more detailed descriptions to most of them are given in Kirwan and Ainsworth (1992).

4. DATA COLLECTION

There are three ways by which task analysis data can be obtained; namely the exploitation of existing data, the acquisition of verbal reports, or the observation of task performance. The data sources must be selected carefully to ensure that the data are representative and to reduce bias. If possible, any findings should be verified by using alternative data sources.

Verbal reports are particularly useful and these range from relatively informal discussions about limited aspects of a task, through to more structured talk-throughs in which a person describes all the steps that would be undertaken in a more
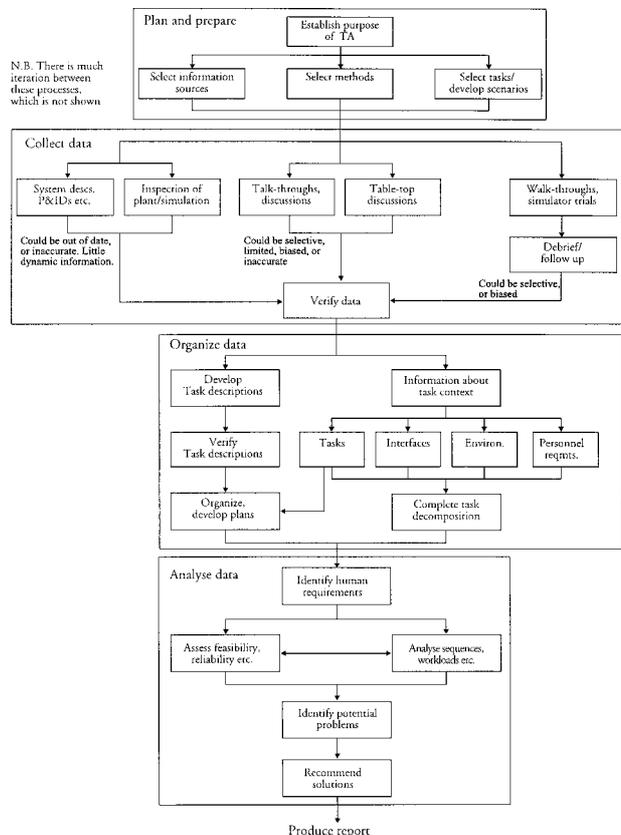
Figure 1.   Task analysis process.

complex scenario or operating procedure, but does not actually undertake any tasks. A slightly higher level of fidelity and realism can be achieved by undertaking a walk-through. This involves asking a person to describe particular tasks whilst walking around a system, or some representation of it, and pointing to the controls and information sources that would be used. The strong visual cues that are available during a walk-through generally mean that the reports can be much more complete and detailed. However, the most comprehensive source of task data will come from direct observations of simulator trials or actual operations.

Another useful source of data for task analyses is a table-top discussion. This is a structured discussion by a selected group of experts, in which an attempt is made to seek out the view of individuals with different technical backgrounds in order to better understanding some of the underlying issues and constraints.

5. DATA ORGANIZATION

During this stage the analyst develops task descriptions and then organizes these and categorizes other task information in ways that are appropriate for systems design or the consideration of human performance. During this process the analyst will apply ergonomics knowledge and past experience from previous analyses to start to develop insights and to identify where additional data may be necessary.

Usually, the main tasks that are being investigated are relatively complex. Therefore, in order to understand them, it is first necessary to break them down into a comprehensive set of smaller tasks that are each more tractable for the analyst to deal with. Thus, data organization begins by developing a set of task descriptions that each describes a specific subtask which is at an appropriate level of complexity for further analysis. These task descriptions can be written by working through the main task and identifying each step in turn, or alternatively hierarchical task analysis can be used.

Hierarchical task analysis (HTA) was originally developed by Annett and Duncan for determining training requirements and a good early account of this approach is given in Annett *et al*. (1971). HTA involves establishing the main goal of a task, then this goal is re-described in greater detail as a small, but all-encompassing set of subordinate goals. This process of re-description continues for these sub-goals until a point is reached beyond which further re-description is deemed by the analyst to be unnecessary. At this point the analyst will have a comprehensive set of task descriptions that have been organized hierarchically. A great advantage of developing the detailed task descriptions in this hierarchical manner is that it assists the analyst to ensure that no important subtasks have been missed, but it does not necessarily mean that the person undertaking the task will develop a mental model of the task that is based upon the same hierarchy. Therefore, if an analyst wishes to show how a set of subtasks might be undertaken, it is necessary to develop an organizing element, or "plan", that specifies the conditions under which each of the sub-goals might be carried out in order to achieve a higher level goal.

When an analyst has broken down a task into a comprehensive set of task descriptions, regardless of how this breakdown was achieved, it is necessary to obtain data for each subtask methodically about specific features that are of particular interest, such as the interfaces that are used, the skills that are required, or the alarms that are given. This process of obtaining specific additional information to supplement each task description is known as task decomposition and the task decomposition categories that are used must be carefully selected to enable an analyst to address the specific issues that are of concern.

Miller's "task analysis" in 1953 was actually a task decomposition in which he proposed that several specific items of information should be collected about each task element by recording particular aspects of task performance or the interfaces that were used. Many other aspects of the tasks can be used as decomposition categories and can be recorded methodically in the same way by directly looking at task performance or various features that are provided. However, decomposition information need not be limited to directly observable features, but can also include inferences about psychological mechanisms that might be made by an analyst, such as estimates of human error or comments about the potential workload.

6. DATA ANALYSIS

When the data associated with each task description have been organized, the analyst can start to identify how well the system supports the person(s) undertaking the tasks and fulfils the human needs that must be met in order to complete the tasks safely and effectively. This will entail first identifying the human requirements, with particular emphasis upon the psychological requirements, such as information processing, decision-making or the likely memory loads. In some cases the analyst will be able to make these judgements directly from the data that are available, but it may also be necessary to transform some of the data in order to identify the salient points.

There are many data analysis methods that can be used during task analysis and most of these have widespread application elsewhere. The approaches to some of the main issues are described briefly below.

6.1 Analysis of the working environment

Where physical factors are of concern, surveys of the environmental conditions or workplace surveys of key dimensions based upon anthropometric requirements, such as desk heights or display positions, can be useful.

6.2 Analysis of information requirements

An analysis of the information requirements can be made to assess the interfaces that are provided, or to identify the knowledge requirements of a task. Such an analysis should consider the information demands placed upon a person and their ability to understand and assimilate the information that they are given. It is also necessary to consider how faults and problems can be detected and diagnosed.

Various charting techniques are useful for examining the information requirements. Where it is necessary to identify the information and control requirements for particular tasks and to establish the interrelationships between tasks, information flow diagrams can be useful. These are also known as decision–action diagrams, because they present complex task sequences graphically in terms of the actions and decisions that are necessary. At the decision elements, these analyses consider all the possible

decisions that may be made, regardless of whether they are correct or appropriate.

Another charting technique that can be useful is a functional flow diagram. These are constructed by identifying the functions and sub-functions that have to be performed and then arranging these as a sequence of blocks that are linked together, using logical gates where appropriate, to indicate the interrelationships between them.

6.3   Analysis of task sequence

Task sequences can be examined in terms of the instruments/ information sources that are used, or communications links, by link analysis. Link analysis was originally used to examine communications or interactions between team members by representing each person on a schematic diagram and then drawing a line between these whenever there is some communication or interaction between particular individuals. This method, which is also known as a frequency link analysis, can be adapted to examine the relationships between different parts of an interface by drawing links between interface items whenever the focus of attention changes. However, this only shows the frequency with which particular interface items are linked, and does not give any information about task sequence.

If information about task sequences is required, it is more useful to undertake a spatial link analysis (which is similar to a spatial operational sequence diagram). In these representations a continuous line is drawn between items as the focus of attention changes. Thus, it is possible to work through the entire task by tracing this line.

Other types of operation sequence diagram can be used to look at other aspects of sequences of task actions. For example, a temporal operational sequence presents different task actions against a timescale, and a partitioned operational sequence diagram shows separate operation sequences for different aspects of a task that might be undertaken together and this can be useful for identifying interactions between them.

6.4   Analysis of temporal requirements and workload

Two important issues that can be investigated by task analysis are how long particular tasks might take and whether a person has the capacity to perform the task reliably. Various timeline techniques can be used to investigate the former, whilst workload assessment is appropriate for the latter. However, in many applications it will be necessary to ensure that a balance is maintained between the time that is taken and the workload that is imposed and so it will be necessary to use the two techniques together.

6.5   Analysis of errors

There are many techniques that can be used to investigate human errors, either by developing quantified estimates of human reliability, or by making entirely qualitative assessments. In either situation, the basic task information will be derived from some form of task analysis.

Task steps where there is a potential for human error are often investigated by using either event trees or fault trees, which show the logical and temporal dependencies between. Another useful technique for identifying potential human errors is a Human HAZOP. This is a table-top method that requires experts to make judgements for each task step about the likelihood of particular types of human errors from a prescribed list.

6.6   Analysis of cognitive processes

It is clear that complex decision-making and other cognitive tasks might be done in different ways on different occasions, even by the same person. Therefore, as well as directly investigating the overt task actions themselves, it is sometimes necessary to examine the underlying cognitive processes that are involved in order to understand the underlying strategies that might be developed to undertake particular tasks.

Several cognitive task analysis methods have been proposed to specifically examine cognitive aspects of tasks. Typical of these are the Generic Error Modeling System (GEMS) of Reason (1987), and Applied Cognitive Task Analysis (ACTA) by Hutton and Militello (1996). Each of these proposes a model of the processes of information handling and the structure of that information in human memory. This is then used as a framework for data organization and interpretation. However, it is also possible to make inferences about the cognitive structure of tasks without using a specific cognitive model. For instance, decision–action diagrams can be supplemented by asking additional questions in order to define the underlying conditions and reasons behind particular decisions.

There is also another central issue concerned with cognitive task analysis, namely how to elicit information about the different approaches that may be taken to a task and the reasoning behind these. These depend upon various knowledge elicitation techniques, such as verbal protocol analysis, which requires operators to verbalize without interruption whilst they undertake a task, or the withheld information technique, which probes into the way that information is used, by forcing persons to make decisions with incomplete information.

REFERENCES

ANNETT J., DUNCAN K.D., STAMMERS R.B. and GRAY M.J., 1971, *Task analysis*. Training Information Paper No. 6 (London: HMSO).

HUTTON , R.J.B. and MILITELLO L.G., 1996, Applied Cognitive Task Analysis (ACTA): A Practitioner's Window into Skilled Decision Making. In D. Harris (ed.) *Engineering Psychology and Cognitive Ergonomics. Vol. II: Job Design and Product Design* (Aldershot: Ashgate Publishing), pp. 17–23.

KIRWAN B. and AINSWORTH L.K. (eds), 1992, *A Guide to Task Analysis* (London: Taylor & Francis).

MILLER R.B., 1953, *A Method for Man–Machine Task Analysis* (Wright-Patterson Air Force Base, Ohio, USA: WADC Tech. Report No. 53, 157).

REASON J.T., 1987, Generic error modelling system: A cognitive framework for locating common human error forms. In J. Rasmussen, K. Duncan, and J. Leplat (eds), *New Technology and Human Error* (Chichester: Wiley).

Training and Acquisition of Complex Tasks

B. P. Goettl

Air Force Research Laboratory, Lackland AFB, TX 78236-5507, USA

1. INTRODUCTION

Today's systems are more complex than ever before posing monumental challenges for training. Computers can be used to improve training of complex tasks and simulate complex scenarios for the purpose of task rehearsal. They have an inherent appeal and can produce immersion into the training environment. Because of these characteristics, computer-based training systems can give the illusion of training, but without sound learning principles they can be ineffective. Training systems, whether computer based or not, must be founded on empirically derived principles of learning. Here a review of metrics for evaluating training systems will be followed by a short review of training principles and a summary of guidelines and recommendations.

2. TRAINING SYSTEM METRICS

Two metrics of training systems are important in the evaluation of training systems. These metrics include training effectiveness and training efficiency.

2.1. Training Effectiveness

Training effectiveness is defined as training that leads to improved performance. It is usually the metric of interest in training research. This is because the primary goal of training research is often to show that one method of training is more effective than another method. In prototypical studies, the experimental group practices on the training system under development while the control group practices on the criterion task. Both groups are then tested on the criterion task and training effectiveness is measured in terms of performance transfer between the two groups.

The simplest measure of transfer, which Gagne *et al.* (1948) called the *raw score* formula, is simply the difference between the experimental group and the control group on the criterion task. For measures on which poor performance results in large numbers (e.g. response time), the performance of the experimental group is subtracted from the performance of the control group. For measures where poor performance results in small numbers (e.g. percent correct), the control group is subtracted from the experimental group. In this way positive values always reflect positive transfer while negative values reflect negative transfer. The major drawback of this formula is that comparisons between different tasks are difficult and often impossible. A popular method used to compensate for this comparison problem is to divide the savings by the performance of the control group. This *percent improvement* formula is given by the equation:

$$T_{\% \text{ improvement}} = (C_b - E_b)/C_b \text{ x } 100. \tag{1}$$

While this formula normalizes for the performance level of the control group, it does not compensate for different amounts of improvement that are possible for different tasks. For tasks in which the range of performance is small, a small difference between the experimental and control group may be quite meaningful even though the percent improvement is not large. Tasks showing large variance in performance may produce large percent transfer scores even though the changes do not represent meaningful learning effects. To avoid these problems, Singley and Anderson (1989) recommend the *percent learning* formula for transfer. In this formula, transfer is the amount of savings divided by the amount of performance gain by the control group. This measure of transfer is given by the formula:

$$T_{\% \text{ learning}} = (C_{b1} - E_{b1})/(C_{b1} - C_{b2}) \text{ x } 100. \tag{2}$$

Here C_{b1} is the initial performance of the control group on the criterion task, C_{b2} is the performance of the control group following practice on the criterion task and E_{b1} is the initial performance of the experimental group on the criterion task following training. With this formula, negative scores represent negative transfer, positive scores represent positive transfer, and scores > 100 represent a sort of "super transfer" in which the experimental group shows greater improvement than does the control group. A training protocol showing positive transfer that is < 100% may be considered successful if it is cheaper or safer than training on the criterion task, or if it produces additional benefits related to other dimensions of the task.

2.2. Training Efficiency

Training efficiency is measured in terms of training costs. An increase in efficiency is realized when the same level of training effectiveness can be achieved with a training system that is less costly than the system against which it is compared. Cost is measured not only in terms of hardware, software, time and instructor support, but also in terms of safety for the trainees and instructors. While training effectiveness is usually the primary focus of basic research on training, from a practical standpoint, efficiency is often just as important as, if not more important than, training effectiveness. This is because operational training facilities must balance both the cost and the effectiveness of a training system. For example, while time-in-aircraft may be more effective than simulators for training piloting skills, simulator-based training is much less expensive in terms of hardware costs and potential loss of life. Training efficiency can also be realized when trainees share instructional resources without a meaningful loss of training effectiveness, such as when partners alternate practice periods on a training simulator.

2.3. Trade-off Between Effectiveness and Efficiency

The practitioner needs to find an appropriate balance between the effectiveness of prospective training systems and their associated costs. In the case of pilot training, for example, one must choose between the effectiveness of flight time and the cost efficiency of simulators. Practitioners are often faced with the task of comparing several training systems that differ in effectiveness and efficiency. Training system fidelity often plays a major role in the trade-off between effectiveness and efficient. High fidelity simulators have the ability to allow workers to train in environments that are nearly identical to the operational environment. These simulators are founded on the assumption that the training environment should match the work environment as closely as possible to maximize the transfer. While similarity between the
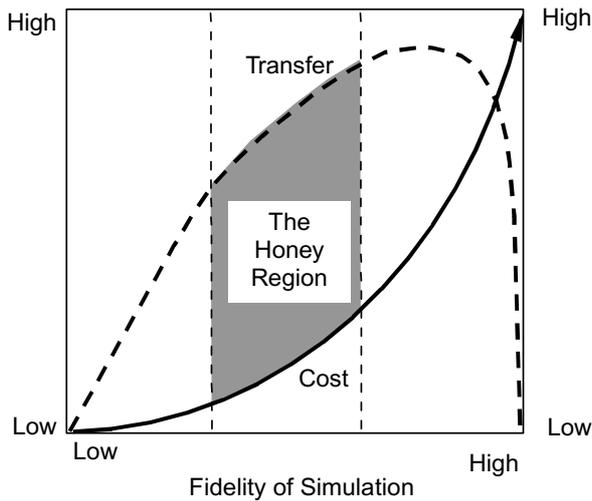
Figure 1. Training effectiveness and training cost as functions of training system fidelity (adapted from Roscose 1980, exhibit 17.3)

training and work environment can improve transfer, there are many situations in which the operational environment is not an optimal learning environment. In these situations high fidelity training can be ineffective as well as inefficient. For example, Schneider *et al*. (1982) showed that use of operational radar screens for training air traffic controllers to perceive aircraft turn radii is ineffective and costly. Typical aircraft require ~4 min to complete a full turn. Human operators cannot integrate perceptual information across the multiple radar sweeps during this time to produce a meaningful perception of the turn radii. Additionally, when the training proceeds at this slow rate, there is a great deal of training time wasted on a few trials. Compressing the time frame of the perceptual events not only makes the events more perceptible (i.e. effective), but also allows more trials to be presented in a given training session (i.e. efficient).

Roscoe (1980) depicted the trade-off between efficiency and effectiveness as in Figure 1. The dashed line shows the relationship between the amount of transfer from a training system and the fidelity of the system, while the solid line represents the relationship between the cost of the training system and system fidelity. Transfer shows a negatively accelerated relationship with fidelity while cost shows a positively accelerated function. Thus, beyond a certain point, additional costs spent on training systems produce diminishing returns on training effectiveness. Roscoe argues that the practitioner must find the "honey region" that produces the most transfer for the least amount of money.

Principled use of computer-based training can expand the size of the honey region. The increased capabilities of computers along with the development of tutor authoring software expand the honey region by lowering the costs of training systems. Principles of learning derived from well-founded research on training and complex-skill acquisition expand the honey region by increasing the transfer effectiveness of training systems. Principles of learning can also increase training efficiency by determining the dimensions of fidelity that maximize transfer. Focusing costs on those dimensions of fidelity that produce the most transfer can lead to improved training effectiveness and efficiency.

3. PRINCIPLES OF TRAINING

As the discussion above suggests, the effectiveness and efficiency of training systems depend critically on the training principles underlying the training systems. Only when based on principles of training derived from sound empirical studies will training systems produce meaningful transfer *and* cost efficiency.

3.1. Cognitive Task Analysis

To identify critical training needs, one must understand the cognitive, perceptual and motor demands of a task. The first step in understanding these demands is an accurate cognitive task analysis. Cognitive task analysis differs from regular task analysis in that it emphasizes the cognitive, perceptual and motor demands of the task. A good cognitive task analysis does not just indicate what an expert operator does, but how and why the operator does it. As such it forms the basis for identifying the most critical aspects of the task. Once the critical aspects are identified, an assessment must be made about whether the components are trainable. This assessment is critical because training should focus on those critical task components that can be trained.

3.2. Spaced Practice versus Massed Practice

There is a mature literature exploring the relative effectiveness of spaced practice versus massed practice. The bulk of the research shows that practice that is scheduled over several training sessions separated by non-practice periods produces greater acquisition, retention and transfer than training sessions that are not separated by the non-practice periods. The effectiveness of spaced practice has been demonstrated for a wide variety of tasks and for a wide variety of spacing intervals. In general, longer and more frequent spacing between training sessions leads to greater training gains.

There is an abundance of theories of the spacing effect. One class of theories suggests a memory consolidation explanation. These accounts hold that spacing, especially across days, promotes neural processes that serve to strengthen relevant cortical connections and weaken irrelevant ones. Another class of theories holds that the benefits of spaced practice are the result of enriched encoding. That is, spaced practice produces a richer set of encoding cues that promote information access during retrieval. A third class of theories holds that spaced practice, by requiring active reloading of information into working memory, promotes the development of control processes that are used to perform the task. Research is only now beginning to provide empirical comparisons between the theoretical explanations. One prediction that the latter two theories share in common is that the interval between practice periods need not be an empty "rest" interval. This research indicates that filling the interval with cognitive tasks can actually be beneficial.

3.3. Component–based Training

Component-based or part-task training rests on the component fluency hypothesis. This hypothesis assumes that complex tasks can be divided into several components and that efficient performance on the complex task depends on proficiency of the component tasks. Part-task training approaches assume that proficiency on the complex task can be achieved through proficiency on the components. The effectiveness of part-task training depends critically on how the complex task is decomposed and

how the task components interact. One of the more effective methods for decomposing the task is to divide it into different time periods, rather than into components that are performed in parallel. One reason for this difference may be that time-based decomposition keeps the interrelationships between components intact. When parallel components of the whole task are divided into separate components that are practiced individually, the relationship between the components may be lost. Special care must be taken to provide trainees with sufficient understanding of how the components combine and interact in the context of the whole task. This re-integration requires extra effort when parallel components are practiced separately. One way of achieving re-integration is by intermixing component task practice with practice on the whole task. Another way to achieve component practice within the context of the whole task is through *Gopher et al.*'s (1989) procedure of multiple emphasis on components. In this procedure, participants are asked to perform the whole task but are given explicit instructions to focus their efforts on one or two components of the task rather than the whole task. The emphasis is shifted between components throughout training. In this way, trainees focus on component practice but experience how the components interact with each other in the context of the complete task. This procedure may have the additional advantage of training time-sharing skills and attention-switching skills.

3.4. Consistent Mapping

One of the major milestones in training research is the empirical demonstration that the consistency between stimuli and their responses plays a critical role in whether a task can be learned. Schneider and Shiffrin (1977) presented a series of experiments showing the importance of task consistency in training. A task or task component is consistently mapped when the stimulus requires the same response whenever the stimulus is presented. In contrast, when a stimulus is associated with different responses at different times or in different contexts, the stimulus is variably mapped. Research on consistent and variable mapping has clearly established that consistently mapped components and tasks show dramatic improvements in response time and accuracy, indicating learning. On the other hand, variably mapped components show little performance gains, indicating little or no learning. This research has had a monumental impact on training research and applications. Several important training guidelines can be directly derived from this work. A short list of these guidelines is presented below.

3.5. Observational Learning

In the push by instructional designers to individualize instruction, observational learning is at risk of becoming overlooked as a useful training method. There is, however, a place for observational learning in modern training programs including computer-based training. Bandura (1986) proposed a multi-process theory of observational learning that holds that the effectiveness of observational learning depends on how well observers can attend to relevant information, represent the information in memory and produce the modeled responses. Thus, the complexity and salience of the modeled activities influence attention, memory,

and production processes involved in learning. Research on observational learning has shown that when components can be easily extracted and coded, internal models are complete and performance is high without hands-on practice. Recent evidence has expanded the scope of this theory to highly complex tasks, computer-based tasks and operational training centers. The practical appeal of observational learning is that it is an inexpensive training intervention. It can overcome limitations in training budget, hardware, software and manpower. For example, by alternating practice and observation periods on a training task, trainees sharing a single piece of equipment can benefit from modeling each others' behavior while realizing factorial gains in training efficiency.

4. GUIDELINES AND RECOMMENDATIONS

Several training guidelines can be derived from the abbreviated review of research on training presented here. The guidelines presented represent an incomplete list. A more complete list of guidelines is in Schneider (1985). First, practitioners need to consider both the effectiveness and efficiency of training and realize the diminishing returns of increased fidelity. Second, care should be taken to identify those dimensions of fidelity that are most relevant to the operational setting. Third, training should be focused on those tasks that are critical to performance. Fourth, cognitive task analysis has shown promise as a technique for identifying the critical components of tasks and thus should be one of the initial steps in developing training systems. Fifth, practice should be spaced to improve transfer. More importantly, filling the interval between practice periods with cognitive tasks may not be harmful but beneficial. Sixth, component-based training programs hinge critically on accurate identification of critical components. Training should promote an understanding of how the components interact in the context of the complete task. Seventh, training should be designed specifically to promote the development of consistent task components. Finally, social aspects of learning, such as observational learning and modeling, should not be overlooked in the push toward individualized instruction.

REFERENCES

BANDURA, A., 1986, *Social Foundations of Thought and Action: A Social Cognitive Theory* (Englewood Cliffs: Prentice-Hall).

GAGNE, R.M., FORSTER, H. and CROWLEY, M.E., 1948, The measurement of transfer of training. *Psychological Bulletin*, 45, 97–130.

GOPHER, D., WEIL, M. and SIEGEL, D., 1989, Practice under changing priorities: an interactionist perspective. *Acta Psychologica*, 71, 147–78.

ROSCOE, S.N., 1980, *Aviation Psychology* (Ames: Iowa State University Press).

SCHNEIDER, W., 1985, Training high-performance skills: fallacies and guidelines. *Human Factors*, 27, 285–300.

SCHNEIDER, W. and SHIFFRIN, R.M., 1977, Controlled and automatic human information processing: I. Detection, search, and attention. *Psychological Review*, 84, 1–66.

SCHNEIDER, W., VIDULICH, M. and YEH, Y., 1982, Training spatial skills for air-traffic control. *Proceedings of the Human Factors Society 26th Annual Meeting*, 10–14.

SINGLEY, M.K. and ANDERSON, J.R., 1989, *The Transfer of Cognitive Skill* (Cambridge, MA: Harvard University Press).

United Kingdom: The Ergonomics Society

S.A. Robertson

Centre for Transport Studies, University College London, Gower Street, London WC1E 6BT, UK

1. THE SOCIETY

The Ergonomics Society is the professional body for ergonomists in the UK. It was founded in 1949 as the Human Research Society while a more appropriate name was chosen. And its aim was to foster the interdisciplinary nature of work, because people had realized that a combination of disiplines could achieve more than any individual discipline. Up to 1949 there was no word in common usage to describe such interdisciplinary collaboration.

One of the tasks of the early committee was to decide on an appropriate name. A number of suggestions were made, including the Society for Human Ecology and the Ecuminological Society. They even coined a new word æergonomics. Figure 1 shows the original notes made by Murrell when he took part in choosing the name. Eventually the options were whittled down to two, and in a ballot the members chose the Ergonomics Research Society. Not only had a new professional body come into being, but a discipline had been born. In 1977 the society changed its name from the Ergonomics Research Society to the Ergonomics
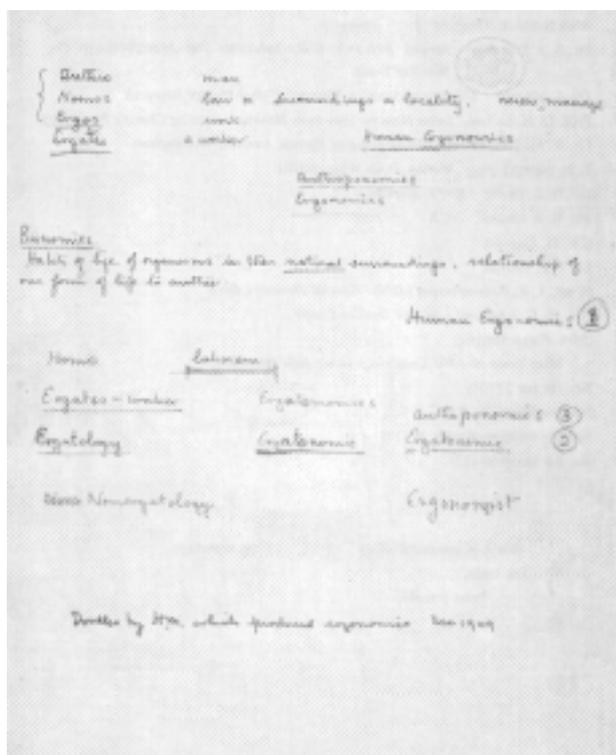
Figure 1. The development of the word "ergonomics" by K.F.H. Murrell in 1949 (Reprinted, with permission, from Edholm, O.G. and Murrel, K.F.H., 1973, *The Ergonomics Research Society: A History 1949-1970*, Loughborough: Ergonomics Society)

Society, for its members were not only fundamental researchers but also applied ergonomists. In 1991 the Ergonomics Society instituted the grade of registered member, so it could identify competent practitioners.

The society gained strength over the years, and after discussions with Taylor & Francis the journal *Ergonomics* was started in 1957 as its official journal. *Ergonomics* rapidly became a highly respected international journal and was adopted as the official publication of the International Ergonomics Association in 1961.

The Ergonomics Society annual conference is a major event and the scientific papers take a very short time to appear in print. It takes about seven months from submission of abstracts to publication of the proceedings. This makes the annual conference an ideal forum for disseminating cutting-edge research and for researchers to exchange ideas at an early stage of a project. The proceedings have been published by Taylor & Francis as *Contemporary Ergonomics* since 1982.

The Ergonomics Society today has a membership of over a thousand ergonomists throughout the world and maintains a register of professional ergonomists and ergonomics consultancies. Its aims are to promote the awareness, education and application of ergonomics in industry, commerce, the public sector and government, and the public domain in general. The society provides its members with a focus for important matters affecting the profession.

Through various participative committees and interest groups, it addresses short- and long-term issues affecting professional status and the development of ergonomics, including education, practice and legislation. The society represents all these issues both nationally and internationally and coordinates its activities to communicate a cohesive and coherent view of ergonomics to the outside world. It has a growing number of awards targeted at members and nonmembers; details may be obtained from the society.

2. MEMBERSHIP GRADES

To maintain academic standards and to encourage participation from practitioners in fields related to ergonomics, the society implemented a new structure in April 1998. The full professional grade of membership is registered membership, giving the designation Member of the Ergonomics Society (MErgS). This is reached through one of two routes. A graduate member will enter the society having successfully completed a course of study approved by the society and will be eligible for registered member status having gained three years' approved experience after graduating. An associate member who has a qualification which is acceptable to the society but not on the list as a qualification for graduate membership, may attain registered membership on the basis of five years' approved experience.

Associate membership is also open to applicants from related disciplines, such as physiotherapy, but these members would not necessarily be expected to seek promotion to registered membership. Registered members who have contributed significantly to ergonomics and the society may become fellows (FErgS), and honorary fellowship may be awarded to distinguished practitioners and other people who have made special contributions to the society.

The society welcomes new members from the UK and other
countries. Here is the contact address:

The Ergonomics Society
Devonshire House
Devonshire Square
Loughborough LE11 3DW
United Kingdom

Tel/fax   +44 1509 234 904
Email   ergsoc@ergonomics.org.uk.
Website   http://www.ergonomics.org.uk/

United States: Human Factors and Ergonomics Society

Although the origins of the human factors and ergonomics field can be traced to earlier dates, World War II was the key factor in the rapid maturation of the discipline known as human factors engineering in the United States. By the end of the war, a number of centers in the US were involved in human factors research, most of which related directly to military programs and dealt largely with aircraft and electronics equipment design problems. After a wartime program was initiated by Arthur W. Melton, the Engineering Psychology Branch of the Aeromedical Laboratory was established in 1945, with Paul M. Fitts as its chief. Other centers in the United States that were early contributors to the human factors field included the Navy Electronics Laboratory Human Factors Division under Arnold M. Small, the Massachusetts Institute of Technology Radiation Laboratory, the laboratory at Johns Hopkins University, and other universities and government research facilities.

There was little human factors activity, as such, within the aircraft industry during the war. However, a number of individuals were engaged in what was later to be considered human engineering. Most of this activity was concentrated in design organizations labeled as aircraft interiors, furnishings, or cockpit equipment groups. Compelled by the growing complexity of postwar military systems, and stimulated by the pioneering work of government and institutional centers, the aircraft industry began the gradual development of in-house human engineering capabilities.

Much of this work centered in the West Coast, where, in the early 1940s, the Aeromedical Engineering Association was formed as a medium for discussions among engineers, pilots, and physicians. Meetings were held on a regular basis in and around the University of Southern California in Los Angeles. By the mid-1950s most of the major weapon system manufacturers, as well as the military service laboratories, had developed or initiated clearly defined human-factors oriented groups charged with human engineering design or design support functions.

A number of other informal student and professional groups had sprung up around the country by this time, including the Chowder and Marching Club at Ohio State University and the San Diego Human Engineering Society, which formed in 1954. During the 1952 convention of the American Psychological Association (APA), Jack W. Dunlap of Dunlap & Associates chaired a meeting of 45 people to discuss the merit of forming a society and publishing a journal devoted exclusively to human engineering topics. A smaller group met again at the APA convention in 1953. Three more years would elapse before a concerted move was made to form a national human factors society based on the needs and aspirations of a much broader segment of the people engaged in human factors work than that represented by psychologists alone.

In the spring of 1956, Arnold M. Small and Wesley E. Woodson, then with the Convair Division of General Dynamics Corporation in San Diego, dispatched Clair E. Cunningham and Donald W. Conover up the California coast to study the organization of human factors groups within other aircraft firms. Each discussion invariably turned to the desirability of having a national human factors organization. Lawrence E. Morehouse, professor of physical education at the University of Southern California (USC) and the 1956 chair of the Los Angeles Aeromedical Engineering Association, issued a memorandum on 19 June calling for a meeting to consider the establishment of a new human factors organization. The meeting was held in Beverly Hills on 28 June 1956, and resulted in the establishment of a working committee. A follow-up meeting of 100 individuals from Southern California was held in Los Angeles on 25 October, at which time a new organization "The Human Engineering Society" was voted into existence. Morehouse was elected chair pro tem.

In the ensuing months a steering committee consisting of Arnold M. Small (Convair), Lawrence E. Morehouse (USC), Donald W. Conover (Convair), Donald Hanifan (Douglas Aircraft), Wesley W. Woodson (Convair), Stanley Lippert (Douglas Aircraft), John Poppen (an MD and consultant to Douglas Aircraft), Max Lund (Office of Naval Research), Jesse Orlansky (Dunlap & Associates), and Paul M. Fitts (Ohio State University) created the framework for a new organization. The seven committee members from Southern California met in Laguna Beach (halfway between Los Angeles and San Diego) every month or two during the next year to work out the details of what, at the time, was to be called the Human Factors Engineering Society of America. The Human Factors Society of America was voted into official existence on 25 September 1957, at the first official meeting in Tulsa, Oklahoma. About 100 human factors practitioners were in attendance. The name was changed at the 1958 annual meeting to The Human Factors Society, Inc., and then, in 1992, to the Human Factors and Ergonomics Society.

The San Diego Human Engineering Society and the Aeromedical Engineering Association of Los Angeles became the first two local chapters; there are now 36 local chapters and 31 student chapters in North America and Europe. Twenty-one technical interest groups have been formed within the umbrella of the Society, providing further communication links for individuals with similar interests and information needs. Membership in HFES has grown to about 5000 since the first official meeting in Tulsa in 1957, and more than 40 annual meetings have been held around the country since that time. Meetings, held each fall, feature an extensive technical program featuring paper presentations, posters, and hands-on workshops, as well as an annual awards ceremony, exhibits, and technical site tours.

The Society's flagship journal, *Human Factors*, was first published in 1958. *Human Factors* presents original papers of scientific merit that contribute to the understanding of human factors and advance the systematic consideration of human factors. The *HFES Bulletin* has been published monthly since 1959, containing news of conferences, elections, and events within and outside HFES; employment opportunities; editorials; and other items of interest to human factors researchers and practitioners. In 1993, the Society began quarterly publication of an applications-oriented magazine, *Ergonomics in Design. EID* features case studies, interviews, book and software reviews, and editorials. *HFES* has also produced monographs, proceedings

paper compilations, indexes, directories, standards, reports, guides, and videos. Some of the publications will be available online at the HFES Web site, http://hfes.org. In addition, HFES operates an on-line job placement service.

The Society found permanent headquarters in 1964 in Santa Monica, California. It has chapters in the United States, Canada, and Europe, which sponsor local meetings and publications. The Society also has student chapters and 21 technical interest groups.

Adapted from "HFES History," by Steven M. Casey, HFES Historian, from 1999—2000 Human Factors and Ergonomics Directory and Yearbook, copyright 1999 by the Human Factors and Ergonomics Society.

Universal Design for the Aging

S. Kose

Housing and Building Economy Department, Building Research Institute,
Ministry of Construction, Tatehara, Japan,

1. INTRODUCTION

It is only quite recently that the impact of population aging has gained much attention. What does it mean to our society, and how can we cope with it? What does the expression "universal design" mean in this context, and how can human factors and ergonomics contribute to solve the problems?

2. POPULATION AGING AND DESIGN

Population aging is caused by two major factors: extended life expectancy and reduced birth rate. The means that in developed countries about 80% of populations live to 65 or older — longevity has become a reality due to improved health and medical conditions. Reductions in birth rates result mainly from social attitudes — i.e. the wish of females to continue in work and to have fewer children. In most developed countries, the birth rate is less than 2 per 1000, which means that the total populations of these countries are destined to decrease. Figure 1 illustrates population trends for 65-year-olds and over in selected countries for the years 1950–2050, and shows clearly that in the twenty-first century 20–30% of the populations of developed countries will be in the 65-and-over age group; moreover, the populations of developing countries are following a similar pattern.

With regard to design, and universal design, since we have to design artifacts, etc., in order to use them, the first question to be asked is: "Who are the users?" Traditionally, we have had average users in mind — both Mr and Mrs Average — and have long assumed that meeting the needs of these average users would suffice. In the past, the few exceptional user groups that designers have specifically targeted have included the disabled. Unfortunately, the aged have been equated with the disabled, and it has been assumed, because only the frail aged were visible, that designs for the disabled would also cover the needs of the aged. But, usually, most of the older members of society who
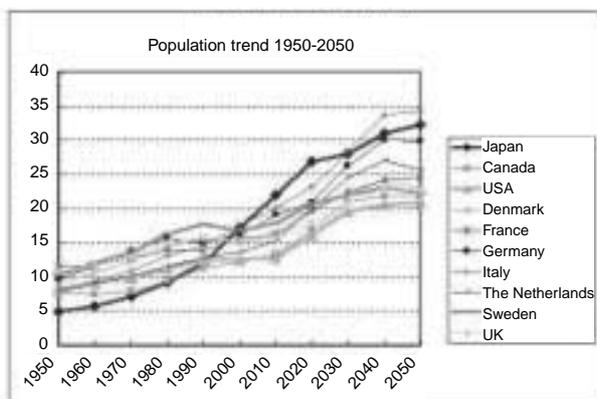
Figure 1. Population ratio of people aged 65 and over in selected countries 1950–2050
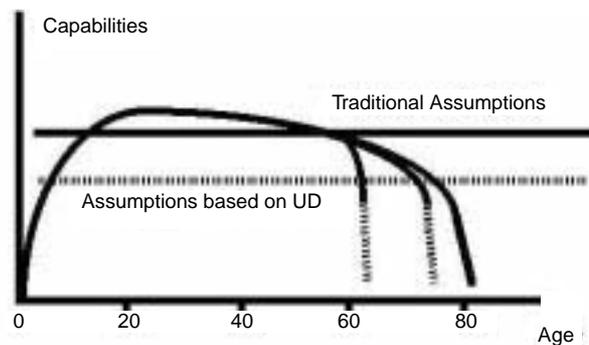
Figure 2. Schematic diagram of the development of human capabilities with age

have none of the apparent problems associated with aging, refuse to admit that they themselves are among the aged.

However, as has already been pointed out, the change in population structure reveals that, by occupying one-fifth or one-quarter of the total population, the aged are no longer a minority group in society but a sizeable segment of it. The problem with this is that the older generation do not necessarily have the capabilities assumed by designers for Mr and Mrs Average, nor can they be equated with the disabled. Figure 2 is a schematic diagram that illustrates reduction of human capabilities with age.

Our capabilities develop from birth onwards so that, between approximately 15 and 54 years of age, we are of "average" or "above average" capability, or at least this is assumed. In figure 2, the upper horizontal line shows what has, in the past, been regarded by designers to be the traditionally accepted level of expectations of capability in the public in general. It does not reveal any critical problems because, until recent years, the 65-year-olds and over occupied only 5% of the total population so that the majority of users were relatively young. Now, however, the situation is completely different, and in order to cover the wider age range, we must lower the line. In other words, we can no longer assume that the majority of people are healthy and robust; the new line has to allow for the changes in the level of capability necessary to accommodate the aging population. How then can we attain this goal of universal design?

3. THE EMERGENCE OF UNIVERSAL DESIGN

The concept of universal design emerged from barrier-free design, which took into account the needs of people with disabilities. The inherent nature of barrier-free design should perhaps have been "design for everybody", but, in reality, barrier-free design provided solutions for people with particular disabilities to offset badly designed products, environments, and systems. Because these have been specific solutions for particular users, they have had no universality. It has been a matter of initial bad designs which then required remodeling — and thus a doubling up on the design contract!

A completely new approach should take into account the needs of *all* users — that is universal design. The idea has long been in existence, but it was not until the Center for Universal Design at the State University of North Carolina issued the "Principles of Universal Design" that the idea received much attention. The message was more evident with the passing of the Americans with Disabilities Act (ADA) in 1990. Although the

rights of people with disabilities are acknowledged, the real pushing force must involve other user groups whose needs differ from those with disabilities.

Design for disabilities sounds too limiting; it also tends to emphasize specific disabilities and can lead to conflict between differing disabilities — for instance, tactile warnings designed to benefit people with visual disabilities can create hazards for people who have difficulty in walking. Universal design tries to solve such contradictions.

The definition of universal design written in The Principles of Universal Design is as follows (Version 2.0, The Center for Universal Design) (please note that no reference is made here to disabilities):

Universal design is the design of products and environments to be usable by all people, to the greatest extent possible, without the need for adaptation or specialized design.

The basis of universal design and barrier-free design therefore is good design — and, in reality, the essence of "design" is to accomplish *good* design.

4. THE ESSENCE OF DESIGN

The essential requirements of good design are: safety, accessibility, usability, affordability, sustainability, and esthetics (Kose 1998, 1999). Among these, barrier-free design emphasizes the first three, while universal design emphasizes the first four. The difference, therefore, is whether or not economic factor is taken into account, and this difference is crucial because, without the support of the general users, designs aimed at all users will not survive in the market. To put it another way, universally designed goods must appeal to the general users, and in particular to the aging population who do not qualify for the legal definition of disabled, or who deny any association with disability, but who suffer from badly designed goods.

Of course, designs suitable for everyone, or to be usable by everyone are goals that are difficult to achieve. What is important is that designers are well prepared from the beginning, and that the effort is sustained to ensure products, environments, and systems are *more* universally applicable in order to cover a larger segment of end-users, with emphasis being given to the aged. It is also important for designers to be prepared to adapt or modify products according to needs.

In addition to this, there will be requests for assistive technologies, designed for special groups. A wheelchair is unnecessary when one can walk; an oxygen container means nothing if one has no respiratory problems (Mace 1998: 25); but these are examples of assistive technologies whose existence are vital for those who need them. However, no one would ever describe them as universal design products. The layers of design policies that are required to meet varying needs are shown in figure 3, and at the pinnacle of the layers lies the support from human resources — personal intervention is often more effective and logical (it would be ridiculous to rely on high-tech devices for everything).

Next, we consider the problem of who will pay and who will benefit. Figure 4 shows the relationships between duration of designed products and environments and their characteristics as to whether they are public goods or private ones. To the bottom left are consumer products and to the upward right are durable goods, then come housing and public buildings, and to the extreme top
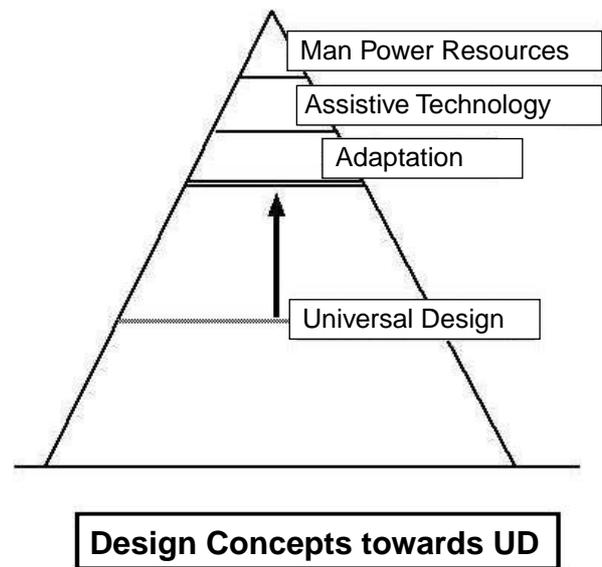
Figure 3. Design concepts: universal design and assistive technology

right lie infrastructures such as urban environments and civil structures. Those in the bottom left are cheaper products of short duration that give the users relative freedom of choice: in other words, one can select the most preferred from among the alternatives — for example, the choice of a ball-point pen that best fits one's gripping capacity or a telephone that one can easily dial/push. As one moves towards the top right, opportunities for choice become reduced to the point where it may be that one set of conditions must cover virtually all users — e.g. the railway station must perhaps provide stairs, escalators, and elevators. Nothing can be spared in the majority of cases. The reason for the differences is partly because of the possibilities of providing alternative means, and partly because of the public nature of the designed environments.

5. EXAMPLES OF UNIVERSAL DESIGN CONCEPTS

What then are the universally designed products, and what specific requirements does aging put upon us? Microwave ovens, remote controls (for TV, VCRs, etc.), mobile phones, automatic operations in automobiles (such as power assisted steering, power
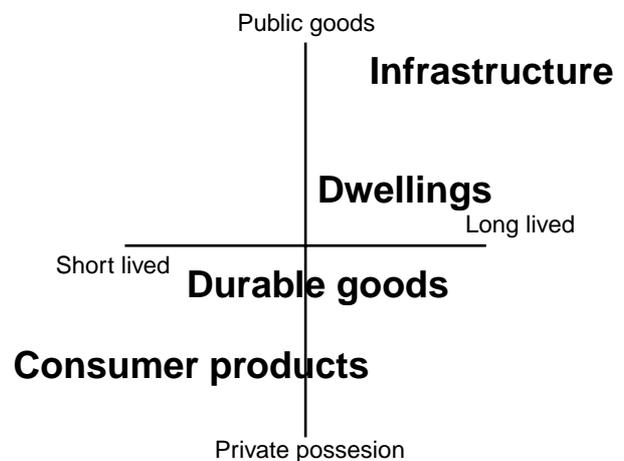
Figure 4. The relationships of duration and ownership with designed products and environment

windows), and personal computers are all recent innovative technologies that have enhanced most people's lives, including those of the aged and the disabled (see Goldsmith 1997). Microwave ovens assist people by speeding up food preparation and eliminating some of the repetitive work involved; remote controls overcome the distance barrier between the user and the appliance, automatic operations in automobiles eliminate the need for physical strength and simplify adjustments for height and position; while personal computers have simplified communications and empowered people to do many things without undue effort. The crucial point is that such new ideas have integrated universal design features that enable them to be used by everyone — regardless of age or physical ability.

Of course, the design solutions of individual products are not necessarily such that they can be described as of universal design for the aged. Not all remote controls are easy to operate; on the contrary, many are notoriously difficult to use, and VCR timers are among the worst examples of this — it is said that the majority of older people have given up using them. Some of these design errors result from complex operating procedures, while others are due to the alien nature of the products.

6. CONCLUSION

Universal design concepts attempt to solve usage problems by gradual improvements made as a result of user feedback. It is expected that an increased knowledge of the capabilities of older users will add greatly to the improvement of designs — which is where human factors research can make an important contribution. In addition, a cognitive science approach to design (Norman 1990, 1998) will also assist in realizing a better future for all groups of users.

REFERENCES

GOLDSMITH, S., 1997, *Designing for the Disabled: The New Paradigm* (Oxford: Architectural Press).

KOSE, S., 1998, From barrier-free to universal design: an international perspective. *Assistive Technology*, 10, 44–50. (This issue, with Molly Follette Story and Ron Mace as guest editors, is devoted to Universal Design.)

KOSE, S., 1999, Universal design: myths, beliefs and realities. In S. Kose, (ed.) *Universal Design: An International Workshop Proceedings* (Tsukuba, Japan: Building Research Institute), pp. 187–90.

MACE, R., 1998, The essence of design. In S. Kose (ed.) *Future of Design: Universal Design of the Environment, Products, and Information* (Tokyo: Toshi-Bunka-sha) (in Japanese), pp. 17—58. (The original text transcription of a lecture given during the International Conference on Universal Design: Designing for the 21st Century can be found at: http://www.adaptenv.org/21century/1998_overview/plenary_speech_mace.htm)

NORMAN, D.A., 1990, *The Design of Everyday Things* (New York: Doubleday).

NORMAN, D.A., 1998, The Plenary keynote lecture to the International Conference on Universal Design: Designing for the 21st Century. In S. Kose (ed.) *Future of Design: Universal Design of the Environment, Products, and Information* (Tokyo: Toshi-Bunka-sha) (in Japanese), pp. 59–100. (The original text transcription will be found at:

http://www.adaptenv.org/21century/1998_overview/plenary_contents.htm

The Center for Universal Design, North Carolina State University, http://www.design.ncsu.edu/cud/pubs/udprinciples.html

WELCH, P. (ed.), 1995, *Strategies for Teaching Universal Design* (Boston: Adaptive Environments Center/Berkeley: MIG Communications).

Part 2

Human Characteristics

Alternative Controls

G. L. Calhoun and G. R. McMillan

Air Force Research Laboratory, AFRL/HECP, 2255 H Street, Bd. 33, WPAFB, OH 45433, USA

1. INTRODUCTION

Typically, human interaction with computers involves a keyboard for text entry and a mouse for pointing and selecting text and icons. Trackballs and touch pads are also employed to perform much the same function as a mouse. Each of these controllers involves a manual input. However, there are many alternative controls that have the potential of providing a hands-free means of interacting with computers or computer-based systems. These hands-free controls would be useful for applications in which manual control is less convenient or impossible for the operator to exercise. Examples of the latter include assistive technologies for persons with physical disabilities and instances where operators are required to wear protective gear in hazardous environments, limiting their manual dexterity. Manual control is also not convenient in anticipated wearable computer applications in which a tetherless mobile computer, head-mounted display, and telecommunications systems will allow workers to access information, communicate with others and store information, all while performing their field operations. Even though one-handed keyboards, special purpose mice, trackballs, dials and finger sensors have been designed for wearable computers, there are instances where it is preferable for an operator's hands to remain engaged in their primary task. Use of speech input is one possibility, but is constrained by noise and may not be applicable in some task environments. Non-conventional controllers that do not require a direct mechanical linkage between the operator and the input device are desirable to provide hands-free interfaces that can substitute for or supplement manual control. These emerging hands-free technologies provide operators with a variety of new channels for interacting with computers.

2. CANDIDATE HANDS-FREE ALTERNATIVE CONTROLS

Alternative controls that use signals from the brain, muscles, voice, lips, head position, eye position, and gestures are candidate hands-free means for humans to interact with computers. US Air Force scientists (McMillan *et al*. 1997) have documented a comprehensive description of non-conventional control technology. A NATO Working Group recently reviewed this technology for aerospace applications (Leger 1998). Before discussing application issues for alternative controls, a brief description of several candidate controls is provided below. Each is addressed in more detail elsewhere in this volume.

2.1. Speech-based Control

Speech-based control uses pattern recognition methods to map an input speech waveform into corresponding text or a discrete output. Commercially available speech systems are fairly mature making their application both feasible and cost effective. Despite optimization techniques that can make speech-based control more robust, a key issue is whether speech systems can perform in high and dynamic noise environments (Jennings and Ruck 1995).

Another challenge is efficient dialogue design such that the vocabulary and syntax are manageable, without imposing a great memory load on the operator. Speech-based control must not interfere with operator communications and verbal inputs that can be heard by others nearby must be acceptable in the user's environment.

2.2. Gaze-based Control

For applications in which the operator views a display during control operations, harnessing the direction of gaze promises to be a very natural and efficient control interface. Careful interface design will be necessary, though, to ensure that the operator's head and/or eye movements during task completion are natural and not fatiguing. Rather, natural head and eye movements should be used to provide a direct pointing capability that can be combined with other control modalities to command activation. The task design also needs to take into account the accuracy limits of the gaze measurement system.

2.3. Gesture-based Control

There are a variety of techniques to read hand and body movements directly. Since the body and hands can be involved in other activities, gesture-based control for hands-free computer operation may best involve detecting defined movements or positions of the operator's face or lips. Optical and ultrasonic sensing technologies have been used to monitor an operator's mouth movement. Two candidate control approaches include processing lip movements to provide "lip reading" and translating symbolic lip gestures into control inputs. A related possibility is to employ special mouthpieces, which enable tongue-operated pointing or keypad input.

2.4. Electromyographic (EMG)-based Control

EMG-based control uses the electrical signals that accompany muscle contractions, rather than the movement produced by these contractions, for control. Electrodes positioned on the surface of skin detect these electrical signals (e.g. produced during an operator's clenching of the jaw) and translate them into control inputs. Continued development is required to optimize the signals employed, to assess the stability of the electrode contact over time and to minimize the effect of operator movement and external electrical activity on signal recordings.

2.5. Electro-encephalographic (EEG)-based Control

EEG-based control translates the electrical activity of the brain into a control signal. In one approach EEG patterns are brought under conscious voluntary control with training and biofeedback. Another approach harnesses naturally occurring brain rhythms, patterns and responses that correspond to human sensory processing, cognitive activity or motor control. Although detection of these signals is easily accomplished with inexpensive components, optimization of this alternative control requires minimizing the time required for signal processing and developing easily donned electrodes.

3. APPLICATIONS ISSUES

With perhaps the exception of speech-based control, hands-free alternative controls need further evaluation to specifically assess

their utility for human–computer interface applications (Calhoun and Macmillan 1998). Most are not commercially available and are not configured to facilitate integration with software designed for keyboard and mouse operations. However, given the potential advantages of hands-free control, further research and development are warranted and eventual commercialization is likely. The following are some issues relevant to their application.

3.1. Application Environment

Hands-free alternative controllers need to be operable in the anticipated application environment. Environmental factors include ambient noise, light, temperature, smoke and industrial contaminants. Some controllers (i.e. speech) may not be appropriate when privacy or covertness is a concern. For mobile applications, the input device needs to be operable regardless of operator movement or position and employ wireless transmitter technology as much as possible. Likewise, collaborative use of computing and networking technologies can impact input device choice.

3.2. Target Operator Population

For hands-free alternative controls to be acceptable to any operator population, components need to be lightweight, compact, comfortable and easy to don and doff. Implementation of these controllers will be impacted by whether the general population is targeted or whether the controller can be customized for an individual operator. If the former, the design should assume that the operator has little computer sophistication. For the general population, procedures to employ the controller need to be obvious, natural and amenable to self-training. Likewise, the need for operator calibration and adjustments should be minimized. However, for more specialized military or technical applications it may be acceptable, and even advantageous, to customize the controller and/or utilize a longer training protocol. For example, tailoring signal detection algorithms to an operator's EMG response or training on a speaker-dependent speech recognition system may produce significant long-term payoffs.

3.3. Task Requirements

To date, the control achieved with most hands-free devices can be described as rudimentary. Any application must take into account the limited dimensionality, accuracy, speed and bandwidth of control afforded by these devices. For some applications and target users (for instance, those with severe physical limitations), speed and accuracy of control may be of less concern, since conventional control options are not accessible. Other applications require more rapid and error-free performance. Since hands-free controls do not require associated limb or hand movement, control inputs are typically very rapid, unless lengthy signal processing is required. Rather, it is the precision and accuracy limits of the fundamental human responses and of the controller hardware and software that constrain application. In light of these limitations, efficient procedures are needed for correcting erroneous entries and for safeguarding the system from hazardous control inputs.

The characteristics of the task must also be considered. If the content of the data input is not known in advance, then the input device needs to support arbitrary input. In this case, keyboard surrogates or speech recognizers are likely candidate devices. If the input content is fairly constrained, and user interaction can be reduced to a selection process, then a more limited input device is acceptable. Concurrent tasking must also be considered. If the operator's visual attention is totally occupied by a task, then the use of gaze pointing for control is not appropriate.

3.4. Task–Controller Mapping

In addition to considering the general adequacy of the candidate alternative controls, the specific mapping of the input device to control functions must be addressed. It is unlikely that a single hands-free input device will be adequate for all control functions required in a particular application. A specific input device will be elegantly appropriate for some control functions and clearly inappropriate for others. It is most likely that alternative controllers will be used in conjunction with conventional manual input devices and other alternative controllers. Thus, task–controller mapping must take into account how best to increase overall functionality by using multiple input devices. The following describe some mapping alternatives. Computer-based systems may incorporate more than one of these mapping techniques in the overall control system design.

3.4.1. Single input device mapped to single control function

With this mapping, the "optimal" input device is assigned to each control function. For example, one might use keyboard control for inputting lengthy text information and a hands-free controller for sequencing through display screens or options. The potential advantages of using optimized input devices needs to be weighed against controller cost and potential operator confusion concerning which device to use for each function.

3.4.2. Single input device mapped to multiple control functions

It is doubtful that any hands-free input device will be capable of performing *all* the control functions required for wearable computer operation. To the extent that this can be achieved, controller cost can be reduced. However, the designer needs to avoid the pitfall of compromising overall efficiency for the sake of using fewer devices. For instance, speech-based control currently offers the most capability for this type of mapping. However, use of a speech command to designate a position on a two-dimensional surface can be cumbersome.

3.4.3. Multiple input devices mapped to single control function

Just as operators with desktop computers can navigate with a variety of controllers (mouse movement, arrow keys, tab key), it is possible to implement other computer operations such that several alternative input devices can be used for a single control action. This mapping approach provides the operator with increased flexibility: (1) the operator may have individual preferences for specific input devices, (2) a temporary task or environmental condition may deem one input device more efficient than another and (3) should one device malfunction, the control action can still be performed with another device. Once again, the designer needs to ensure that overall efficiency is not compromised. To map one, less optimal, input device to a control function may require procedures (e.g. a hierarchical menu

approach) that are not optimal for another device (e.g. single speech command). This requires the operator to use different procedures, depending on the active input device, for the same control function. There is a limit, though, to an operator's ability to remember or manage multiple procedural steps.

3.4.4. Mappings that truly integrate multiple input devices

In certain cases a combination of two or more input devices can perform a control function better than either one operating alone. Although little research has been done with integrated mappings, it seems clear that these types of designs will best capitalize on the capabilities of hands-free alternative controls.

One method is to map the input devices to subcomponents of the control action. For example, if hands-free function selection is required, eye gaze can be used to designate a desired control function and a purposeful facial muscle signal can serve as a consent response, commanding the system to act on the function last designated by the eye gaze. It would be difficult to use the individual input devices for this overall function. The use of both input devices capitalizes on the ability of eye gaze to rapidly designate position on a two-dimensional surface and a muscle signal to quickly send a discrete command.

A second method integrates multiple input devices to increase the accuracy or reliability of a control action. For example, the use of lip movement data, together with acoustic signals recorded during voice commands, has been found to enhance speech recognition accuracy in a noisy environment. Similarly, a controller design might require a simultaneous eye blink and muscle signal to minimize the chance of spurious activations.

A third method uses one input device to improve the performance of a second device. For example, eye gaze data might be used to enhance speech processing by restricting the vocabulary search to the most probable verbal commands associated with the current gaze point.

3.4.5. Implicit control and inferencing

The mappings described above focus on applications where controls are used to support conscious and explicit operator actions. Alternatively, these technologies can be used to support implicit actions in which control is largely transparent to the operator. One example is the use of head tracking to change the imagery presented on a head-mounted display in virtual reality environments. Here the controller is responding to an essentially subconscious act of the operator. A suite of alternative controls, coupled with inferencing software, may one day support even higher level implicit control. The combined information available from operator monitoring systems may allow a computer-based system to infer an operator's intent or requirements, such that

information presentation and task allocation can be automatically optimized.

3.5. Frequency of Control Input

Application of an alternative controller should take into consideration the anticipated input frequency. For example, just as extended manual keyboard entry can cause carpal tunnel syndrome, frequent use of jaw clenches to activate EMG-based control can aggravate TMJ (tempero-mandibular joint) disorders.

3.6. Control Evaluation

Laboratory evaluation, together with field testing using the target operators, are essential for the design of effective alternative controls and task–controller mappings. Objective measures of performance (e.g. time to activate the device, task completion time and accuracy, and performance on concurrent tasks) should be recorded, in addition to subjective ratings on the usability, satisfaction and learning associated with the candidate controller.

4. CONCLUSIONS

Creative human–computer dialogue design, together with further development of enabling technologies, human factors research, and field testing are required to evolve a system design that exploits the combined potential of conventional and alternative controls. With careful design, there are many advantages to using alternative controls, either individually or as part of a multi-modal interface. Hands-free operation can be realized for many tasks. Given the variety of available input modalities, there will useful options for any special user population (e.g. those with severe physical limitations). Properly implemented, the human machine interface will be more natural and require less training. In many cases, inputs can be made with more speed and accuracy. Given these advantages, alternative controls have great potential to assist operators in benefiting from the full capabilities afforded by computer-based systems.

REFERENCES

CALHOUN, G.L. and MCMILLAN, G.R., 1998, Hands-free input devices for wearable computers. In *Proceedings of the Fourth Annual Symposium on Human Interaction with Complex Systems* (New Jersey: Lawrence Erlbaum Associates) 118–23.

JENNINGS, D.L. and RUCK, D.W., 1995, Enhancing automatic speech recognition with an ultrasonic lip motion detector. In *Proceedings of the 1995 IEEE International Conference on Acoustics, Speech and Signal Processing*, 868–71.

LEGER, A., 1998, Synthesis- and expected benefits analysis. *RTO Lecture Series 215 Alternative Control Technologies: Human Factors Issues*, 9-1-9-10.

MCMILLAN, G.R., EGGLESTON, R.G. and ANDERSON, T.R., 1997, Nonconventional controls. In G. Salvendy (ed.) *Handbook of Human Factors and Ergonomics* (2nd ed) (New York: Wiley), 729–71.

Anaerobic Threshold

J. L. Smith

Department of Industrial Engineering, Texas Tech University, Lubbock, TX 79409-3061, USA

Anaerobic threshold (AT) has been defined in several ways, with the most accepted definition being "that point at which blood lactate begins to increase beyond the basal level." This definition implies that AT is determined through analysis of a series of blood samples taken as the exercise or work bout progresses. Some purists might argue that lactic acid accumulation should be examined in the muscle, rather than waiting for lactate to appear in the blood. At the other end of the spectrum, are the non-invasive measurement advocates who recommend that AT be defined by changes in respiratory parameters. The most commonly observed respiratory parameter used to determine AT is the non-linear increase in pulmonary ventilation. Software programs are available to predict AT from respiratory parameters, or the respiratory data can be plotted and determined by fitting two lines to the data and looking at their intersection point (figure 1). There are inaccuracies associated with the graphical analysis of respiratory parameters, but such a method is much less risky than one requiring blood sampling. AT is normally reported as a percentage of VO_{2max}.

Why is AT of interest to the ergonomist? It represents the point at which physiological fatigue is to be expected and where some sort of work–rest scheduling might be required. For untrained individuals, AT typically occurs at ~50–60% of VO_{2max}. With training, not only does the threshold to lactic acid increase, but also the onset of lactic acid accumulation occurs at a higher percentage of VO_{2max} (up to 80–85% of VO_{2max} in highly trained athletes).

Most ergonomics references recommend that physiological limits for work activities be set at 33% of VO_{2max}. If such recommendations are heeded, then a "safety margin" should exist between the recommended limit of 33% of VO_{2max} and AT, which
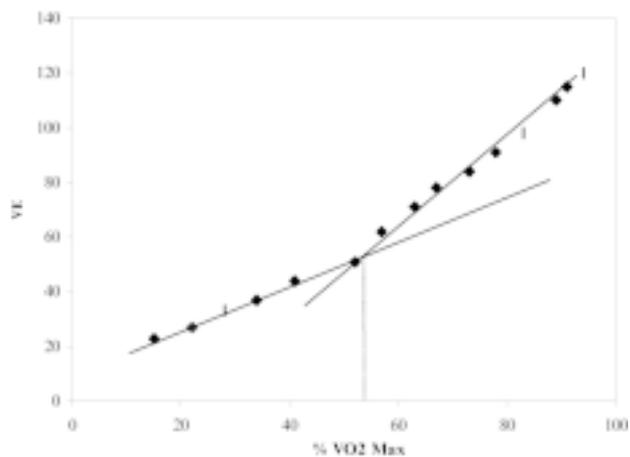
Figure 1. Graphical determination of anaerobic threshold

could be expected to occur ~50–60% of VO_{2max}. For most whole-body type activities, where large muscle masses are utilized, the worker would likely exceed the 33% of VO_{2max} recommendation before reaching his/her anaerobic threshold. Proper ergonomics guidelines should call for a "rest break" for workers exceeding 33% of their physiological capacity.

AT is probably of more interest to the ergonomics researcher than to the ergonomics practitioner. For the researcher interested in the development of work–rest schedules, AT provides information related to the physiological limits of work that require no special "rest breaks." Therefore, the ergonomics researcher would like to have knowledge of AT to insure that physiological fatigue is not accumulating, resulting in performance decrements of the worker.

REFERENCES

ASTRAND, P. and RODAHL, K., 1986, *Textbook of Work Physiology: Physiological Bases of Exercise* (New York: McGraw-Hill).

MCARDLE, W.D., KATCH, F.I. and KATCH, V.L., 1991, *Exercise Physiology: Energy, Nutrition, and Human Performance* (Philadelphia: Lea & Fèbiger).

Anthropometric Databases

R.E. Herron

Department of Ergonomics, Colorado State University, Fort Collins, CO 80523, USA

1. INTRODUCTION

Anthropometry literally means the measurement of humans. Although one can measure humans in many different ways, anthropometry focuses on the measurement of bodily features such as shape and size ("static anthropometry"), body motion, use of space and physical capacities such as strength ("functional anthropometry"). A database is simply a compilation of information, generally used for reference purposes.

There are several types of database in ergonomics. They include: (1) much of the published literature itself, (2) bibliographic aids containing citations, abstracts and indexes to the literature and (3) manuals, guides, handbooks and other compilations that summarize the basic data and methods of the field. Anthropometric databases include all of these formats, but the most widely used are in the form of manuals, guides, handbooks and statistical compilations that attempt to summarize the variability in anthropometric characteristics of different human populations or groups. A new type of anthropometric database — the computer-generated human model — is a rapidly growing phenomenon that is leading to major changes in how anthropometric information is used for ergonomic design and other purposes.

2. HISTORY OF ANTHROPOMETRIC DATABASES

Interest in applying anthropometric data to the design of tools, instruments, equipment and systems has a long history. Early human societies employed folk norms for the design of digging, cutting and shaping tools. Natural units of measurement first developed around the hand, fingers and foot and were not replaced with physical units of measurement until many years later. The systematic collection of anthropometric data on large groups of individuals did not occur until the 18th century with the work of Linne and Buffon. The industrial revolution increased the demand for pertinent body dimensions to meet the needs of such mass markets as furniture and clothing. In 1883 Sir Francis Galton undertook a detailed, quantitative survey of anthropometric and other characteristics of over 9000 visitors to the London Health Exposition. However, it was not until World War II, when the human–technology interface reached new levels of complexity, that the need for better anthropometric data for workstation and equipment design became a major focus of attention. Today, there is no longer any doubt that the wide and constantly growing variety of ergonomic design problems calls for more comprehensive, versatile, parsimonious, valid and reliable anthropometric data sets for use in representing the shapes, sizes and functional capacities of target user populations.

3. HOW ARE ANTHROPOMETRIC DATABASES USED IN ERGONOMICS?

In the majority of real-world ergonomic design applications involving the interaction of humans with tools, instruments and systems, there is a need for information about the population of users. Very often the need is for anthropometric data. Differences in body shape, size and physical capacities are among the most obvious manifestations of human variability. It is clearly apparent that people differ in height, weight and arm length, as well as in the other myriad ways in which body form and function can be measured. A well-designed or selected anthropometric database can provide vitally important information for optimizing the human–machine–equipment or system interface.

Creating or selecting and using an anthropometric database effectively is not a trivial matter. There are many hazards, which can undermine the best intentions of the unwary practitioner. First, one has to define which anthropometric parameters are relevant and necessary. Determining which bodily dimensions are important starts with a clear understanding of the task being performed. Then, the user population must be identified and characterized, in terms of homogeneity (e.g. highly selected astronauts) or the lack of it — heterogeneity (e.g. more diverse assembly line workers), as well as other relevant demographic variables, such as age and sex. An initial search should be made to locate a database that matches the target user population as closely as possible. Unfortunately, although hundreds of anthropometric databases are now available, the prospects of finding a good match (from a valid and reliable source) are not good — due to the wide diversity of today's working population, in terms of racial, ethnic, sex, age, nutritional and other variables. This ever-changing diversity is the source of numerous human factors problems, due to the difficulty in defining exactly what the make up of the civilian population really is. Wide-ranging differences in body shape and size of various ethnic groups are of special significance in the design and manufacturing of products for the global marketplace.

But the current utility of anthropometric databases is not all doom and gloom. Notwithstanding the low probability of finding a close match between existing databases and various target populations, informed and useful compromises can be made. This involves supplementing or adjusting the best available data set, using sophisticated methods for estimating or forecasting, to arrive at a description of the desired design criteria — the range of relevant dimensional measurements which would be expected in the target population. While these strategies can be productive in generating usable and useful anthropometric data for design purposes, a well-conceived and executed methodology is essential. Further details about such matters can be found in the references.

4. WHAT ANTHROPOMETRIC DATABASES ARE AVAILABLE?

There is a vast supply of anthropometric databases on military populations, but unfortunately those on civilian populations are fewer. The main sources of military data are publications of government agencies and specialized centers such as the Air Force, Army, NASA and CSERIAC (Crew Systems Ergonomics Information Analysis Center). Many such publications can be obtained from the National Technical Information Service (NTIS) or the Defense Technical Information Center.

The scarcity of good anthropometric databases on the US civilian population is due to a number of limiting factors: (1) inadequate measurement technologies, (2) costs and (3) logistical

problems — such as difficulty in mobilizing and accessing representative populations. Fortunately, the situation is slowly improving. In the meantime, one can locate the best extant anthropometric databases via the references.

5. WHAT NEW ANTHROPOMETRIC DATABASES ARE UNDER DEVELOPMENT?

Better civilian anthropometric databases, which are more representative of today's diverse population, are on the horizon, largely due to a growing demand for anthropometric data and the continuing evolution of more practical computer-based anthropometric measurement instruments.

These new compilations will go far beyond traditional linear (2-D) dimensional data and include the much more comprehensive and informative multidimensional (3-D, 4-D, etc.), biostereometric data.

The use of a biostereometric approach permits the storing of digital geometric descriptions of entire body forms, i.e. 3-D spatial replicas). These archival databases can be mined for an infinite variety of anthropometric data to meet a wide range of user needs. Another advantage is the ease with which such computer-based compilations can be maintained and up-dated to ensure continuing representative fidelity of the total population and various subpopulations.

6. COMPUTER MODELS OF HUMAN FUNCTIONAL CAPACITIES

Another consequence of using a biostereometric approach to the measurement of body geometry is the development of more realistic, animated human models. Such multidimensional models constitute another type of anthropometric database which is changing how anthropometric data are used for ergonomic design applications. However, the biostereometric data on which these models are based do not yet adequately represent the variations in body geometry and motion patterns which are present in the civilian or military populations. This is largely due to the continuing inadequacies of 3-D scanning methods for gathering the basic body geometry and motion data. As these methods become more cost-effective, portable, faster, user-friendly and technologically sophisticated, the current deficiencies will be overcome.

Ultimately, highly representative computer models can be installed in computer-aided design (CAD) platforms and accessed instantaneously to serve a wide variety of ergonomic design functions. The growing availability of such "virtual humans" can overcome many of the problems associated with current anthropometric databases and help promote the proactive use of anthropometric information for improving the human–machine and human–environment design interface.

7. FURTHER READING

There is an extensive literature related to the subject of anthropometric databases. The references constitute but a small sample of the extant sources. However, they should at least provide a point of entry for obtaining further information about an area of ergonomics which is entering its most exciting phase — when anthropometric databases of realistic, multidimensional, human models are found in the computers of every designer who is concerned with ergonomic design for human use.

REFERENCES

CHAPANIS, A., 1996, *Human Factors in Systems Engineering* (New York: John Wiley).

HERRON, R.E., 1973, Biostereometric measurement of body form. In *Yearbook of Physical Anthropology 1972*, vol. 16 (New York: American Association of Physical Anthropologists), 80–121.

PHEASANT, S., 1996, *Bodyspace: Anthropometry*, *Ergonomics and the Design of Work* (2nd ed.) (London: Taylor & Francis).

ROEBUCK, J.A., 1995, *Anthropometric Methods: Designing to Fit the Human Body* (Santa Monica: Human Factors and Ergonomics Society).

VAN COTT, H.P., KRAMER, J.J., PEZOLDT, V.J., PORTER, L.G., FRIED, C., FECHTER, J.V. and PERSENSKY, J.J., 1978, *A Standard Ergonomics Reference Data System: The Concept and Its Assessment*. NBSIR 77-1403 (Washington, DC: Institute for Applied Technology, National bureau of Standards).

Anthropometry of Children

E. Nowak

Institute of Industrial Design, Warsaw, Poland

1. INTRODUCTION

From the design point of view, ergonomics can be divided into three areas, which are as follows:

- environment of work
- environment of the home
- environment of the leisure activities

When we wish to adjust these environments to meet human needs, it is necessary to take into account the age of the person or persons using them. In all these environments anthropometric data of children and juveniles play an important role in the design of toys, interior furnishings and furniture, kindergartens, schools, hospitals, playgrounds, etc.

A young and not completely developed organism is extremely susceptible to the influence of the external environment. Furniture, equipment, and toys, wrongly designed and not adjusted to the characteristics of a young child can result in defective body posture and the establishment of pathological states. In order to prevent this, it is necessary to make use of various remedial measures. Besides providing children with proper medical care and correct nourishment, it is necessary for their development to shape consciously the surroundings, including the world of objects. Correct design, taking note of ergonomic criteria, can facilitate good physical development in a child, while the use of harmonious forms and rational solutions helps to form aesthetic feelings and improve the quality of everyday life. Ergonomic values should be apparent in every object designed and intended for the youngest generation.

Anthropometry is one of the important elements that has an effect upon the ergonomics of the child's environment. It provides data concerning the physical development of children and the changes that take place in body proportions with age. It also provides information about the inter-generation differences in somatic features observed in a given population. This is the so-called phenomenon of secular trend. When this phenomenon occurs in a given population, a designer should use current anthropometric data only. In this respect, the Institute of Industrial Design, among others, specializes in the environmental ergonomics of children and the young people, and conducts research to develop anthropometric data characterizing the basic design requirements of children.

This article describes successive phases of investigation involving Polish children and presents their actual somatic characteristics. It also introduces two-dimensional manikins that illustrate to scale the changing physical shape of a child's body with age.

2. CHARACTERISTICS OF SOMATIC DEVELOPMENT OF CHILDREN AND JUVENILES: DATA FOR DESIGNING

Before a proportionally shaped figure of an adult individual is finally formed, it undergoes many intermediate states, which are characterized by different body proportions. This results from the irregular development of the body systems, such as the skeletal and muscular systems, in particular developmental periods, and different growth rates of individual parts of the body. In relation to adults, newborns and, to a lesser degree, infants have large heads, short necks, long trunks and short limbs, especially the lower ones. In the course of development, from the moment of birth to adulthood, the trunk, already proportionately long in infants, increases to three times its original length. At the same time, the mass of muscles also increases. It is estimated that muscular system increases by as much as 42 times its original size.

Apart from this basic information concerning changes in the structure of the human body, designers need to become acquainted with some of the biological phenomena that are directly connected with physical development. These are secular trend phenomenon, the acceleration process, and the pubertal spurt of body growth.

Secular trend is defined as the sequence of changes of human physical features, which occur gradually from one generation to another, tending in a constant direction. These changes concern mainly stature dimensions — e.g. children are taller than their parents and have bigger body size.

Acceleration of development is understood as the intergeneration speeding-up of biological development and puberty, and thus the earlier arrival of consecutive phases of development. The acceleration of development of the whole organism in every successive generation by comparison to the preceding one initially concerns puberty. One of the manifestations of this period is the so-called pubertal spurt, its characteristic being that the yearly increase in height is greater just before the onset of puberty than previously in both boys and girls.

The observations of Polish anthropologists (Waliszko *et al.* 1980; Charzewski 1984; Hulanicka 1990) prove that in Poland both the phenomena of secular trend and developmental acceleration continue to occur. In Poland, the physical development of children and youth is differentiated and depends mainly on social and economic living conditions. According to Polish anthropologists significant differences in this field can be noticed for particular regions of Poland (Hulanicka 1990). Researches carried out by the Polish Academy of Science indicate that there are significant differences between the physical dimensions of children living in city centers, small towns, and in villages, and that these differences are greatest during the time of the pubertal spurt. For example, 14-year-old boys from city centers are, on average, about 50 mm taller than their peers from small towns, and nearly 90 mm taller than those living in villages. In Poland, the greatest environmental differences are seen in the comparison of physical dimensions of children of the Warsaw intelligentsia to those of children living in villages whose parents have elementary education. At the same time, the results of investigations carried out by Polish anthropologists confirm that the biggest development indices are obtained by children and the young people living in the center of Warsaw. Contrary to West European and American cities, where the intelligentsia live mainly on the outskirts of urban areas, in Warsaw the intellectual elite tends to live in the city center rather than the suburbs. This is the most urbanized area of the city, having the highest number of utility premises (commercial, cultural, and medical). In

Table 1. Annual growth values of stature (B-v) in boys and girls from the big town population in 1966, 1978 and 1992 (in mm)

Age	Boys			Girls		
(years)	1966	1978	1992	1966	1978	1992
7–8	51	44	57	49	54	59
8–9	47	52	46	56	45	56
9–10	57	55	45	57	54	48
10–11	34	44	47	55	73	75
11–12	52	62	88	57	70	63
12–13	64	68	56	66	45	44
13–14	47	88	79	29	38	45
14–15	100	49	51	27	14	18
15–16	43	34	37	6	17	12
16–17	22	15	21	5	10	5
17–18	9	14	12	3	14	5

comparison with other districts and regions of Poland, the center of Warsaw is characterized by having the lowest number of people living in a household, and this contributes to their better living conditions.

Taking this into consideration, the Institute of Industrial Design conducted an anthropometric investigation embracing a selected group of children and young people aged 6–18 years. They represented the environment with the highest living standard and were characterized by the highest index of physical development (Nowak 1993). After comparing the results of investigations obtained in 1992 with those of the years 1966 and 1978 (table 1), it can be seen that in Poland the phenomenon of secular trend, manifesting itself by the regular increase of stature dimensions, continues to occur in both boys and girls.

The greatest differences in the body height of boys between the years 1966 and 1978 were observed at the age of 14 and amounted to 95 mm. The difference between the years 1978 and 1992 was lower (31 mm) and was seen at 12 years of age. However, comparison of the years 1966 and 1992 shows big difference in body height. The greatest increase occurs at the age of 14 and amounts to as much as 105 mm. The increase of stature characteristic in girls show a similar tendency, though it is less marked. The analysis of the yearly increase of stature (B-v) and pubic height (B-sy) defining the length of lower extremities, shows that both characteristics confirm the evident *process of pubertal spurt acceleration*. In boys investigated in 1966 this process was observed between the ages of 11 and 12 years. In girls, the pubertal spurt occurred about two years earlier. Based on these investigations, it is possible to indicate directions for designing:

- since secular changes and developmental acceleration occur in the Polish population, anthropometric data based on investigations conducted in earlier years cannot be used;
- the greatest growth acceleration and activity is seen between the ages of 11 and 14. While designing for this age group, one should be extremely careful and take into consideration up-to-date data.

In order to make up-to-date data accessible to Polish designers the Institute of Industrial Design has developed the so-called target standard. It is based on the anthropometric data of children of the highest indices of development. It is assumed that in the

future children from other environments will draw nearer this standard when their living conditions improve. Table 2 presents the dimensions of the target standard of selected stature characteristics of children and young people aged 6 years and 18 years.

3. TWO-DIMENSIONAL MANIKINS OF CHILDREN: MODELS FOR DESIGN

The following criteria were taken into consideration while preparing the manikins: conformability of functional measurements with current anthropometric data, simplicity of the use of models, of their removal and storing, and the lowest costs of production. In order to meet these criteria it was necessary to reconcile many factors important both for anthropometry and design. It is obvious that even the best model is a static arrangement — it cannot present fully the dynamics of the human body, and especially the dynamics of a child's body changing in the ontogeny. To achieve data synthesis the idea of differentiating manikins in respect of sex was given up. The analysis of the development of body height as well as of other height features of boys and girls indicates that the development of these features of boys and girls up to 10 years of age is similar and that differences of somatic features are statistically insignificant (up to 8 mm) and can be neglected for design purposes (Nowak 1989).

The manikins were not prepared for each age class but were divided into groups consisting of several age classes. It was very difficult to separate the above groups since it was impossible to take into account the dynamics of development with the division resulting from the necessity of designing furnishings separately for children in kindergartens, primary schools and secondary schools simultaneously.

The period between 11 and 15 years of age was the most controversial one. This is a period of rapid changes in a child's body due to puberty. This period differs slightly in the case of boys and girls. The pubescent leap is a sign of puberty and results from the hormonal, functional and morphological changes that are taking place in a child's body in preparation for puberty. Although the pubescent leap is common to both boys and girls, its effects differ in intensity and persistence in individual children.

In the case of boys, the leap occurs on average at the age of 14; it is about two years earlier in girls and less intensive. The differences in dimensions between adult males and females result to a certain extent from the differences of their pubescent leap. Since the girls' pubescent leap starts earlier, it is less intensive and as a result the process of body development is completed earlier. On average, girls reach puberty at the age of 16 and, at the same time, their bodies cease growing; in the case of boy, their development process can continue up to 21 years of age.

In view of this, the groups of boy and girl manikins were divided differently, although the divisions were made according to successive stages of child development. These covered post-infantile age (1–3 years), pre-school age (4–6 years), younger school age (7–10 years), puberty (11–14 years), and juveniles (15–18). This division corresponds with the ages of children attending nursery, kindergarten, primary, and secondary schools. Thus, the following seven groups were selected according to age: I: 1–3-year-old boys and girls; II: 4–6-year-old boys and girls; III: 7–10-year-old boys and girls; IV: 11–14-year-old boys; V: 15–18-year-old boys; VI: 11–13-year-old girls; VII: 14–16-year-old girls.

Table 2. Anthropometric measured data in (cm) of Polish boys and girls aged 6 and 18

Dimension	Boys				Girls			
	6 years		18 years		6 years		18 years	
	percentiles							
	5	95	5	95	5	95	5	95
1. Stature	112	127	168	189	111	127	157	178
2. Eye height	101	114	156	175	100	115	147	161
3. Suprasternal height	89	101	137	156	87	101	127	146
4. Elbow height	69	81	109	122	68	79	101	113
5. Pubic height	57	66	88	101	57	66	83	95
6. Head and neck height	23	25	31	33	24	25	30	32
7. Trunk height	30	38	47	56	29	37	43	51
8. Thigh length	25	33	41	51	25	33	37	47
9. Knee height	28	36	45	55	28	35	42	50
10. Upper extremities length	43	57	71	86	42	56	64	79
11. Arm length	17	23	30	37	17	23	27	33
12. Forearm length	14	20	23	37	17	23	27	33
13. Hand length	11	14	18	21	11	14	16	19
14. Arm overhead reach*	130	152	209	232	126	148	188	214
15. Arm reach down*	42	52	69	82	41	51	65	77
16. Arms span*	92	115	146	176	94	116	134	159

* Measurement taken with the hand clenched.

The set of manikins was prepared using the values of 5th and 95th percentiles, to allow the application of threshold percentiles in design. The set consists of 26 models presented in two views: a side view, where figures are in the sagittal plane, and a top view, where figures are in the transverse plane. Each model gives information on sex, age, and its percentile value. The manikins are made of a stiff material, plexiglass, and in a scale of 1:5, and have holes which enables the placing of the figure in a given position or the movement of a given segment of the body. The manikins are fixed by sticking a compass leg or a pin into a hole on the top of head or on a foot. Movement capacities are possible thanks to holes representing the axis of rotation of the following joints: shoulder, elbow, wrists , hip, knee, ankles, as well as head motions (upper and lower joints of head), neck motions (junction of neck and thoracic segments of spine) and trunk motions (junction of thoracolumbar segment). Flexion and extension movements in all these joints can be obtained by means of the manikin developed in the sagittal plane, while the manikin developed in the transverse plane allows the determination of the abduction and adduction movements of limbs and head movements. The model in transverse plane is a simplified shape of a child's figure seen from above. The left side consists of the shoulder, upper limb and breast; the right side consists of the thigh, hip and abdomen. By rotating the manikin along the axis of symmetry, one can obtain the contour of the whole body. Contours of upper limb and reach dimension are obtained by putting the model into the sagittal plane.

Figure 1 shows the application of manikins in the ergonomic analysis of furniture. If the analysis is undertaken while still in the design stage, this reveals any problems in the functional dimensions of furniture to be revealed and corrected.

REFERENCES

CHARZEWSKI, J.,1984, *Social Conditioning of Physical Development of Children Living in Warsaw* (Warsaw: Academy of Physical Education) (in Polish).

HULANICKA, B., 1990, Body height of boys from Wroclaw in comparison with body height of boys from Warsaw. *Anthropological Materials and Works,* 111, 37–41 (in Polish).
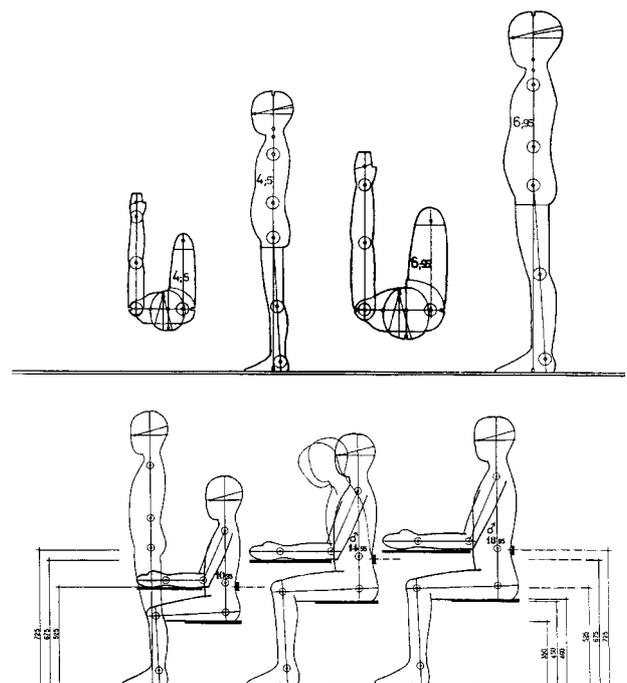
Figure 1.

NOWAK, E., 1989, Two-dimensional manikins of children: models for design. *Applied Ergonomics,* 20.2, 136–9.

NOWAK, E., 1994, Physical development of children and the youth from the center of Warsaw. *Human Biology,* 25, 401–8.

WALISZKO, A., *et.al.*, 1980, *State of physical development of children and the young of school age in Poland* (Wroclaw: Department of Anthropology of the Polish Academy of Science) (in Polish).

Anthropometric Terms

K. H. E. Kroemer

Industrial And Systems Engineering, Virginia Tech, Blacksburg, VA, and Eri, Inc., A Consulting Company, PO Box 3019, Radford, VA 24143-3019, USA

- Anterior — in front of, toward the front of the body.
- Breadth — straight-line, point-to-point horizontal measurement running across the body or a body segment.
- Curvature — point-to-point measurement following a body contour; this measurement is neither closed nor usually circular.
- Circumference — closed measurement following a body contour; hence, this measurement is usually not circular.
- Coronal — same as frontal.
- Depth — straight-line, point-to-point horizontal measurement running fore–aft the body.
- Distal — away from the center of the body; opposite of proximal.
- Distance — straight-line, point-to-point measurement between landmarks of the body.
- Dorsal — toward the back or spine; opposite of ventral.
- Height — straight-line, point-to-point vertical measurement.
- Frontal — in a plane that cuts the boy into fore–aft (anterior–posterior) sections; same as coronal.
- Inferior — below, toward the bottom; opposite of superior.
- Medial — in a plane that cuts the boy into left and right halves; same as mid-sagittal.
- Mid-sagittal — in a plane that cuts the boy into left and right halves; same as medial.
- Lateral — to the side, away from the middle (medial).
- Posterior — behind, toward the back of the body; opposite of anterior.
- Proximal — toward or near the center of the body; opposite of distal.
- Reach — point-to-point measurement following the length of an arm or leg.
- Sagittal — parallel to medial (occasionally used as medial).
- Superficial — on or near the surface of the body.
- Superior — above, toward the top; opposite of inferior.
- Transverse — in a plane that cuts the boy into upper and lower (superior and inferior) sections.
- Ventral — toward the abdomen (occasionally used like anterior).

Anthropometry

K. H. E. Kroemer

Industrial And Systems Engineering, Virginia Tech, Blacksburg, VA, and Eri, Inc., A Consulting Company, PO Box 3019, Radford, VA 24143-3019, USA

While all humans have heads and trunks, arms and legs, the body parts come in various sizes and are assembled in different proportions. The science of measuring human bodies is called anthropometry. It uses special terms and procedures described in the relevant literature. The results of anthropometric surveys are described in statistical terms.

In statistical terms, most body data appear in a normal (Gaussian) distribution. Such data set can be explained by the statistical descriptors *mean* (same as *average*), *standard deviation*, and *range*, if the sample size is large enough. Misunderstanding and misuse have led to the false idea that one could "design for the average" (see the chapter on Engineering Anthropometry). Yet, the mean value is larger than half the data, and smaller than the other half; consequently, the "average" does not describe the ranges of different statures, arm lengths, or hip breadths. Furthermore, one is unlikely ever to encounter a person who displays mean values in many or all dimensions. The mythical "average person" is nothing but a statistical phantom.

For more information see Roebuck (1995) and Kroemer (1999).

REFERENCES

KROEMER, K.H.E., 1999, Engineering anthropometry. In W. Karwowski and W. S. Marras (eds) *The Occupational Ergonomics Handbook* (Boca Raton: CRC Press), 139–65.

ROEBUCK, J.A., 1995, *Anthropometric Methods* (Santa Monica: Human Factors and Ergonomics Society).

Body Sizes of US Americans

K. H. E. Kroemer

Professor Emeritus of Industrial and Systems Engineering, Virginia Tech, Blacksburg, VA, USA

The most recent and complete set of information on the body sizes of US Americans, 19 to 60 years of age, is presented in the following table. It contains measurements in mm (weights in kg) taken by Gordon, Churchill, Clauser and co-workers in 1988 (reported 1989) on US Army soldiers, male and female. Their dimensions are close to those of the civilian population which has not been measured comprehensively for decades.

The commentary identifies the measurements and their design implications. The numbers in brackets are the same as used by Gordon et al. (1989), who also provide exact definitions of the anthropometric terms. However, in order to apply this information for design, the data must be converted to reflect actual body positions and motions instead of the standardized postures (all body angles at either 0, 90, or 180 degrees) which the subjects assumed for the measurements. Furthermore, allotments must be made for clothing, shoes, gloves, and other ''real world'' conditions (see this volume, Engineering Anthropometry).

1. Stature

The vertical distance from the floor to the top of the head, when standing. [99]

A main reference for comparing population samples. Relates to the minimal height (clearance) of overhead obstructions. Add height for more clearance, hat, shoes, stride.

2. Eye height, standing

The vertical distance from the floor to the outer corner of the right eye, when standing. [D19]

Origin of the visual field. Reference point for the location of vision obstructions and of visual targets such as displays; consider slump and motion of the standing person.

3. Shoulder height (acromion), standing

The vertical distance from the floor to the tip (acromion) of the shoulder, when standing. [2]

Starting point for arm length measurements; near the center

Table 1. Anthropometric measured data of US adults, 19 to 60 years of age (in mm).

Dimension	Men				Women			
	5th %ile	Mean	95th %ile	SD	5th %ile	Mean	95th %ile	SD
1. Stature [99]	1647	1756	1867	67	1528	1629	1737	64
2. Eye height, standing [D19]	1528	1634	1743	66	1415	1516	1621	63
3. Shoulder height (acromion), standing [2]	1342	1443	1546	62	1241	1334	1432	58
4. Elbow height, standing [D16]	995	1073	1153	48	926	998	1074	45
5. Hip height (trochanter) [107]	853	928	1009	48	789	862	938	45
6. Knuckle height, standing	na	na	na	na	na	na	na	na
7. Fingertip height, standing [D13]	591	653	716	40	551	610	670	na
8. Sitting height [93]	855	914	972	36	795	852	910	35
9. Sitting eye height [49]	735	792	848	34	685	739	794	33
10. Sitting shoulder height (acromion) [3]	549	598	646	30	509	556	604	29
11. Sitting elbow height [48]	184	231	274	27	176	221	264	27
12. Sitting thigh height (clearance) [104]	149	168	190	13	140	160	180	12
13. Sitting knee height [73]	514	559	606	28	474	515	560	26
14. Sitting popliteal height [86]	395	434	476	25	351	389	429	24
15. Shoulder–elbow length [91]	340	369	399	18	308	336	365	17
16. Elbow–fingertip length [54]	448	484	524	23	406	443	483	23
17. Overhead grip reach, sitting [D45]	1221	1310	1401	55	1127	1212	1296	51
18. Overhead grip reach, standing [D42]	1958	2107	2260	92	1808	1947	2094	87
19. Forward grip reach [D21]	693	751	813	37	632	686	744	34
20. Arm length, vertical [D3]	729	790	856	39	662	724	788	38
21. Downward grip reach [D43]	612	666	722	33	557	700	664	33
22. Chest depth [36]	210	243	280	22	209	239	279	21
23. Abdominal depth, sitting [1]	199	236	291	28	185	219	271	26
24. Buttock–knee depth, sitting [26]	569	616	667	30	542	589	640	30
25. Buttock–popliteal depth, sitting [27]	458	500	546	27	440	482	528	27
26. Shoulder breadth (bicromial) [10]	367	397	426	18	333	363	391	17
27. Shoulder breadth (bideltoid) [12]	450	492	535	26	397	433	472	23
28. Hip breadth, sitting [66]	329	367	412	25	343	385	432	27
29. Span [98]	1693	1823	1960	82	1542	1672	1809	81
30. Elbow span	na	na	na	na	na	na	na	na
31. Head length [62]	185	197	209	7	176	187	198	6
32. Head breadth [60]	143	152	161	5	137	144	153	5
33. Hand length [59]	179	194	211	10	165	181	197	10
34. Hand breadth [57]	84	90	98	4	73	79	86	4
35. Foot length [51]	249	270	292	13	224	244	265	12
36. Foot breadth [50]	92	101	110	5	82	90	98	5
37. Weight (kg), estimated by Kroemer	58	78	99	13	39	62	85	14

Note: According to Gordon, Churchill, Clauser et al. (1989), who used the numbers in brackets.

of rotation of the upper arm (shoulder joint), reference point for hand reaches; consider slump and motion of the standing person.

4. Elbow height, standing

The vertical distance from the floor to the lowest point of the right elbow, when standing, with the elbow flexed at 90 degrees. [D16]

Reference point for height and distance of the work area of the hand and the for location of controls and fixtures; consider slump and motion of the standing person.

5 Hip height (trochanter), standing

The vertical distance from the floor to the trochanter landmark on the upper side of the right thigh, when standing. [107]

Starting point for leg length measurement; near the center of the hip joint; reference point for leg reaches; consider slump and motion of the standing person.

6 Knuckle height, standing

The vertical distance from the floor to the knuckle (metacarpal bone) of the middle finger of the right hand, when standing.

Reference point for lowest location of controls, handles and handrails; consider slump and motion of the standing person.

7. Fingertip height, standing

The vertical distance from the floor to the tip of the index finger of the right hand, when standing. [D13]

Reference point for lowest location of controls, handles and handrails; consider slump and motion of the standing person.

8. Sitting height

The vertical distance from the sitting surface to the top of the head, when sitting. [93]

The vertical distance from the floor to the underside of the thigh directly behind the right knee; when sitting, with the knee flexed at 90 degrees.

Relates to the minimal height of overhead obstructions. Add height for more clearance, hat, trunk motion of the seated person.

9. Sitting eye height

The vertical distance from the sitting surface to the outer corner of the right eye, when sitting. [93]

Origin of the visual field; reference point for the location of vision obstructions and of visual targets such as displays; consider slump and motion of the seated person.

10. Sitting shoulder height (acromion),

The vertical distance from the sitting surface to the tip (acromion) of the shoulder, when sitting. [49]

Starting point for arm length measurements; near the center of rotation of the upper arm (shoulder joint), reference point for hand reaches; consider slump and motion of the seated person.

11. Sitting elbow height

The vertical distance from the sitting surface to the lowest point of the right elbow, when sitting, with the elbow flexed at 90 degrees. [48]

Reference point for height of an arm rest, of the work area of the hand and of keyboard and controls; consider slump and motion of the seated person.

12. Sitting thigh height (clearance)

The vertical distance from the sitting surface to the highest point on the top of the right thigh, when sitting, with the knee flexed at 90 degrees. [104]

Minimal clearance needed between seat pan and the underside of a structure, such as a table.; add clearance for clothing and motions.

13. Sitting knee height

The vertical distance from the floor to the top of the right knee cap, when sitting, with the knees flexed at 90 degrees. [73]

Minimal clearance needed below the underside of a structure, such as a table; add height for shoe.

14. Sitting popliteal height

The vertical distance from the floor to the underside of the thigh directly behind the right knee; when sitting, with the knees flexed at 90 degrees. [86]

Reference for the height of a seat; add height for shoes, consider movement of the feet.

15. Shoulder—elbow length

The vertical distance from the underside of the right elbow to the right acromion, with the elbow flexed at 90 degrees and the upper arm hanging vertically. [91]

A general reference for comparing population samples.

16. Elbow-fingertip length

The distance from the back of the right elbow to the tip of the middle finger, with the elbow flexed at 90 degrees. [54]

Reference for fingertip reach when moving the forearm in the elbow.

17. Overhead grip reach, sitting

The vertical distance from the sitting surface to the center of a cylindrical rod firmly held in the palm of the right hand. [D45]

Reference for height of overhead controls to be operated by the seated person. Consider ease of motion, reach and finger/hand/arm strength.

18. Overhead grip reach, standing

The vertical distance from the standing surface to the center of a cylindrical rod firmly held in the palm of the right hand. [D42]

Reference for height of overhead controls to be operated by the standing person. Add shoe height. Consider ease of motion, reach and finger/hand/arm strength.

19. Forward grip reach

The horizontal distance from the back of the right shoulder blade to the center of a cylindrical rod firmly held in the palm of the right hand. [D21]

Reference for forward reach distance. Consider ease of motion, reach and finger/hand/arm strength.

20. Arm length, vertical

The vertical distance from the tip of the right middle finger to the right acromion, with the arm hanging vertically. [91]

A general reference for comparing population samples. Reference for the location of controls very low on the side of the operator. Consider ease of motion, reach and finger/hand/arm strength.

21. Downward grip reach

The vertical distance from the right acromion to the center of a cylindrical rod firmly held in the palm of the right hand, with the arm hanging vertically. [D43]

Reference for the location of controls low on the side of the operator. Consider ease of motion, reach and finger/hand/arm strength.

22. Chest depth

The horizontal distance from the back to the right nipple. [36]

A general reference for comparing population samples. Reference for the clearance between seat backrest and the location of obstructions in front of the trunk.

23. Abdominal depth, sitting

The horizontal distance from the back to the most protruding point on the abdomen [1]

A general reference for comparing population samples. Reference for the clearance between seat backrest and the location of obstructions in front of the trunk.

24. Buttock—knee depth, sitting

The horizontal distance from the back of the buttocks to the most protruding point on the right knee, when sitting with the knees flexed at 90 degrees. [26]

Reference for the clearance between seat backrest and the location of obstructions in front of the knees.

25. Buttock—popliteal depth, sitting

The horizontal distance from the back of the buttocks to back of the right knee just below the thigh, when sitting with the knees flexed at 90 degrees. [27]

Reference for the depth of a seat.

26. Shoulder breadth, biacromial

The distance between the right and left acromion. [10]

A general reference for comparing population samples. Indication of the distance between the centers of rotation (shoulder joints) of the upper arms.

27. Shoulder breadth, bideltoid

The maximal horizontal breadth across the shoulders between the lateral margins of the right and left deltoid muscles . [12]

Reference for the clearance requirement at shoulder level. Add space for ease of motion, tool use.

28. Hip breadth, sitting

The maximal horizontal breadth across the hips or thighs, whatever is greater, when sitting [66]

Reference for seat width. Add space for clothing and ease of motion.

29. Span

The distance between the tips of the middle fingers of the horizontally outstretched arms and hands. [98]

Reference for sideway reach.

30. Elbow span

The distance between the tips of the elbows of the horizontally outstretched upper arms with the elbows are flexed so that the fingertips of the hands meet in front of the trunk.

Reference for "elbow room."

31. Head length

The distance from the glabella (between the brow ridges) to the most rearward protrusion (the occiput) on the back, in the middle of the skull. [62]

A general reference for comparing population samples. Reference for head gear size.

32. Head breadth

The maximal horizontal breadth of the head above the attachment of the ears. [60]

A general reference for comparing population samples. Reference for head gear size.

33. Hand length

The length of the right hand between the crease of the wrist and the tip of the middle finger, with the hand flat. [59]

A general reference for comparing population samples. Reference for hand tool and gear size. Consider changes due to manipulations, gloves, tool use.

34. Hand breadth

The breadth of the right hand across the knuckles of the four fingers. [57]

A general reference for comparing population samples. Reference for hand tool and gear size, and for the opening through which a hand may (or may not) fit. Consider changes due to manipulations, gloves, tool use.

35. Foot length

The maximal length of the right foot, when standing. [51]

A general reference for comparing population samples. Reference for shoe and pedal size.

36. Foot breadth

The maximal breadth of the right foot, at right angle to the long axis of the foot, when standing. [50]

A general reference for comparing population samples. Reference for shoe size, spacing of pedals.

37. Weight

Nude body weight taken to the nearest tenth of a kilogram.

A general reference for comparing population samples. Reference for body size, clothing, strength, health, etc. Add weight for clothing and equipment worn on the body.

For more information, see:

GORDON, C.C., CHURCHILL, T., CLAUSER, C.E., BRADTMILLER, B., MCCONVILLE, J.T., TEBBETTS, I., and WALKER, R.A., 1989, *1988 Anthropometric Survey of U.S. Army Personnel: Summary Statistics Interim Report.* Natick-TR-89/027 (Natick, MA: US Army Natick Research, Development and Engineering Center).

KROEMER, K.H.E., 1999, Engineering anthropometry. Chapter 9 in W. Karwowski and W.S. Marras (eds) *The Occupational Ergonomics Handbook* (Boca Raton, FL: CRC Press), pp.139—65).

ROEBUCK, J.A., 1995, *Anthropometric Methods* (Santa Monica, CA: Human Factors and Ergonomics Society).

Control of Rapid Actions: Motor Programming

R. A. Schmidt†, T. D. Lee‡ and D. E. Young†

†Exponent Failure Analysis Associates, Inc., 5401 McConnell Avenue, Los Angeles, CA 90066, USA

‡Department of Kinesiology, McMaster University, 1280 Main Street West, Hamilton, Ontario, Canada L8S 4K1

In the study of movement control, it is often convenient to subdivide actions into two, broadly defined classes: (1) those with brief execution times and (2) those with longer execution times. Brief actions have movement durations of a few hundred milliseconds or less (such as a throw, button press, pedal movement), whereas the longer durations have movement times (MT) > ~5 s and may actually require several minutes for their execution (e.g. juggling). These two classes of actions are sometimes termed "fast" (or "rapid") versus "slow," and this terminology is adopted here. But it should be clear that the defining characteristic is the duration (in seconds), and is not their actual speed (cm/s). While many brief actions are also of very high velocity (e.g. a batswing), they need not be (pressing an accelerator pedal).

These two movement classes follow very different principles. The major difference is that the so-called rapid movements are essentially organized in advance and run off as a unit. The slower actions tend to be more controlled by interactions and compensations for sensory information delivered during the conduct of the action (termed "response-produced feedback," or simply "feedback"), and hence they do not rely (so) heavily on advance organization. Of course, this distinction between the processes in rapid and slow movements blurs at their border. Movements that are in the region near this border (with MT of from ~400 to ~1500 ms) have components that are both pre-organized and feedback controlled (see, e.g., Principles of Simple Movement, Chapter xx in this volume). The factors that determine the relative contributions of these two kinds of control processes have been examined and modeled recently (Meyer *et al*. 1990). But, at the polar extremes, rapid and slow movements appear to be quite differently controlled. We will be dealing with rapid actions in this section. Other sections in this volume deal with those processes in slower actions (see, e.g., Tracking, Chapter yy).

1. FEEDBACK CONTROL IN RAPID ACTIONS

The rationale supporting the idea that rapid actions are organized in advance and run off as a unit are both simple and compelling. Probably the most important reason is the inadequacy of sensory information, or response-produced feedback, to serve as the basis for the action's control (as in a closed-loop system). Feedback-control viewpoints have come in many different forms throughout the 20th century (for a review, see Schmidt and Lee 1999). However, all require that the response-produced sensory information be transduced, travel to the CNS (or at least to the spinal cord), be integrated with other information, be delivered to the muscles, result in a different pattern of contraction, and eventually produce some effect on the movement of bones and joints. The problem is that these processes are generally too slow to be a part of the control of these brief actions, and a 200-ms action is essentially over before the feedback can have any effect on the action.

About 20–30 years ago there was considerable debate about this issue, as it is at least possible for some "reflex"-like activities to be delivered to and from the CNS in from 30–80 ms after a perturbation to a movement was applied, providing some encouragement that the movements were controlled via feedback. More recent evidence and thinking, however, has revealed that these-lower level feedback processes are mainly involved in the maintenance of the already organized actions, and are not the basis (or the origin) of the pattern. Various levels of feedback involvement can be identified, each with various contributions to the motor output (Schmidt and Lee 1999). But the fundamental question has centered on the issue of what was being modified by the feedback information. The answer usually given is that some prestructured, centrally generated pattern of activity is produced, and the feedback modifies it in various ways.

Several other lines of evidence also converge on the idea that at least rapid movements are organized in advance. Earlier work with deafferented monkeys, for whom sensory information from the responding limb was eliminated surgically, revealed that these animals moved very well, and almost normally in some situations. They had particular problems with fine-controlled finger actions, however, suggesting that feedback is a relatively large part of these slower actions. In humans, electromyographic (EMG) patterns from the responding limb in rapid tasks that were unexpectedly blocked mechanically are essentially normal (i.e. like the unblocked movements) for ~150 ms, revealing that the patterning was presumably organized in advance. Importantly, this finding suggests that neither the feedback from, nor dynamics of the responding limb can be responsible for the EMG patterning of brief actions, as such patterning would surely have been massively disrupted by blocking the limb at the beginning. In a different paradigm with humans, the time to prepare an action during reaction time is dependent on the complexity (and/or duration) of the movement to be made, as if the organization of a more complex action required more time for organization before the movement actually starts. Finally, research on animal locomotion uses an analogous concept of central pattern generators (CPG) to account for patterns of action. This concept is quite similar to the idea of motor programs for human motor control. All of this thinking has led to the concept of the motor program, the name most often used for the structure that is presumably organized in advance. See Schmidt and Lee (1999) for a review.

2. GENERALIZED MOTOR PROGRAMS

Given that some sort of motor program is organized in advance and "run off" as a unit, considerable research and thinking has focused on the nature and organization of these structures, particularly using rapid actions in relatively stereotyped environments. One important line of thinking concerns the observation that, at least in its original form, the motor program is very inflexible, so that essentially countless movement programs would be required for all of the motor actions humans can do. This, of course, raised issues about where (and how) in the CNS all of these programs could be stored (the so-called storage

problem). Second, there was concern that such programs could not generate novel behavior (the novelty problem), where the production of an action is "new" and not precisely like any of the similar actions done previously. These ideas have suggested the notion that programming is flexible in various ways, leading to the idea of the generalized motor program (GMP). According to this idea, a given GMP consists of a rigidly defined (or invariant) fundamental pattern that determines the form of an action, but this pattern can be varied — or scaled — in several, relatively superficial ways, each of which allows the action to meet slightly different environmental demands, as described below.

2.1. Relative Timing Invariances

A key observation has been that most actions can be sped up or slowed down easily (e.g. via instructions) without changing the fundamental pattern of action. When speed is scaled, several features remain constant, or invariant. One of these is the order of events. But the most important is relative timing, or the fundamental temporal structure of the action. For relative timing to be invariant, the durations of any subpart of the action must remain in constant proportion to the overall movement time as the action is changed in duration. It is as if the whole action is sped up as a unit, with all parts of the action changing their speed in the same proportion. Relative timing stays roughly invariant over large changes in movement time for tasks that are initially rapid and of short duration. Relative timing is not perfectly invariant, however, and small but systematic deviations from invariance have been documented. However, many have felt that the conspicuous tendency toward invariance provides a good approximation of these processes. If the actions are longer in duration (e.g. juggling, reaching), relative timing sometimes breaks down completely, as one would expect if feedback-based processes were intervening to modify the originally programmed movement pattern. Often, though, relative timing remains roughly invariant even in long-duration tasks of several seconds, such as in typing, certain speech tasks, stair climbing, etc. Such tasks presumably can be programmed, but of course they need not be; and their conduct can be interrupted easily if the action goes awry.

These relative timing invariances have been considered as the fundamental structure, or "signature" of a GMP. Each different class of activity (throwing versus kicking, for example) is argued to have its particular GMP, each with its unique and invariant relative timing. Therefore, as we discuss later, this observation provides a basis for deciding when a particular GMP is being produced (its unique relative timing pattern is observed), or when it has stopped or been replaced by another GMP (with a different relative timing pattern). Extensions of this idea show that two GMP running sequentially can be distinguished using the identification of the relative timing invariances of each (Schneider and Schmidt, 1995).

2.2. Parameters of the GMP

According to this view, when actions are controlled by GMP, they must be parameterized (or scaled) before initiation so that they meet the environmental demands. Research and common observation suggests that actions can be essentially linearly scaled in several ways. The most important of these parameters determines the rate, or the duration of the action, mentioned

above. The changes in movement speed are such that the durations of the subparts remain in constant proportion to the overall movement duration. Another is an amplitude parameter, by which the amplitude of all of the features of an action are increased in constant proportion so that the movement can be varied in size, or can be varied in terms of the masses or frictions encountered, without changing the fundamental temporal structure. Large and small handwriting has been modeled in this way by using a force parameter to scale the sizes of all of the muscular contractions in an action. Finally, many have argued that the GMP requires a limb parameter so that the action can be produced in either the right or left hand, or perhaps even with other effector systems (the mouth, foot, etc.), a phenomenon known as motor equivalence. Because the GMP can be involved in the control of various limbs, it is regarded as an abstract representation of the action patterning, not containing muscle- or effector-specific instructions.

3. WHAT IS CONTROLLED IN MOTOR PROGRAMMING?

Two general models of what is controlled by the motor program have been presented. One view argues that the program controls impulses, while the other argues that the system controls limb position.

3.1. The Control of Impulses

Considerable evidence suggests that a GMP controls impulses. From physics, the impulse is the integral of the force–time curve (or the area "under" the curve), and it has the property that the velocity of an object (starting from rest) after an impulse has stopped acting on it is proportional to the impulse size. Therefore, by adjusting the size of the impulse — either in force or in duration, both of which are physiologically plausible — the system can control velocity simply and indirectly. Further, the GMP under this view controls the temporal distribution of forces to the various musculature, defining when the particular forces are to be applied, for how long and in what order. This viewpoint, wherein a pattern of impulses is distributed over time, is called the impulse-timing view. Furthermore, this pattern of impulses over time can apparently be scaled almost linearly by the various parameters mentioned in the previous section, preserving the fundamental patterning.

3.2. Equilibrium–Point Control

A possible alternative viewpoint for the nature of limb control has been termed the equilibrium-point model (or the mass-spring model). According to this view (e.g. Feldman and Levin 1995), the control of single-joint actions is produced by programming an equilibrium point; the spring-like properties of the muscles on the two sides of a joint then achieve that position for which the torques are equal and opposite (the so-called equilibrium point). In many ways, this is an attractive alternative to impulse-timing views, as no central timing is required for action. However, it appears to have limitations (1) when two-joint actions (e.g. elbow–shoulder) are required, (2) when the action requires sequential control of rapid movements with reversals in direction, or (3) when the action involves the coordination among separate limbs. It appears that some higher-level timing of the contractions is important for control in these situations (for a discussion, see

Schmidt and Lee 1999), and that the equilibrium-point view does not have sufficient generality at present for it to be taken as an effective account of movement control.

4. THE CONTROL OF COORDINATION

Coordination, or the patterning of various limbs together to form a coherent whole action, also appears to differ markedly depending of the duration of the actions. When the actions are slow and continuous, such as swimming or juggling, coordination appears to be related strongly to interactions among the limbs, perhaps with a feedback basis. Presumably, sensory information plays a strong role in on-going control, and the actions are not programmed. On the other hand, rapid coordinated action of the two hands has been modeled recently in terms of the GMP viewpoint described earlier. Here, the GMP is a single structure that provides impulse-timing information for both limbs, so that speeding up the coordinated task speeds both limbs (and all of their subparts) proportionally, and the relative timing both within- and between-limbs remains invariant (Heuer *et al.* 1995, Schmidt and Lee 1999). Such a view has potential generalization to far more complicated tasks involving the upper and lower extremities, where complex actions can be thought of as programmed in the same general way as simpler ones.

5. PRACTICE FOR GENERALIZED MOTOR PROGRAMS AND PARAMETERS

Practice, of course, has strong influences on the control of actions, and this is also understandable in terms of the changes in the GMP and parameters. When we learn such actions, we appear to learn at least two, separate things. First, we acquire some fundamental pattern for the action; in throwing, we learn the coordination patterns among arm, trunk, and legs that leads to an effective throw. Theoretically, we need to acquire a GMP for the class of actions. Second, we learn to scale the pattern so that we can achieve various environmental goals, such as throwing heavy versus light objects, long versus short distances or rapidly versus slowly. That is, we need to learn how to parameterize the GMP. Recent research has suggested that the fundamental learning of these two facets of motor responding are distinct, and that their acquisition follows different principles (e.g. Wulf *et al.* 1993). Some of these differences are highlighted next.

We know from common experience, and also from empirical data, that the acquisition of the fundamental patterning (or the GMP) is the most difficult and time-consuming aspect of practice. When the tasks are new, different from other actions already learned, and complex, it may require considerable time and effort for learners to be able to acquire the fundamental patterning. In a similar way, it requires considerable practice for the learner to change the fundamental patterning once it has been learned, as

in eliminating a "bad habit" in one's golf swing once acquired. And, under conditions of stress or fatigue, the "old" pattern can emerge to replace the "new" one that was just learned. The GMP appears to be acquired most effectively using randomized (or spaced) practice sequences, where feedback is withheld frequently to avoid the learner becoming dependent on it. See Skill Learning: Conditions of Training, Chapter yy, elsewhere in this volume for a summary of the most important learning methods.

Once the pattern is learned, it is almost trivial to scale it to match the environmental demands. On the one hand, it is very easy to change the speed or forcefulness of an already learned GMP. The problem for the learner is that s/he must acquire the capability to scale the parameters of the GMP to match the particular environmental demands. This kind of learning has been termed schema learning, where the individual acquires a relationship, or rule (the schema), defined between the environmental conditions (say, the distance that the object is to be thrown) and the parameter for the GMP required to produce a throw of that distance. Research has shown that the most effective practice leading to schema learning uses many repetitions of the same GMP, but under varied environmental conditions that require different parameters (so-called variable practice). As might be expected, the goal is to practice many variants of the action so that the system can build an effective and accurate rule between the environmental demands and the parameters provided to the GMP. Also, unlike GMP learning, frequent feedback does not seem to be a limitation for parameter learning.

REFERENCES

FELDMAN, A.G. and LEVIN, M.F., 1995, The origin and use of positional frames of reference in motor control. *Behavioral and Brain Sciences*, 18, 723–806.

HEUER, H., SCHMIDT, R.A. and GHODSIAN, D., 1995, Generalized motor programs for rapid bimanual tasks: a two-level multiplicative-rate model. *Biological Cybernetics*, 73, 343–56.

MEYER, D.E., SMITH, J.E.K., KORNBLUM, S., ABRAMS, R.A. and WRIGHT, C.E., 1990, Speed–accuracy trade-offs in aimed movements: toward a theory of rapid voluntary action. In M. Jeannerod (ed.) *Attention and Performance XIII* (Hillsdale: Erlbaum), 173–226.

SCHMIDT, R.A. and LEE, T.D., 1999, *Motor Control and Learning: A Behavioral Emphasis* (3rd ed.) (Champaign, IL: Human Kinetics).

SCHNEIDER, D.M. and SCHMIDT, R.A., 1995, Units of action in motor control: role of response complexity and target speed. *Human Performance*, 22, 805–17.

WULF, G., SCHMIDT, R.A. and DEUBEL, H., 1993, Reduced feedback frequency enhances generalized motor program learning but not parameter learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 19, 1134–50.

Dynamic Muscle Strength

J. L. Smith

Department of Industrial Engineering, Texas Tech University, Lubbock, TX 79409-3061, USA

Since most industrial activities are dynamic in nature, dynamic strength testing has a great deal of appeal to the ergonomist. Dynamic strength tests are somewhat difficult to perform due to the variety of factors that must be accounted for or controlled. For example, the speed of movement, range of movement and the body posture will all affect the strength test results. Historically, dynamic strength tests have been classified as: isokinetic, isotonic and iso-inertial. Although some have argued over the appropriateness of the names, they are widely used and understood as defined below.

Isokinetic tests refer to testing performed under conditions of moving at a constant velocity. Isokinetic devices restrict the speed of movement. Therefore, when a subject is instructed to move as fast and as hard as possible, the device will limit the speed of movement and the additional muscle force will be recorded typically as a torque curve throughout the range of the motion. Isokinetic devices can be relatively expensive, require trained personnel to operate, and have long set-up times for the subjects, but they can provide an excellent "strength profile" along the path of motion at a fixed velocity of movement.

Isotonic strength tests refer to testing performed under conditions of constant muscle tension. Such devices alter the tension or resistance of the device at different points along the motion path, resulting in a constant tension required by the subject performing the test. Muscles generate different tensions (or strengths) at different lengths of the muscle. For example, if a muscle is stretched or compressed, it will be able to generate less tension than when it is at its resting length. Therefore, the iso-inertial devices attempt to alter the resistance of the strength-testing device in order to require the subject to exert a constant muscle force throughout the range of motion. Isotonic devices do not control speed of movement. Isotonic devices tend to be relatively expensive and are not widely used for industrial strength testing.

Iso-inertial devices refer to testing conducted on constant external loads. Typical iso-inertial testing devices are free weights and devices that use weights to change the external load. The US Air Force developed an incremental 6 foot lifting machine (figure 1) that was an iso-inertial device that restricted motion of the load to the vertical plane, by having handles mounted to a carriage that rode within two vertical rails (Ayoub *et al*. 1987). Adding or subtracting 10-lb weights on the carriage changed the external load. The subject was instructed to lift a load to a height of 6 feet and then lower the load. The test administrator then added more weight and the subject repeated the task until he/she reached a point where the load could no longer be successfully lifted to a height of 6 feet. The heaviest load successfully lifted to the 6-foot height became that subject's strength score. Test scores from the incremental 6-foot lift were found to be good predictors of manual materials handling capabilities in a variety of tasks. The incremental 6-foot lifting machine is relatively inexpensive, the tests are easy and fast to conduct and requires very little training for the test administrator.
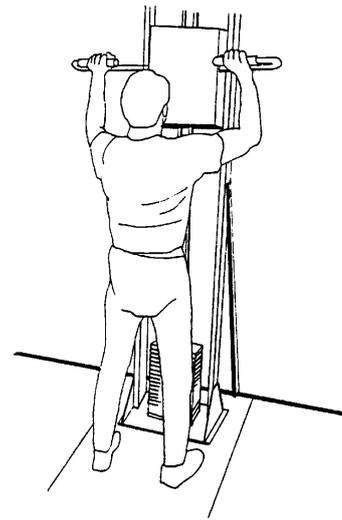
Figure 1. Incremental 6-foot weight lifting machine.

In cases where engineering and administrative controls are not feasible, strength training of workers may be a consideration if their capabilities are exceeded by job demands and the chances of injury are unacceptable. A sound strength training program will adhere to two basic physiological principles: specificity of training and the overload principle. Specificity of training states that the best training is actually using the activity that the worker is being trained for. As an example, a worker involved in lifting objects should be trained on a lifting task under conditions similar to those encountered on the job. Thus, the specific muscle groups and energy systems that will be utilized on the job are the systems being trained. The overload principle states that adaptations will take place and that to see any further increase in strength, more stressful workloads must be encountered. For example, a worker can be trained to lift 20 kg, but will soon adapt to that load and receive no further training benefits until the load is increased.

For dynamic muscle strength, a popular strength training protocol utilizes the repetition maximum (RM) approach. Under such an approach, the maximum amount of weight that can be handled for a fixed number of repetitions is established. For example a 10 RM protocol would require that the maximum amount of weight that can be handled 10 times under a given set of task conditions be determined. The training session would then be for the worker to perform 10 repetitions at one-half of the 10 RM load, followed by 10 repetitions at three-quarters of the 10 RM load, and finally 10 repetitions at the 10 RM load. Such a protocol allows a "warm up" by starting the worker at a lower load and increasing to the maximum load. When a subject can perform more than 10 repetitions at the 10 RM load, a new 10 RM should be established. For greater training effects a 3 or 5 RM protocol utilizing heavier weights could be utilized. By the same account a less stressful 12 or 15 RM protocol would be more conservative and start subjects with a lighter load.

Whether a strength training program would be appropriate depends on the industry and the worker population. However, some sort of strength training incorporated along with skill training could be beneficial in reducing the risk of injury for new employees.

REFERENCES

AYOUB, M.M., JIANG, B.C., SMITH, J.L., SELAN, J.L. and McDANIEL, J.W., 1987, Establishing a physical criterion for assigning personnel to U.S. Air Force jobs. *American Industrial. Hygiene Association Journal*, 48, 464–70.

CHAFFIN, D.B. and ANDERSSON, G.B.J., 1991, *Occupational Biomechanics* (2nd ed.) (New York: John Wiley).

MITAL, A., GARG, A., KARWOWSKI, W., KUMAR, S., SMITH, J.L. and AYOUB, M.M., 1993, Status in human strength research and application. *IIE Transactions*, 25, 57–69.

SMITH, J.L. (ed), 1990, "Human Strength, Special Issue". *International Journal of Industrial Ergonomics*, 6.

Dynamic Properties of Human Limb Segments

M. K. Lebiedowska

Southern Illinois University School of Medicine**,**

Division of Orthopedics and Rehabilitation, PO Box 19652, Springfield,

IL 62794-9652, USA

1. INTRODUCTION

Every living system is under the influence of a variety of physical fields. The information about the environmental changes (external and internal) detected by various receptors is transferred to different levels in the nervous system, and the decision to react or not to react is taken. Any human activity in the external environment involves the motor system. The utilization of the motor system allows humans to change their position in space (locomotion) and to interact with the environment in a controlled manner (work). The human motor system is complex, with about 280 biomechanical degrees of freedom (number of coordinates necessary to describe the body's motion in space) and thousands of neural channels. Such a complex system contains many subsystems, inputs, and outputs; it is a multilevel system.

The motor system has the ability to learn and to change its structure. It can act in a very wide range of external conditions while changes to the internal conditions are kept within the small ranges necessary to maintain the basic living functions (homeostasis). There exists no consistent theory to analyze these systems. The properties of the human motor system are described by its biomechanical parameters. The central nervous system (CNS) monitors biomechanical variables such as force, position, and velocity. Intensive studies are currently investigating the manner in which the CNS maintains these variables constant, stable, and reproducible during a variety of motor tasks.

Even though the control rules are unclear, quantification of the human motor system is under intensive development. There are two basic subsystems: structural and functional. The structural system contains morphological passive and active musculoskeletal elements and can be described with a large but limited number of parameters. The functional system describes the spatial and temporal ordering of the structural elements and is practically infinite. Quantitative description of the system can be achieved using its structural and functional subsystems.

The basic structural properties of the motor system include segment dimensions, ranges of motion, segment dynamics (moment of inertia, stiffness, damping), maximum voluntary and reflex activation. At the range of the functional subsystem, the most often quantified human functions include walking, upright standing, initiation of walking, reaching, climbing stairs, and grasping. This article looks at how to quantify human segment dynamics using moment of inertia, stiffness, and viscous damping; then it examines how these parameters are affected by factors such as activation, angular position, and human growth.

2. DYNAMIC PROPERTIES OF HUMAN LIMB SEGMENTS

Human ability to perform any motor task depends on many factors, including biomechanical properties of the body. Segments (rigid parts of the body) operate around their joints through the muscles which are under neural control. The segment endpoint is usually the final executor of any motor task. The segment endpoint moves in space with different tasks, depending on the criteria (time, precision, etc.), or it generates force on the external environment. In return, forces generated by the environment affect the execution of the motor task. The inertial and viscoelastic properties of human limb segments describe the dynamic properties of the system, which is under neural control.

Neural control signals are transformed into motor tasks through the mass, inertial, and viscoelastic properties of joint segments. The active and passive elastic structures were originally defined at the level of the individual muscle by Hill (1940). In human bodies the biomechanical properties can be evaluated at the level of a limb segment and its joint. The viscoelastic passive properties are characteristics of muscle tissue, joint capsule, tendons, and skin (Johns and Wright 1962). The active dynamic properties are related to the state of the segment's muscle activity: relaxation or activation (voluntary or reflex). Knowledge about these properties is important in designing any devices that interact with the environment (ergonomics) and designing compensatory control loops for patients (rehabilitation engineering).

The dynamics of a human limb segment are defined by its generalized equation of motion:

$$Ja'' + (B_p + B_a)a' + (K_p + K_a)a + MgR = Tq \qquad (2.1)$$

or

$$a'' + ba' + wa + g = tq \qquad (2.2)$$

where

$$b = (B_p + B_a)J^{-1} \qquad (2.3)$$
$$w^2 = (K_p + K_a)J^{1} \qquad (2.4)$$
$$g = MgRJ^{1} \qquad (2.5)$$
$$f = FJ^{1} \qquad (2.6)$$

J = moment of inertia around the axis of rotation. Moment of inertia defines the mass distribution along the segment long axis. It is equal to a sum of the element masses multiplied by the square of the radius of rotation

K_p, B_p = passive viscous damping and stiffness of the muscles acting on the joint when muscles are relaxed (nonactive). In static conditions the stiffness may be defined as a ratio between the change in force and the change in motion which was caused by it. Viscous damping can be defined as a ratio between the change in force and the velocity caused by it

K_a, B_a = active stiffness and damping of the muscles acting on a joint; the changes of stiffness are due to muscle activation (voluntary or reflex)

M = segment mass

R = radius of mass

a, a', a'' = angular motion, velocity, and acceleration, respectively

F = external torque (force multiplied by the radius of action)

w = natural free frequency of the segment, $w = (K/J)^{1/2}$

x = dimensionless damping ratio, $x = B/2(KJ)^{1/2}$

2.1 Measuring moment of inertia, stiffness, and viscous damping.

The generalized equation of motion (2.1) is usually reduced in the measurement techniques. The reduction concerns the biomechanical degrees of freedom, the viscoelastic properties of

Table 1

Segment	Axis of rotation	Moment inertia (N m)	Stiffness (N m rad^{-1})	Damping (N m s rad^{-1})	Free frequency, ω (rad s^{-1})
Finger	II metacarpophalanger	–	0.1–0.8	0.005–0.008	30
Hand	Wrist	0.015	4–88	0.04–0.1	16–76
Forearm	Elbow	0.037–0.082	1–12	0.1–0.3	6–16
Upper limb	Shoulder	0.29	10	1.1–1.7	5.8
Foot	Ankle	0.027–0.07	17–58	0.2–0.42	26–46
Lower Leg	Knee	0.32–0.58	10–120	0.35–1.85	5–14

[a]Around proximate axis in flexion /extension plane.

the muscles (relaxation or steady effort) and gravitational forces (unloading or constant loading condition). The methods for identifying the parameters of the reduced equation depend on the choice of structural elements, the complexity of the model, the identification procedure, and the technical methods of force or motion analysis being applied. The identification procedures are based on the analysis of the system response to different types of external stimuli.

Methods based on the geometrical segment dimensions and appropriate anthropometric measures were used in measuring the moment of inertia of different segments in cadavers and in living subjects. Moment of inertia may be predicted from these studies (Conwey 1972) or measured with different techniques. Here are some of them: peak acceleration after quick release of a forearm following a fixed force; acceleration and force of a limb segment undergoing sinusoidal displacement (Allum and Young 1976); and calculation of lower leg moment of inertia from the period of free segment oscillations with additional mass elements.

Static stiffness, defined as a coefficient between the change of force and displacement, has been measured in the elbow, knee, ankle, metacarpophalangeal, and hip joints with the aid of complex torque devices or hand-held force transducers (MacKay *et al.* 1986). An apparatus supplying constant velocity displacement (the inertial properties are negligible in the equation of motion) and measuring force (or vice versa) has been used in ankle joint stiffness measurements. The resistance to sinusoidal movement, or the movement as a result of sinusoidal forces was used to measure interphalangeal, elbow, wrist, ankle, hip, and knee muscle stiffness (Johns and Wright 1962). Force, displacement (step, pulse) or random perturbation functions were used to identify the viscoelastic properties of interphalangeal, wrist, elbow, finger, and ankle joints (Hayes and Hatze 1977). A popular method involves driving the limb system with increasing frequency sinusoidal force to determine its natural frequency.

Dynamic stiffness can be calculated from the free frequency with an estimated or measured moment of inertia. This method was used to identify the stiffness of muscles acting on the ankle, wrist, elbow, knee, and interphalangeal joints (Lakie *et al.* 1981).

3. DYNAMIC PARAMETERS OF DIFFERENT LIMB SEGMENTS.

3.1 Dynamic parameters of selected body segments in adult populations

Table 1 shows dynamic parameters of different limb segments based on results available in the literature. All parameters were measured for joint motion with one degree of freedom, in the flexion/extension direction around the proximal axis of rotation. The discrepancy of the data reported by different authors is re-

lated to the fact that the data sets were collected using different measurement methods and at different ranges of motion. When a joint is stretched from any resting position by external force, its force increases much more rapidly at the beginning. It is reported in many studies (MacKay *et al.* 1986) that the stiffness (K) rapidly decreases with the amplitude (A) for movements up to 0.1 rad (when $K = A^{0.8}$) and then gradually decreases. This observation might be related to the breakdown of a short-range elastic component (Hufschmidt and Schwaller 1987). Segment masses, centers of mass, and moments of inertia can also be predicted from the anthropometric measures (Conwey 1972).

3.2 Changes of viscoelastic properties related to the joint position

Many studies (Hayes and Hatze 1977) have shown that the viscoelastic properties, when studied over the whole range of joint motion, exhibit highly nonlinear properties. Both passive stiffness (figure 1a) and passive damping (figure 1b) increase with the angular position until the passive joint and bone structures limit the range of motion. At the midrange, both parameters exhibit minimal values. The passive stiffness exhibits hysteresis, i.e., the passive moment is different when moving from flexion to extension than when moving from extension to flexion; this indicates energy loss during the displacement cycle. It might be related to plastic changes within the muscle (yielding property, thixotropy).

3.3 Dynamic active parameters of some body segments in adults

In an isolated animal muscle preparation, the stiffness increases proportionally with muscle activation. The dynamic properties of different segments were measured during muscle relaxation and voluntary and reflex activation. During muscle activation, the stiffness and damping of the limb segment increase as a result of structural changes within the muscle. The increase in stiffness with the muscle activation was found for hand, forearm, foot, leg, and lower leg segments (Kearney and Hunter 1990). The joint exhibits elastic behavior up to yield stress, and viscous behavior above yield stress. The increase in the stiffness (figure 2a) and damping (figure 2b) during muscle activation might be related to the increasing number of elastic cross-bridges formed between myofilaments and the increase in the active areas of sliding myofilaments. However, more data is needed to describe the active dynamics of different joint segments at higher levels of muscle activation.
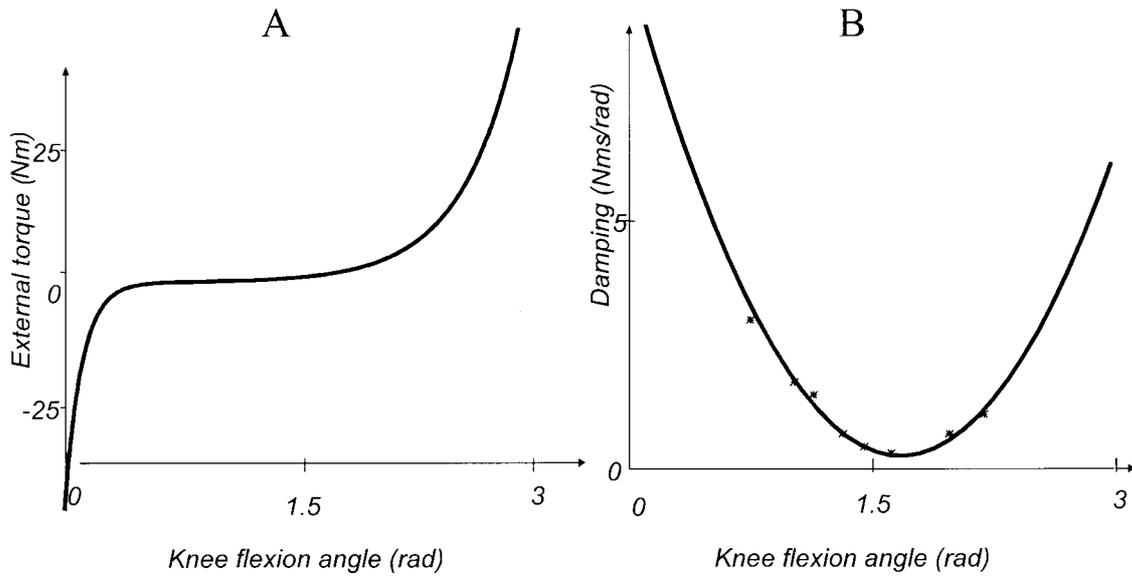
Figure 1. Knee joint: (a) passive stiffness, (b) passive damping.
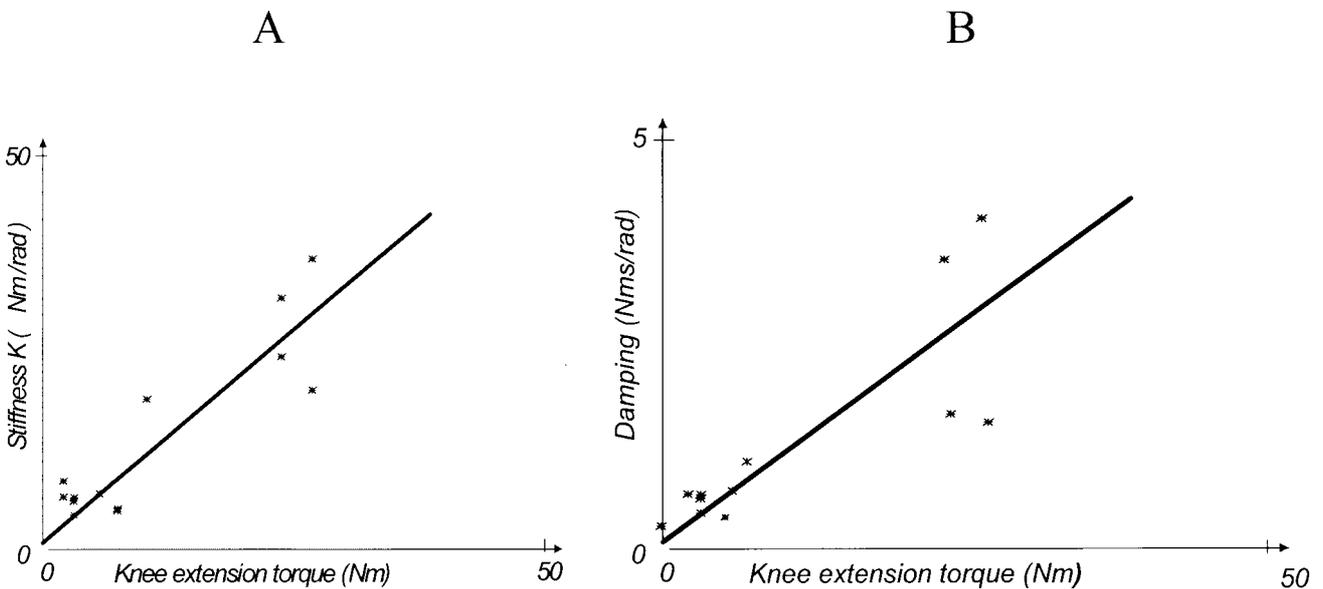
Figure 2. Knee muscles: (a) active stiffness of knee flexors and extensors, (b) active damping of knee extensors.

3.4 Dynamic passive parameters of some body segments in children

The parameters of human motor systems undergo changes during the lifetime as a result of ontogenetic development, aging processes, and diseases. There is very limited data on changes in early development (ages 0 to 6) and aging (above 55 years) (Schneider and Zernicki 1992, Jensen and Fletcher 1994). For lower leg (Lebiedowska and Fisk 1999) and forearm, the moment of inertia (J), stiffness (K), and viscous damping (B) increase as the fifth power of body height in children aged 6 to 18 years.

	Moment of inertia	Passive stiffness	Passive damping
Lower leg	$J = 0.03H^5$	$K = 4.59H^5$	$B = 0.06H^5$
Forearm	$J = 0.0049H^5$	$K = 1.16H^5$	$B = 0.032H^5$

These figures confirm the proportional growth of human body segments in children aged 6 to 18. When the segment's motion

(a) is restricted to the horizontal plane and/or the torque generated by the muscles is much larger than the torque needed to overcome gravity (fast movements), the segment dynamics may be described by the following equations:

Lower leg: $a'' + 2a' + 165a = T_{ext} / (0.03H^5) + aTq_{max} / (0.03H^5)$ (3.1)

Forearm: $a'' + 5.8a' + 219a = T_{ext} / (0.0049H^5) + aTq_{max} / (0.0049H^5)$ (3.2)

Thus, the dynamics are an invariant property of the growing human motor system. Similar motor activity with respect to the shape, amplitude, and duration can be expected as a result of the proportional propelling of a segment. The same activation patterns (3.2) can be used in children during development (age 6 to 18).

Other studies found that the maximal isometric knee torque Tq_{max} (at the midrange) in children changes as the fourth power of body height. Thus, to obtain the same angular joint motion, proportional scaling of activation to body height can be applied. The same linear displacement of the segment endpoint can be

obtained with the same activation. Although not experimentally confirmed, this should also be true in subpopulations characterized by different average body sizes (as long as the body shape is geometrically similar).

On the other hand, the sensitivity of a human operator to external torque (Tq_{ext}) is scaled proportionally to segment inertia or the fifth power of body height. This means that the same torque applied to the same body segment of a smaller person would cause larger motion than in a taller person (scaled by the fifth power of the ratio between their body heights).

3.5 Changes in passive dynamic parameters caused by disease

An increase in passive stiffness was documented in patients with rheumatic diseases and Parkinson's disease. There were several studies investigating the changes in passive stiffness and viscous damping in human spastic paresis. The inconsistency of findings might be related to the application of different measurement methods and different patient populations (Douglas *et al.* 1989, Lebiedowska and Fisk 1999). However, for differential diagnosis of spasticity, it is better to use the changes of active stiffness evoked by application of external force with increasing velocity of stretch.

4. RECOMMENDATIONS

4.1 How to choose a measurement method

Different measurement methods are available at the moment: some of them involve expensive, complex torque devices, and some of them use simple hand-held transducers. When choosing the measurement method, researchers should analyze the aim of the study carefully. In basic research studies choose the more accurate and more complex methods; whereas in clinical applications choose the simpler and less time-consuming methods. The viscoelastic properties exhibit highly nonlinear properties, and special attention should be paid to the range of motion for each measurement method. It is well known that viscoelastic results obtained with a small range of motion can be 10 times larger than results obtained with a large range of motion.

4.2 How segment dynamics limit the maximal voluntary velocity

The range of segment motion is restricted (limited) by the anatomical ranges of motion (joint capsule, bones, skin). The maximal segment velocity (a'_{max}) is limited by the level of muscle activation and segment dynamics. These relations depend on the trajectory of the motion. For motion (a) which is sinusoidal (flexion/extension motion in the same range as A), the relation can be estimated as $a'_{max} = Aw$. This means that the segment free frequency w (2.4) and the range of the motion limit the maximal velocity which can be exerted by the joint. The ballistic movements (from one position to another) at the initial part (generation of maximal velocity) might be considered to be governed by the same rule. Other task criteria (e.g., precision) also modify the maximal velocity of a segment.

4.3 How to use the dynamic parameters of limb segments

The dynamic parameters are necessary to quantify the properties of the human motor system in ergonomics, rehabilitation, and sports medicine. The dynamic parameters collected in different populations might be used in ergonomics to design the optimal workplace. Rehabilitation ergonomics deals with the optimization of different designs, targeting selected patient populations, again characterized by their biomechanical parameters. Quantitative methods are needed in medical rehabilitation to make objective evaluations about the effectiveness and efficiency of different rehabilitation methods, and about the progress of rehabilitation for an individual patient; they also help to choose the best methods of improvement. Rehabilitation engineering uses the motor system parameters to design the compensatory control loops to regain or improve the patient's performance.

REFERENCES

ALLUM, J.H. and YOUNG, L., 1976, The relaxed oscillation technique for the determination of the moment of inertia of limb segments. *Journal of Biomechanics*, 9, 21–25.

CONWEY, B.A., 1972, Development of Skylab Experiment T–013: Crew/Vehicle Disturbances. NASA Technical Note D–6584.

DOUGLAS, A.J., WALSH, E.G., WRIGHT, G.W., and EDMOND, P., 1989, Muscle tone around the human knee in paraplegia. *Quarterly Journal of Experimental Physiology*, 74, 897–905.

HAYES, K. and HATZE, H., 1977, Passive visco-elastic properties of structures spanning the human elbow joint. *European Journal of Applied Physiology*, 37, 265–74.

HILL, A.V., 1940, The dynamic constant of human muscles. *Proceedings of the Royal Society B*, 128, 263–74.

HUFSCHMIDT, A. and SCHWALLER, I., 1987, Short-range elasticity and resting tension of relaxed human lower leg muscles. *Journal of Physiology*, 391, 451–65.

JENSEN, R.K. and FLECHER, P., 1994, Distribution of mass to the segments of elderly males and females. *Journal of Biomechanics*, 27, 89–96.

JOHNS, R.J. and WRIGHT, V., 1962, Relative importance of various tissues in joint stiffness. *Journal of Applied Physiology*, 17, 824–28.

KEARNEY, R.E. and HUNTER, I.W., 1990, System identification of human joint dynamics. *Critical Reviews in Biomedical Engineering*, 18, 55–87.

LAKIE, M., WALSH, E.G., and WRIGHT, G.W., 1981, Measurement of inertia of the hand and the stiffness of the forearm using resonant frequency methods with added inertia or position feedback. *Journal of Physiology*, 310, 3P–4P.

LEBIEDOWSKA, M.K. and FISK, J.F., 2000, Passive dynamics of the knee joint in healthy children and children affected by spastic paresis, *Clinical Biomechanics*, forthcoming.

MACKAY, W.A., CRAMMOND, D.J., KWAN, H.C., and MURPHY, J.T., 1986, Measurement of human forearm viscoelasticity. *Journal of Biomechanics*, 19, 231–38.

SCHNEIDER, K. and ZERNICKE, R.F., 1992, Mass, center of mass and moment of inertia estimates for infant limb segments. *Journal of Biomechanics*, 25, 145–48.

Engineering Anthropometry

K. H. E. Kroemer

Professor Emeritus of Industrial and Systems Engineering, Virginia Tech, Blacksburg, VA, USA

People come in a variety of sizes, and their bodies are not assembled in the same proportions. Thus, fitting clothing, equipment or workstations to suit the body requires careful consideration; design for the statistical "average" will not do. Instead, for each body segment to be fitted, the designer must determine the critical dimension(s).

Design for fitting clothing, tools, workstations, and equipment to the body is usually done in steps:

Step 1: Select those anthropometric measures that directly relate to defined design dimensions. Examples are: hand length related to handle size; shoulder and hip breadth related to escape-hatch diameter; head length and breadth related to helmet size; eye height related to the heights of windows and displays; knee height and hip breadth related to the leg room in a console.

Step 2: For each of these pairings, determine whether the design must fit only one given percentile (minimal or maximal) of the body dimension, or a range along that body dimension. Examples are: the escape hatch must be big enough to accommodate the largest extreme value of shoulder breadth and hip breadth, considering clothing and equipment worn; the handle size of pliers is probably selected to fit a smallish hand; the leg room of a console must accommodate the tallest knee heights; the height of a seat should be adjustable to fit persons with short and with long lower legs. (See below how to determine percentiles.)

Step 3: Combine all selected design values in a careful drawing, mock-up, or computer model to ascertain that they are compatible. For example: the required leg-room clearance height, needed for sitting persons with long lower legs, may be very close to the height of the working surface determined from elbow height.

Step 4: Determine whether one design will fit all users. If not, several sizes or adjustment must be provided to fit all users. Examples are: one extra large bed size fits all sleepers; gloves and shoes must come in different sizes; seat heights of office chairs are adjustable.

Determination of percentiles can be done either by estimation or by calculation.

Estimation is used when the data set is not normally distributed or too small. In this case, the data point is estimated by counting, weighing, or sample measurement according to the best possible judgment.

Calculation is used based on statistical considerations.

A normally distributed set of n data is described by two simple statistics:

The 50th percentile is by definition the same as the mean m (also commonly called average)

$$m = Sx/n$$

where Sx is the sum of the individual measurements.

The Standard Deviation SD describes the distribution of the data:

$$SD = [S(x - m)^2/n-1]^{1/2}$$

It is often useful to describe the variability of a sample by dividing the standard deviation SD by the mean m. The resulting Coefficient of Variation CV (in percent) is:

$$CV = 100 \, SD / m$$

To calculate a percentile value p of a normal distribution you simply multiply the standard deviation SD by a factor k, selected from the table below. Then subtract the product from the mean m if p is below the mean:

$$p = m - k \, SD$$

Table 1. Factor for computing percentiles from mean x and standard deviation s.

Percentile p associated with X		
K	X = m – KS	X = m + KS
2.576	0.50	99.5
2.326	1	99
2.06	2	98
1.96	2.5	97.5
1.88	3	97
1.65	5	95
1.28	10	90
1.04	15	85
1.00	16.5	83.5
0.84	20	80
0.67	25	75
0	0.50	0.50

If p is above the average add the product to the mean:

$$m = m + k \, SD$$

Examples:
To determine 95th percentile, use k = 1.65
To determine 20th percentile, use k = 0.84

For more information see:

KROEMER, K.H.E., 1999, Engineering anthropometry. Chapter 9 in W. Karwowski and W.S. Marras (eds), *The Occupational Ergonomics Handbook* (Boca Raton, FL: CRC Press), pp.139—65.

KROEMER, K.H.E., KROEMER, H.B., and KROEMER-ELBERT, K.E., 2000, *Ergonomics: How to Design for Ease and Efficiency* (2nd ed.) (Englewood Cliffs, NJ: Prentice-Hall).

ROEBUCK, J.A., 1995, *Anthropometric Methods* (Santa Monica, CA: Human Factors and Ergonomics Society).

Ergophthalmology: The Visual System and Work

B. Piccoli

University of Milan – ICP Hospital, Vin San Barnaba 8, 20122 Milan, Italy

1. INTRODUCTION

For quite some time people have been showing interest in the health aspects of the relationship between work and vision. In the past such interest was mainly orientated towards the etiology and treatment of chronic and degenerative diseases, either ocular or visual, connected to the presence in the working environment of large amounts of chemical (lead, sulfur-dioxide, solvents, etc.) and physical (ionizing radiation, UV, IR), pollutants. Consequently, medical intervention in this field has been mainly directed towards diagnosis and treatment of impairments of the visual system.

Over the last 50 years working conditions have greatly improved in technologically advanced countries. Illness and injuries, while not completely eradicated, have been noticeably reduced. However, many new problems have arisen. Among these are aspects of "visual functionality", which are much less worrying in terms of immediate impairment, but are most important in terms of wellbeing and efficiency and are destined to become progressively more widespread.

In truth, this side of the matter is not completely new. The first author to note the possible relationships between refraction and work was Bernardino Ramazzini in Italy in the seventeenth century. In his famous *De Morbis Artificum Diatriba*, he wrote a chapter, No. XXXVI, entitled "De lepturgorum morbis (About illnesses of small-object workers)*,* in which he mentioned specific visual disorders for some categories of workers (watchmakers, copyists, goldsmiths, etc.) and pointed out, for the first time, the possible relationships between near work and ocular/visual disturbances. The topic kindled the interest of other authors like Duke Elder and Kuhn, in the early twentieth century, but it is only in the last 20 years that its importance has been fully appreciated.

Indeed the many and varied interventions of transformation and reconversion implemented in the workplace with the introduction of new techniques, mainly computer-based, have noticeably changed the traditional relationship between man and machine. In the workplace, compared with the past, the effort required of the musculo-skeletal apparatus is clearly reduced, while the demand made upon the perceptive sensory system in general, and the visual system in particular, are correspondingly increased.

In this context, a new scientific field emerged in the early 1980s — *ergophthalmology*. According to the Scientific Committee on Work and Vision of the International Commission on Occupational Health (ICOH), its definition is the following:

Ergophthalmology is a scientific field aimed at analyzing, evaluating and designing simple or complex working systems pertaining to the relationship between work and visual performance. Ergophthalmology makes use of established knowledge derived from ophthalmology, industrial hygiene and occupational medicine, as well as from technology (physics, engineering, architecture, etc.) and social disciplines (psychology, sociology, etc.). The purposes of Ergophthalmology are mainly the prevention and management of discomfort and disease in order to obtain maximum efficiency and effectiveness of visual function in organized work.

2. AN OUTLINE OF THE PHYSIOLOGY OF VISION

The visual system can easily be divided into three well-defined groups of components: a peripheral component (eyeball) with its sensory functions; a central component (visual pathways and the primary visual cortex) with its functions of transmission, elaboration, and integration of the luminous input; a number of accessory components (eyelids, eyebrows, lacrimal film, conjunctiva, etc.), also called *adnexa,* with their mechanical, trophic, and protective functions.

In synthesis, it can be said that light emitted or reflected from the objects present in the surrounding space penetrates the eyeball producing, by means of a photochemical process, nervous impulses which give rise to the visual sensations.

The structure dedicated to ocular motility (extrinsic muscles) and focusing (accommodative system) achieve a wider and more detailed exploration of the surroundings as well as a clearer perception of objects, whether stationary or moving.

Besides the three sensorial functions mentioned above, in human beings the visual system also plays an essential "metabolic" role. Visually perceived light stimulates, via a specific retinal pathway, the synthesis of melatonin, a hormone essential for many physiologic processes. Because of its importance in ergonomics and occupational health, this peculiar function will also be considered.

2.1. The Eyeball

The eyeball is a spheroid (average diameter 24 mm front to back, 23 mm across and 23.6 mm top to bottom) situated in the orbit, a pyramid-shaped cavity of the cranium. About one-third of the eyeball, the "ocular surface", protrudes from the orbit and is in direct contact with the environment. The prominent bony structure (frontal, maxillary, and zygomatic bones) of the external part of the orbit, plus the eye lids, provide excellent protection from injury. These structures also help to reduce visual disturbances such as those caused by bright or undesired light.

Because of its structure and functions the eyeball is often compared to a photographic camera: this example, with a few qualifications, can be usefully employed to explain how it works. It is possible to distinguish three parts in this organ, dividing them according to their function.

The first part, which corresponds to the optical system of a photographic camera, is composed of the structures found in the anterior segment: the cornea, the iris, and the crystalline lens with its associated systems (the ciliary body). Contraction of ciliary body muscles increases the convexity (and consequently the refracting power of the lens), allowing a sharp vision of near objects.

The second part is formed by the vitreous, a gelatinous, transparent, colorless mass which fills the posterior segment of the eye: its function is essentially morphostatic. In a photographic camera this part is represented simply by an empty space, the dark room (*camera obscura*).

Lastly, there is a membranous structure, the retina, which corresponds to the sensitive element of a photographic camera, i.e. the film. The retina can be schematically divided into four components: photoreceptors (100 000 "rods" for nocturnal and twilight vision and 6000 "cones" for daylight vision), elements for the conduction of nervous impulses (bipolar and ganglion cells), elements with associative functions (horizontal and amacrine cells) and, finally, support elements (Mueller cells).

For an adequate understanding of human eye functions that suits ergonomics, it is, however, necessary to clarify a few points.

In the optical system of a photographic camera, focusing is achieved by moving the lens backwards and forwards: in the eye this is obtained, as mentioned above, by a contraction of the ciliar muscle, which allows the lens, by decreasing the tension on its capsule, to achieve a more convex shape, mainly in the central portion. This mechanism, supported by a complex neuromuscular system, is a reflex action and is *automatically* activated whenever the image is not sufficiently sharp. Consequently, every time objects or images appear in the visual field within six meters in an emmetrope subject, (i.e. not affected by refractive defects), focusing occurs *involuntarily* and *unconsciously*. The response to environmental light behaves in the same way, i.e. by a reflex mechanism.

The pupil, mesopically adjusted, has an average diameter of about 3.5 mm. In extreme conditions it can go from 1 mm (maximum miosis) to 9 mm (maximum midriasis), but under normal circumstances it oscillates between 2.5 and 6 mm. Pupillary size changes (dilatation and constrictions) brought about by the action of the sphincter and dilatatory muscles of the pupil are activated in response to the environmental light fluctuations (pupillary reflex). Experimental studies have shown that following repeated light stimulation (50 to 60), the pupillary reflex shows *clear signs of fatigue*, evidenced by prolonged latency, reduced contraction and poor redilatation.

While observing objects placed within one meter, pupillary constriction can also occur regardless of the lighting conditions but following accommodation and convergence changes (synkinesis near response), in order to eliminate as much as possible stray light not pertaining to the object/image under observation from visual field.

Another important difference which distinguishes the eye from a photographic camera is the structure of the sensory elements. Indeed a film has "receptors" that are all the same and uniformly distributed. Its sensitivity is not variable and its resolution power is homogenous in all points.

In the retina, however, the sensorial elements have different characteristics according to type, density, and nervous connections, from which is derived a vast range of retinal sensitivities (from full sunlight to minimum nocturnal light there is a ratio of 1 billion to 1) and a remarkably variable resolution power (from 1 minute of arc to several minutes of arc at the periphery).

The highest retinal resolution power area, called "fovea" or "macula" — which, in the presence of adequate light and perfect focusing enables us to identify the smallest details of an object — is located approximately at the center of the retina and has a diameter of only 5.5 mm. The remaining part of the retina has a function which is similar to that of radar, i.e. very sensitive to moving objects in the visual field but with poor discriminative capacities.

An important consequence for ergonomics of this heterogeneous receptor distribution is that stray light produces *much greater discomfort and visual disturbance* when the *foveal areas* rather than the *peripheral areas* of the retina are involved.

2.2. Ocular Motility

Among the various structures found around the eyeball there is a muscular system which carries out all the eye movements. This system is composed of six small muscles (extraocular muscles) responsible for all the movements of the eyeball. These movements, which are always harmonious and measured, can be very fast (saccadic, up to 400 degrees /sec.) or slow and accurate (pursuit, 30 degrees/sec.), and are able to perform complete and rapid explorations of the surrounding environment as well as a detailed identification of objects of interest.

It is important to point out that the motorial activity of the eyeball is a consequence of, and depends upon, the sensory component of visual processes.

Indeed when we fix an object placed within the binocular field (i.e. the visual field common to both eyes) which captures our interest, two foveal images are sent to the visual cortex but we perceive only one (sensory fusion). This is possible only if the two images sent from the foveas, identical in size, color, and luminosity, simultaneously reach the cortical areas. Consequently, the realization and maintenance of a state of "bifoveal fixation", particularly during the observation of moving objects, implies the activation of very complex nervous and muscular synergistic processes. It is therefore possible to speak of a motor-sensory unit in which the sensory system elaborates and transmits the information received from the environment, while the motorial system, completely at the service of the first, performs the exploration of the surrounding space, allowing maximum visual detail and tridimensional perception of the objects.

Ocular movements can be divided into voluntary and involuntary.

The first are the result of a decision made by the subject and are completely conscious (for instance to turn the eyes to the right or left, up or down, on command).

In the second group some (vestibular) movements established by unconditioned reflexes and coordinated by sub-cortical structures leave the subject completely unaware (maintaining eye position with respect to changes in head and body posture).

Others (saccadic, pursuit, vergence, and position maintenance movements) are activated by the coordination of specific brain structures (supra-nuclear centers), when an image arouses the interest of the subject. By means of fast and precise rotations of the eyeball, the image is then placed onto the fovea and followed if moving.

These involuntary eye movements are not completely "unconditioned" in that to be activated they require a state of "visual attention" (interest) which is only possible with the cooperation of the subject. All this has important implications for staff who work in confined spaces, particularly small offices. Indeed, when an operator is not directly engaged in observing the task-objects, which in offices are very frequently placed within one meter, there are often few possibilities, due to the presence of walls, dividers, cupboards, curtains, etc. to use as far vision.

In fact, in such a situation, practically nothing is placed at a sufficient distance to permit complete disactivation of convergence and accommodation, which remain active, obviously at varying intensities, *for the whole time the operator performs that task*, possibly causing overloading of the functions involved and consequently "visual fatigue".

2.3. The Optical Pathways

The fibers (axons) coming from the ganglion cells of the internal layer of the retina are united into a single bunch (optic nerve)

that emerges from the eyeball and crosses the back part of the orbital cavity penetrating into the cranium. Shortly after, the two optic nerves merge partially together (optic chiasm). Those coming from the nasal hemiretina of each eye pass into the optical tract of the opposite cerebral hemisphere where they reunite with the uncrossed fibers coming from the other eye (semidecussation). Thus, when an individual fixes an object with interest, images from the right (or left) side of the visual field, projected onto the nasal hemiretina of one eye and onto the temporal hemiretina of the other, simultaneously reach the same cortical areas where they are fused into one image (binocular vision).

After, the optic chiasm axons continue on (optic tract) to terminate within the lateral geniculate body where, by very complex synaptic interactions with the geniculate cells, a number of elaborations take place that are not yet completely understood.

From this site a pathway (optic radiations) reaches the striate cortex of the occipital lobe: it is here that the most important and sophisticated elaborations and the most complex visual perceptions are achieved in association with several other cortical areas.

2.4. The Metabolic Function

The effect that visually perceived light can produce on the pineal gland via a specific retinal pathway (retinohypothalamic tract) is now well known.

The pineal gland is located in the brain and synthesizes a specific hormone, melatonin, which is found in blood with high/low levels (with remarkable inter-individual variability), according to the daily dark/light cycle. Melatonin is considered an internal pace-maker ("zeitgeber") which translates light information from the environment to the central nervous system via high affinity binding sites. The exact role of melatonin in humans is not yet fully understood, but the presence of melatonin receptors in the human circadian clock, as well as the ability of humans to respond to melatonin signal with a well-defined circadian phase-response curve, is evident.

The rhythmic melatonin blood-level changes appear to be implicated in a number of physiological phenomena such as body temperature, cortisol blood levels, sleep–wake cycle regulation, and physiopathological situations such as jet lag and seasonal affective disorder syndrome (SADS).

Most circadian rhythms, including that of melatonin production, are driven from an endogenous circadian pacemaker located in the suprachiasmatic nuclei (SCN) of the hypothalamus. Since it is active perinatally and continues its activity throughout life, the SNC generates an approximate 24-hour rhythm that plays a role in most, if not all, the body processes. In subjects experimentally kept away from time cues, the circadian rhythm is still present, but slightly longer (free-running circadian rhythm). Humans seem to require a more intense natural light than other species to activate their physiological day–night melatonin cycle, maybe because of an adaptation to prolonged artificial indoor light exposure.

Pineal sensitivity, which varies in the different species, in humans seems to require *an intensity of 2500 lux* (i.e. the illuminance normally present by a window on a sunny day, but rarely found in indoor work places) to be fully inhibited. Dim light (approx. 300–500 lux) also suppresses nocturnal melatonin production, but to a lesser extent.

A light wavelength of 530 nanometers (green) appears to be the most effective, while 435 (blue) and 660 (red) have poor or no suppression effect. The degree of suppression produced by light exposure duration, other than intensity and wavelength, is so far not completely clear. Nevertheless, the effects caused by prolonged and marked reduction in daylight exposure, in particular for occupational reasons and in young subjects, have to be carefully considered in ergonomics.

3. THE OCCUPATIONAL VISUAL PERFORMANCE

The main function of human vision is to explore the surrounding space and gather detailed information about the objects and images most attracting the subject's interest.

In modern work organization, the visual system is used progressively less as a sensory organ able to satisfy an individual's interest and increasingly as a "work instrument" obliged to make predetermined and repetitive visual efforts for prolonged periods of time. In work places, in fact, the visual function is subject to many constrictions of various nature imposed by organizational necessities (environment features, time and duration, instruments and equipment, procedures, etc.), which can be responsible for discomfort and/or disorders.

Occupational risk for the visual system could be schematically subdivided into: (i) task-related and (ii) environment-related.

3.1. Task–related Risk

The occupational visual effort can be divided into two main categories: (i) requiring mainly *far vision*, and (ii) requiring mainly

Table 1. **Examples of some common near-visual tasks.**

Sector	Task
Aviation	air traffic controller
Bank and insurance press Telephone companies Airport service Design activities in industry	VDU work
Ceramics	decoration
Clothing	quality inspection and mending
Research Laboratories Microsurgery Electronic components industry	work with microscopes
Pharmaceutical industry Food industry	impurity controls
Sheet metal industry	precision mechanics engravers welders
Jewellery industry	bench work
Photographic industry	photo compositors and binders cartographers topographers
Craftsmen (artisans)	decorators sculptors engravers wood carvers lithographers
Movie industry	film editing, splicing, cutting, montage

near vision. From a physiological point of view, the effects that can be induced by different efforts are clearly different.

While far vision is achieved by a natural and well-balanced involvement of the different components of the visual system (nervous, muscular, and refractive), near vision burdens the system particularly with two specific functions: accommodation and convergence. Their activation implies a complex muscular and nervous effort (see 2.1, 2.2)

It is well known that tasks which require near visual effort can cause ocular and visual disturbances (occupational asthenopia) often associated with conjunctive hyperemia and over-lacrimation, rarely found in far-visual tasks and in non-occupational activities.

Studies were carried out mainly in VDUs operators, but it should be mentioned that many other workers are exposed to similar conditions (see table 1).

From an ergophthalmological point of view all these job categories have a common element: they all require a *near*, *prolonged*, and *fixed* observation.

The term *near* is used because the objects and instruments to be observed are usually placed within 1 meter, with an average daily observation distance ranging from 50 to 70 cm.

The term *prolonged* refers to situations where types of tasks and/or techniques used demand near point fixation to be maintained for many hours a day.

The term *fixed* is related to the physical and structural limitations of indoor work places where walls, curtains, wardrobes, dividers, etc., frequently do not allow far vision (i.e. over 6 meters), therefore greatly reducing the physiological alternation of activation and disactivation of accommodation and convergence to their full extent and consequently increasing the amount of contraction, which is mainly isometric, of the muscles involved.

For these reasons in a vast number of operators occupational tasks can produce an overloading of accommodation and convergence which normally represents a major cause of transitory disturbances. The risk of permanent alteration, particularly for long-term near work exposure (over decades), is still matter of dispute. To date the literature reports no clear-cut studies able to confirm or deny this hypothesis.

3.2. Environmental Risks

3.2.1 Chemical pollution

Studies carried out on the Sick Building Syndrome (SBS) and on health effects linked to Indoor Air Quality (IAQ), have demonstrated the presence of a high number of airborne substances in industrial and non-industrial working environments. Indeed, these researches quite often report "eye irritation" as a common or very common complaint in some categories of workers, particularly office operators.

Among the many potential chemical agents found in these workplaces, those that most affect ocular mucosa are: aldehydes (especially formaldehyde, acetaldehyde and acrolein), oxides and nitrogen (NOx), Volatile Organic Compounds (VOCs), Environmental Tobacco Smoke (ETS, i.e. a mixture which includes respirable suspended particles, CO, nicotine, nitrogen oxides, acrolein, nitrous compounds and benzo(a)pyrene), ozone, dust, and fibers.

The source of these pollutants, which are frequently present in higher concentrations in indoor rather than outdoor environments, can be carpets, furniture, painted surfaces, and cleaning agents, as well as photocopying machines, laser printers, facsimile machines and combustion sources (kerosene and unvented gas heaters).

All these chemical agents are potentially able to produce ocular irritation and disturbances when in contact with the outermost part of the eye, i.e. the "ocular surface". The ocular surface is an anatomo-functional identity that comprises lacrimal film, corneal and conjunctival epithelia. Thanks to its versatility and adaptability it is able to sustain adverse indoor air quality conditions, obviously up to a certain limit. Over that limit, which can vary considerably even in normal subjects in relation to age and gender, ocular disorders, such as decreased foam formation, reduced lacrimal film stability (abnormal break-up time test), conjunctival and corneal epithelium alterations can be observed.

3.2.2 Microclimate and Illumination

Microclimate and illumination are the most important environmental sources of ocular and visual disorders. Regarding the first, relative humidity, particularly if below 40%, and air velocity, i.e. eye hit by flows originating from fan-coils, air conditioners, office equipment cooling systems (VDUs, printers, photocopiers), etc., are possible causes of eye irritation. In these circumstances, a corneal flogosis and sufferance can arise due to an excessive evaporation and consequent lack of lacrimal film on the ocular surface.

Environmental lighting today is much more important than in the past due to two groups of interrelated factors:

i. Work activities are carried out increasingly in indoor environments.

ii. Fine work, particularly "computer tasks" are becoming more and more widespread. (In this respect, it is appropriate to note that according to the European Foundation of Living and Working Conditions, in 1997, out of 147 million workers in the European Union, 55 million were using computers.)

There are numerous researches where office lighting is linked to visual discomfort and fatigue (see table 2). Moreover, studies on VDU work, where illumination is unanimously recognized as a major contributing factor, often report a frequency of complaints varying from 25% to 85 % , with an average frequency of 62% (see table 3).

Table 2. Illumination and visual complaints in VDU operators.

Authors	Year	Population studied	Frequency of complaints
E. Gunnarson and O. Ostberg	1977	German VDU operators	76%
A. Cakir *et al.*	1979	Swedish VDU operators	68–85%
L. Ghiringelli	1980	Swiss VDU operators	76%
L. Scullica and C. Rechichi	1980	Italian VDU operators	25%
B.L. Cole, I.D. Breadon *et al.*	1986	Australian VDU operators	50%
R.I. Wibon and L.W. Carsson	1987	Scandinavian VDU operators	70%
H. Lindner	1994	VDU operators	58%
B. Piccoli	1996	Italian VDU operators	70%

Table 3. Office lighting and visual disturbances.

Authors	Year	Detail
A. Cakir, D. Hart *et al.*	1980	Factors able to affect the incidence of visual fatigue. . . The level of illumination
Human Factor Soc.	1981	Incidence of eye troubles is associated with excessive luminance contrast...
ILO	1989	Increased visual disturbances attributed to glare and inadequate lighting....
P. Boyce	1991	... visual enviroment created by lighting can cause discomfort in VDT users..
Sheedy	1992	The most common complaint of VDT users is glare.
H. Lindner	1994	Asthenopic complains frequently occur in conjunction with illumination...
IES	1995	Discomfort may be caused by viewing a light source or a reflection...

There could be several causes for this high prevalence of occupational visual disturbances reported as illumination-related:

(a) a wide distribution of self-compiled questionnaires where the data are biased by subjectival factors;

(b) a synergistic link with overloading of visual functions and/or with non-lighting environmental-related aspects;

(c) inadequate indoor work-place lighting design, both in theory and in practice.

Regarding the last point a few observation are in order.

The main parameter referred to when a lighting system is being designed is almost always the *illuminance,* measured by the luxmeter. This kind of photometer does not seem adequate to furnish measurements congruent with human anatomy and physiology for two reasons:

1. Its measurements are strongly influenced by the incidence angle of straylight ("cosine effect"), while this does not happen in the eye (figure 1);

2. The photovoltaic cell of the luxmeter detects light over a large angle (i.e. includes straylight originating from a wide portion of space), while for most operators, particularly when involved in near work, due to the characteristics of ocular structure and functions, the light which actually affects the eyes, and namely the fovea, comes from a very limited area of the surrounding space (figure 2).

Actually, a photometric evaluation able to detect light over small angles (< 5°) can be carried out by using a luminance photometer (telephotometer or video-based photometer).

Both approaches have their validity. The first is appropriate for lighting control strategies (energy saving, minimum maintained levels, daylight factor component, esthetic applications, etc.). The second seems more adequate for a reliable evaluation of light actually involving the operator's visual system.

3.2.3 Microbiological agents

Bacteria, fungi, and viruses are often found in indoor environments and, according to the numerous SBS investigations reported in the literature, they are believed to have a role in the

Figure 1.
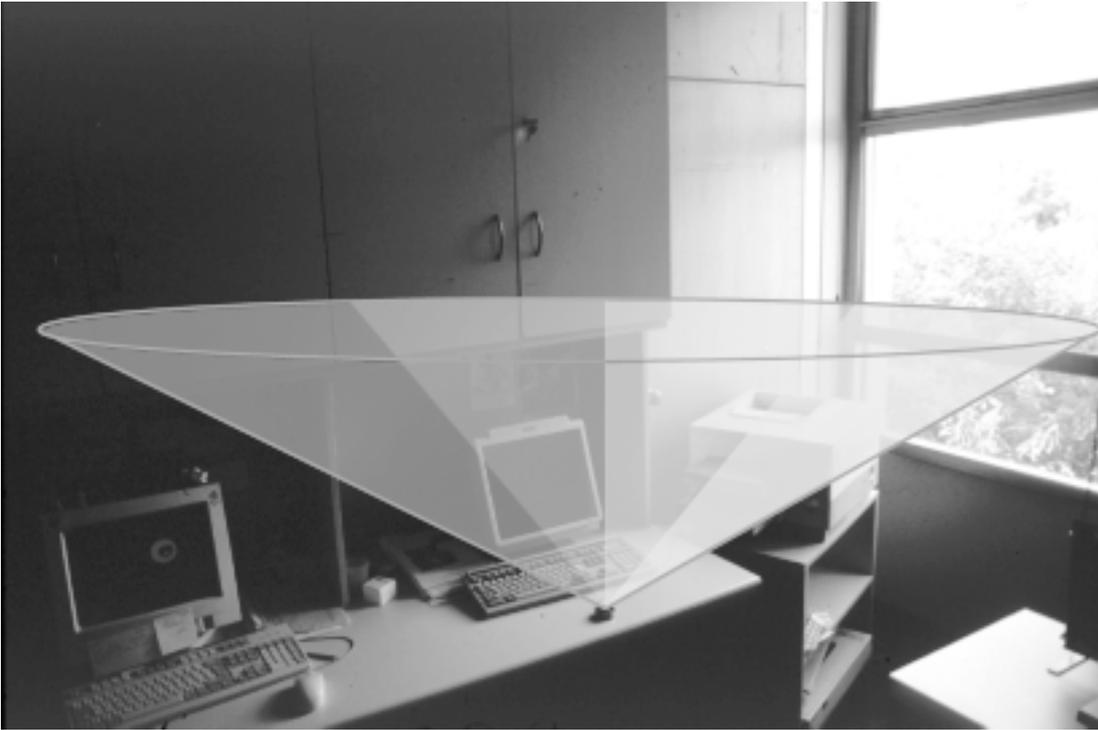
Figure 2.

development of certain respiratory and skin disease of the occupants.

The origins of microbial contamination are usually outdoor air (bacteria and fungal spores) and humans. They are sustained or worsened by poor maintenance of air conditioners, fan coils, and humidifier systems, as well as by carpets and office equipment, including keyboards and computer mice.

We should also consider the numerous microrganisms normally present in the working environment, which can play a pathogenic role if they come into contact with specific organs and mucosa of the human organism (for instance *Pseudomonadaceae*).

The ocular conjunctiva and cornea, due to their external position, can be easily contaminated by microrganisms present on periocular tissues and hand skin. In everyday clinical practice, infective diseases of the ocular surface (conjunctivitis, cheratitis, dacriocystitis, etc.) are quite common and symptoms normally include burning eye, photophobia, lacrimation, gritty feeling. Due to the fact that these symptoms are also commonly reported by office operators ("VDU asthenopia"), it is quite likely that, unless specific ophthalmological examinations are carried out, the origin of these infections will be misinterpreted.

General preventative measures can be recommended:

(a) worker education (good personal hygiene and restriction of food consumption at work);

(b) cleaning procedure using an appropriate antimicrobial product;

(c) environmental investigations and ophthalmological check-ups in cases of high prevalence of ocular surface infections in a specific working population or recurrent ocular infection in individual operators.

3.2.4. Preventive measures

The following essential primary and secondary preventive measures, concerning job organization and environmental and medical issues, can be recommended:

(a) implement working procedures and task sequences where near (< 1 m) vision sessions are interrupted by breaks or far-vision activities;

(b) make it possible to look into the distance from each workstation and wherever possible with outside view;

(c) keep luminance ratios in the environment generally, and inside the "occupational visual field" particularly, to a minimum with specific regard to computer-based equipment operators;

(d) avoid relative humidity being lower than 40% and eliminate constant air-flows towards the operator's eyes;

(e) minimize ocular-irritative airborne substances;

(f) provide antimicrobial cleaning procedures when multi-user work equipment (particularly keyboards, mice, telephones, etc.) is present;

(g) carry-out appropriate eye examinations before commencing work in order to evaluate the compatibility of the operator's visual capacity with the performance required (job-fitness evaluations);

(h) activate health surveillance programs for: (i) prolonged near-work operators (computer-aided design, precision mechanics, goldsmiths, etc.) particularly if under 18 years old; (ii) operators involved in high attention and intense cognitive performances (airline pilots, crane drivers, air traffic controllers, nuclear power station operators, etc.); (iii) hypersusceptible subjects (refractive, binocular, and ocular surface alteration carriers, frequent task-related asthenopia cases).

The application of these measures will not only reduce injury and health risk, but will also have a positive influence on the morale of employees, thus becoming ethically appropriate and highly cost effective.

Event related Potentials

A. Yagi

Department of Psychology, Kwansei Gakuin University, 1 Uegahara, Nishinomiya, Hyogo, Japan

1. INTRODUCTION

Some brain potentials (e.g. EEG, spontaneous brain wave; ERP, event-related brain potential) can be measured from electrodes placed on the scalp. ERP is a sequence of electrical changes elicited by sensory or perceptual stimuli, and cognitive events. ERP technology has some advantage compared with other psychological methods such as questionnaire and performance indices. ERP changes can be measured objectively, quantitatively, in real time and temporally. The potentials to an unconscious (subliminally) stimulus and to an ignored stimulus can be measured in the study of attention. Aims of research on ERP in ergonomics are classified into two categories. The first is basic laboratory studies on the nature of ERP in relation to human information processing. The second is studies aimed at the application to field settings. Studies have been conducted that are applied to the assessment of environment and mental workload in ergonomics. At an early period (the late 1960s) in the history of ERP, the potential was called the evoked potential (EP) or the evoked response, because the potential was recognized as evoked by sensory stimuli. However, since similar potentials are elicited not only by sensory stimuli, but also by such subjective events as attention, decision-making and language process, and by the voluntary movement, the potential is now referred to as the event-related potential (ERP).

2. MEASUREMENT OF ERP

Since ERP is a very small change in electrical activity of the brain, direct observation of ERP is difficult since it is intermingled with EEG. EEG must be recorded to obtain the ERP. Regular EEG amplifiers can be used but an amplifier with a long time-constant is preferred because the ERP includes lower frequency components. To obtain an ERP, successive stimuli or a sequence of events have to be presented to a subject. The EEG is triggered by the onset of a stimulus or event and EEG epochs associated with the onsets are averaged (Figure 1). Because the spontaneous EEG is assumed to be a random process, the averaged EEG gradually approaches zero. On the other hand, ERP linked to the events become increasingly clear as the number of trials averaged increases. Visual, auditory and tactile stimuli have been mainly used in ERP studies. Therefore, to obtain an ERP, signals associated with repeatable stimuli, responses or events are required to trigger EEGs for averaging. Some of the ERP occur not after the stimulus onset but before the event onset (e.g. the readiness potential and the motor potential).

3. VARIETY OF ERP
3.1. Endogenous and Exogenous Components
ERP is classified into several groups based on the stimulus modality used, the paradigm of a task, latency (the time from stimulus onset to the peak of the potential), and the distribution
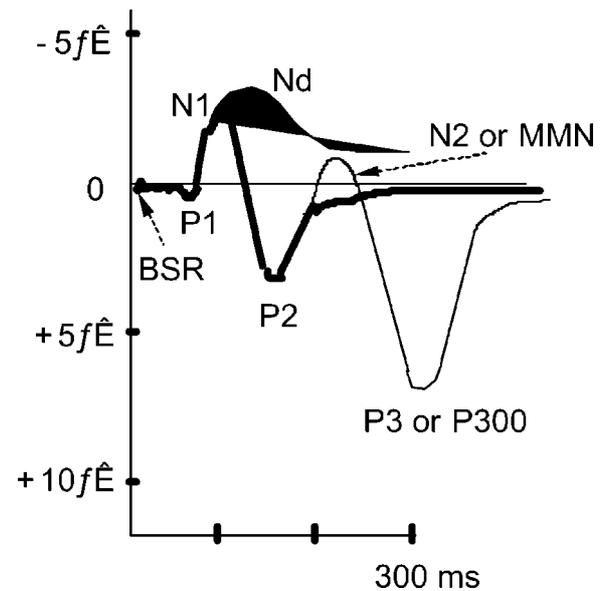
Figure 1. Model of auditory ERP. ERP consists of several components. Upward negativity.

of the potentials on the scalp. Further, each component can be classified into subordinate components. Some well-known ERP related to ergonomic issues will be discussed here.

A unique component, one which is specific to the modality and which changes with the physical property of the stimulus, is called an exogenous component. A component elicited by a psychological or physiological event is called an endogenous component. For instance, when a stimulus is unexpectedly omitted in a sequence of stimuli presented regularly, a large amplitude-positive deflection appears, in spite of the absence of an external stimulus. This component appears at the moment when the stimulus was expected. This is a typical endogenous component. An exogenous component may be used in the assessment of sensory or perceptual events (such as sensation and discrimination) in ergonomics. An endogenous component is expected to be most useful in ergonomic studies concerning mental workload, attention, decision-making, vigilance, etc. The Nd wave, the mismatch negativity, P300, the eye fixation-related potential, the contingent negative variation, the readiness potential and the motor potential are slow potential changes occasionally applied in ergonomics studies.

3.2. ERP by Auditory Stimuli
Figure 2 shows a model ERP recorded at the vertex of the head evoked by auditory stimuli. The waveform consisted of several components. Upward deflection in the potential is traditionally used to demonstrate increases in negativity in both EEG and ERP studies. The components, which appear very early (< 10 ms) after onset of auditory stimuli, are called the brain stem response (BSR). They reflect the progress of neural signal through the auditory nervous system. The responses used for the diagnosis of deficits in the auditory system. The components with latency of > 100 ms are related to psychological events. N1 with latency of ~100 ms, is an exogenous component, which changes with intensity of the auditory stimulus. When a subject attends to a given stimulus, the Nd wave extends beyond N1. When attention

is directed to an unexpected stimulus, a large amplitude positive component (P300) occurs after the stimulus onset (see below).

3.3. Visual Evoked Potential (VEP)

VEP, which is observed predominantly at the occipital region, is a transient potential evoked by a flash, pattern and color. VEP changes with such stimulus conditions as luminance, contrast, the spatial frequency, the wavelength, stimulus duration and the visual field. The slow negative wave and/or P300 may, as is true of auditory stimuli, also occur at the vertex of the head depending upon the nature of the event. The VEP has been principally applied to studies in clinical settings rather than in ergonomics,

Event onset

visual

stimulus

auditory

voluntary movement

response

eye fixation

EOG

saccade

EEG

Average

+

ERP
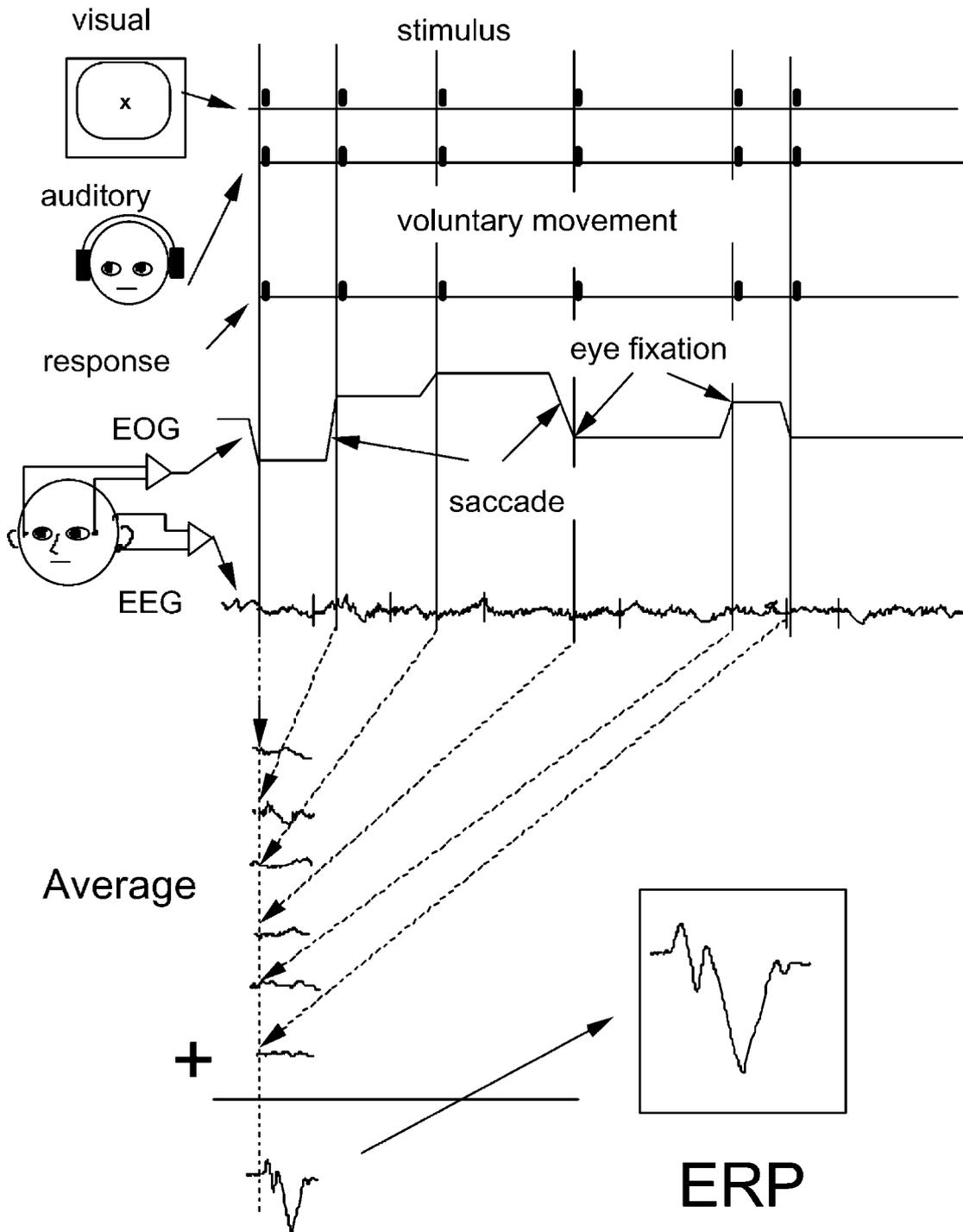
Figure 2. Prosedure of ERP detection. EEG epochs are triggered and averaged at onset of visual stimuli, auditory stimuli, responses and eye fixation pauses to obtain the visual ERP, the auditory ERP, the motor potential and the EFRP respectively.

because movements of eyes have to be restricted in measurement of these VEP.

3.4. P3 or P300

When auditory or visual stimuli are presented after the subject had to guess the modality of the upcoming stimulus, a late positive component with a peak latency of ~300 ms occurs in ERP to both auditory and visual stimuli (Sutton *et al.* 1967). The component is called P300 or P3, which is the third major component of the ERP. P300 is affected by psychological parameters such as stimulus probability, target discrimination (Figure 3), relevance, discriminability, response requirement, refractory period, age, performance task and etc. A widely used paradigm in the ERP (P300) study requires the subject detect infrequently occurring target or 'oddball' stimuli occurring unpredictably in a sequence of standard stimuli. Therefore, many researchers deal with P300 as an index of mental workload, attention, etc. in ergonomics.

3.5. MMN or N2

If, following a series of auditory stimuli, the stimulus suddenly changes, a negative wave at ~200 ms occurs. This component is called the mismatch negativity (MMN) or N2 (Naatanen *et al.* 1978). MMN occurs with any change in the physical property of the repetitive stimulus. MMN occurs to stimulus change even if the subject is not aware of the change because attention is directed elsewhere. The amplitude of MMN varies with the degree of the difference from the repetitive standard stimuli. When the subject is aware of the change, MMN is followed by P300.

3.6. N400

When a subject is asked to read silently a word completing a sentence (e.g. 'He spread the warm bread with socks'), which is

NT ——— TG ·······

Oz

LR

P3

5 μV

0   400   800

Time (ms)

Figure 3. Example of the lambda response (LR) and P3 in EFRP. The lambda response is followed by P3, when a target signal is detected in a situation requiring visual search. NT: EFRP to eye fixations at non targets. TG:EFRP to eye fixations at detected targets. OZ: occipital region on the scalp. Time zero (0) shows onset of eye fixations. Upward negativity.

presented one word at a time on a CRT, a prominent negative component peaking ~400 ms (N400) is elicited by an unexpected and semantically anomalous word. However, words that are semantically appropriate but physically deviant from the rest of the sentence elicit a late positive component similar to P300. Thus, the processing of any semantic stimulus that is unpredictable and unprimed may be associated with N400.

3.7. Eye Fixation-related Potential (EFRP)

When a subject looks at something, the eye movement record shows step-like pattern identified as saccadic eye movements with eye fixation pauses between such saccades. Information concerning the nature of the visual object is sent from the retina to the brain during the fixation pause. When EEGs are averaged time-locked to fixation pause onset, i.e. offset of saccades, the eye fixation related potential (EFRP) is obtained (Figure 2). EFRP is an ERP measurable in situations requiring eye movements. EFRP like ERP consists of several components. The most prominent component with latency of ~100 ms is called the l response. Some components of EFRP, especially the l response, change as a function of stimulus properties, e.g. spatial frequency, contour and the brightness of the stimulus, as well as subjective factors, e.g. attention, signal detection and language processing. The l response is followed by P300, when a signal is detected in a situation requiring visual search (Figure 3). It is thus the possible to apply the EFRP in ergonomic studies.

3.8. Contingent Negative Variation (CNV) and Readiness Potentials (RP)

CNV is a relatively long-lasting negative shift, which develops between the presentation of a warning signal (S1) and a second stimulus (S2) that requires a response. The CNV begins within 200 to 100 ms after S1 and reaches its peak within 100–900 ms if the S1–2 interval is 1000 ms. The amplitude drops suddenly when S2 is presented. A simple pairing of stimuli without response requirements results in no appreciable negative shift. The CNV reflects a state of expectancy. CNV indicates an intention to act, subject motivation or attention. The readiness potential (RP) begins at ~500–1000 ms before a voluntary movement and peaks at the moment of response (Figure 1). The last part of RP includes the motor potential (MP) which initiates a voluntary movement.

4. QUANTIFICATION OF ERP

Indices for quantification used in ERP studies are: peak amplitude of a specific component from the base line, the peak-to-peak amplitude between neighboring components, the area of a component, and peak latency from stimulus onset. Another index of ERP in cognitive studies is the similarity (or inverse variability) that can be calculated by measuring the correlation between two ERP waveforms. The distribution of those values is displayed topographically to observe mutual relations of ERP recorded from many regions on the scalp. Further, new procedures have been developed to identify electric dipoles of ERP in the brain to identify the location from which the electrical event is triggered.

5. COGNITIVE FUNCTIONS AND ERP

5.1. Attention and ERP

There are studies that deal with the relationship between attention and ERP. When the subject pays attention to a specific auditory
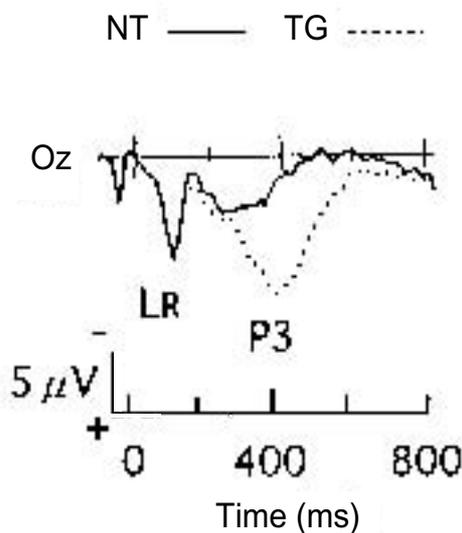
stimulus, the N1–P2 component increases in amplitude. However, attention is an ambiguous term that has many meanings, such as selectivity, concentration, vigilance, search, arousal level, etc. The effect of selective attention on ERP has been an issue of controversy for a long time. The criticism suggests that an increase of N1–2 amplitude at the vertex reflects a nonspecific arousal effect produced in response to the task relevant stimulus rather than selective attention. In a selective attention study, it is required that ERP should be measured under the condition where non-specific arousal does not affect ERP. For example, the subject is asked to attend to one channel and to respond to a stimulus change under conditions where auditory stimuli are delivered rapidly and to the two ears (Hillyard *et al.* 1973). Nd is obtained by subtracting ERP in an unattended channel from the ERP to those same stimuli while they are being attended. Nd is a neural sign of stimulus processing that follows stimulus set selection. When the task requires visual attention, the subject is asked to fixate on a fixation point in the center of the visual field. The subject attends to flashes presented to either side of the light or to only one side to detect target flashes that deviated in brightness, size or color from the non-targets. In the occipital region, positive and negative components with latency between 100 and 300 ms are elicited to the flashes in the attended field. The Nd-like slow negative wave is observed at the frontal region in the visual task. Thus, in a visual ERP study, the subject is asked to fixate eyes on a special point and not to move eyes.

5.2. Mental Workload and ERP

Recording the ERP to stimuli in an unattended channel can provide a measure of how fully those stimuli are being processed, without disrupting the subject's attentional strategy. In the early years of the 1970s the first application of ERP (to auditory stimuli) to the study of mental workload started in dual task situations on the assumption of the single channel theory. In 1980, it was found that the P300 varied with the demands made on perceptual rather than response resource (Isreal *et al.* 1980). An auditory oddball sequence as a secondary task has been used to probe cognitive processing during joystick tracking of visual target. Increasing the perceptual demands of the tracking task by increasing the number of visual targets to be tracked attenuated the P300 to the oddball stimulus. P300 reflects an allocation of resources to stimulus evaluation. There are many studies on P300 in the dual task method. For instance, the technique is applied to assess the workload in a radar operation and an instrument board.

5.3. Signal Detection and ERP

In signal detection task, the subject makes decision concerning the presence or absence of a stimulus. Vigilance is one of the first themes on ERP studies. The ERP waveform to detected stimuli is different from that to missed signals. When the subject reports the confidence of detection judgment, the P300 and a negative wave with peak latency of ~100–200 ms increases in amplitude and decreases in latency as a function of the level of confidence. The ERP is measured to assess responsiveness to the warning signal.

5.4. Application of EFRP

EFRP is applicable as an index of a visual task under eye movement situations. The waveform of the EFRP changes with the type of screen (positive or negative) on the CRT. When subjects enjoy performing a computer graphic task, EFRP increases in amplitude. On the other hand, the amplitude decreased when subjects are tired or when the arousal level decreases. Variations of EFRP are related to attention and arousal level during task performance.

Intermittent visual stimuli are required to measure ERP related to a visual environment. However, such intermittent lighting is different from usual illumination. EFRP can be obtained under the majority of lighting conditions by measuring EOG and EEG. EFRP is a useful index for assessing changes in the lighting environment. Latency of the lambda response in EFRP changes with the luminance of the work area. The stability (inverse variability) of the waveform of EFRP reflects the level of concentration of attention to the tasks under the various lighting conditions (Yagi *et al.* 1998).

Early components in the EFRP change in amplitude with complexity of textile patterns. The result suggests that EFRP may be applicable as an index to evaluate complexity of textile patterns and visual environments. Thus, EFRP has major possibilities for application in ergonomic studies.

6. CONCLUSION

Because EEG is a very small electrical signal, noise or artifacts seriously compromise results in field studies. To reduce such noise the worker's body and eye movements are often restricted. Therefore, ERP as perceptual or cognitive indices is mostly studied in the laboratory. Current developments of very small EEG amplifiers and a telemetry system make it possible to measure EEG in industrial settings and in outdoor situations. Since new analysis procedures are being developed to allow for the extraction of usable signal from noisy data. This will allow for the application of ERP technology to many more areas in ergonomics.

REFERENCES

HILLYARD, S.A., HINK, R.F., SCHWENT, V.L. and PICTON, T.W., 1973, Electrical signs of selective attention in the human brain. *Science*, 182, 177–80.

ISREAL, J.B., CHESNEY, G.L., WICKENS, C.D. and DONCHIN, E., 1980, P300 and tracking difficulty: evidence for multiple resources in dual-task performance. *Psychophysiology*, 17, 259–73.

NAATANEN, R., GAILLARD, A. and MANTYSALO, S., 1978, The N1 effect of selective attention on evoked potential reinterpreted. *Acta Psychologica*, 42, 313–29.

SUTTON, S., TUETING. P., ZUBIN, J. and JOHN, E.R., 1967, Information delivery and the sensory evoked potential. *Science*, 155, 1436–9.

YAGI, A., IMANISHI AKASHI, Y. and KANAYA, S, 1998, Brain potentials associated with eye fixations during visual tasks under different lighting systems. *Ergonomics*, 41, 670–7.

Force Exertion: Pinching Characteristics and Strengths

S. N. Imrhan

Industrial and Manufacturing Systems, Engineering Department, University of Texas, Arlington, TX 76019, USA

1. INTRODUCTION

Grasping and squeezing objects and manipulating them with the fingers are common workplace activities. However, it is widely accepted that they cause or contribute to hand overexertion disorders, known as cumulative trauma disorders (CTDs). Three main factors have been implicated in the prevalence and severity of these disorders: great muscular contractions in squeezing, repetitive muscular contractions in maintaining high rates of work (even if they are not of great magnitude), sharp bending of the wrist, and inadequate rest. In order to prevent or minimize CTDs, these factors must be carefully analyzed. This article examines the strength of the fingers in pinching. Like hand gripping, pinching has been deemed one of the main factors responsible for CTDs of the hand, such as carpal tunnel syndrome, trigger finger, and other forms of tendinitis. This article discusses pinching. No attempt is made to explain the details of CTDs of the hand since that topic is discussed elsewhere.

In this article, the term "strength" is defined as the maximum force or torque that a person can generate at the hand-handle (object) contact while gripping and squeezing the handle or object. Such strength is measured from a maximum voluntary muscular contraction (MVC); that is, the maximum a person is willing to exert during a 4-6 s maximal effort without suffering from pain or significant muscular discomfort. This technical definitions is widely accepted within the discipline of ergonomics but other definitions and interpretations have been used in other disciplines (e.g., physiological studies and exercise and physical fitness studies) and by the general public.

The terms "grasp" and "grip" will be used interchangeably. Existing definitions of grasping have identified two types — prehensile and nonprehensile grasps. Both types are defined in my article on handgrip characteristics and strength. In pinching, an object is grasped between the pad of the distal phalange of the thumb and the pad of the distal phalange of one or more of the other fingers. In finger torquing, the object being torqued (e.g., a small jar lid or electrical connector) is pinched and a rotary (turning) force applied either in a tightening or loosening mode.

2. TYPES OF PINCH

The pinch terminology in the clinical field (Smith 1985) has been adopted by ergonomists. It refers to one-handed pinching. Work tasks typically use the following types of pinch:

- *Lateral pinch*: the thumb opposes the lateral aspect of the index finger (which acts as a buttress) with the fist clenched.
- *Pulp pinch*: the thumb opposes one or more of the other fingers.
- *Chuck pinch*: the thumb opposes the index finger and middle finger simultaneously.
- *Four-finger pinch*: the thumb opposes the index, middle, and ring fingers.
- *Five-finger pinch*: the thumb opposes the other four fingers.

There are different types of pulp pinch depending on which finger opposes the thumb: pulp 2 pinch uses digit 2 (index finger); pulp 3 uses digit 3 (middle finger); pulp 4 uses digit 4 (ring finger); and pulp 5 uses digit 5 (little finger). The pulp 4 and 5 pinches are uncommon for work tasks.

3. PINCH STRENGTH TESTS

The strengths of some of these pinches have been measured in several experiments by different researchers. The typical method of measurement is the clinical one in which the two lips of a pinch gauge are gripped with the thumb and other finger(s) with all finger joints in low to moderate flexion. The hand is held at elbow or shoulder height with the forearm in mid-orientation. The lips of the pinch gauge are then squeezed by the fingers maximally for 4–6 s and the peak or middle 3 s mean force recorded. Most published pinch strength data uses peak values. Mean values are about 10–15% lower than corresponding peak values for the same pinch. There is very little motion of the lips when squeezing on the pinch gauge. So for practical purposes, the exertion may be considered as static.

Dynamic pinching has yet to be investigated. For all except the lateral pinch, pinching involves flexion of all the fingers. There is a limited amount of data using a variation of this method in which the thumb is hyperextended (Imrhan and Rahman1995); this reflects what happens in many work tasks. It is wise to check these points in order to avoid misinterpretation or misapplication of published data.

The method of pinch testing described above has disadvantages for ergonomics because much pinching in the workplace is performed differently and on objects that are of different physical characteristics (size and shape) from those of the typical pinch gauge. Some pinch gauge lips have limited area (2.0–1.4 cm), each smaller than the area needed to accommodate more than one finger comfortably in a large percentage of the population; and these lips are separated by a fixed width (separation distance between thumb and the other fingers) of 1.6–1.8 cm. These limitations preclude the measurement of pinch strength at all except this fixed width. It is also important to remember that almost all published pinch strength data is based on one-handed exertions using the clinical method of pinching. My two-handed pinch data will soon be published in the ergonomics literature.

4. PINCH MAGNITUDES

Because of differences in pinch conditions and subject samples in different experimental studies, pinch strength magnitudes vary widely among the different studies. But if we omit data from studies which show much greater or smaller magnitudes (somewhat extreme) than the other studies, the following generalizations may be made about static pinch strength magnitudes.

4.1. Traditional (Clinical) Pinching Method

For traditional methods the lateral pinch is 8.2–10.1 kgf (80.4–99.0 N) in males and 6.4–8.0 kgf (62.8–78.5 N) in females. The other pinches may be expressed as ratios of the lateral pinch as

follows. Handgrip strength is about 4.5 times lateral pinch strength.

Five-finger	1.3
Chuck	1.0
Pulp 2	0.7
Pulp 3	0.7
Pulp 4	0.45
Pulp 5	0.3

4.2. Nontraditional Pinching Method

For the nontraditional pinching methods in which the thumb is allowed to hyperextend, pinch strengths are significantly greater. For this method, at large pinch widths, chuck pinch is 16.9–18.4 kgf (165.7–180.4 N) for 2.0–5.6 cm width range, lateral pinch is 12.8–14.2 kgf (125.5–139.3 N), and pulp 2 pinch is 11.0–11.7 kgf (107.9–114.7 N). At greater widths all three pinches decrease steadily from these values. Chuck and pulp 2 decrease to 9.0 kgf (88.3 N) and 5.7 kgf (55.9 N), respectively, at 14.0 cm width, and lateral pinch to 8.0 kgf (78.5 N) at 9.2 cm width.

5. FACTORS INFLUENCING PINCH STRENGTH

MVC pinch strength magnitudes are strongly influenced by the method of pinching and the characteristics of the object being pinched. These two factors include wrist position (or wrist bending), forearm orientation, forearm position in the coronal plane, forearm support, elbow angle, body posture, finger joint position, position of inactive fingers, use of gloves, and personal factors such as age, sex (or gender), and anthropometry. The nature and extent of these these influences are discussed below. Unless otherwise stated, pinch strength below refers to static exertions using the traditional (clinical) method.

5.1. Types of Pinch

The various types of pinch commonly used for work tasks have been listed above. People know from experience that pinches are not all of equal strength, and several studies have been conducted specifically to determine which pinches are greater or less than others under similar pinching conditions. Swanson *et al.* (1970) and Mathioweitz *et al.* (1985) are two of the earliest large-sample studies. Most of these studies have concentrated on the lateral, chuck, pulp 2, and pulp 3 pinches since they are the most commonly used in the workplace, and they are customarily tested for clinical applications. The results have indicated that pinch strength magnitudes may be arranged in the following order, from strongest to weakest:

Five-finger and four-finger
Chuck and lateral
Pulp 2 and pulp 3
Pulp 4
Pulp 5

The individual pinch strengths within the pairs, such as five-finger and four-finger pinches, are not significantly different from each other. Some studies have found that the chuck pinch is stronger than the lateral pinch whereas others have found near equality of the two pinch strengths. The disparity is most likely due to variations in pinching conditions. However, the accumu-

lated evidence indicates near equality of strength for lateral and chuck pinches.

5.2. Types of Test

Pinch strengths measured by traditional (clinical) methods have yielded different results from those measured by nonclinical methods, both in the absolute magnitudes and in relative magnitudes among different pinches. When people pinch (i) with greater finger-object contact area, (ii) with freedom to hyperextend the thumb (for chuck and pulp 2 pinches), and (iii) with the hand at about worktable surface height (instead of at elbow or shoulder height), then they achieve greater MVC strengths and the chuck pinch is stronger than the lateral pinch at any pinch width. Moreover, at great pinch widths (greater than or equal to 6.8 cm) the pulp 2 pinch is stronger than the lateral pinch. Only pulp 2, chuck, and lateral pinches have been tested at large pinch widths.

5.3. Wrist Bending

The strongest pinches may be obtained when pinching with the wrist in its natural position — the hand at 15–35° dorsiflexion (also called extension, bending toward the back of the hand) relative to the forearm, depending on the person. Slight bending of the wrist away from this wrist position does not change strength significantly but sharp bending degrades strength appreciably (Imrhan 1991, Dempsey and Ayoub 1996). The amount of decrease depends on the direction of bending, with palmar flexion (flexion toward the palm of the hand) producing the greatest decrease, followed by dorsiflexion, ulnar deviation, and radial deviation in decreasing order of effect. In other words, up/down wrist bending is worse than sideways bending. Degradation due to flexion is about 24–43% and degradation due to extension is 9–31%, depending on the research study. Considering that wrist bending has been implicated as one of the major causes or contributing factors in carpal tunnel syndrome, tool and job designers should design such that sharp bending of the wrist is avoided.

5.4. Finger Joint Position

One study has found that the pulp 2 pinch is much stronger when the nonpinching fingers are flexed than when extended, with greater differences in females than in males.

5.5. Anthropometric Variables

Ergonomists have always been fascinated by the hope of developing a predictive equation for strength from one or more other variables that are easy to measure, such as anthropometric variables. Several of these variables, such as stature, body weight, and hand length are significantly correlated with various pinch strengths. However, attempts have been unsuccessful to find models that can predict pinch strengths from anthropometric variables accurately enough for ergonomic applications. Variables such as height, body weight, hand length, hand breadth, forearm circumference, biceps circumference, and grip diameter have all been explored. This means that pinch strength values for ergonomic engineering applications must be obtained from tables. Anthropometry-strength relationships are stronger in children than in adults and the elderly; this is most likely due to the parallel growth in body dimensions and muscle size.

5.6. Forearm Orientation and Position, Elbow Angle and Glove Use

A few studies have tested how pinches are affected by forearm orientation (pronated, supinated, and mid-orientated), forearm position in the coronal plane (from left to right of the body), external forearm support, elbow angle (90° and 180°), and glove use. Though differences were reported for these variables, they were either too small (2.5% and 3% for elbow angle, and 6–7% for forearm support) or too inconsistent across studies (forearm orientation and body posture) for these variables to be deemed influential; that is, some studies have shown an increase, decrease, or no effect for one or more of these variables. Moreover, biomechanical theory does not indicate that these variables should exert significant influences.

The single study on forearm position difference reported that the strongest position is with the forearm in front of the body with a decrease on either side; that is, an increasing/decreasing trend across the body. Use of gloves has been tested on the chuck pinch but it appears to be uninfluential; other pinches have not been investigated but it is unlikely the results will be different as long as the glove is not bulky enough to obstruct the pinch grip. Also, it is likely that pinch strength on sharp edges would be painful and weak, and it would be enhanced by using gloves of appreciable thickness to prevent pains but not thick enough to obstruct proper gripping.

5.7. Relationships with Other Manual Strengths

As expected, pinch strengths are strongly correlated with other types of hand strength, such as handgripping, pinch/pulling, and manual torquing, all of which use pinching muscles partly. Lateral pinch strengths have high correlation coefficients with torque strength on large-diameter cylinder handles, and lateral and chuck pinches have strongly correlations with pinch/pull strengths.

6. CONCLUSIONS

There are different types of pinch depending on which fingers are used. Their strengths vary considerably and any reference to pinch strength at work should specify clearly the type of pinch. Many factors influence pinch strengths and it would be simplistic and misleading to make reference to "pinch strength" as if it were a single entity. The conditions under which pinching occurs, such as pinch width, wrist angle, and use of gloves must be specified. Almost all the pinch strength data available in the literature is for static pinching with one hand using a restrictive method (clinical method) that does not reflect many pinching tasks in the workplace. Caution must therefore be exercised in applying this data to the design of work or hand tools for the workplace, and in assessing the severity of tasks. A few small data sets exist for methods of pinching resembling real tasks more closely. Since pinch strengths among studies vary widely (due to differences in subject samples and test methods), caution must also be employed in accepting data from any single study.

REFERENCES

DEMPSEY, P.G. and AYOUB, M.M, 1996, The influence of gender, grasp type, pinch width and wrist position on sustained pinch strength. *International Journal of Industrial Ergonomics*, 17(3), 259–73.

IMRHAN, S.N., 1991, The influence of wrist position on different types of pinch strength. *Applied Ergonomics*, 22(6), 379–84.

IMRHAN, S.N. and RAHMAN, R., 1995, The effects of pinch width on pinch strengths on adult males using realistic pinch-hand coupling. *International Journal of Industrial Ergonomics*, 16, 123–34.

MATHIOWETZ, V., KASHMAN, N., VOLLAND, G., WEBER, K., DOWE, M., and ROGERS, S., 1985, Grip and pinch strength: normative data for adults. *Archives of Physical Medicine and Rehabilitation*, 66, 69–74.

SMITH, R.O., 1985, Pinch and grasp strength: standardization of terminology and protocol. *American Journal of Occupational Therapy*, 39(8), 531–35.

SWANSON, A.B., MATEV, I.B., and DE GROOT, G., 1970, The strength of the hand. *Bulletin of Prosthetics Research*, 9, 387–96.

Force Exertion for (Consumer) Product Design: Information for the Design Process

B. J. Daams

Delft University of Technology, Department of Industrial Design Engineering, Jaffalaan 9, 2628 BX Delft, The Netherlands

Presently at: Daams Ergonomie, Mauvezand 17, 1251 JC Laren, The Netherlands

1. INTRODUCTION

First, a summary is given of design activities and accompanying ergonomic aspects. Next, a number of rules of thumb are given and some examples of application. Then, more precise data on force exertion are given. Finally some tips on measuring force exertion for design purposes are given.

2. DESIGN PROCESS

In a nutshell, the process of incorporating force exertion in design is the following.

First information should be gathered and analyzed. Existing products should be investigated. The function of the product, the target user group, the situation and the behavior of the users should be established or estimated. For design purposes, the weakest users are often more relevant than the strongest, and beginners more than experienced users. The result of the information and analysis phase is a program of requirements.

Then the implications for the design should be considered. Some elementary knowledge of physics and human force exertion, and some logical thinking are indispensable. Rules of thumb may be very useful in this stage. This may result in a schematic idea for the product with the most comfortable, efficient or maximal force exertion.

After the choice of the best concept the chosen idea will be detailed. Then it can be inferred which detailed information on force exertion is needed, to check whether the rules of thumb are right in this case, and to learn how high the exerted forces will be. This information must then be gathered from somewhere. A private investigation into force exertion with the right subjects and the desired postures is advised, but if there is not enough time, literature can sometimes give an indication of the range of forces involved. Now, knowing the limits within which the product should function, the program of requirements can be completed with operational requirements concerning force exertion.

Then the product can be designed. The first prototypes should be tested with subjects and evaluated, and the product should, if possible, be improved accordingly. This is important, especially if the information forming the basis for the program of requirements is obtained from literature.

If the required information on force exertion cannot be found, the probable forces involved cannot be predicted, but only very roughly estimated, and the design can neither be adapted beforehand to the needs of the users, nor be adequately tested and evaluated on the basis of a prototype. In most design projects, neither time nor funds are available to conduct experiments, and sometimes experience and enthusiasm are also lacking (except perhaps in cases where force interaction is a key topic for the design). So in general, designers have to rely on existing literature, which is so scarce that it takes much time to find details on the required force, if such information is to be found at all.

Thus, the information that is needed for the design consists of two parts:

- general information (rules of thumb) for the early stage of design; and
- detailed information (exact forces) for the detailing stage of design.

These rules of thumb are important, because they force the designer to think about a good design concept. If no effort is put in this first stage, it is no use to detail an ergonomically basically bad design.

3. RULES OF THUMB

The following rules are deducted from literature and physics. The concerning literature can be found in Daams (1994).

3.1. Users

- Women can exert on average two-thirds of the maximal force of men.
- For the lower extremities the ratio female/male maximal force is relatively larger, for the upper extremities it is relatively smaller.
- Maximal force is exerted between ~20 years and ~60 years of age.
- For different ethnic groups, maximal force may differ.
- People with rheumatism or arthritis (a common condition of old age) prefer push buttons over twist buttons. They do not like to hold buttons any longer than necessary.

3.2. Posture

- With two hands, more force can be exerted than with one hand (but not as much as twice).
- With the whole hand, more force can be exerted than with a (few) finger(s).
- With a support or a bracing possibility, more force can be exerted than without.
- With push and pull, most force can be exerted between shoulder and elbow height.
- With torsion and grip force: the closer to the body, the more force can be exerted.
- Use of body weight can lighten a heavy task.
- Pulling and pushing vertically down, maximal force is limited to the body weight of the user.
- Neutral joint position is to be preferred. Joints should not be held in extreme positions for a long time, and heavily loaded joints should not be held in extreme positions anyway.
- Moment = Force x Arm. A shorter arm with the same force thus results in a lower moment and vice versa. These simple calculations can help designing a lighter task.
- Maximal torque, but not maximal force, increases with increasing diameter of knobs and lids.
- Maximal forces exerted while standing with the feet close together are lower than those with one foot in front of the other.

● In free posture, more force can be exerted than in standardized or restricted posture.

3.3. Force, Time and Velocity

● Heavy work is lightened by rest periods.
● The longer a force is maintained the lower the maximal force.
● With low forces that are endured for a long period, the load caused by posture (especially the weight of the limbs) should be taken into account.
● The larger the movement, the more tiring it is.
● Large force exertion and precise positioning or handling do not go well together.
● With dynamic forces, eccentric forces are larger than concentric forces (eccentric is when the muscle lengthens during force exertion, concentric is when the muscle shortens during force exertion).
● Dynamic concentric forces are lower than maximal static forces in the same situation. The faster the movement, the lower the possible maximal force is.
● Endurance time of comfortable forces range from 0.4 (at 80%) to 0.2 (at 15%) of endurance time of maximal forces, when endurance time of comfortable force is defined as the "time to the first change of hand" (Daams 1994).

3.4. Grip

● Good diameter of a grip for lifting and carrying is 3–4 cm.
● Good grip span to exert grip force is ~4.5–5.5 cm wide.
● Grip should have no curvature, no sharp corners, edges, ridges, finger grooves or protrusions.
● Surface should be soft and non-slip.
● Grip force should not be a prerequisite for torsion, so the grip for torsion is preferably not round but has protrusions. The larger the protrusions, the less grip force is required for torque. A paddle or handle is best.

4. EXAMPLES OF THE APPLICATION OF RULES OF THUMB IN PRODUCT DESIGN

Questions concerning the forces that could be exerted on a product can never receive a standard, ready-to-use answer. Rules of thumb can help to configure a basic design that is easy to operate or to use. Some examples are given to illustrate this scope.

The first example concerns a large professional cheese slicer, as used in supermarkets (Figure 1). Present cheese slicers are fitted with a handgrip at the end of a blade that rotates around a pivot. The larger the blade (or the arm of the moment), the less force is needed to slice, but also the larger the movement of the hand. The handgrip is positioned in the same direction as the blade, so that the wrist is in an uncomfortable position when exerting force. The slicer is usually positioned on a table or bench, so that the force is exerted on the handle from about shoulder height to about elbow height. The users are women and men aged between 18 and 65. Instead of trying to find out how much force can be exerted in such a situation, thought is given to a more comfortable way to slice. Suggested improvements include lower positioning of the equipment, so that the force is exerted with the hand at elbow height and lower and body weight can be utilized, and a change of the handgrip so that the wrist need not be flexed to extreme degrees. If possible, the movement should be a translation instead of a rotation, so that force needs to be
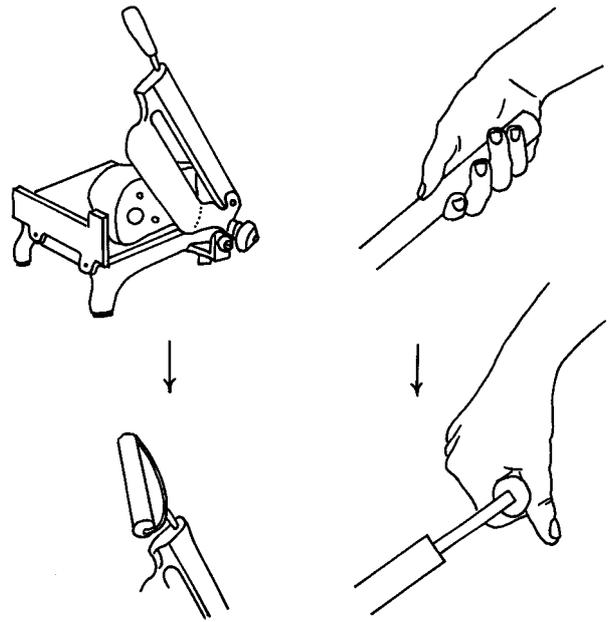
Figure 1. Suggestion for improvement of a professional cheese slicer (Daams 1994).

exerted in one direction only. The optimal length of the arm, weighing the length of the stroke against the force needed to operate the cheese slicer, may be determined experimentally.

The second example concerns a wheelchair for children in Sri Lanka who suffer from the consequences of polio. Their legs are paralyzed, so the vehicle has to be moved by using the arms. How much force can they exert? For some mechanical reason, force was to be exerted in one direction only. Deciding on a force direction based on the direction in which the greatest force can be exerted is in this case not advisable. When pulling in a horizontal direction, the child will tend to pull itself out of the seat, which is uncomfortable and prevents maximal force exertion, even if the subject were to be strapped to the back of the seat. On the other hand, when pushing, the child will be able to brace itself against the back of the seat, which is more comfortable, and allows more force to be exerted. Therefore, if a choice has to be made, pushing against the back of the seat is preferred. It should be noted, however, that to exert force in both directions in a cyclic movement is better for the development of the muscles. Uneven development of muscle groups may lead to incorrect loading of the joints and related problems in the future. The use of two arms instead of one will, of course, allow more force to be exerted, and will also stimulate better physical development in the subject. To know the maximal forces, they should be measured for the actual children concerned. Although there are some data on maximal pushing and pulling abilities of European children in literature, this information can not be used, as it can not be assumed that children who are disabled and children from different ethnic groups exert equal maximal forces.

Third example: how much force is exerted on a paper punch? This can easily be measured with a weighing scale on a table. Place the punch on top of it and the exerted force can be read from the scale of the weighing scale. The maximal force that can be exerted is limited by the weight of the user. The results of a

small experiment showed that this theoretical maximum is not attained by any of the subjects, their maximal force being slightly lower.

The fourth example: for the design of a portable or rolling easel, information on maximal exertable push, pull and carry forces were asked. The easel is intended to be used for outdoors painting and should be easily transportable by a person on foot. The target group consists of elderly people, so the required force forms an important aspect of the design. Carrying is no option, because it will certainly require more force and energy than simply rolling the easel along. Rather than pushing it, a wheeled object is preferably pulled along because this makes it easier to negotiate ramps and kerbs (this is everyday experience and can be backed up with statics). The force necessary to stabilize and maneuver the easel should be as small as possible, because energy and attention should not be diverted from the main activity, pulling the easel along. The question then arises how much force elderly people can exert when pulling something along. If the lowest part of the distribution (say the P_1) of elderly women is included in the target group, the force that can be exerted is near to zero, for these people have barely sufficient strength to walk about unsupported, and will not have much force left to pull easels around. Therefore, the less force is needed to pull the easel, the better.

The fifth example concerns a hand-held device to suck venom from insect bites. The device had to be operated with two hands. Much force had to be exerted pulling vacuum, while at the same time the device had to be positioned on the skin. When operated by the person who was stung, it could only be applied to half of the body. Advice was given to separate force exertion and positioning in time, for heavy force exertion does not go well with precise positioning. The resulting product was constructed according to this advice. The vacuum is first built up, and it is released by pressing the device down with one hand on the skin. Both force exertion and positioning are now easier. Added advantage is that the product can be positioned single handed, which improves the accessibility of the body when used by the person who was stung.

The conclusion is that, with some rules of thumb, no measuring is needed to come to a design principle which is ergonomically sound concerning force exertion.

5. DETAILED DATA ON FORCE EXERTION

After the concept phase of the design, the details should be looked into. How much force can be exerted by the envisaged user group? Or better, how much force can be exerted by the weakest users?

This question must preferably be answered by research, but usually there is lack of time, money and/or interest to do so. Therefore the question about the force that can be exerted can only be answered by literature. Adequate information on force exertion in literature is very scarce, because a situation in which force exertion is investigated seldom corresponds with the situation in which a product is expected to be used, and even less with the real situation. An attempt to gather design-relevant information and to present it in a way useful for the designer can be found in Daams (1994). A choice from these tables is given in

the articles "Push and Pull data" and "Torque data". They comprise pushing and pulling force of adults, children, elderly and disabled.

6. SOME TIPS ON MEASURING FORCE EXERTION FOR DESIGN PURPOSES

For research, intended to provide information for (consumer) product design in general: measure in free posture. Measure discomfort of hand forces by change of hands (Daams 1994). Measure not only young and healthy males, but also children and elderly disabled women and everything in between.

For research, applied to one specific product: measure in the situation of envisaged product use. Measure in free posture if that is appropriate. Have subjects from the envisaged user group (preferably the weakest/smallest, etc. because then possible problems will show easier) Measure only what you need to know.

Provide the subjects with clear instructions on the purpose, duration, expectations etc. of the experiment. During the actual exertion of force, allow them a few seconds to increase their force to the required level. If maximal force for a short time is measured, follow the Caldwell regimen (2 s increase, 4 s maximal, the average of the four seconds is the maximal force; Caldwell et al. 1974). Do not encourage the subjects before or during the measurements and do not provide them with any feedback (except during measurement of endurance times). For submaximal forces at a low level (40% of maximum, and less), measurement of maximal endurance can take up to > 30 min.

If research on force exertion is published, relevant information should be mentioned in the publication. Indispensable information include, for example, description of subjects (number, sex, age, anthropometric data, and health status and occupation if relevant), direction of exerted force, shape and size of handle or contact surface, position of handle in relation to the subject, posture (if possible with picture), bracing possibilities and method of measurement. All this information is indispensable.

7. CONCLUSION

It must be emphasized that the general way of considering and approaching force exertion and using rules of thumb is even more important than specific and accurate data, as proper, logical reasoning will in itself lead to better designs. Probably even the fact that designers are giving the problem serious thought would lead to some improvement. The right data can provide the finishing touch in realizing a good design principle.

REFERENCES

BORDETT, H.M., KOPPA, R.J. and CONGELTON, J.J. 1988, Torque required from elderly females to operate faucet handles of various shapes. *Human Factors*, 30, 339–46.

CALDWELL, L.S., CHAFFIN, D.B., DUKES-DOBOS, F.N., KROEMER, K.H.E., LAUBACH, L.L., SNOOK, S.H. and WASSERMAN, D.E. 1974, A proposed standard procedure for static muscle strength testing. *American Industrial Hygiene Association Journal*, 35, 201–6.

DAAMS, B.J. 1994, *Human Force Exertion in User–Product Interaction. Backgrounds for Design* (Delft: Delft University Press).

Force Exertion for (Consumer) Product Design: Problem Definition

B. J. Daams

Delft University of Technology, Department of Industrial Design Engineering, Jaffalaan 9, 2628 BX Delft, The Netherlands

Presently: Daams Ergonomie, Mauvezand 17, 1251 JC Laren, The Netherlands

1. INTRODUCTION

Ergonomics is the science and the art of adapting a product, machine or environment to its users. The product should be comfortable, safe, and effective and efficient to operate, handle or use. It should meet these requirements in a number of ways, one being the force required to use the product.

This article focuses on forces exerted on consumer products. As an introduction, the problems with force exertion on products are illuminated. In two other articles by the same author attention is given to the relation between design, ergonomics and research, and to information for the design process.

2. CONSUMER PRODUCTS AND THEIR USERS

Characteristic for consumer products is that they are used voluntarily, by a variety of non-specialist users with different backgrounds, different levels of education and of varying strength. Generally speaking, they are used less than 8 hours a day, 5 days a week, all year round. The relative importance of consumer products for the user ranges from "essential" to "just for fun." Usually the person buying a product is also one of its users. If the consumer is not satisfied with the product, it may be replaced or just remain unused, and it will certainly not be bought again.

3. IMPORTANCE OF DESIGN FOR FORCE EXERTION

Users, even the physically weakest, should be able to operate, use or handle a product designed for them. A designer should therefore take into account the physical capabilities of weaker users, like the elderly, women, children and the handicapped, if they are targeted as future users. This is especially important in view of the "gray wave" of elderly people that is now flooding industrialized countries, and will continue to do so for the next few decades. The ability of the elderly to handle everyday products in and around the house determines how long they will remain able to live on their own. This ability depends not just on the people themselves, but also on the proper design of the products in their immediate surroundings. About half of the people in The Netherlands who move into a home for the elderly do so because they become can not to do their own housekeeping (De Klerk and Huijsman 1993). Products designed for this special group would help them to remain self-sufficient longer if they so choose, from which both they and the community would benefit. Even if a product is not especially intended for the elderly, a designer should realize that physical strength varies widely among individuals and that products intended for the general public will be used by different people, varying from those who are almost incapable of exerting any force at all, as is the case with the ill and elderly, to those with more than three times the average strength, as with young, healthy and well-trained men. Many individuals belong to relatively weak categories, and the lack of information on the characteristics and abilities of these groups is conspicuous.

On the other hand, very strong users should not be able inadvertently to damage or break a product. If this happens, the product may still be functioning, in which case the damage will be no more than a minor annoyance. However, it may happen that the product is broken and cannot be repaired. In the worst scenario, depending on the product and the situation, a breakdown will cause serious injury. This, of course, is something a designer must try to avoid. For some products, safety standards and regulations are defined; for others the designer has to find out what the acceptable limits are.

It can be argued that in these modern times products are becoming increasingly mechanized, servo-assisted and equipped with electronics. Less and less force is required for interaction with a product. Soon, one will require hardly any force at all to use the products the modern world offers. Why bother then about the forces people can exert? Because this reduction of force needed to operate or use products does not come about spontaneously, but has to be brought on by the designers of new products. It is therefore necessary to know about the capabilities of future users. Furthermore, if a product is designed with a more sophisticated view to force exertion, servo-assistance may not always be needed, which consequently makes it less complicated, cheaper and possibly less damaging to the environment.

Although large mechanical resistance and the corresponding necessity of large force exertion should be avoided in products, some force will always be required. For most products, this is inevitable as some functions inherently require force. For some products, resistance is indispensable for proper positioning. Without resistance, positioning would be awkward and overshooting would easily occur. All continuously variable controls, for example those on electronic equipment, therefore require some built-in resistance. For other products, resistance is indispensable to provide feedback, as for example with rotating dimmer switches, where increased resistance indicates the point where moving the knob any further will switch off the lighting completely.

In itself, it is a good thing that products require some force to operate them, as it is very healthy to do some physical exercise in daily life. In a domestic environment in which everything can be controlled at the flick of a switch, the inhabitants through lack of exercise will run the risk of losing their physical condition due to atrophy of the muscles and the cardiovascular and respiratory systems. One example of this is the use of the increasingly popular power-assisted bicycle. For the elderly, the ability to open a can with a normal can-opener can decrease through the use of an electric machine. This deterioration of health due to the use of products that require little or no force mainly applies to people who do not take regular exercise. This may eventually result in deterioration of health on a national or even international scale.

4. EXAMPLES

Some products are renowned for their resistance to consumer

efforts. These cases indicate the consequences of bad design-for-force-exertion. The list of problematic products includes a large diversity of products, among which the nutcracker, the train door and the jam jar feature prominently.

An interview with eight Dutch elderly ladies who do their own housekeeping gives some insight in the kind of products they can not operate (Daams 1987). Some of the women have problems turning water taps or radiator valves not fitted with knurled knobs, medicine jars, toothpaste tubes, and hand-operated can-openers because they require too much force. None can open jam jars or bottles of lemonade by hand. They must ask someone stronger than they to open these products, or resort to a variety of tools and tricks (although the use of tools does not always guarantee success). With some corkscrews, extracting the cork from a bottle is very difficult, but the double-lever type of corkscrew will easily do the trick. Most of these women had electric watches and alarm clocks, but those with mechanical clocks had problems winding them. Winding a watch is mainly difficult due to the small size of the crown.

Jam jars that are almost impossible to open will be a familiar phenomenon to most people. It is the traditional example of a form of packaging that is difficult to open. It is referred to as such in newspapers and popular magazines. A patient can even attempt to open a jar of peanut butter, and in the struggle strain the muscles of an arm, as experienced by Dorrestein (1993: 128–9). Problems with opening jam jars also received attention in research (Berns 1981, Daams 1987, Imrhan and Loo 1988) and design (Schoorlemmer 1989).

The nutcracker is a popular though problematic product. Many different mechanisms have been tried, none perfect. Cracking a nut takes too much force, not only for weak persons, and certain types of nut will resist any force applied to them. With several nutcracker systems the result is often nut pulp mixed with pieces of shell. Some nutcrackers, however, require so much force to operate that even persons of average strength will not succeed in using them properly.

Door-closing devices are designed to keep doors shut against draughts or to comply with fire regulations. Some perform so well that it becomes extremely difficult to open the door. Kanis (1991) describes a door in a train passage that automatically closes with such force that the elderly are hurt.

In many cases, operating the push button to flush a lavatory requires not only a minimum force, but also a minimum speed to succeed. This severely increases the degree of difficulty of the operation. In one case, the only way a healthy young woman could get the flusher to work was by applying a karate kick.

Opening the lid of some large photocopiers ought to be prohibited by law. The lid is very heavy and must be opened by pulling a handle positioned so that the wrist is supinated maximally and the hand is in near-maximal dorsal flexion. Pain in the wrist is the price to pay for a few good copies.

Another interesting product of which use is complicated by a lack of feedback is the car handbrake. It is hard to tell whether the brake is pulled sufficiently. Weaker subjects are stuck with the fear that their efforts were not sufficient and they will find their vehicle several meters from where they parked it (and hopefully undamaged). A stronger but not exceptionally strong subject is known to have broken off a brake handle with a seemingly gentle pull.

Disabled people who require the temporary aid of crutches can, after some exercise, move around quite swiftly and can sometimes even continue business as usual. Only then do they find out how great a force is applied by most door-closing devices. It is hard to open automatically closing doors while standing on one leg, and it requires some dexterity to keep them open long enough to pass through. Doors with less powerful closing devices are much easier to cope with, and apparently perform equally well in all other respects.

Latchkeys can be a problem too. An old woman had to make herself a tool for turning the key to her front door — something she could not do otherwise by hand.

Lighting a gas fire or a geyser can be quite tricky. To get the pilot flame going a button must be pushed or turned and held for as long as five minutes, or even more. This is certainly not an easy task with a spring-loaded button that requires much force.

Child-resistant closures are another category of products in which the required force is an important factor determining the rate of success. On the one hand, children up to a certain age should not be able to open the closure, while on the other hand the elderly must open it without problems. A complicating factor is that the maximal grip force of most 4-year olds exceeds that of many of the elderly. As the closures should be easy to open by the latter, the resistance to opening by children is often further improved by using a combination of two forces (like pushing and turning at the same time) to open the closure. Nevertheless, this still requires too much force of the weakest elderly person who must wrestle regularly to obtain their daily medicines or to open bottles of household detergents.

The examples of man struggling with various types of packaging, apart from the notorious jam jar, are many, notably with coffee-creamer cups (Kanis 1989a), milk cartons (van Putten *et al.* 1990), complimentary sachets of bath foam in hotels (den Uyl 1979) blister packaging, plastic wrapping, certain juice cartons, shrink-wrapped cucumbers (Stephan and Daams, personal communication, 1999) and slices of luncheon meat hermetically sealed in plastic. In those cases the packaging material is so smooth that the friction between packaging and hand is very small, and in addition the shape is often such that it is hard to get a good grip on it. Consequently, although a subject may exert sufficient force, he or she can not transfer it to the packaging.

Some products are not as easy to handle or operate as they could be. A good design may require investments, and when the benefits are not directly visible such improvements find little support. The cost aspect is especially important in packaging and may explain why a relatively large number of complaints originate from that area.

For some products, the cause of the problems can be bad maintenance (like those with the latchkey), or bad adjustment (like the toilet push button). This does not make the problem less serious for the user. Designers should allow for bad maintenance and bad adjustment in the design of their products.

5. CONCLUSION

It is of vital importance for designers to take the physical capacity of the users into account when designing a product.

REFERENCES

BERNS, T. 1981, The handling of consumer packaging. *Applied ergonomics*, 12, 153–61.

DAAMS, B.J. 1987, *Onderzoek naar torsiemomenten in de huishouding met betrekking tot ouderen*. Internal report (Delft: Delft University, Faculty of Industrial Design Engineering).

DORRESTEIN, R. 1993, *Heden ik* (Amsterdam: Contact).

IMRHAN, S.N. and LOO, C.H. 1988, Modeling wrist-twisting strength of the elderly. *Ergonomics*, 31, 1807–19.

KANIS, H. 1989, *Het gebruik van koffiemelkkuipjes*. Internal report (Delft: Delft University, Faculty of Industrial Design Engineering).

KANIS, H. 1991, Ergernomie in de trein. *Tijdschrift voor Ergonomie*, 16, 22–4.

KLERK, M.M.Y. DE and HUIJSMAN, R. 1998, *Ouderen en het gebruik van hulpmiddelen*. Report no. 93.24 (Rotterdam: Instituut voor Medische Technology Assessment, Erasmus Universiteit).

PUTTEN, VAN, R., KANIS, H., MARINISSEN, A.H. and WEEGELS, M. 1990, *Usable Does Not Mean Handy*. Video (Delft: Delft University, Faculty of Industrial Design Engineering).

SCHOORLEMMER, W. 1989, Het openen van verpakkingen met schroefsluiting. Thesis (Delft: Delft University, Faculty of Industrial Design Engineering).

UYL, B. DEN 1979, *Vreemde verschijnselen* (Amsterdam: Querido).

Force Exertion for (Consumer) Product Design: Research and Design

B. J. Daams

Delft University of Technology, Department of Industrial Design Engineering, Jaffalaan 9, 2628 BX Delft, The Netherlands

Presently: Daams Ergonomie, Mauvezand 17, 1251 JC Laren, The Netherlands

To clarify the discussion on force exertion on consumer products, some characteristics of their relation are given, as is their respective relationship with ergonomic research. For more information on these subjects, see Daams (1994).

1. FORCE EXERTION AND CONSUMER PRODUCT DESIGN

The two main differences between consumer products and professional products are that consumer products are used voluntarily and by a broad group of users, with much variation in characteristics and background. Users of professional products are aged between 18 and 65, are generally healthy and for some occupations are predominantly male. Consumers include children, the elderly, the physically disabled and the world's largest minority group — women. Users also range from the mentally handicapped to the highly gifted. With regards to the force required to operate a product, the capacities of all groups must be taken into account. Everyone should be able to operate the product — or even better, enjoy operating it.

To the user, the relative importance of consumer products varies from "essential" to "just for fun." Many products range somewhere between those two extremes. For essential products, the most important thing is that the product can be operated, whereas for fun products enjoying their use is the main requirement. If the consumers do not like the product, they will not buy it again. Usually a consumer who buys a product is one of the users, whereas in the professional world different individuals often handle the decision to purchase, the purchase, payment and use. It could well be that satisfaction of the user contributes more to the sales figure for consumer products than for professional ones, for which, for example, efficiency and safety may be more important to the decision-makers. Therefore, the subjective experience of the consumer is an important factor in the assessment of what is an acceptable force for operating a product. The information on the maximally exerted force is in this view less important than information on the experience of the user during exertion of submaximal forces. The time of exertion must also be taken into account. Forces on consumer products can be exerted for any period, but they are usually not operated with the intensity and frequency of a full working day (8 hours a day, 5 days a week, all year round). Therefore, the necessity to prevent injury through excessive exertion during work is not the first priority in the research into force exertion on consumer product design.

Force can be exerted both statically and dynamically. It is estimated that with products in general, dynamic force exertion occurs probably more often than static force exertion. Mostly, force is exerted to create a certain effect, like opening, turning on, fastening, propelling or adjusting. By moving a handle, a knob or even the whole product, the desired effect may be obtained. Some forces are initially static, like those exerted on the lids of jam jars, but in the end the idea is that the lid should move.

2. FORCE EXERTION AND ERGONOMIC RESEARCH

Research on physical strength and force exertion is found in several areas of ergonomics.

The first place is industrial ergonomics. Emphasis in research is put on lifting and carrying, or "manual materials handling" as it is called. The aim is to establish the maximum workload, so that laborers will survive an 8-hour working day, 5-day working week, without physical problems. Research is geared to prevent injury and improve efficiency. Research on cyclic work is necessary. Subjects are aged between ~18 and 65.

Second is military ergonomics. Research for military purposes focuses mainly on the control and maintenance of vehicles and aircraft. Subjects are predominantly young and male.

Third is healthcare. Forces are measured to establish the progress of disabled people. Static grip force is a favorite in this area, and research is often done to establish standards for various groups of people, to which the patient's grip force can then be compared. The maximal force is the maximum score of a short, maximal exertion. Subjects are recruited from all age groups, varying states of health and both sexes.

The last group is sports. The aim of research here is to monitor the condition of athletes and to establish the effect of training schemes. In this case, dynamic force and staying power are investigated rather than static force exertion. The measures used to determine these are the maximum oxygen uptake (aerobic capacity) and power output. Subjects are usually younger people, both athletes and others.

Every area has its own research purposes, methods and subject groups. Research of forces on products is scarce, and in general is too specific, and directed towards one product or a select group of products. Research on force exertion according to a more general product-oriented principle, with a large variety of subjects, could make the results more valuable for designers.

3. PRODUCT DESIGN AND ERGONOMIC RESEARCH

The relation between design and research is a controversial and tricky one. This applies not only to ergonomic aspects, but also to other areas of industrial design engineering. The cause is sometimes attributed to bad communication between researchers and designers, but at close inspection it is clear that there are many discrepancies between their activities. Researchers are often interested in the cause of a phenomenon, whereas designers need data and care little where these come from or what the reasoning behind them is. Researchers extract information from the surrounding world. Designers analyze information to generate a new idea for a product. Thus, to researchers information is the goal of their efforts, whereas for designers it is the starting point.

To designers, the gathering of information should be quick, easy and to the point, in order to get on as quickly as possible with the design work. They prefer not to indulge in statistical analyses or to linger at the deeper causes of the results. To researchers, experiments carried out that way are considered "quick and dirty." Although researchers and designers may agree that research should be to the point, researchers are usually not familiar with the way a product is designed or the problems encountered by the designer. Consequently, they are ill-equipped to judge the relevancy of their research to design practice.

If a way can be found to show researchers how to gear research for design purposes and to show designers how to interpret data correctly and with the right degree of caution, it will greatly improve their ways of communication and mutual understanding and, consequently, the results of their work.

4. DESIGN AND RESEARCH ON FORCE EXERTION

Research on force exertion is usually conducted in one specific situation, with only a few variables varied each time. The daily use of products, on the other hand, has the potential of an enormous variety of situations, of which the designer can anticipate only the most common. The number of possible combinations of postures and force directions is nearly unlimited, and the research that has been done in this area lacks system and is rather thin on the ground. Furthermore, the usefulness to designers is not taken into consideration when the research is set up. A situation in which force exertion is investigated seldom corresponds with the situation in which a product is expected to be used, and even less with the real situation. Postures are standardized and mainly static, maximal forces are measured for ~4 s, whereas in practice only the position of a handle or control is known and subjects will exert static or dynamic submaximal force for any time and in any posture they feel like, and to some extent with different muscle groups. And even if the usefulness to designers is taken into consideration, usually only one (type of) product is investigated so the results are too specific to be of much use for other product areas.

Research should be geared to generate those results useful to designers. Designers know the product they are going to make, they can target certain user groups and can estimate the situation in which the product is likely to be used. Armed with this information, they should find out how much force can be exerted.

REFERENCE

DAAMS, B.J. 1994, *Human Force Exertion in User–Product Interaction: Backgrounds for Design* (Delft: Delft University Press).

Gaze-based Control

G. L. Calhoun

Air Force Research Laboratory, AFRL/HECP, 2255 H Street, Bd. 33, WPAFB, OH 45433, USA

1. OVERVIEW

Human beings naturally look at objects that they want to manipulate or use. For applications in which an operator is controlling a computer-based function or device, harnessing the direction of the operator's gaze promises to be a very natural and efficient control interface. With such control, the computer initiates a predefined action once it receives an input based on the operator's point-of-gaze. Based on the assumption that an operator looks in the general direction that the head is pointing, "gaze-based" control has already been employed for several applications. These systems translate the position of an operator's head into a system input, enabling "head-based control." More recently, "eye-based control" has become commercially available, enabling the eye point-of-gaze to be translated into a system input. Unless the head is stationary (or, with some tracking systems, held within a small motion box), determining the eye gaze point also involves tracking the head position/rotation. For gaze-based control to be useful, it is important that the operator's eye and/or head movements remain natural and not involve any unusual blinking, lengthy fixations or fatiguing inputs.

One common approach to implementing gaze-based control is to combine line-of-sight position data with line-of-sight dwell time criteria. The operator selects an item on a display simply by looking at it for the criterion time. Using dwell time to initiate the control action is particularly useful if the operator's gaze is only being utilized to call up additional data. In this manner, the operator's sequential review of a series of icons can be made more rapidly, with detailed information popping up, as the gaze briefly pauses on each icon. Typically, required dwell times range from 30 to 250 ms. Longer dwell times tend to mitigate the speed advantage of gaze-based control. However, shorter dwell times increase the likelihood of a "Midas touch," with commands activating whereever the operator gazes.

One solution to this potentially distracting problem is to require a consent response. Such a response can also be used to reduce dwell time requirements and minimize the inadvertent selection of objects. With a consent response step, gaze-based control is similar to the operation of a computer mouse and button press. The gaze position (or mouse) indicates the response option on a display and the consent response (or button press) triggers the control action. Different types of consent responses have been evaluated and it is recommended that a dedicated, conveniently located button be employed as a universal consent response, unless the application environment dictates a hands-free option.

Another possible implementation enables object selection either by satisfying the dwell time requirement *or* by pressing a button, whichever occurs first. Alternatively, dwelling on an object for about 30 ms can cause it to be highlighted, while a button press or voice command is used to select it. This highlighting may be maintained, without requiring constant gaze on the object, so that the operator can make the consent response while performing another task.

2. HEAD-BASED CONTROL

2.1. Applications

Since humans can position and hold their head with relatively good accuracy and stability, head-based control can be used to place a cursor on a selectable item or object. For instance, the helmet-mounted sights in aviation environments use head position and orientation in this manner. Fully integrated sights allow the pilot to lock radar and weapon systems onto targets without maneuvering the aircraft. The sighting reticle provides a fixed line-of-sight with respect to the head and this feedback is critical for precise head pointing. The pilot performs these operations simply by looking (i.e. pointing the head) to align the sight on the target, and executing appropriate manual consent commands. In rehabilitation applications, physically limited operators who still have head mobility can interact with a computer-based system by aiming a pointing device with head positioning. In these applications, head-based control is useful for tasks short in duration or when traditional input devices cannot be employed. However, for typical human computer interactions, head-based control is not ideal. For these applications, able-bodied operators would find the frequent and precise head movements tiring and unnatural to perform.

Head-based control is well suited to non-command-based applications (Adams and Bentz, 1995). For example, the head-mounted display systems used in simulated and virtual environments seamlessly update the displayed scene based on the user's head position. In "visually coupled" airborne systems, helmet-mounted display components present task-critical information, regardless of where the pilot is looking. In these examples the operator does not produce an explicit command to change the display viewpoint or information. The system senses the change in gaze direction as an implicit signal to take this action.

2.2. Head Tracking Methodology

Mechanical, acoustic, electro-optical, and magnetic technologies have been employed in head tracking systems. Two of these approaches, electro-optical and magnetic tracking systems, have been used most extensively. Optical trackers generally use light-emitting diodes located on the object to be tracked. Multiple cameras are used to measure the location of each light source. Magnetic trackers include a transmitter that is made up of three coils radiating electromagnetic fields in a radius of a few meters. The three coils in the sensor (located on the tracked object) receive varying signals depending on its position relative to the transmitter. Both of these approaches employ sensors mounted on the head and externally mounted transmitters. Magnetic trackers require the mapping of ferromagnetic and metal conductive surfaces in the user's environment. Some of the performance parameters, as well as the strengths and weaknesses of each approach are summarized in Kocian and Task (1995). Magnetic systems (accuracy ranging from 0.75 to 2.5 mm translation and 0.15–0.5° orientation) have probably seen the widest use and can be considered a relatively mature technology.

More recently, advancements in micro-fabrication techniques have made accelerometers and gyroscopes available for consumer markets. These inertial "transmitterless" sensors utilize the earth's gravity and magnetic fields as reference and do not require external receivers/transmitters. They do, though, require frequent

drift correction. Also, inertial systems cannot monitor an operator's absolute position in space, as only linear acceleration and rotational information are sensed. With their small size and modest cost, application of inertial trackers in computer systems will undoubtedly increase, as already seen in a variety of cordless mouse controllers with embedded inertial sensors.

The ideal choice of a head tracking approach for a specific application depends on a number of factors including the operational environment and performance requirements (e.g. accuracy, precision, resolution, range, and delay). The last factor, in particular, is difficult to quantify, as the operator experiences the overall delay in the head tracking process which is a function of a several variables (e.g. number of update cycles required to obtain an adequate measurement, type of movement pattern being tracked, and degree to which time is required for graphics processing and image display). Considering all of these factors, system throughput delays with magnetic trackers can be as high as 400 ms. Usually, the delay is ~40 ms as most trackers sample at 120–240 Hz. To improve overall system performance, head position prediction algorithms can be employed. Another promising approach is to augment magnetic or optical system data with information from inertial sensors which require only a fraction of the processing time (~1 ms) since they measure changes directly. Moreover, a hybrid system consisting of inertial sensors and one or more other head tracking systems also provides absolute positional information.

3. EYE–BASED CONTROL

3.1. Applications

The visual system is the primary channel for acquiring information. Because of its speed, accuracy, and stability, it may be advantageous to use the direction of eye gaze as a control input. In other words, if the operator is looking at a control-related function, it is more efficient to use the operator's eye line-of-sight to inform the system, rather than align the head or manually slew a displayed cursor, over the function. In this manner, eye-based control can increase the speed of control operations. However, moving one's eyes is usually a subconscious act, and it is relatively difficult to control eye position consciously and precisely at all times. Therefore, there is an increased likelihood of unintentional activity in the eye-line-of-sight signal for control. For this reason, a fully "eye-driven" mouse is not practical. Rather, natural eye movements should be used to provide a direct pointing capability that can be combined with other interface modalities to command activation. Application of eye-based control also needs to take into account the accuracy limits of the eye-tracker (Calhoun, Janson and Arbak, 1986).

The majority of commercial eye trackers are used in laboratories or in the home as a communication device for individuals with motor control disabilities. It is anticipated that once developers capitalize on advances in optics, sensor and processing technologies, eye-tracking systems will become affordable for a variety of home and work environments. Moreover, efforts toward miniaturizing eye tracking components will facilitate their integration into head mounted displays.

3.2. Eye Tracking Methodology

Methods for tracking the eye (Jacob 1995) can be classified into those that: (1) measure the electrical potential of the skin around the eyes, (2) involve image processing of one or more features that can be optically detected on the eye and (3) employ a special contact lens that facilitates eye position tracking. The latter method is too intrusive for most applications. The first method is probably the least expensive and easiest to implement. However, the drift inherent in electro-oculography (EOG) measurements makes this technology more suitable for measuring eye velocity and acceleration profiles rather than measuring eye point-of-gaze. It is possible that EOG tracking would suffice if the eye-based control operations only required detecting whether the operator is looking generally left, right, up or down.

The most practical line-of-sight measurement technique involves illuminating the eye with a near infra red source, capturing an image of the eye with a solid state video camera, and tracking one or more features that can be optically detected on the eye. The most often used features include the limbus (boundary between the iris and sclera), the pupil, movement of the lower eyelid, the reflection of a light source from the cornea (1st Purkinje image or corneal reflex) and the reflection from the rear surface of the eye lens (4th Purkinje image). However, these video-based methods, besides being more complex and costly to implement, are very sensitive to the placement of tracking components.

Tracking a single feature on the eye does not enable one to discriminate between eye rotation and eye translation caused by movement of the head. This ambiguity can be resolved by tracking two features, which are at different radii from the eye center of rotation. The two sets of features most commonly employed are (1) the pupil center and 1st Purkinje image and (2) the 1st and 4th Purkinje images. Two-feature eye trackers thus permit a small amount of head motion, assuming that the user stays within the field of view of the video camera observing the eye. Some table-mounted systems also employ a servo-controlled camera or mirror, which follows the operator's eye as the head moves. Such systems permit head motion within a volume of ~0.03 m^3. To permit essentially unlimited head motion, head-mounted eye tracking is required. The eye tracker data are used to determine eye position with respect to the head and are combined with head position data to compute line-of-sight in the environment.

Commercially available systems permit eye-tracking accuracies of ~1° of visual angle. At the typical 61 cm viewing distance, this translates into a resolution of ~1 cm on the surface of a computer monitor. Eye tracking, therefore, provides much coarser operation than a mouse or trackball. Unless a head-mounted system is employed, the range over which eye movements can be tracked is limited to about the area of a 48-cm monitor. Another important constraint is the temporal resolution of the eye tracker. The feature-tracking systems are video-based and typically operate at 60–120 Hz frame rates. Total throughput delays of 25–50 ms can be expected which includes updating the eye position every 8–17 ms, filtering the data and computing the eye position.

4. FUTURE DIRECTIONS

In summary, head-based control is not the best substitute for manual inputs in human–computer interaction. Head mass, head-tracker delays, and the need for frequent, precise head movements all tend to degrade operator performance. However, there are several applications where head-based control is beneficial,

especially for non-command based applications. Eye-based control is a promising alternative for numerous applications in which high control resolution is not a requirement. Ware and Mikaelian (1987) evaluated eye line-of-sight control for computer inputs and found that object selection and cursor positioning tasks were performed approximately twice as fast with an eye tracker as with a mouse.

While some applications will continue to employ head- and eye-based control in an explicit command mode, it is clear that this does not fully exploit the capability of gaze-based control. Seamless interpretation of head and eye movements to infer operator interest, needs and desires appears to be a much more powerful, albeit challenging, path.

Both head and eye trackers would benefit from enhancements that enable faster measurements and transformation of these measurements into input signals for human computer interactions. Future improvements in the accuracy and resolution of head and eye trackers are not likely to be major factors in enhancing control operations. Head trackers are already quite good in this regard and adequate for most applications. Eye trackers are approaching the 1° useful limit imposed by the size of the fovea. Improvements are needed, though, in the stability and repeatability of eye tracking so noisy eye signal data or momentary signal loss does not impact eye-based control. Additionally, the video-based eye tracking systems are not particularly robust in high or variable illumination environments.

REFERENCES

ADAMS, K.D. and BENTZ, B.D., 1995, Headpointing: the latest and greatest. In *Proceedings of the RESNA 18th Annual Conference*, Vancouver, Canada, 446–8.

CALHOUN, G.L., JANSON, W.P. and ARBAK, C.J., 1986, Use of eye control to select switches. In *Proceedings of the Human Factors Society 30th Annual Meeting*, 154–8.

FREY, L.A., WHITE, K.P. and HUTCHINSON, T.E., 1990, Eye-gaze word processing. *IEEE Transactions on Systems, Man, and Cybernetics*, 20, 944–50.

JACOB, R.K., 1995, Eye tracking in advanced interface design. In W. Barfield and Furness (eds) *Virtual Environments and Advanced Interface Design* (New York: Oxford University Press), 258–88.

KOCIAN, D.F. and TASK, H.L., 1995, Visually coupled systems hardware and the human interface. In W. Barfield and Furness (eds) *Virtual Environments and Advanced Interface Design* (New York: Oxford University Press), 175–257.

LONGINOW, N.E., 1994, Predicting pilot look-angle with a radial basis function network. *IEEE Transactions on Systems, Man, and Cybernetics*, 24, 1511–18.

WARE, C. and MIKAELIAN, H.T., 1987, An evaluation of an eye tracker as a device for computer input. In *Proceedings of Human Factors in Computing Systems and Graphics Interface Conference*, Toronto, Canada, 183–8.

Gesture-based Control

G. R. McMillan and G. L. Calhoun

Air Force Research Laboratory, AFRL/HECP, 2255 H Street, Bd. 33, WPAFB, OH 45433, USA

1. INTRODUCTION

Hand and body gestures are an important component of interpersonal communication. Gesture-based control seeks to exploit this channel for human–machine communication (Baudel and Beaudoin-Lafon 1993). Most currently available input devices only make use of discrete pieces of data produced by an operator's movements. This sometimes stems from the use of intrinsically discrete devices, such as a keyboard or numeric keypad. Even with continuous input devices, such as mice, only specific events and data points (e.g. the coordinates of the pointer when the operator clicks) are taken into account by most applications (Bryson and Levit 1992). Gesture-based interaction attempts to take advantage of the continuity and dynamics of an operator's movements, instead of extracting only discrete information from these movements (Brooks *et al.* 1990). Although the terms are sometimes used interchangeably, *gesture* formally refers to dynamic hand or body signs, while *posture* refers to static positions or poses.

2. GESTURE MEASUREMENT TECHNOLOGY

Gestures can be measured using a variety of hardware devices. Contact devices, besides classical ones such as mice, trackballs and touch screens, include a variety of more exotic items such as spaceballs and three-dimensional (3-D) mice. The head, hands and body can be localized in space using trackers, video techniques, gloves or suits. Trackers are devices that allow one to measure the position and orientation of a body part in space. Video techniques use image processing in order to identify a specific body part and then reconstruct its position, orientation and posture from two-dimensional (2-D) video images. Gloves and suits allow one to measure the relative positions and angles of body components. A comprehensive directory of manufacturers of gesture capture devices is available in Buxton (1998).

2.1. Contact Devices

This category includes devices such as light pens, touch screens and mice. They typically move in 2-D and operate through a straightforward translation into 2-D screen space. Such devices can, however, be used for more advanced purposes, such as 2-D gesture or handwriting recognition. The most serious limitation of these devices is that operator input is quite constrained. One hand is generally completely involved with the contact device.

2.2. Trackers

These devices allow one to measure, in real time, the position of an object in space. They can be used for tracking the head, hands or other body components. They have also been used for body posture recognition, although the number of devices affixed to the body can make such systems awkward.

- *Mechanical* systems connect the tracked object to potentiometers, or other sensors, using rods or cables. This allows very high update rates and very low latencies. It is also an inexpensive approach. On the other hand, the usable range is small and the apparatus impairs free movement.

- *Electromagnetic* trackers include a transmitter made up of three coils radiating electromagnetic fields in a radius of a few meters. The sensor, located on the tracked object, is also made up of three coils. It receives varying signals depending on its position relative to the transmitter. These trackers are moderately expensive, but on the whole they are the most precise and popular of non-contact techniques. They also offer a large operating range. One must account for metallic objects and other sources of electromagnetic interference in the tracking region. Also, there must be an electrical connection between the sensor element and the electronic unit. This can limit free movement.

- *Ultrasonic* trackers make use of ultrasonic pulses to compute distances based on time propagation measurements. The main advantage of these trackers is that they work seamlessly in metallic environments. They also tend to be less expensive than electromagnetic systems. However, ultrasonic trackers must have a direct line-of-sight between the emitter and sensor. The latency is greater than with other trackers since it includes the propagation of ultrasonic waves.

- *Optical* trackers generally use light emitting diodes or reflective dots located on the object to be tracked. Multiple cameras are used to measure the location of the light source or reflective marker. The 2-D marker positions detected by each camera are correlated to compute the 3-D coordinates. These devices face the same line-of-sight constraint as ultrasonic trackers, and the usable range is similarly limited.

Recently, a variety of low-cost *transmitterless* trackers have begun to appear using principles similar to aircraft and missile inertial guidance systems. The sensing components may include inclinometers, Hall-effect compasses, gyroscopes or accelerometers. Some of these devices use telemetry to communicate with the electronic control unit, so no cables encumber the operator. While they are often not as accurate as the devices discussed above, many are inexpensive and have satisfactory capabilities for selected gesture measurement applications.

2.3. Computational Vision Systems

These systems use classical image recognition techniques to find silhouettes of the hands or body and to identify gestures. The computational problems are even more challenging than with optical tracking systems, if real-time operation is required. Limited camera resolution necessitates a compromise between adequate recognition of small elements, such as fingers, and the large field of view necessary for free movement. Obstruction of the fingers by the hand or other body segments is another problem. Nevertheless, this approach is inherently unobtrusive and has been successfully employed.

2.4. Gloves

Gloves measure hand and finger angles and movements of the fingers relative to the hand. Most can be equipped with a position tracker in order to measure global hand position. Numerous sensors are needed and the resulting data rate can be high. Various measurement technologies have been used including optic fibers, Hall effect sensors, resistance variation or accelerometers. The main problems encountered are repeatability, precision and reliability. Almost every glove needs calibration before each use,
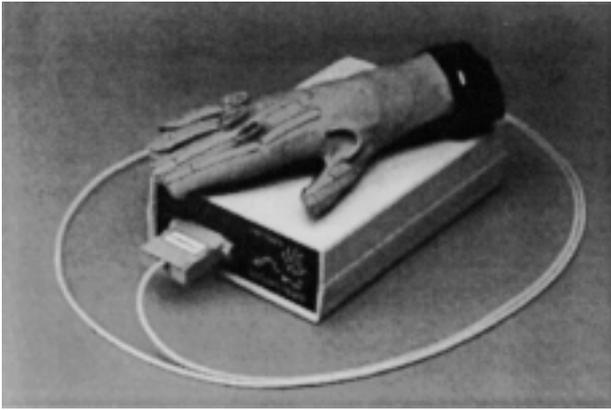
Figure 1. The CyberGlove. Photo: courtesy Virtual Technologies, Inc., Palo Alto, CA, USA.

since the manner in which it is worn greatly affects the measurements. A sophisticated example, the CyberGlove, employs 18- or 22-foil strain gauges to measure finger, thumb and palm relationships (Figure 1).

2.5. Some Issues in Measurement Device Selection

The optimal gesture measurement device depends not only on the task, but also on the environment in which the device is used. A positional hand tracker might be the preferred device in a relatively benign environment, but might become unusable under the acceleration and vibration levels found in a fighter aircraft or helicopter. A contact device would have the advantage of supporting the hand, but the space needed to integrate such a device must then be considered. Glove-based gesture recognition might be incompatible with safety equipment or might interfere with other tasks requiring fingertip sensitivity. System lags that are tolerable in a laboratory experiment might become problematic when the operator has to attend to several tasks at once. Environmental and integration issues such as these must guide the choice of gesture measurement devices.

3. GESTURE RECOGNITION

The algorithms for determining fundamental positions and angles, based on the sensor inputs, are provided with the systems described above. With glove-based systems, some individual operator calibration is required (Fels and Hinton 1993). Magnetic trackers require the mapping of ferromagnetic and metal conductive surfaces in the operator's environment, but no individual operator calibration. For interactive applications that employ rapid body movements, one may need to add or modify movement prediction algorithms to compensate for system delays. Although general-purpose posture recognition software is becoming available for glove-based systems, the development of algorithms to recognize specific postures or gestures is often left up to the system designer. These algorithms tend to be application specific, although some general approaches are common. For example, the recognition of a fixed set of hand postures is often based on look-up tables that contain acceptable ranges of values for each position and joint measurement.

Gesture recognition is a much more challenging problem since pattern analysis must be performed on moving body

segments. Many approaches compare the motion vectors for each body segment to reference vectors representing the target gesture. This match must be within error tolerances and these tolerances are weighted by the contribution of each motion vector to gesture discrimination. For robot control applications one must address the kinematic differences between the human hand and the robot hand. Algebraic transformations have been employed to perform this human-to-robot mapping. Alternatively, the kinematic differences can be resolved by determining the 3-D position of the operator's fingertips and driving the robot's fingertips to match. It is also a challenging problem to identify the beginning and end points of a gesture. Typical solutions require the operator to take a "default" posture between gestures, which serves as an anchor for the system.

Still another issue to be addressed is the "immersion problem." If every movement is subject to interpretation by the system, the operator can be deprived of interpersonal communication for fear that a movement will be acted upon by mistake. The solution is to provide an effective and unobtrusive way of detecting whether a gesture is addressed to the recognition system or is a part of normal interpersonal communication. "Charade," a system for control of computer-based presentations to an audience, addressed this issue by creating an active zone in which gestures were effective. Gestures performed in a defined area in front of the screen were recognized, while gestures directed toward the audience were ignored.

4. MODES OR STYLES OF CONTROL

Given that a specific gesture can be reliably discriminated from other activity, one still must determine how it will be used for interaction with a system. First, one can choose to use discrete or continuous features of a gesture. Within each of these categories, three styles of input may be considered:

- *Direct* — Features of the gesture or posture generate kinematically similar actions in the task domain. One-to-one control of a robot hand is a good example of this style of interaction.
- *Mapped* — Features of the gesture or posture are mapped in some logical fashion to actions in the task domain, but there may be no kinematic similarity between the features and actions. For example, the number of raised fingers might indicate which level of force should be applied by a hydraulic press.
- *Symbolic* — Features of the gesture or posture are interpreted as commands to the system. While this may be similar to the mapped style, it may also be significantly more abstract and may employ knowledge-based reasoning to determine the intention or emotional state of the operator. Most uses of facial expressions would fall into this category.

5. FEEDBACK REQUIREMENTS

In many applications the only feedback that is provided, or required, is the system's response to the recognized gesture. Examples include simulated movement in the direction toward which the operator is pointing and synthesized speech following recognition of a sign language gesture. Feedback requirements for simulated object manipulation, vehicle control and robot operations are still the subject of research and development. In each of these cases tactile and kinesthetic feedback play an

important role in normal human–system interaction. These cues are absent in most gesture-based systems. Significant progress is being made in the development of force-reflection and tactile stimulation systems that can provide this feedback through normal sensory modalities. In addition, there is evidence that substitute feedback can be provided using vibratory, auditory and electrical stimulation. The importance of feedback depends on the specific task, experience of the operator, availability of substitute visual and auditory cues, and implementation of the artificial feedback. Until additional parametric studies are performed, it is difficult to provide specific guidelines. Nevertheless, it seems clear that some form of tactile and kinesthetic feedback will be required for object manipulation and tool operation tasks in many robotic applications.

6. APPLICATIONS

6.1. Sign Language Interpretation

Sign language interpretation continues to be a significant area for gesture research and development. Scientists at the University of Toronto, for example, developed a hand gesture to speech system using a neural network. Performance of a single "speaker" with a vocabulary of 203 words was evaluated following a network training phase. With near real-time speech output, the system successfully translated the gestures > 90% of the time.

6.2. Robot Control

In one study sponsored by the National Aeronautics and Space Agency (NASA), a glove system was used to control a robot arm in a task that required retraction, movement and insertion of a block in a test panel. The researchers compared these results to another study that used a conventional 6°-of-freedom hand controller as the input device. They concluded that robot control performance with the glove compared favorably with the "standard" device and that it provided a natural and intuitive operator interface.

6.3. Virtual Reality

Several examples are provided by computer generated displays in which the operator can touch, grab and move objects by pantomiming these activities with glove-based sensors. In these applications the operator actually sees a rendering of their hand performing the object manipulations. Researchers at the NASA/Ames Research Center have used this approach in a virtual wind tunnel to explore simulations of computational fluid dynamics. Aeronautical engineers can put their hands and head into a simulated fluid flow and manipulate the patterns in real time. The GROPE project at the University of North Carolina was among the first applications to use force feedback for interacting with a computer simulation. GROPE displays the interactions between complex molecules using a force-feedback-manipulating rod. When the operator modifies the position of a molecule by moving the rod, the simulation computes intermolecular forces and reflects them back through the feedback system.

6.4. Aircraft Control

A series of experiments carried out by the US Air Force required true 3-D selection of targets from a head-down, 3-D tactical map. An electromagnetic tracker located on the back of the hand sensed hand position. The tracking volume was remote from the actual map so that hand movements were actually made in a space close to the aircraft controls. The hand tracker worked well and was, in general, a faster and more accurate way to select targets than a 3-D joystick.

7. FUTURE DEVELOPMENT

A significant disadvantage of current gesture measurement devices is that most of them limit the operator's freedom of movement (Hale 1992). This results from limited sensor range, wires attached to sensors, or the need to grasp a sensor component. Progress in component miniaturization and telemetry continues to reduce this problem. Static posture recognition has made great progress and allows reasonably high recognition rates, provided that the operator performs a standard procedure such as pointing at a target area or assuming a default posture BEFORE issuing a command. This is not yet true for dynamic gesture recognition and software techniques are still developing in this field. The main difficulty is segmentation, i.e., detecting gesture beginning and end points. Despite these challenges, gesture-based applications are beginning to take advantage of the dexterity and natural coordination of the human body, and to permit less constrained forms of operator input.

In addition to using this technology for explicit control applications, gesture may also play an expressive role in future human–machine communication (Reising *et al.* 1993; Sturman and Zeltzer 1993). Interface designers are beginning to explore the recognition and application of facial expressions and other emotive inputs. Here unconstrained gesture recognition will be required, and an effective system will undoubtedly integrate the inputs from a variety of operator monitoring systems.

REFERENCES

BAUDEL, T. and BEAUDOIN-LAFON, M., 1993, Charade: remote control of objects using free-hand gestures. *Communications of the ACM*, 36, 28–35.

BROOKS, F.P., OUH-YOUNG, M., BATTER, J.J. and KILPATRICK, J., 1990, Project GROPE: haptic displays for scientific visualization. *Computer Graphics*, 24.

BRYSON, S. and LEVIT, C., 1992, The virtual wind tunnel. *IEEE Computer Graphics and Applications*, 12, 25–34.

BUXTON, B., 1998, A directory of sources for input technologies. [http://www.dgp.utoronto.ca/people/BillBuxton/InputSources.html]

FELS, S.S. and HINTON, G.E., 1993, Glove-talk: a neural network interface between a data-glove and a speech synthesizer. *IEEE Transactions on Neural Networks*, 4, 2–8.

HALE, J.P., 1992, *Anthropometric Teleoperation: Controlling Remote Manipulators with the Dataglove* (NASA-TM-103588) (NTIS: N92-28521/2/XAB).

REISING, J.M., LIGGETT, K.K. and HARTSOCK, D.C., 1993, Exploring techniques for target designation using 3-D stereo map displays. *International Journal of Aviation Psychology*, 3, 169–87.

STURMAN, D.J. and ZELTZER, D., 1993, A design method for "whole-hand" human–computer interaction. *ACM Transactions on Information Systems*, 11, 219–38.

Hand Grip Strength

S. K. Adams

Department of Industrial Engineering, Iowa State University, Ames, Iowa, USA

1. INTRODUCTION

Hand grip strength is an important limiting factors in many occupational tasks as well as in general activity. Grip strength is affected by many factors. Some of these are tasks related including grip span requirements, grip surface geometry and friction hand orientation, force–frame requirements, movement of the hand during the grip effort, task repetition, and the use of gloves. Other factors relate to individual subject or worker strength characteristics as affected by age, gender, hand size and muscular strength acquired over time through occupational, domestic, and recreational activity. As is true for any type of strength, hand grip strength is highly individualistic. Population means and standard deviations should be developed for a given class with workers or product users whenever this is feasible to assure human/ equipment compatibility.

2. FACTORS AFFECTING HAND GRIP STRENGTH

2.1. Importance of Analyzing Hand Capabilities

Most occupational work, as well as many domestic and recreational activities, are performed with the hand, generally employing the hand–wrist–forearm system in combined and coordinated motion. Understanding the strength and motion capabilities of the hand and wrist, as well as methodologies for determining them, are important in designing tasks to minimize the frequency and severity of work-related upper extremity disorders. These cumulative injuries and functional losses, defined as cumulative trauma illnesses by the US Bureau of Labor Statistics, account for 11% of all work-related musculoskeletal disorders (illnesses) in the US. The National Safety Council (1998: 56) listed the total cost of injuries and cumulative trauma to the arm and wrist in the US at $12 billion for the two-year period covering 1995 and 1996. Hand and finger injury and trauma costs were $7.46 billion for the same period.

The versatility of the human hand in performing tasks is remarkable, including activities that involve large loads as well as very fine and delicate manipulations with the fingers. When large forces and torques are required to perform a task that includes the hands, these must be absorbed and transmitted through the hands and wrist, often resulting in high biomechanical stresses that may occur many times. Only the legs and feet transmit more force than the hand and wrist as working limbs. Since the hand is a delicate and complex biomechanical structure, designing tools, equipment and tasks to protect its many structures and systems from cumulative trauma is important.

2.2. Anthropometric Structure of the Hand and Wrist

The skeletal structure of the hand consists of three segments: the carpus or wrist bones, the metacarpus or bones of the palm, and the phalanges or bones of the digits. There are eight bones in the carpus, arranged in two rows. The upper row, progressing from the radial (thumb) to the ulnar (little finger) side includes the scaphoid, semilunar, cuneiform, and pisiform. The lower row (fingers down) includes, in the same order, the trapezium, trapezoid, os magnum, and unciform. These are as small as the fingertip segments and have complex shapes. The metacarpus comprises five long cylindrical bones, one for each digit. These connect the digits with the carpus. The phalanges comprise the fourteen bones of the fingers. There are two in the thumb and three for each finger. The finger phalanges are designated as being proximal, medial or distal according to their position. Gray's classic book on human anatomy provides complete descriptions and drawings of the anatomical and muscular structures of the hand, wrist, and forearm (Gray: 1977).

Numerous muscles are attached to the lower row of carpal bones as well as to the metacarpals and phalanges. Space does not permit an enumeration of their names and points of attachment. These muscles provide forces for precision movements in the hand. Larger forces required for power gripping are provided through tendons connecting strong flexor muscles located in the forearm to the fingers and thumb. Of special importance in work task design and analysis is the carpal tunnel bounded by the carpal ligament on the palmar side through which these tendons pass, along with the median, ulnar and radial nerves. The median nerve activates the muscles which flex the forearm, wrist and fingers and also those which pronate the forearm. The median nerve also provides control feedback from the thumb, first two fingers, and half of the feedback for the third finger. The ulnar nerve controls the fourth and fifth fingers. The radial nerve supports the ulnar (toward the little finger) side of the thumb and also the fingers. Repeated or sustained overstress through torque (high-grip tension), or rapid rotations of the wrist can interfere with the lubrication of tendons in their sheaths causing in swelling and a compression of the median nerve resulting in carpal tunnel syndrome, estimated to occur at an incidence rate of from 0.5% to 3.0% among adults (Moore 1992).

2.3. Hand and Wrist Mobility and Strength

2.3.1 Mobility

The hand, operating through motions of the wrist and forearm, can rotate in three planes. These include: extension and flexion (upward and downward motion of the hand with the palm down), adduction and abduction (rotation of the hand toward and away from the trunk with the palm down), and rotation of the wrist about the major axis of the forearm (as in turning a doorknob) caused by rotating the forearm. These rotational capabilities for adults are summarized in table 1 from Kroemer *et al.* (1994: 57).

Significant differences exist between the 5th and 95th percentiles for both genders for all three planes of rotation. Data are given in degrees of rotation. Gender-related differences (female–male) are also significant and quite large in a few instances. Significance in this case is defined at an alpha level of 0.05 and is based on 50th percentile values.

Neese and Konz (1989: 698) investigated the effects of age (16–88) on wrist mobility for a general adult population (40 males and 83 females) and concluded that overall, females have a 5% to 15% greater range of motion than males, averaging approximately 10%. Maximum flexion was 57 degrees for females and 56 degrees for males. Maximum extension was 53 degrees and 50 degrees for females and males, respectively. Adduction

Table 1. Comparison of elbow-forearm and wrist mobility data for college-aged males and females

Joint	Movement	5th Percentile		50th Percentile		95th Percentile		Difference*	
		Female	Male	Female	Male	Female	Male	Female	Male
Elbow–forearm	Flexion	135.5	122.5	148.0	138.0	160.5	150.0	+10.0	
	Supination	87.0	86.0	108.5	107.5	130.0	135.0	NS	
	Pronation	63.0	42.5	81.0	65.0	99.0	86.5	+16.0	
Wrist	Extension	56.5	47.0	72.0	62.0	87.5	76.0	+10.0	
	Flexion	53.5	50.5	71.5	67.5	89.5	85.0	+4.0	
	Adduction	16.5	14.0	26.5	22.0	36.5	30.0	+4.5	
	Abduction	19.0	22.0	28.0	30.5	37.0	40.0	–2.5	

* Listed are differences at the 50th percentile at an alpha significance level <0.05
Data adapted from Staff, K. R., *A Comparison of Range of Joint Mobility in College Females and Males*, 1983, unpublished M.S. Thesis, Department of Industrial Engineering, Texas A & M University, College Station, TX. Data are in degrees of rotation from the standard neutral orientation for each limb axis and plan.

was 24 degrees for both genders. Abduction was 55 degrees for females and 57 degrees for males. Pronation was 139 degrees and 130 degrees for males versus females. Supnation was 89 degrees for females and 78 degrees for males. Pronation capability increased significantly with age according to the formula:

Pronation angle = 89.2 + 0.94 (age in years)

2.3.2 Grip strength

Many factors affect hand and wrist strength as measured in laboratory or field studies. These include, but are not limited to:

- orientation to the work surface
- plane of rotation of the hand
- working position
- resisting force dynamics
- repetition
- obstructions (barriers to motion)
- grasp interference
- grasping method (many partial- and full-grip variations)
- grip diameter and surface area
- surface roughness or coefficient of friction
- use of gloves
- gender
- worker

Since only a few of these can be varied in a given study (typically 2 or 3), experimental results from laboratory or field studies are limited to the application and conditions tested. Generalizations can often be made despite these limitations from functional relationships and the prioritization and interactions that occur in a given set of experimental results. Grip strength resulting from a wrap around or power grip results from forces generated by all of the fingers acting together. Fransson and Winkel (1991: 881) conducted a study that included both traditional and reversed (inverted, jaws backward) grip on a pair of multiple slip joint pliers on which they measured individual finger forces together with full grip force for males and females. While the overall optimal hand grip space for maximum force at approximately 750N occurred at approximately 6 cm, optimal grip span varied for different fingers, suggesting a new approach to handle profiling for hand tools, based on individual finger strengths. Radwin and Oh (1991: 843) measured finger and palmar forces independently in determining optimum grip span length for maximum grip for a simulated pneumatic power tool. Exertion level (defined as actual exertion force divided by maximum voluntary contraction strength) was related by a U-shaped function to handle span and was minimized for a span between 5 cm and 7 cm. An extended trigger resulted in

lower finger and power grip forces. Handle size preferences for subjects were directly related to their hand lengths.

2.3.3 Wrist flexion and extension strength

The flexion and extension force capability of the wrist muscles (unassisted by the forearm) varies with the flexed or extended position of the wrist. Hallbeck (1994: 379) performed a study using 30 male and 30 female right-handed seated subjects ranging in age from 20 to 30 years in which horizontal flexion and extension hand forces generated by wrist muscles only (the forearm being held in a sagittal plane direction (forward) by a padded clamp) were measured over a range of plus and minus 90 degrees from the neutral (forward) position.

Flexion forces varied from approximately 90N (105N for males, 75N for females) at a 90 degree flexion angle (toward the body) to approximately 45N (50N for males, 40N for females) at a 75 degree extension angle (away from the body) with a slight increase of 10N at 90 degree extension angle. Flexion forces decreased over the entire range from the maximum flexion to maximum extension position. Extension forces remained fairly constant, averaging approximately 55N (65N for males, 50N for females) with an increase to 67N (75N for males, 60N for females) at an extension angle of 90 degrees. Female flexion strength averaged 76.3% of male flexion strength while female extension strength was 77.4% of male extension strength. Extension forces averaged 83.4% of those generated in flexion for the entire group of subjects.

2.3.4 Repetitive wrist flexion and extension

Snook et al. (1995: 1488) used psychophysical methods to determine maximum acceptable forces for several types and frequencies of repetitive wrist motion, employing 15 female subjects. Four adjustable work stations were built to provide simulated repetitive wrist flexion using a power grip, wrist flexion with a pinch grip, and wrist extension with a power grip. There were no significant differences in maximum acceptable torque on a day-to-day basis. A 36.3% decrement on average maximum acceptable torque did occur for a five-days per week seven hours per day exposure as compared with a two-days per week seven hours per day per week exposure. The five-days per week experiment lasted 23 days. The two-days per week experiment also lasted 23 days. Under the assumption that other repetitive motions are affected similarly with respect to maximum acceptable torques tables were developed for female wrist flexion (power grip), wrist flexion (pinch grip) and wrist extension (power grip). These are presented in simplified form in tables 2, 3 and 4, which

Table 2. Maximum acceptable forces for female wrist flexion (power grip) (N) (Newtons)

Percentage of Population	Repetition rate				
	2/min	5/min	10/min	15/min	20/min
90	14.9	14.9	13.5	12.0	10.2
75	23.2	23.2	20.9	18.6	15.8
50	32.3	32.3	29.0	26.0	22.1
25	41.5	41.5	37.2	33.5	28.4
10	49.8	49.8	44.6	40.1	34.0

Adapted with permission from Snook, S.H., Vaillancourt, D.R., Ciriello, V.M., and Webster, B.S. (1995). *Ergonomics*, 38(7), 1488–1507

Table 3. Maximum acceptable forces for female wrist flexion (pinch grip) (N) (Newtons)

Percentage of Population	Repetition rate				
	2/min	5/min	10/min	15/min	20/min
90	9.2	8.5	7.4	7.4	6.0
75	14.2	13.2	11.5	11.5	9.3
50	19.8	18.4	16.0	16.0	12.9
25	25.4	23.6	20.6	20.6	16.6
10	30.5	28.3	24.6	24.6	19.8

Adapted with permission from Snook, S.H., Vaillancourt, D.R., Ciriello, V.M., and Webster, B.S. (1995). *Ergonomics*, 38(7), 1488–1507

Table 4. Maximum acceptable forces for female wrist flexion (power grip) (N) (Newtons)

Percentage of Population	Repetition rate				
	2/min	5/min	10/min	15/min	20/min
90	8.8	8.8	7.8	6.9	5.4
75	13.6	13.6	12.1	10.9	8.5
50	8.9	18.9	16.8	15.1	11.9
25	24.2	24.2	21.5	19.3	15.2
10	29.0	29.0	25.8	23.2	18.3

Adapted with permission from Snook, S.H., Vaillancourt, D.R., Ciriello, V.M., and Webster, B.S. (1995). *Ergonomics*, 38(7), 1488–1507

show the effect of work rate (exertions per minute) as well as duration.

The most frequently reported health-related symptom in this study was muscle soreness (55.3%). The most frequently reported locus of pain was in the hand or wrist (51.8%). In addition, numbness in the palmar side of the fingers and thumb was frequently reported (69.1%) as well as stiffness on the dorsal (top with palm down) side of the fingers and thumb (30.6%). The number of symptoms increased consistently over the seven-hour, five-day work shift and were 2.3 times as numerous after the seventh hour of work than after the first hour of work. The incidence rate was 2.4 times as high after the seventh hour of work than after the first hour of work in the seven-hour, two-day work-shift.

3. METHOD OF MEASUREMENT

Hand grip strength is typically measured using an adjustable length hand dynamometer which measures force mechanically, hydraulically, or as electrical voltage change from one or more load cells containing strain gages. Conversion of signals to provide direct digital readout is often done for convenience in recording data. Graphic output is desirable when the distribution of force over time (linear impulse) is of interest as, for example, determining force or grip strength requirements or capabilities when a sustained exertion is required or when the maximum peak grip force is an important factor.

Hand grip strength measurement of a given individual performing a given task is a function of two important basic types of variables: those associated with the type of strength being measured and the method of measuring it, and those associated with the type of grasp employed along with factors that affect biomechanical effectiveness in exerting the muscular contractive effort required.

3.1. Types of Muscular Strength

Chaffin and Andersson (1991: 40) describe five types of muscular contraction: isometric, isotonic, eccentric isometric, and isoinertial. An isometric contraction is one in which the muscle length does not change during the contraction. No motion (or negligible motion) of the hand occurs and the task is therefore static (another term for this type of motion). Static hand grip or torque apply to this definition. The grip length of the dynamometer or simulated hand tool handle does not change with the applied grip force. No work or power can be calculated since there is no movement of the applied force. An isotonic (or concentric) contraction is one in which the internal force within the muscle is constant and the muscle shortens. In this case work is performed against a constant resistive force (e.g. friction). In actual work settings, the resistive force changes with limb segment position, so this type is rare.

In general, isotonic strength is approximately 80% of isometric strength, but the ratio between the two varies with the rate of decrease in muscle length. An eccentric muscle contraction is one in which the external resistance force is greater than the internal muscular force. Therefore, the muscle increases in length while maintaining tension. In doing this, the muscle merely controls the movement and the muscular effort is in resisting work done on the muscle in contrast to work performed by a muscle in overcoming the resistance force in concentric contraction. An isokinetic contraction is one in which joint motion occurs at constant velocity, representing a task in which the speed of motion is controlled. While isometric, isotonic (concentric) and eccentric motions occur commonly in all human activity, isokenitic contractions do not normally occur although they can be produced in the laboratory using an isokinetic dynamometer.

Lehman *et al.* (1993: 715) used an isometric dynamometer to measure grip strength under forearm flexion and extensions at angular velocities from zero to 175 degrees per second. Mean grip force for the hand decreased for all angular velocities above 5 degrees. Forces varied from 311 N under no motion (data averaged for five males and five females) to 178N at 5 deg/sec and to 133N for angular velocities of 25 deg/sec or higher. Male and female grip forces averaged approximately 170N and 98N respectively, a difference attributable to the combined effect of gender and hand size. Larger grip forces were provided during forearm flexion than during extension. Grip forces were also higher during radial to ulnar motion (abduction) than during ulnar to radial motion (adduction) of the forearm. Average grip

force decreased from approximately 356N under no motion to 280N for angular velocities ranging from 5 to 115 deg/sec. Although this study shows important dynamic properties of grip strength, it should be considered a pilot study because of its small size.

Isoinertial contractions are those in which the muscle contracts against a constant load. Acceleration of the body segment and any object being moved will occur if the force generated by the muscle is greater than the resistance. In this case, muscular performance evaluation requires the additional measurement of acceleration and velocity. While nearly all research studies of hand grip and hand grip torque strength have involved isometric or static muscle contractions, isotonic (concentric) and eccentric muscle strength can also be determined when this is of interest. Isokinetic strength has no known application in hand grip or grip torque strength testing. The evaluation of isoinertial contractive strength requires extensive instrumentation and has little or no practical value in ordinary applications.

3.2. Types of Grasp

Kroemer *et al.* (1994: 386) list ten types of grasp or coupling mechanisms between the hand and the object being held. These are illustrated and described in figure 1. Grip strength depends upon the type of contact made between the hand and the object being grasped. Numerous factors affecting overall grip strength are listed in section 1.3.2, including grasping method. Biomechanically, the grip span surface area of contact with the hand flexion or extension position and the ability to employ the large flexor muscles in the forearm are major factors affecting grip strength. The number of fingers employed is also, obviously, a major factor. Only a few of the ten types of grip have been investigated to determine strength capabilities.

3.3. Grip Force Dynamics

Whenever strength is measured by the exertion of muscular force, the force delivered to the handle or other contact surface in a given tool varies over the duration of the exertion. Thus, many measures of strength are possible including peak force, average force, linear impulse (force multiplied by its duration), sustained force over a given interval, or any other definition involving force and time. (Caldwell *et al.* (1974: 201) developed a general procedure for defining static strength which has become a standard used by many investigators. Minor modifications of the procedure are often done to adapt the definition to particular applications and instrumentation. The Caldwell standard static strength measurement procedures are described in terms of data recording; instructions to subjects; experimental conditions including body position, support and coupling; subject description; and data reporting including the mean, median or mode, standard deviation, skewness, maximum and minimum values, and sample size.

The proposed standard measurement of static strength is as follows:

i   Static strength should be measured during a steady exertion lasting four seconds.

ii The transient periods of one second preceding and one second following the four-second exertion should be disregarded.

iii The measure of strength for the exertion is the mean score recorded during the first three seconds of the four-second exertion.

This procedure should be followed, with possible minor

Coupling #1.   Digit touch:
One digit touches an object.

Coupling #2.   Palm touch:
Some part of the palm (or hand) touches the object.
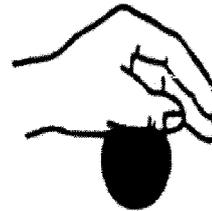
Coupling #3.   Finger palmar grip (hook grip):
One finger or several fingers hook(s) onto a ridge, or handle. This type of finger action is used where thumb counterforce is not needed.
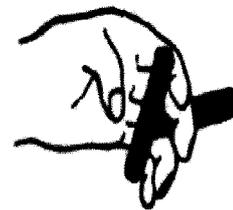
Coupling #4.   Thumb–fingertip grip (tip pinch):
The thumb tip opposes one fingertip.
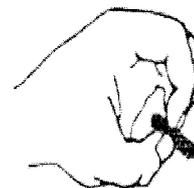
Coupling #5. <u>Thumb–finger palmar grip (pad pinch or plier grip):</u>
Thumb pad opposes the palmar pad of one finger (or the pads of several fingers) near the tips. This grip evolves easily from coupling #4.
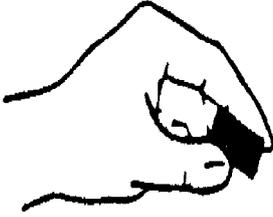
Coupling #8. <u>Thumb–fingertips enclosure (disk grip):</u>
Thumb pad and the pads of three or four fingers oppose each other near the tips (objects grasped does not touch the palm). This grip evolves easily from coupling #7.

Coupling #9. <u>Finger-palm enclosure (collect enclosure):</u>
Most, or all, of the inner surface of the hand is in contact With the object while enclosing it. This enclosure evolves easily from coupling #8.

Coupling #6. <u>Thumb–forefinger side grip (lateral grip or side pinch):</u>
Thumb opposes the (radial) side of the forefinger.

Coupling #10. <u>Power grasp:</u>
The total inner hand surfaces is grasping the (often cylindrical) handle which runs parallel to the knuckles and generally protrudes on one or both sides from the hand. This grasp evolves easily from coupling #9.

Coupling #7. <u>Thumb–two- finger grip (writing grip):</u>
Thumb and two fingers (often forefinger and middle finger) oppose each other at or near the tips.
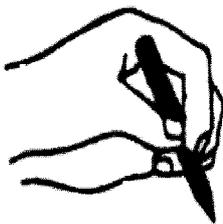
Figure 1. Couplings between hand and handle. Adapted from Kroemer, K.H.E, Coupling the hand with the handle: An improved notation of touch, grip and grasp, *Human Factors*, 28(3), 1986. Reprinted with permission. Copyright © 1986 by The Human Factors and Ergonomics Society. All rights reserved.

modifications to meet specific testing or data processing requirements, in measuring static grip or grip torque strength.

3.4 Selected grip strength data

Grip strength is a function of several basic task and tool design variables. Among the more important ones whose effects on grip strength have been studied are: handle width or span, forearm rotation (palmar-dorsal), wrist flexion or extension, elbow flexion,

and contact area of the hand (including the number of fingers used). Chaffin and Andersson (1991: 423) citing Greenberg and Chaffin (1976: 55) present typical grip span data derived from 50 men and 50 women employed in a large electronics component manufacturing plant shown in figure 2. The optimum grip span resulting in maximum grip strength (measured in Newtons) is approximately 8 cm. Fransson and Winkel (1991: 881) describe the effects of grip span, grip type, and the contributions of

individual finger strengths on total hand grip strength (discussed in section 1.3.2). Terrell and Purswell (1976: 28) found significant decreases in grip strength for a grip span of 5.25 cm as a function of forearm orientation and wrist flexion–extension. For forearm orientation, grip strength was highest for supine (palm up) orientation, and lowest for prone (palm down) orientation. Mid-position strength was slightly lower than supine strength. Grip strength for prone orientation was 87% of strength measured in the supine position. Decrements from the neutral position (hand parallel with the forearm) for wrist flexion (45 degrees), extension (50 degrees), radial flexion (20 degrees) and ulnar flexion (20 degrees) were 30%, 22%, 18% and 15%, respectively. Grip strength (force) in the neutral position varied with forearm position from 356N (supine position) to 422 N (prone position).

Marley and Wehrman (1992: 791) investigated grip strength as a function of forearm rotation and elbow posture for 10 male and 10 female subjects using 90 and zero degree elbow flexion and seven positions of forearm rotation ranging from 90 degree (complete) pronation to 90 degree (complete) supination. Maximum grip strength occurred with no forearm rotation (thumb up) with the elbow flexed at 90 degrees and at a 60-degree supination angle with no elbow flexion (zero degrees, arm straight) for males. Female subjects produced similar data except that maximum grip strength was produced at a supination angle of 30 degrees. Severe decrements in grip strength occurred for all prone or supine deviations in forearm rotation angle from the neutral (thumbs up) position as a function of the angle. Less severe decrements occurred for deviations from the maximum strength forearm rotation position with zero degree elbow flexion. Data from this study were gathered and tabulated in detail by the authors. Radwin and Oh (1991: 843) investigated the effect of

Figure 2. Maximum linear grip strengths for various handle openings (maximum grip span) in a random population sample. The grip span was measured at the illustrated grip axis – the middle of the hand at the beginning of closure of the test fixture. From Greenberg, L. and Chaffin, D.B. *Workers and Their Tools: A Guide to the Ergonomic Design of Hand Tools and Small Presses,* **1976, Pendell Publishing Co., Midland, MI, p.55 (adapted in Chaffin, D.B. and Andersson, G.B.J.** *Occupational Biomechanics, 2nd ed.***, 1991, p.423). Reproduced with kind permission of the authors.**

handle grip span and trigger length upon grip strength as it applies to power tool operators.

3.5. The Effect of Gloves on Grip Strength

Whether the use of gloves results in an increase or decrease in grip strength has not been clearly indicated. Gloves affect friction and grasp closure. These factors affect the interaction of biomechanical forces with the test handle or object being grasped. Moore *et al.* (1995: 582) found in a pilot study that latex medical examination gloves had no effect on three-jaw chuck pinch strength or power grip (full grasp) strength. Bronkema-Orr and Bishu (1996: 702) in another pilot study found that the amount of surface friction of a glove with the grasped surface affects the amount of force that a person feels is necessary to grasp an object. A high frictive force results in a lower grasping force. Their study also indicated that thicker gloves with a coefficient of friction similar to that of the bare hand tend to result in grip-related task performance similar to that occurring when using the bare hand. Special types of gloves that must be pressurized, such as space suit gloves, produce a significant decrement in grip strength. Roesch (1987: 786) compared hand grip strength with the bare hand to strength using a pressurized space suit glove with a 0.5 psi differential and one with a 4.3 psi differential. The smaller pressure differential resulted in a 35% grip strength decrement compared with bare hand. For the larger pressure differential, the strength decrement was 42%. Strength for the bare and gloved hand was positively correlated with hand size, body weight, height, and forearm circumference. Strength decrements were larger for larger-handed subjects because of the pressure differential acting over a greater area thus creating a higher resisting force against the grip.

4. RECOMMENDATIONS

i   The 4-second hand grip force exertion measure with sampling of grip force over the last 3 seconds of exertion as recommended by Caldwell *et al.* (1974: 201) should be used as a procedural guide in measuring hand grip strength.

ii   A hand grip span between 5 and 7 cm should be used when high grip forces are required. The span providing maximum grip strength for males is approximately 6 cm and for females approximately 5.5 cm.

iii The forearm should be kept in the neutral position (thumb up) or in the supine position (palm up) for tests requiring large grip forces. Grip strength is maximized in the supine position. With the forearm pronated (palm down), static grip strength is only 87% of the level attainable with the forearm supinated.

iv   The wrist should not deviate from the neutral position (parallel with the long axis of the forearm) in a radial or ulnar direction or be flexed or extended from the neutral position for maximum static grip strength.

v   A wrist flexion angle between 75 and 90 degrees is recommended when high wrist flexion forces are required.

vi   A wrist flexion angle between 45 and 60 degrees and a wrist extension angle between 45 and 75 degrees should be avoided in tasks requiring high wrist extension forces.

vii The hand should be stationary when high grip forces are required since isokinetic grip strength is lower than isometric grip strength.

viii When feasible, an adjustable grip span should be provided to accommodate a broad range of hand sizes.

ix An extended trigger handle permitting grip by several fingers (all if feasible) is desirable to minimize repeated exertions using only the forefinger.

x A high level or coefficient of friction should exist between the hand or glove and the handle or object being grasped.

xi When thick gloves are used, their coefficient of friction with the handle or object being grasped should be similar to that of the bare hand.

xii When latex examination gloves are used, they should fit the hand closely. Latex examination glove usage does not affect three-jaw chuck pinch strength, power grip strength or manual dexterity.

xiii Significant allowances should be made in hand grip strength and flexion–extension torque requirements based on task duration per day and number of repetitions of grip exertion per minute. A decrease in grip strength and flexion–extension torque up to 36% from maximum single effort strength can be expected for high repetition and prolonged task duration. The data and recommendations of Snook et al. (1995: 1488) should be used as a guide in establishing standards for tasks that include grip exertion.

REFERENCES

BRONKEMA-ORR, L. and BISHU, R.R., 1996, Effects of glove frictional characteristics and load on grasp force and grasp control. *Proceedings of the Human Factors and Ergonomics Society, 40th Annual Meeting*, pp. 702–6.

CALDWELL, L.S., CHAFFIN, D.B., DUKES-DUBOS, F.N., KROEMER, K.H.E., LAUBACH, L.L., SNOOK, S.H., and WASSERMAN, D.E., 1974, A Proposed Standard Procedure for Static Muscle Strength Testing. *American Industrial Hygiene Association Journal*, pp. 201–6.

CHAFFIN, D.B. and ANDERSSON, G.B.J., 1991, *Occupational Biomechanics* (New York: Wiley), pp. 40–1.

FRANSSON, C. and WINKEL, J, 1991, Hand strength: the influence of grip span and grip type. *Ergonomics* 34(7), 881–92.

GRAY, H., 1901, *Gray's Anatomy* (New York: Bounty Books), pp. 158–71, 389–409.

GREENBERG, L. and CHAFFIN, D.B., 1976, *Workers and Their Tools: A Guide to the Ergonomic Design of Hand Tools and Small Presses* (Ann Arbor: University of Michigan), p. 55.

HALLBECK, M.S., 1994, Flexion and extension forces generated by wrist-dedicated muscles over the range of motion. *Applied Ergonomics 25* (6), 370–85.

KROEMER, K.H.E. and GRANDJEAN, E., 1997, *Fitting the Task to the Human* (5th ed.) (London: Taylor & Francis), p. 50.

KROEMER, K.H.E., KROEMER, H.B., and KROEMER-ELBERT, K.E., 1994, *Ergonomics:How to Design for Ease and Efficiency* (Englewood Cliffs: Prentice-Hall), pp. 57, 386.

LEHMAN, K.R., ALLREAD, W.G., WRIGHT, P.L., and MARRAS, W.S., 1993, Quantification of hand grip force under dynamic conditions. *Proceedings of the Human Factors and Ergonomics Society, 37th Annual Meeting*, pp. 715–19.

MARLEY, R.J. and WEHRMAN, R.R., 1992, Grip strength as a function of forearm rotation and elbow posture. *Proceedings of the Human Factors Society, 36th Annual Meeting*, pp. 791–5.

MEESE, R.M. and KONZ, S., 1989, Ranges of motion in the human wrist. *Proceedings of the Human Factors Society, 33rd Annual Meeting*, pp. 698–702.

MOORE, B.J., SOLIPURAM, S.R. and RILEY, M.W., 1995, The effects of latex examination gloves on hand friction: a pilot study, *Proceedings of the Human Factors and Ergonomics Society, 39th Annual Meeting*, pp. 582–5.

NATIONAL SAFETY COUNCIL, 1998, *Accident Facts* (Chicago: National Safety Council).

RADWIN, R.G. and OH, S., 1991, Handle and trigger size effects on power tool operation. *Proceedings of the Human Factors Soceity, 35th Annual Meeting*, pp. 843–7.

ROESCH, J.R., 1987, Handgrip strength with bare hand and in the NASA spacesuit glove. *Proceedings of the Human Factor Society, 31st Annual Meeting*, pp. 786–90.

SNOOK, S.H., VAILLANCOURT, D.R., CIRIELLO, V.M., and WEBSTER, B.S., 1995, Psychophysical studies of repetitive wrist flexion and extension. *Ergonomics 38*(7), 1488–1507.

TERRELL, R. and PURSWELL, J.L., 1976, The influence of forearm and wrist orientation on static grip strength as a design criterion for hand tools. *Proceedings of the 6th Congress of the International Ergonomics Association*, pp. 28–32.

Hand-grip Torque Strength

S. Keith Adams

Iowa State University, USA

1. INTRODUCTION

Hand-grip torque strength is important over a broad range of manual tasks that involve the tightening and loosening of fasteners using powered or un-powered hand tools. This strength is also a necessary capability in assembly tasks that include the tightening and loosening of threaded parts or connectors by unaided hand torque. A number of studies have been conducted to model, measure and predict hand-grip torque strength under specified conditions. These studies have revealed ways in which task and hardware design variables interact with biomechanical action and effective torque strength.

2. FACTORS AFFECTING HAND-GRIP TORQUE CAPABILITY

2.1. Factors affecting Hand-grip Torque Strength

All of the factors that affect grip strength (listed in the article on 'Hand-grip Strength' by the same author) also affect hand-grip torque strength. Every task requiring torque based on hand-grip contact contains some combination of these factors. Only a few factors can be studied in a single laboratory investigation. Since the hand must rotate in a dynamic task or exert rotational torque in a static task, factors affecting friction as well as grip force are important. These include the diameter of the object grasped, surface coefficient of friction, knurling or indentations, shape, type of grip (full wrap-around, two-finger pinch, three-jaw chuck), orientation of the hand, and interference with segments of the hand involved with the grasping or torque effort. Direction of torque (clockwise or counter-clockwise) can also affect strength. The use of gloves is another factor. As in any task involving hand strength, hand size and gender have significant effects with larger hand sizes and males typically producing higher torques, provided that adequate clearance exists to permit a high grip force and sufficient contact area with the surface.

2.2. Hand-grip Torque Strength for Circular Electrical Connectors

Adams and Peterson (1988: 733) investigated hardware and task design factors affecting hand-grip torque strength in tightening or loosening circular electrical connectors used to attach and detach multiwired electrical cables on US Air Force aircraft using simulated connectors. Circular electrical connectors consist of a male and female section. After the respective electrical terminal pins that connect the circuits are aligned with and inserted into their receptacles, a threaded female connector ring is hand-tightened onto the corresponding male threaded shaft on the other half of the connector thus securing the connection. The connector ring must fit tightly to prevent inadvertent disconnection from vibration or acceleration forces. Thus, tightening and loosening require a secure grip and high hand torque. The use of hand tools as an aid when torque requirements exceed the hand torque strength of the technician is undesirable since their misalignment or misuse could damage the connector casing or cause rotation of two connector halves in opposing directions thereby breaking connecting pins. This results in very costly and time-consuming replacement and repair. It is, therefore, important that hand torque capabilities be known for this task for a variety of connector sizes and task configurations. Such knowledge permits the proper design of connectors and task configuration.

Adams and Peterson (1988: 733) investigated the effect of connector size (2.3, 3.8, 5.1 cm tightening ring diameter), grip type (full thumb and forefinger wrap-around, thumb tip and forefinger tip), grip orientation (frontal plane, sagittal plane (right side) and reverse frontal plane facing away from the subject), height (60 and 80% of maximum reach height), glove use (no glove, work glove, and chemical defense glove), and direction of torque (clockwise, counter-clockwise) for 20 male and 11 female

Table 1 Mean hand-grip torque (NM) and standard deviation for males and females as a function of independent variables

Variable	Condition	Male(N=20)				Female (N=11)			
		Loosening		Tightening		Loosening		Tightening	
		Mean	s.d.	Mean	s.d.	Mean	s.d.	Mean	s.d.
Connector size	Large(5.1cm)	−3.31	1.72	3.56	2.01	−2.31	0.91	2.56	1.23
Connector size	Medium(3.8cm)	−1.92	0.83	1.84	0.86	−1.50	0.54	1.46	0.64
connector size	Small(2.3cm)	−0.61	0.20	0.53	0.18	−0.52	0.15	0.43	0.15
Grip type	Full	−2.51	1.90	2.61	2.16	−1.77	1.10	1.91	1.37
	Fingertip	−1.38	0.81	1.34	0.89	−1.12	0.63	1.06	0.77
Orientation of	Right	−2.19	1.78	2.14	1.92	−1.60	1.07	1.60	1.28
connector	Front	−1.80	1.36	1.81	1.47	−1.35	0.83	1.37	1.03
	Back	−1.74	1.43	1.97	1.97	−1.30	0.91	1.52	1.29
Use of gloves	Work	−2.01	1.64	2.08	1.90	−1.52	0.99	1.58	1.26
	Chemical	−1.96	1.55	2.15	1.95	−1.41	0.93	1.58	1.26
	None	−1.91	1.53	1.84	1.60	−1.42	0.95	1.39	1.11
Height of connector (percentage of reach height)	85	−1.96	1.55	2.00	1.77	−1.48	0.97	1.52	1.23
	60	−1.93	1.58	1.96	1.78	−1.41	0.94	1.46	1.16

right-handed subjects in standing position. Results of the study are summarized in Table 1. Minus signs on torque values in the table denote loosening or counter-clockwise torque. Data are given in Nm. A modified Caldwell procedure (Caldwell *et al.* 1974: 201) was used as a measure of strength with data obtained from a torque measuring load cell and converted from analog signals acquired in 0.1 sec. intervals to direct digital readout and computer disc storage. Grip torque strength increased with connector size and was higher for right-side grip and reverse frontal orientation than for frontal orientation. The use of chemical defense gloves and work gloves resulted in slightly higher torque strength than the use of the bare hand. Strength gradients among varied conditions were approximately equal for tightening and loosening torques. Torque strength for males was significantly higher than for females.

Adams and Ma (1988: 642) investigated the effects of interference with hand and finger grip and torque in a study employing 18 male and 16 female standing subject, all right-handed. Torque measurement methods were identical to those used by Adams and Peterson (1988:733). In this study, the effects of four configurations of obstructing adjacent electrical connectors or flat surfaces upon hand-grip torque strength were determined for six behaviorally defined levels of interference. The effects of electrical connector size (2.3, 3.8, 5.1 cm diameter) and the use of work gloves were also measured in the study. The four conditions of interference were as follows:

i.   Interference by an identical connector located to the right and left of the grasped connector.
ii. Interference by an identical connector located to the right and left and also above and below the grasped connector.
iii. Interference from a flat rigid surface obstruction to the right of the grasped connector.
iv. Interference from a flat rigid surface obstruction below the grasped connector.

The six behaviorally defined levels of interference were as follows:

i.   Unobstructed
ii. First noticeable effect
iii. Moderate obstruction for full grip
iv. Loss of full grip
v.   Moderate obstruction for fingertip grip
vi. Limit of fingertip grip

The respective behavioral criteria for these interferences are as follows:

i.   No noticeable (by subject) interference with full grip (full wrap-around grasp)
ii. First detectable interference (by subject) with full grip; first contact with obstruction
iii. Very noticeable interference with full grip; some compromising necessary on full wrap-around grasp
iv. Subject forced to change grip to fingertip grasp using thumb and forefinger
v.   Very noticeable interference with fingertip grasp using thumb and forefinger
vi. Minimum clearance for which a fingertip grasp is possible using thumb and forefinger

In all cases, the grasped connector was located at 60% of the subject's reach height (maximum fingertip reach height above the floor). Its forward position was ahead of the right foot at a

Table 2 Hand-grip torque and separation distance as a function of level of interference

Level of Interference	Male		Female	
	Torque in NM	Separation (mm)	Torque in NM	Separation (mm)
1 Noninterference	1.53	34.5	1.01	29.8
2 First noticeable full grip	1.28	28.9	0.89	24.7
3 Significant, full grip	1.01	23.4	0.74	20.8
4 Switch to fingertip grip	0.74	18.8	0.57	16.7
5 Significant fingertip grip	0.68	14.8	0.52	13.9
6 Limit of fingertip grip	0.62	11.3	0.46	11.1

* not significant

Adapted with permission from Adams S.K and Ma, X. 1988, Hand-grip torque for circular electrical connectors: the effect of obstructions. Proceedings of the Human Factors Society, 32nd Annual Meeting pp. 642–46.

distance equal to one-half of the maximum grip distance. The maximum grip distance is defined as the displacement range between the center lines of two cylindrical handles, one held by a full grip at arm's length and the other held by a full grip against the sternum. Hand-grip torque was sampled in 0.1 sec. intervals over a period of 4 seconds and measured over the final 3 seconds according to a modified adaptation of the Caldwell procedure (Caldwell *et al.* 1974:201) as used previously by Adams and Peterson (1988:733).

Results are presented in Tables 2 to 5. Table 2 shows the effect of level of interference on hand-grip torque in Nm and the separation distance between connectors in mm necessary to achieve the indicated torque. All strength level differences shown in Tables 2 to 5 are significantly different at an alpha level of .05 using the Duncan test except those connected by a vertical line on the right side. The data in Table 2 are averaged over condition of interference (type of obstruction), connector size and the use or nonuse of work gloves. A significant difference averaging approximately 18% is shown between male and female hand torque strength, female strength averaging 72% of male strength. Strength differences are high under the higher torque conditions. A large compromise to hand torque strength occurs between levels 2 and 3, suggesting a minimum spacing of 25–30 mm. Higher male torque capability does not permit significantly greater torque under close crowding of connectors because of more limited grip contact area for larger hands. The relationship between torque strength and separation distance varies with the type of obstruction.

Condition of interference (type of obstruction) also produced significant differences in hand-torque strength and separation distance effects when averaged over the six levels of interference, three connector sizes and the use or nonuse of work gloves. The effect of obstructions on hand torque is shown in Table 3. Differences between male and female strength are evident, but do not vary as much by type of obstruction as by level of interference (Table 2). Under conditions 3 and 4, providing the highest torque strength, male and female strength are in reverse order as a function of condition. These differences are not statistically significant. Interference from a flat rigid surface located to the right side or below the grasped connector caused a much smaller decrease in torque strength than interference from

Table 3 Hand-grip torque and separation distance as a function of condition (type of obstruction)

Condition	Male Torque in Nm	Separation (mm)		Female Torque in Nm	Separation (mm)
3 Surface on right	1.08	27.4	4 Surface on bottom	0.76	25.2
4 Surface on bottom	1.07	23	3 Surface on right	0.75	20.6
2 Connect on L, R	0.93	21.5	2 Connect on L, R	0.69	18.8
1 Connect on on L, R, T, B	0.83	15.8	1 Connect on L, R, T, B	0.60	13.5

Adapted with permission from Adams, S.K. and Ma. X. 1988, Hand-grip torque for circular electrical connectors: the effect of obstructions. Proceedings of the human Factors Society, 32nd Annual Meeting, pp. 642–6.

adjacent connectors. Connectors located on the right and left side produced a smaller strength decrement than connectors located on all four sides (right, left, top, bottom). For design of equipment, a minimum clearance from any flat rigid surface of 23–25 mm is desirable. Somewhat less clearance is acceptable between connectors, but only with a significant compromise in torque strength.

Table 4 shows the very significant effect of connector size on torque strength, averaged over all other conditions. This effect is to be expected since torque is directly related to diameter and a higher surface contact area results in a higher surface friction force. Male–female strength differences are smaller for small- and medium-sized connectors than for large connectors. Higher strength provides limited benefit when grasping is difficult. An important principle for equipment design that can be derived from this table is that minimum separation distance increases as connector size decreases. Nearly 40% more surface-to-surface separation distance was required for 2.3 cm diam. than for 5.1 cm diam. connectors. Nearly 20% more separation was required for 3.8 cm diam. than for 5.1 cm diam. connectors. The data in Table 4 suggest a minimum surface separation distance between connectors of 25 mm.

The use of work gloves increased hand-grip torque strength slightly for all subjects, as shown in Table 5. Work gloves also required significantly more clearance for all conditions of interference. These data also suggest a minimum surface separation distance between connectors of 25 mm.

Table 4 Hand-grip torque and seperation distance as a function of connector size

Connector Size	Male Torque in NM	thing Seperation(mm)	Female Torque in Nm	hiinggng Seperation(mm)
Large (5.1 cm diam.)	1.55	18.3	1.06	16.5
Medium (3.8 cm diam.)	0.95	22.0	0.72	19.4
Small (2.3 cm diam.)	0.42	25.5	0.32	22.6

Adapted with permission from Adams S.K. and Ma, X., 1988, Hand-grip torque for circular electrical connectors: the effect of obstructions. Proceedings of the Human Factors Society, 32nd Annual Meeting. pp. 642–6.

Table 5. Hand-grip and separation distance with and without work gloves

Glove use	Male Torque in Nm	Separation (mm)
Work glove	1.01	24.7
No glove	0.94	19.2
	Female Torque in Nm	Separation (mm)
Work glove	0.74	22.5
No glove	0.66	16.6

Adapted with permission from Adams, S.K. and Ma., X., 1988, Hand-grip torque for circular electrical connectors: the effect of obstructions. Proceedings of the Human Factors Society, 32nd Annual Meeting. pp. 642–6.

2.3. Torque Strength for Cylindrical Handles

Replogle (1983: 412) investigated the relationship between hand torque strength and the diameter of a smooth phenotic fiber cylinder with a diameter ranging from 0.95 to 8.89 cm. and developed a mathematical model predicting torque as a function of handle diameter, gripped area, and relative grip force (proportional to gripped circumferential angle). Torque strength increased with the square of handle diameter up to approximately 2.5 cm, the "grip-span diameter" at which the fingers just reach the base at the palm in encircling the handle. Beyond 2.5 cm, torque strength increased at a decreasing rate to a maximum diameter of approximately 5 cm. The hand torque at this point was approximately 1.5 times the torque exerted on the 2.5 cm handle.

Female torque strength was 40% of male torque strength. Grip span and maximum torque diameters did not differ significantly between males and females, suggesting that the same size handles can be used for both genders for tools such as screwdrivers for which hand torque constitutes the primary effort required. A deviation in handle diameter at plus or minus 1 cm from the 5 cm optimum (for maximum torque) does not reduce torque strength significantly. For diameters below 2.5 cm or above 7.5 cm, hand torque strengths significantly reduced. The proposed mathematical model was verified accurately in a laboratory study using 20 subjects.

2.4. Torque Strength for Hand Tools

Pheasant and O'Neill (1975: 205) performed research on the effect of the diameter, shape, and amount of hand contact area of screwdriver handles on grip torque strength. They also compared screwdriver handles with smooth and rough steel cylinders for this effect. Handle shape was found to be unimportant. Handle diameter and the effect of contact area and friction were important factors. Hand torque strength normalized against each subject's mean torque strength over seven polished steel handles ranging in diameter from 1 to 7 cm in 1 cm increments increased linearly up to a diameter of 5 cm, at which point the rate of increase dropped although linearity was maintained up to 7 cm. A total of 24 subjects served in this study. Normalized thrust force and shear force (equal to torque divided by handle diameter) were both maximum for a handle diameter of 4 cm. The contact area between the handle and the hand was maximized for a handle diameter of 5 cm. A knurled cylindrical handle 5 cm in diameter is recommended as a design for maximum torque strength.

Mital and Sanghavi (1986: 283) compared maximum voluntary torque in the use of five different hand tools for 55 subjects (30 male, 25 female), for 2 postures (seated and standing), 3 work heights, 3 reach distances and 6 angles relative to the mid-sagittal plane (plus or minus 15, 45 and 75 degrees). Moment arm effect among the five tools (short and long screwdriver, socket wrench spanner (crescent) wrench and adjustable vice-grip) produced widely differing torque capabilities. The mean torques exerted for the latter three tools were 27.8, 23.6 and 20.9 Nm, respectively. Mean torque values for the long and short screwdrivers were only 2.7 and 3.1 Nm. For screwdrivers, torque strength was a function of hand-grip contact area while for wrenches, torque strength depended upon the length of the wrench. The socket wrench enabled higher torque than the spanner wrench (27.8 versus 23.6 Nm).

Higher torques were produced in a sitting posture for screwdrivers and in a standing posture for wrenches. Peak torque exertion for female subjects averaged 66% of peak male torque. Work height and angle effect were of no practical significance. Higher torque strength occurred in the standing position (23.5–33.3 Nm) for the socket wrench, spanner wrench and vice-grip) than in the sitting position (17.6–27.4 Nm). Reach distance also affected torque strength with maximum torque (18.5 Nm) exerted at a distance to the work surface of 33 cm and minimum torque (13.1 Nm) at a distance of 71 cm. The decrease in torque strength resulting from reach distance was the same for sitting and standing postures (2.2 versus 2.4 Nm) for an increase of 25 cm in reach distance.

Torque strength in this study was measured using the basic recommended procedure (Caldwell 1974: 201) with a 3-sec. build up in torque exertion followed by a 1-sec. hold at the peak exertion. Torque produced during this final second was used as the measure of torque strength. Among several isometric strength measures taken, shoulder strength appeared to limit torque exertion capability of subjects in this study.

2.5. Torque Strength for Small Knobs

Swain, Shelton, and Rigby (1970: 201) conducted a study in which 96 civilian and 24 military personnel in two age groups, one 29 years of age or under, the other over 29 years of age, exerted grip torque with and without gloves in a standing position on small diamond-knurled knobs 3/8, 1/2, and 3/4 inches (0.95, 1.27, 1.90 cm) in diameter located in front of and on the side of the preferred hand. Maximum single effort exertions were used as a measure of torque strength. No significant differences in torque strength were attributable to age or civilian versus military subject group. Mean and standard deviation data were developed for each of the knob sizes with and without the use of gloves. For the 3/4 in. knob, mean frontal bare-handed torque strength was 0.52 Nm with a standard deviation of 0.127 Nm with the 5th percentile equal to 0.31 Nm. The side mean torque strength for the 3/4 in. knob was 0.55 Nm with a standard deviation of 0.124 Nm and a 5th percentile value of 0.34 Nm. The use of gloves reduced the mean torque values slightly to 0.45 and 0.47 respectively for front and side torques but did not affect the 5th percentile value.

For the 1/2 in. knob, mean torques for the front and side were 0.65 and 0.66 Nm with standard deviations of 0.150 and 0.159 Nm respectively. The 5th percentile values were both equal

to 0.40 Nm. The use of gloves reduced torque strength to 0.57 and 0.59 Nm with standard deviations of 0.104 and 0.116 for front and side torque exertions respectively. The 5th percentile value remained at 0.40 Nm for both orientations.

For the 3/4 in knob, mean torques for the front and side were 1.11 and 1.14 Nm with standard deviations of 0.260 and 0.266 Nm respectively. The 5th percentile torque values were 0.68 and 0.70 Nm respectively. The use of gloves reduced mean torque strength for front and side exertions to 0.92 and 0.93 Nm with standard deviations of 0.170 and 0.178 Nm respectively. The 5th percentile value for both front and side torques was reduced to 0.64 Nm.

These data show increased torque capability as a function of knob diameter, a slight reduction in unobstructed knob torque strength resulting from the use of gloves, and significant variability among subjects not attributable to age. Torque strength also did not differ significantly between civilian and military personnel. The variation in strength among subjects suggests the use of the 5th percentile values as a general design upper limit value for torque requirements when specifying the sizes of knurled knobs used in the study.

3. RECOMMENDATIONS

i. A 4-second hand-grip torque exertion measure with sampling of torque over the last 3 seconds of exertion as recommended by Caldwell *et al*. (1974: 201) should be used as a procedural guide in measuring hand-grip torque strength.

ii. Hand tools requiring high torque exertions should have a handle diameter of approximately 5 cm.

iii. Hand tools requiring high torque should not have handles with diameters below 2.5 cm or above 7.5 cm.

iv. Since grip span diameter (the diameter at which the fingers just reach the back of the palm in encircling the handle) and maximum torque diameter do not differ significantly between males and females; the same sized handle for hand-grip torque can be used for both genders.

v. Hand contact area with the handle should be maximized in the design of screwdrivers.

vi. For small knurled knobs requiring hand torque exertion, strength should be considered to be related to knob diameter.

vii. In specifying hand torque strength requirements for tasks, it should be assumed that women can exert approximately 66% of male hand torque strength.

viii. Right-handed operators can exert stronger hand-grip torque on the right side than in a forward or rearward work surface orientation because of more effective use of flexors and extensors affecting wrist and forearm motion and because of additional torque provided by the upper arm. Based on this fact, components requiring high torque should be accessible from a right-handed operator's right side when feasible.

ix. The height of an electrical connector or other object to which hand torque is to be applied does not affect grip torque strength provided that its height does not affect the type of grasp used. Therefore, adjustments in hand-grip torque strength requirements based on task height are, in most cases, unnecessary.

x. Gloves often increase hand-grip torque strength provided there is sufficient clearance to accommodate the glove. Work gloves or chemical defense gloves are effective in increasing

hand-grip torque strength. Whether gloves are used or not used, good surface friction with the grasped object should be provided.

xi. For the case of circular electrical connectors or other grouped objects such as knobs, torque strength depends upon adequate clearance, which varies inversely with connector or object size. Therefore, wider spacing is needed for smaller connectors or objects to provide the conditions for adequate grip torque. When gloves are used, sufficient clearance for their use should be provided.

xii. The minimum clearance between connectors, knobs or other objects, or between these and a nearby rigid surface should be at least 25 mm when a full grip is required. For connectors or objects 3.8 cm or less in diameter, a surface separation distance of 30–35 mm is recommended for a full grip.

xiii. For a fingertip grip torque, large connectors or objects (5.1 cm diameter or larger) should be spaced at least 15 mm apart, surface to surface. Medium connectors or objects (approximately 3.8 cm diameter) should be separated, surface to surface, by a distance of at least 20 mm. Small connectors or objects (2.3 cm diameter or less) should be separated, surface to surface, by at least 25 mm.

xiv. When interference in applying hand-grip torque to tighten or loosen circular electrical connectors occurs at or beyond Level 2 (first detectable interference with a full grip), required torque should be no more than 1.35–1.70 Nm for a large connector (5.1 cm diameter), 0.90—1.10 Nm for a medium connector (3.8 cm), and 0.35—0.45 Nm for a small connector (2.3 cm).

REFERENCES

ADAMS, S.K. and MA, X., 1988, Hand grip torque for circular electrical connectors: the effect of obstructions. *Proceedings of the Human Factors Society, 32nd Annual Meeting*, pp. 642–6.

ADAMS, S.K. and PETERSON, P.J., 1988, Maximum voluntary hand grip torque for circular electrical connectors. *Human Factors,* 30 (6), pp. 738–45.

CALDWELL, L.S., CHAFFIN, D.B., DUKES-DUBOS, F.N., KROEMER, K.H.E., LAUBACH, L.L., SNOOK, S.H., and WASSERMAN, D.E., 1974, A proposed standard procedure for static muscle strength testing. *American Industrial Hygiene Association Journal*, pp. 201–6.

MITAL, A. and SANGHAVI, N., 1986, Comparison of maximum volitional torque exertion capabilities of males and females using common hand tools. *Human Factors* 28 (3), pp. 283–94.

PHEASANT, S. and O'NEILL, D., 1975, Performance on gripping and turning—a study in hand/handle effectiveness. *Applied Ergonomics 64*, pp. 205–8.

REPLOGLE, J.O., 1983, Hand torque strength with cylindrical handles. *Proceedings of the Human Factors Society 27th Annual Meeting*, pp. 412–16.

SWAIN, A.D., SHELTON, G.C. and RIGBY, L.V., 1970,

Maximum torque for small knobs operated with and without gloves. *Ergonomics* 13 (2), pp. 201–8.

Handgrip Characteristics and Strength

S. N. Imrhan

Industrial and Manufacturing Systems, Engineering Department, University of Texas, Arlington, TX 76019, USA

1. INTRODUCTION

Of all human physical activities none is more pervasive and perhaps more important than grasping and manipulating objects with the hand. While grasping was necessary in early human development for fashioning crude hunting tools and for accomplishing other basic necessities of life, today it is a precursor to more sophisticated tasks, especially in the workplace. The ways objects are grasped and manipulated largely determe the quality of task performance and the associated level of strain on the body.

For convenience early classifications of grasping (Napier 1956) have identified two main types(prehensile and nonprehensile. In prehensile grasping an object is seized partly or wholly within the compass of the hand, whereas in nonprehensile grasping the object is pressed with the flat surface of the hand or fingers. Two major types of prehensile grasp have been identified:

- *The power grasp*: the object is held in the palm of the hand with the force of the thumb opposed by the combined forces of the other fingers.
- *The precision grasp*: the object is held between the flexor aspects of one or more fingers and the opposing thumb of the same hand. This article deals with the strength of the power grasp (power grip).

2. TYPES OF HANDGRIP

Common usage of the term "handgrip" is synonomous with "power grasp" and will be used in this context here. Other synonyms are hand grasp, hand squeeze, and grip. Power grasping may be accomplished in different ways, depending on the size and shape of the object being grasped. Here is a practical classification of the main types of power grasp:

- *Spherical grasp*: the hand grasps a spherical object in the palm of the hand. The object is large enough to allow contact with the palm and fingers, as in holding a baseball or spherical doorknob.
- *Disk grasp*: the fingers encircle the perimeter of a disk-shaped object such that the thumb lies along the grasp surface (perimeter), as in trying to open or tighten a lid on a jar with great force.
- *Cylindrical grasp*: the fingers are wrapped around the surface of a cylinder or similarly shaped object such that the long axis of the object runs along the width of the hand, as in grasping a hammer handle.
- *Hook grasp*: the handle of the object is hooked by the index, middle, ring, and little fingers at the region of the knuckles, as in carrying a briefcase, but the thumb does not apply counter force to these fingers.

3. PROBLEMS WITH STRONG GRIPPING

Power gripping with great forces is common during task performance; great grip forces, especially when applied repetitively, may produce significant muscular strains in the hand and should be avoided as a task requirement. One solution has been to use powered tools, but the advent and proliferation of powered hand tools have not eliminated overexertion problems of the hand. On the contrary, in some cases, new problems have been created, such as the need to maintain more powerful grips on tools to stabilize them from impact motions and vibrations, or to maintain them at specific positions in space against gravitational pull (their weight). Moreover, developing countries routinely use manually operated hand tools and their populations experience many ergonomic problems. The use of manual pliers and screwdrivers is ubiquitous in developing countries. One needs, therefore, to understand the nature of hand gripping in order to prevent or eliminate manual muscular overexertions and enhance task performance. This kind of understanding may also lead to improvements in hand tool design, work design and evaluation, and personnel task assignment.

One of the most important aspects of gripping is strength — (the maximal force that can be achieved by squeezing voluntarily on an object with the hand as hard as possible without feeling pain or significant discomfort (MVC exertion)). This article examines the nature of MVC grip strength. Applications of grip strength will not be discussed except to help readers comprehend a particular aspect of strength.

4. GRIP STRENGTH MEASUREMENT

The standardized method of measuring grip strength involves gripping and squeezing the handle of a handgrip dynamometer. The typical handle comprises two straight or slightly curved parallel bars (ostensibly shaped to fit the contours of the enclosed hand) which are squeezed toward each other but which can accommodate only one hand. Some dynamometers allow a slight movement of one of the bars whereas others show no perceptible movement. For practical purposes, these grip forces are considered as static forces. Grip strength is taken as the peak or middle 3 s mean grip force during a maximal voluntary effort (MVC force) that lasts 4–6 s. Usually the subject is allowed to adjust the separation distance of the two bars over a narrow range to obtain their most comfortable grip. Most adults select a width of about 5.0 cm. The result is a power grip with some resemblance to a hook grip. This method of measuring handgrip strength has its origins in the clinical field in which the data is used mainly for evaluating the effects of injuries and the progress of rehabilitation of injured hands rather than for work task applications.

5. FACTORS INFLUENCING HANDGRIP STRENGTH

Grip strength, like almost any other kind of strength, is not a constant. It varies according to the conditions in which it is exerted. Many of the variables that define these conditions have been investigated experimentally. The results show that among the most influential are grip size, posture and joint angle, type of grasp, use of gloves, anthropometry, and sex (or gender). These factors are discussed below. Unless otherwise stated, the

discussions pertain to one-handed (the preferred hand) static strength measured on a standard handgrip dynamometer as described above.

5.1. Grip Size

Grip size refers to how much the fingers are opened in gripping an object. It may be measured as the diameter or perimeter of the handle being gripped or, for the bars on a handgrip dynamometer, as the distance between the surfaces of the bars in contact with the hand. Grip strength seems to be highest at grip widths in the range 5.0–6.5 cm for males and 4.5–6.0 cm for females. It is lower outside these ranges. This constitutes an increasing/decreasing trend. Research results also suggest that similar optimal ranges exist for other types of one-handed grip: 5.0–6.0 cm for pistol grip; and 6.4–8.9 cm for dynamic contractions on angulated handles (such as those found on sheet-metal shears) and 5.0–6.5 cm for static contractions (Fransson and Winkel 1991). However, not all types of grip have shown an increasing/decreasing trend. One study on gripping cylinders has shown a steady decrease as the diameter of the cylinder increases (within the 5.0–7.0 cm range). One study on two-handed gripping (Imrhan 1999) also found that, for flat surfaces, grip strength followed a linear trend with grip width (distance between the surfaces).

These ranges seem to be the grip sizes where the fingers are about midway between full flexion and full extension and the finger flexor tendons are at about their greatest mechanical advantages; that is, where the torques generated by the finger flexor muscles about finger joints are maximized. These torques translate into grip forces at the hand(handle contact.

5.2. Body Posture and Joint Angle

There seems to be no credible reason to expect a difference in grip strength among sitting, standing, and lying down postures, but one study has reported that grip strength is 6% stronger when standing compared to sitting and 4% stronger when sitting compared to lying supine.

Given the anatomical nature of gripping, it is reasonable to expect that grip strength is influenced by biomechanical variables such as wrist angle and wrist orientation. Experimental data has shown this is indeed the case: grip strength is significantly weaker when the wrist is sharply deviated away from its natural position or when the forearm is sharply rotated (Terrell and Purswell 1976). Thus hand tools and workplaces that impose these constrained wrist and forearm positions can cause people to exert unduly great muscular efforts to produce required grip forces for various tasks. This loss of grip strength or increase in muscular effort is most likely due to (i) excessive shortening or lengthening of the flexor muscles and consequent loss of some of their tension-producing capabilities, (ii) pressure of the tendons on the carpal wall within the bent wrist resulting in pain in the wrist, and (iii) loss of grip effectiveness.

Experimental data has confirmed that handgrip is strongest when the wrist is in its natural position (hand dorsiflexed 15–35°), followed by ulnar and radial deviation, dorsiflexion (hyperextension), and palmar flexion, in decreasing order of magnitude. The losses of strength in these positions compared to the natural position are about 15% for ulnar deviation, 18% for radial deviation, 22% for dorsiflexion, and 30–40% for palmar flexion. This ordering assumes the wrist in any of its unnatural positions

is bent as sharply as possible while still maintaining a reasonably good power grip on a dynamometer handle. These strength–angle relationships do not necessarily hold under all gripping conditions. One experimental study has shown that, for a hook grip, the 30° ulnarly deviated position was stronger than other deviated positions. Gripping with the forearm in pronation is about 12% weaker than in supination or mid-orientation (natural orientation). Supination and mid-orientation seem to be about equally strong.

5.3. Type of Grasp

From a biomechanical viewpoint, grip types that vary sharply should yield significantly different grip strengths, but there is little empirical data for comparing different types of grasp under similar conditions. One study found that a forward grasp on an angulated metal shear handle was stronger than a reverse grasp for grip spans of 4.1–6.7 cm, but at greater spans there was no difference in strength.

5.4. Use of Gloves

Gloves are used for protecting the hands from mechanical injuries (cuts and abrasions), corrosive chemicals, dirt and grease, heat and cold, and vibration. In many cases the gloved hand must grip and squeeze hand tools and workpieces but the extra effort that may be required to stretch and crumple thick, stiff, or close-fitting gloves while flexing the fingers may lead to some loss of grip strength. Research has shown that grip strength is reduced by about 7–30% when wearing regular working gloves, and by about 50% when wearing pressurized gloves used for extravehicular space activities.

5.5. Anthropometry

Body dimensions, especially body weight, height, hand length, forearm circumference, and various other arm dimensions, are significantly correlated with handgrip strength. But either alone or in combination, they cannot predict handgrip strength (standardized or task related) accurately enough for ergonomic designs.

5.6. Sex or Gender

On average, female grip strength is about 50–60% of male grip strength. This is smaller than the percentages for the strengths of most other body parts. Because of the smaller size of the female hand, an optimal hand size for grip strength should be smaller in females than in males. However, experimental results have not established this male/female difference unequivocally; some results indicate that both sexes have the same optimal range whereas others indicate that the female optimum is smaller. Perhaps this apparent contradiction is due to the characteristics of the subject samples in these studies; it is likely that the male/female hand size difference was not great enough to result in different optimal ranges.

5.7. Gripping and Torquing

The nature and strength of gripping is of paramount importance in manual torquing. In many tasks, gripping is not an isolated activity but is often combined with other kinds of muscular exertion. In some cases, such as turning a screwdriver, the necessary output is the torque generated at the tool (object

interface. This torque, however, is a function of the grip force (in addition to other variables) on the screwdriver handle (Pheasant and ONeill 1975). Other influences being equal, torque increases proportionally with grip force. In other tasks, such as pushing/pulling a cart, grip strength may also make an important contribution to push/pull strength, but its importance here is not as great as in torquing. Designs that prevent gripping overexertion or that enhance the magnitude of grip force for a given level of muscular exertion can help to prevent strains and enhance task performance; for example, cylindrical handles of 4.5–6.5 cm diameter. Because other tool/object or task factors (e.g., surface texture, shape, orientation) interact with handle size, it is not always true that the optimal range of handle diameter (or perimeter) for torquing is also optimal for gripping and squeezing.

6. RELEVANCE OF HANDGRIP STRENGTH MEASUREMENTS

Handgrip strength measurement in the field of ergonomics has adopted the clinical method, often without justification. For many manual tasks in industry, for activities of daily living, and for leisure activities the gripping methods may differ significantly from the clinical test grip. At work people often grip and squeeze with two hands; the shape and contours of objects vary considerably and the types of grip employed vary accordingly; the varying sizes of object may lead to widely varying grip widths. Because of these differences in grip strength between real tasks and standardized tests, one must be careful in applying published grip strength data to the design of tasks and equipment for the workplace as well as for living and leisure activities. One study has shown that when gripping and squeezing flat surfaces, grip strength declined linearly as the width increased from 3.8 cm to 16.5 cm; and two-handed grip strength in females was 1.6 times as strong as one-handed grip strength.

7. GRIP STRENGTH MAGNITUDES

Reported one-handed grip strengths have mostly fallen in the range 450–600 N for adult males and 250–300 N for females. Two-handed strengths have been reported as about 1.6 times as strong in females and 1.3 times in males.

8. CONCLUSIONS

Handgrip strength information is important for designing hand tools and other work equipment, and for designing and evaluating work tasks. Among the most important information are strength magnitude, how other variables influence strength, and the nature of gripping objects. Users should exercise care in applying handgrip data because most data is based on a method of measurement which differs significantly from the way people grip real task objects. Even for similar tests, published data varies widely.

REFERENCES

NAPIER, J.R., 1956, Prehensile movement of the human hand. *Journal of Bone and Joint Surgery*, 38B, 902–13.

TERRELL, R. and PURSWELL, J.L., 1976, The influence of forearm and wrist orientation on static grip strength as a design criterion for hand tools. Proceedings of the 20th Annual Meeting of the Human Factors Society, Santa Monica CA, pp. 28–32.

FRANSSON, C. and WINKEL, J., 1991, Hand strength: the influence of grip span and grip type. *Ergonomics*, 34, 881–92.

IMRHAN, S.N., 1999, Gripping, pinching and squeezing. In Kumar (ed.) *Biomechanics in Ergonomics* (London: Taylor & Francis).

PHEASANT, S.T. and O'NEILL, D., 1975, Performance in gripping and turning: a study in hand/handle effectiveness. *Applied Ergonomics*, 6, 205–8.

Human Muscle

K. H. E. Kroemer

Industrial And Systems Engineering, Virginia Tech, Blacksburg, VA, and Eri, Inc., A Consulting Company, PO Box 3019, Radford, VA 24143-3019, USA

In the human, there are three types of muscle. *Cardiac muscle* brings about contractions of the heart. *Smooth muscle* works on body organs, such as by constricting blood vessels. *Skeletal muscle* is under voluntary control of the somatic nervous system (Kroemer *et al.* 1997, 2000) and serves two purposes. One is to maintain the body's postural balance; the other is to cause local motion of body segments by pulling on the bones to which the muscles are attached, thereby creating torques or moments around the body joints which serve as pivots.

There are several hundred skeletal muscles in the human body, identified by their Latin names. The Greek words for muscle, *mys* or *myo*, are often used as prefixes.

Muscles actively perform their functions by quickly and temporarily developing internal lengthwise tension. This is often, but not always, accompanied by shortening, therefore the action is generally called contracting. Muscle my also be lengthened passively beyond its resting length by a force external to it; in response, muscle tissue develops internal tension both by elastic resistance and by attempted active contraction.

The main components of muscle besides water (~75% by weight) are proteins (20%). Collagen, an abundant protein in the body, constitutes the insoluble fiber of the binding and supportive substance in muscle tissue. The proteins actin and myosin form rod-shaped polymerized molecules that attach end-to-end into thin strands.

These actin and myosin filaments are the contracting microstructure of the muscle. As seen in a crosscut through the muscle, several actin molecules surround each myosin molecule. Lengthwise, the thin actin strands are wound around the thicker myosin in form of a spiral (double helix). At rest, separate actins and myosins are separate. For a contraction the actin strands temporarily connect with the myosin via so-called cross-bridges that serve as temporary ratching attachments by which the actins pull themselves along the myosin rod.

Between 10 and 500 muscle filaments are packed tightly into a bundle known as fibril. Along the length of each fibril, lighter and darker stripes appear in a repeating pattern mostly indicating the density of actin and myosin molecules. This banding or striping has led to the name "striated" muscle. Fibrils, in turn, are packed into bundles wrapped by connective tissue called fascia. Bundles of fibrils packed together constitute the muscle fiber. This is a cylindrical unit, 10^5–10^6 Angstrom in diameter and 1–50 mm long. Each is a single large cell with several hundred cell nuclei located at regular intervals near its surface. They contain the mitochondria, the "energy factories of the muscle" where chemically stored energy is converted for use by the muscle to generate physical energy.

Between 10 and 150 muscle fibers constitute the so-called primary bundle (fasciculus). Several primary bundles are packed into a secondary bundle, again wrapped in connective tissue. Secondary bundles, in turn, are grouped into tertiary bundles, and so forth until the structure called a muscle is formed. It usually has several hundred thousand fibers in its middle and tapers off towards its ends, traditionally called origin (proximal) and insertion (distal).

The bundles and the total muscle are wrapped in perimysium, tough and dense collagenous membranes which, at origin and insertion, develop into tough, bendable, slightly elastic tendons. On its other end, the tendon usually attaches to the outer membrane of a bone. Many (but not all) tendons run through slippery tissue tubes, called tendon sheaths, which keep the tendon in position and guide it around a bend while it slides within the sheath, with synovial fluid serving as lubricant for smooth gliding.

For more information see any book on physiology or kinesiology, or Kroemer *et al.* (1997, 2000).

REFERENCES

KROEMER, K.H.E., KROEMER, H.J. and KROEMER-ELBERT, K.E., 1997, *Engineering Physiology* (3rd ed.) (New York: van Nostrand Reinhold–Wiley).

KROEMER, K.H.E., KROEMER, H.J. and KROEMER-ELBERT, K.E., 2000, *Ergonomics: How to Design for Ease and Efficiency* (2nd ed.) (Englewood Cliffs: Prentice-Hall).

Information Processing

C. M. Carswell and E. C. Stephens

Department of Psychology, University of Kentucky, Kansas, KS, USA

1. INTRODUCTION

Information processing (IP) is one approach used to describe how minds (human and otherwise) operate. Its core assumption is that basic mental operations select and transform incoming information in goal-consistent ways. The IP approach evolved in the mid-twentieth century in part due to the work of early military human factors researchers. Today, IP concepts and models are widely used by human factors and ergonomics specialists, especially those interested in enhancing human comprehension, decision-making and problem-solving.

2. INFORMATION PROCESSING MODELS

2.1. From Black Box to White Box Models

Psychology in the first half of the twentieth century, especially in North America, was dominated by models of human behavior that included no descriptions of hypothetical mental activities. Researchers working in this behaviorist tradition objected to such concepts either because they were unobservable, and thus unscientific, or because they were considered unnecessary adequately to predict human performance. Such models described how human responses were adapted to environmental stimuli and often included elaborate and careful measurements of stimulus–response relationships. Organismic events occurring between the stimulus and response were said to take place within a 'black box' that was assumed to be impenetrable by the methods of science. Stimulus–response models were, thus, sometimes called "black box models."

By the 1950s, some researchers were exchanging the black box in their models for a series of intervening mental operations. Their models were often presented as flow charts, usually a series of labeled boxes connected by arrows that indicated the proposed flow of information. These "white box models" became the standard format for presenting proposals about information processing. One example from the human factors literature is shown in Figure 1.

2.2. Central IP Concepts

IP models similar to the one shown in Figure 1 have been used to describe human performance in a variety of complex tasks such as reading, maintaining situation awareness, wayfinding and selecting from computer menu options. Regardless of the specific application, most IP models share a number of features (Lachman *et al.* 1979).

2.2.1. Computer metaphor

IP diagrams and terminology owe much to the emergence of the digital computer. The traditional IP approach assumes that there are strong similarities between human thought and the way
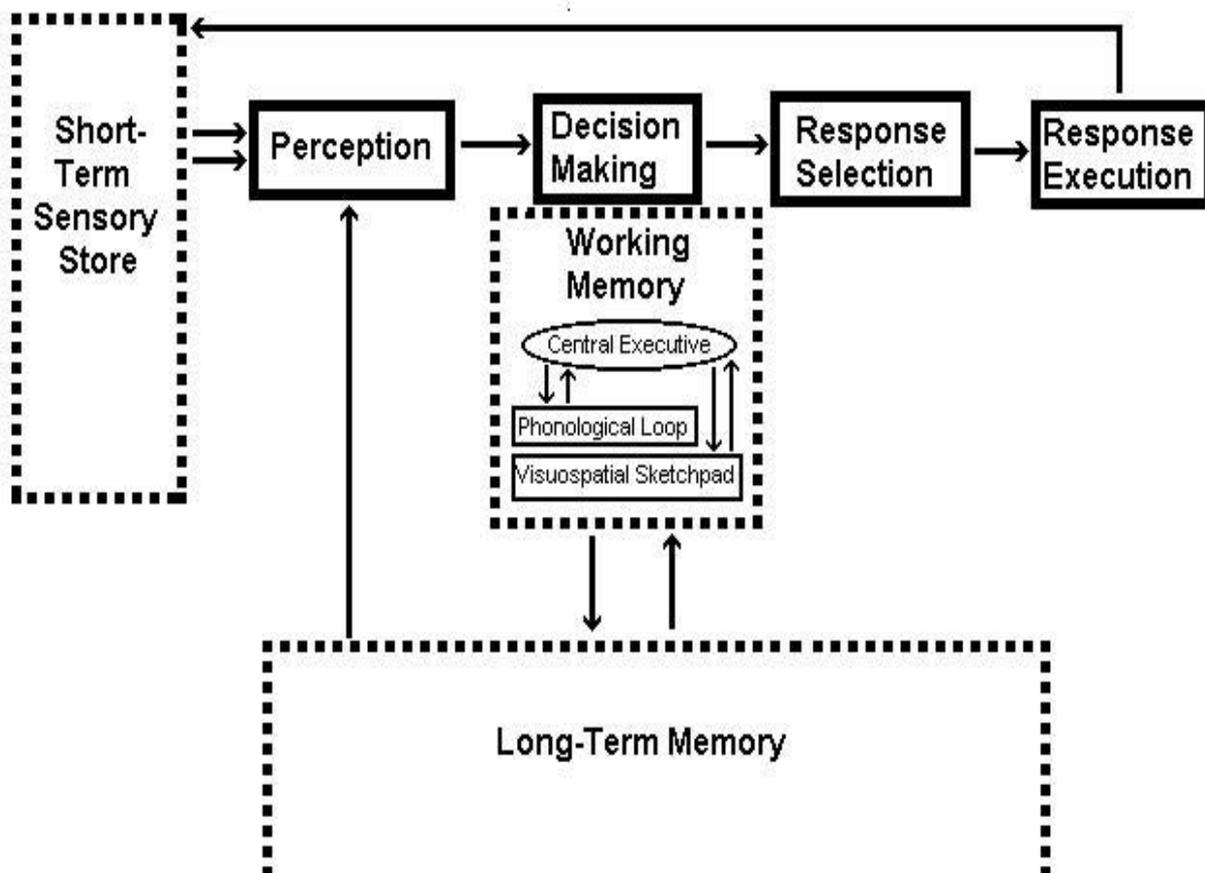
Figure 1. An example of a typical information processing model, combined from the work of Atkinson and Shiffrin (1968), Baddeley (1997), and Wickens (1992)

computers operate. For example, computers, like humans, have different types of memory stores that can be accessed or searched in different ways. Computers also have specialized input and output devices that might be comparable with the specialized sensory and response modalities of humans.

2.2.2. Processing stages

Most traditional IP models propose a sequence of stages through which information passes. At each stage, information from previous stages is manipulated, transformed or combined with information from memory before being passed to the next stage. The overall cognitive performance of a person on a task is determined both by the number and type of stages required as well as by the efficiency with which the operations at each stage are accomplished. Thus, IP models are often used in human factors as a framework for isolating the "location" of cognitive performance problems. For example, researchers may ask whether a person is having difficulty recognizing a critical stimulus or holding this stimulus information in memory long enough to act on it.

2.2.3. Mental representations

All IP models must deal to some degree with the way information is coded and organized. For example, information relevant to a specific task could be stored as a list, a network, a hierarchy, a series of auditory or visual images, or some combination of the above. Research on display design and decision-support systems, in particular, often attempts to develop displays that are compatible with the probable mental representations of the users. The more similar the physical (display) representation is to the user's mental representation of information, the fewer the number of transformations that are necessary to convert incoming information into a usable form.

2.2.4. Capacity limits

An appealing aspect of IP models for human factors applications is their emphasis on the processing limits of the operator. Just as computers are limited in both their storage capacities and processing speeds, humans too must be limited in how much information they can handle in a given interval. Designers need to be aware of these limitations to ensure that their product, device or procedure does not exceed the operator's capacity. While early IP models tended to focus on identifying the location of processing bottlenecks (e.g. Broadbent 1958), more recent research has focused on training and design techniques that reduce the impact of such limits (for a review, see Wickens 1992).

3. INFORMATION STORAGE IN IP MODELS

Most IP models make use of one or more information storage units. Figure 1 uses a tripartite storage system similar to the model originally proposed by Atkinson and Shiffrin (1968) and incorporated in many other IP models since then. The three units vary in terms of how much information each can hold, the durability of the information once stored, and the effort required to maintain stored information.

3.1. Sensory Registers

The initial memory storage unit for incoming stimuli is referred to as the sensory register. In this register, a large amount of sensory information is stored, but this information can only be held for < 1 s. Presumably a sensory register exists for each of the senses, although the iconic (visual) and echoic (auditory) registers have been the most widely studied. Information in the sensory register decays rapidly, and if attention is not devoted to this information it is lost and replaced by new information.

Functionally the sensory registers are important because they may serve as a pre-attentive system that alerts attention-demanding perceptual processes to select stimuli for further consideration. The sensory register can thus be seen as a brief 'second chance' window on a rapidly changing world. As a result, designers sometimes find that periodic breaks in a stream of information can enhance subsequent processing by allowing the operator to make use of information currently in the sensory register.

3.2. Working Memory

When sensory information is attended it is then processed by working memory (WM). Unlike the sensory register that receives information from the environment, WM receives information internally. The traditional view of WM claims it is a memory buffer of limited capacity (five to nine units of unrelated information) and brief duration (~20 s if not rehearsed). The classic example of using WM is looking up a phone number and keeping that number active in memory until dialing. If the number is not rehearsed, the ability to recall it decays rapidly.

WM has also been described as having two storage subsystems called the visuospatial sketchpad and phonological loop. These two storage units serve a central executive system that coordinates information (Baddeley 1997). The visuospatial sketchpad is dedicated to maintaining information in WM for use as mental images and for spatial reasoning. The phonological loop is dedicated to maintaining verbal information in an acoustical form by actively rehearsing this information vocally or subvocally. Both subsystems have a limited capacity and duration.

The limits of WM are generally assumed to be among the most severe bottlenecks in information processing. As such, WM poses a particular challenge to human factors design. Many researchers suggest that no more than four unrelated units of information (whether visuospatial or phonetic) should be presented to users for short-term recall. For example, voice mail menus should be kept to four or less options. If more options are needed, designers should take advantage of familiarity, identifying options by meaningful letters rather than unassociated numbers. Thus, saying "For savings account information press S" is preferable to saying "For savings account information press 1." Further, designers should avoid requiring users to hold information in WM while waiting for the opportunity to use it. Rather than forcing voice mail users to wait until the end of the menu list to respond, they should be allowed to enter their choice as soon as they know it.

3.3. Long-term Memory

Long-term memory (LTM) is where information is stored for later use. LTM has little or no capacity or duration constraints. However, failures of LTM do occur and are generically called 'forgetting.' There are two main causes of forgetting. First, forgetting may occur because the desired information was never successfully encoded. Some information is encoded with little

effort, whereas other information requires active and effortful memorization. Strategies such as placing the to-be-remembered information in the context of familiar information can facilitate encoding. However, even when information has been successfully encoded in LTM, it can still be forgotten. The second reason for forgetting involves retrieval problems, including interference from irrelevant information and lack of associations with easily retrieved memory cues.

Human factors professionals should design products, instructions, and procedures to minimize both encoding and retrieval failure. To minimize encoding failures, products should be designed to accommodate effective encoding strategies. For example, instructions should make use of pictures of familiar objects where relevant because pictures lead to more elaborate encodings than words alone. To minimize retrieval failure, research indicates that recall is enhanced when information available at the time of retrieval is similar to that available at the time of encoding. Thus, an accident investigation team is likely to get more accurate reports from victims returned to the scene of the accident.

4. COMMON IP STAGES AND OPERATIONS

4.1. Perception

In this stage, information from the sensory register is parsed to form separate objects that can then be identified. The arrow extending from LTM to the perception box in Figure 1 indicates that information stored in permanent memory is matched to the visual representation being processed by the perceptual subsystem. If the long-term information matches this representation, it is recognized as a familiar feature, pattern or object.

Perception is, as discussed above, guided by incoming sensory information. However, higher-order or 'top-down' processes including expectancies, beliefs and motives also guide perception. Human factors specialists must acknowledge the influence of top-down processes by attempting to create designs that accommodate user expectancies. Owing to top-down processing, responses to unexpected events are known to be slower and may even go unnoticed by operators. Therefore, unexpected events should be made salient to the operator. If a restaurant often changes its menu, then menu designs should incorporate factors that facilitate accurate sensory processing (e.g. larger fonts, greater contrast). These factors will be somewhat less important for a more predictable (rarely changed) menu.

4.2. Decision-making

After a stimulus has been perceived, it must be integrated with other task-relevant information, and possibly transformed, to allow the operator to select a response. This stage of processing is often called decision-making or thinking and is conceptualized as the manipulation of information in WM. Sometimes called the 'central executive' or 'central processing unit,' this stage of processing is the one most often associated with consciousness.

The central executive must coordinate the integration of many types of information. A mechanic troubleshooting an aircraft malfunction must integrate numerous pieces of evidence. A consumer deciding to buy a computer must weigh the pros and cons of several makes and models. An interior designer may mentally move a sofa in a room before deciding to expend the

effort physically to relocate it. Because WM, and especially the central executive, is constrained by severe capacity limits, the task of the human factors engineer is to make some of the required integrations easier or even unnecessary. For example, when considering two or more brands of over-the-counter medication, the consumer may need to combine information about dosage, number of doses per container, and overall price per container for each brand. However, some of this information could be combined for the consumer by presenting price per dose or number of daily doses per container.

4.3. Response Selection and Execution

As shown in Figure 1, once a decision has been made, the operator must decide how to respond based on this consolidated information. This requires an understanding of the response alternatives which, in turn, draws upon current perception or stored information. For example, when approaching a yellow traffic light, experienced drivers understand that their response options include stopping (pressing the brake pedal) or speeding up (pressing the gas pedal). They will make their response selection based on a decision, integrating their perception of how long the light has been yellow and the probable presence of law enforcement officers. Assuming that they decide to stop, they must still execute the response. This requires knowledge of the location and force required to use the braking system.

The distinction between response selection and execution is an important one for designers. Users frequently complain that they know what they want to do but may not understand how (or do not have the skill) to do it. Novice computer users may know the option they want to select, but they may not yet have the motor skills necessary to execute a 'double click.' Designers should pay attention to both making obvious the alternatives available to the user as well as reducing the difficulties in carrying out the response (e.g. reducing the motor precision required, adhering to population stereotypes).

5. TRENDS, CHALLENGES AND CONTRIBUTIONS

While the computer metaphor, with its associated flow charts, still influences most IP models found in human factors, cognitive psychology, and related disciplines, an alternative metaphor based on neuronal units and brain physiology has been achieving increasing acceptance. These parallel distributed processing (PDP), connectionist or neural network models propose that information processing takes place as the result of the activation and inhibition of networks of neuron-like units (e.g. McClelland and Rumelhart 1986). These models provide a more detailed explanation of how storage units and processing stages might actually operate.

In addition to the development of IP theories based on alternative metaphors, the IP approach as a whole has met with several challenges. One of the predominant limitations of IP models seems to be their focus on serial, discrete processing of stimuli. Various authors have indicated the importance of the continuous fluctuation of stimuli in a natural, as opposed to a laboratory, environment. Both ecological and control theoretic models have emphasized responses to the continuous flow of stimuli, a flow brought about both by changes in the environment as well as the continuous movement of the operator (see Wickens and Carswell 1997 for comparisons with the traditional IP approach).

Despite the search for an appropriate IP metaphor and challenges from other theoretical approaches, the traditional IP approach still provides a useful overall framework for considering the limitations of thought processes within the operator. Contributions of the IP approach involve its use to perform detailed task analyses that may highlight particular performance problems and suggest design solutions. A more detailed account of IP models and their applications to human factors and ergonomics is found in Wickens (1992) and Wickens and Carswell (1997).

REFERENCES

ATKINSON, R.C. and SHIFFRIN, R.M., 1968, Human memory: a proposed system and its control processes. In Spence and Spence (eds) *Advances in the Psychology of Learning and Motivation: Research and Theory*, vol. 2 (New York: Academic Press).

BADDELEY, A., 1997, *Human Memory: Theory and Practice* (Hove: Psychology Press).

BROADBENT, D., 1958, *Perception and Communications* (New York: Pergamon).

LACHMAN, R., LACHMAN, J.L. and BUTTERFIELD, E.C., 1979, *Cognitive Psychology and Information Processing: An Introduction* (Hillsdale: Erlbaum).

McCLELLAND, J.L., RUMMELHART, D.E. and PDP Research Group, 1985, *Parallel Distributed Processing: Explorations in the Microstructure of Cognition* (Cambridge, MA: MIT Press).

WICKENS, C.D., 1992, *Engineering Psychology and Human Performance* (2nd ed.) (New York: HarperCollins).

WICKENS, C.D. and CARSWELL, C.M., 1997, Information processing. In G. Salvendy (ed.) *Handbook of Human Factors and Ergonomics* (2nd ed.) (New York: Wiley), 89–129.

Lifting Strategies

D.H.K. Chow

Jockey Club Rehabilitation Engineering Centre, Hong Kong Polytechnic University Hung Hom, Kowloon, Hong Kong

1. INTRODUCTION

Low back pain (LBP) is one of the most serious and persistent problems in occupational medicine. Epidemiological studies have shown that about 60 to 80% of the population experiences LBP at some time during their active life. The associated costs of LBP in term of loss of productivity, replacement training, and individual suffering are unacceptably high. LBP can develop suddenly or gradually, with or without an initiating event. Although the majority of LBP has no particular etiology, the incidence of reported pain, injury, loss of work, and disability is higher among those who are employed in occupations with a high level of exposure to physical loading. It is therefore essential to understand the ergonomic factors as well as the basic biomechanics related to LBP.

In human beings the spine is held in a vertical position during walking and working hours. In the erect posture, the loads on the spine increase progressively down to the lumbosacral region. The loads produce mechanical, physiological, and psychological responses such as tissue deformation, altered metabolism, and altered circulatory patterns. Depending on the duration and intensity of the loads and recovery time, the responses may cause discomfort and impair performance. If the loads are large enough or last long enough, acute tissue damage will occur. Although damaged tissue stimulates a healing response, healing may not occur if the loads continue to be applied. As the lumbar spine is the main load-bearing area and the most common site of pain, studies of spine loads have focused on this region.

2. BASIC BIOMECHANICS

Manual material handling is a common situation wherein loads applied to the vertebral column may be so high as to damage the spine. The loads on the lumbar spine are produced primarily by body weight, muscle activity, and externally applied loads (figure 1). Although the biomechanics of manual material handling and lifting are based on Newton's laws of motion, a number of biological factors complicate their application. For example, the distance between a muscle's line of action and the joint about which it acts changes with joint position; at the extremes of joint movement the active muscle group is necessarily shortened while its stretched antagonist exerts maximal passive tension.

Further difficulties arise with human motion, since the body is seldom quite still, especially at the start or completion of a lifting or handling task. Moreover, coordination of a handling task involves the constraint of many components. Consider the number of muscles involved in a task, and the need to coordinate their torque production with changing joint dynamics. The neuromuscular system experiences particular difficulty in heavy manual lifting as the trajectory of the load is required to be precisely controlled and any stress to the joints has to be minimized. The difficulty in meeting these requirements is reflected in the high incidence of injury to the back when lifting moderate to heavy loads.
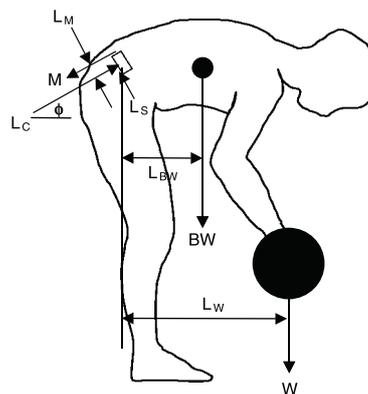
$$L_C = \frac{BW \times L_{BW} + W \times L_W}{L_M} + (BW + W)\sin\phi$$
$$L_S = (BW + W)\cos\phi$$

where

L_C	=	Compressive load on the lumbosacral junction
L_S	=	Shear load on the lumbosacral junction
BW	=	Body weight above the lumbosacral junction
L_{BW}	=	Distance between the centre of gravity of BW and the lumbosacral junction
W	=	Weight of the lifted object
L_W	=	Distance between the centre of gravity of W and the lumbosacral junction
M	=	Tension in the back muscle
L_M	=	Distance between the muscle's line of action and the lumbosacral junction
ϕ	=	Inclination of the lumbosacral junction

Figure 1. **"Manual Material Handling" as the third index phrase.**

Despite this complexity, the loads on the lumbar spine in performing most lifting and handling tasks can be roughly estimated using a simplified free body technique for coplanar forces. Several factors influence the loads on the lumbar spine during lifting activities: (1) the position of the object relative to the center of motion in the spine; (2) the size, shape, weight, and density of the object; and (3) the degree of flexion or rotation of the spine. In general, holding the object close to the body instead of away from it reduces the loads on the lumbar spine.

3. GENERAL PRINCIPLES OF LIFTING

It is generally recommended that lifting should be done with the knees bent and the back relatively straight to reduce the loads on the spine, but this technique must be followed correctly. Lifting with bent knees and a straight back allows smallish objects to be held closer to the trunk and to the center of motion in the spine, closer than when lifting with the trunk flexed forward and the knees straight. However, the loads are not reduced if the object is held out in front of the knees.

Loading of the spinal tissues is not harmful unless the endurance of the tissues is exceeded. This excess may be caused by sudden overload or fatigue in repeated loading. Gradual increase in loading has a training effect; and muscles, tendons, and also vertebrae adapt to endure heavier loads. Weight lifters have less degenerative changes in the lumbar spine than men in heavy physical work. However, LBP can be caused by prolonged overstretching of the ligaments and soft tissues of the lumbar spine. Lifting excessive weights can cause overstretching and damage to the supportive spinal ligaments. This type of injury cannot easily be avoided as it can occur unexpectedly and without any

warning signs. Usually it is the ligaments that become damaged first and give rise to pain.

Overstretching of the surrounding ligaments and retaining walls for the intervertebral disk may affect the disk's ability to act as a shock absorber between the vertebrae. When the ligaments surrounding the disk are injured to such an extent that the disk bulges outwards or even bursts through the outer ligament, the disk may lose its ability to absorb any shocks and it can lead to severe pain. Worst of all is when the disk bulges far enough backwards to compress the sciatic nerve, causing severe pain, numbness, or even weakness.

Although many studies have indicated that LBP is related to heavy physical work, there are also studies in which this relationship has not been detected. It has been argued that for those studies with positive correlation, work was classified as heavy on the basis of general impression but without detailed analysis of the loads on the lumbar spine. In some studies the data on the heaviness of work was based on subjective perception. This may be influenced by the occurrence of LBP.

4. SYMMETRY VERSUS ASYMMETRY

For lifting tasks, conditions of asymmetry have been studied in vitro, by simulation, and in vivo; it has been clearly demonstrated that asymmetry may be harmful to the lumbar spine. Asymmetric posture combining forward bending and twisting of the trunk increases the loads on the lumbar spine. In vitro studies predicted that twisting and bending are likely to cause disk prolapse and possibly damage to the capsular ligaments. Dynamometric studies revealed that the worker's trunk extensor capability decreases with asymmetric trunk angle. These results demonstrate the importance of symmetry in relation to handling tasks; heavy lifting with asymmetric posture should be avoided.

5. HABITUAL EFFECTS

Although people have been trained to lift with a straight back (leg lift) instead of a rounded back (back lift), there are still many back injuries caused by improper manual lifting. One explanation is that leg lift is infrequently used because it requires a greater expenditure of energy than back lift, due to greater body weight being displaced vertically. People tend to be reluctant to lift in a way that they feel unnatural and always try to minimize their energy expenditure. As fatigue is induced earlier in using leg lift, this may encourage people to adopt a poor posture in lifting, especially after a period of continuous work. Moreover, people with degenerative changes in their knees often find difficulty with the high loads that leg lift places on the knees. Although there is no clear indication that the degenerative changes in their knees are actually caused by leg lift, it is obvious that a single lifting method does not suit each individual in all situations.

On the other hand, it is important to consider the habit that people adopt during lifting. In general, movement may be produced by two basic systems of control: feedback (closed-loop control) and feedforward (open-loop control). Movements requiring feedback are usually complex and discrete, requiring a degree of accuracy, such as a visual-motor tracking task. Feedback control is required during early skill acquisition and probably accounts for the slow movement execution during this stage. Once the task is well learned, it can be performed accurately without sensory feedback. The control of movement fluctuates between feedback and feedforward modes.

Rapid movements and well-learned movements are performed with feedforward control. Although the sensory information is available during a rapid movement, this information cannot be used because the movement proceeds faster than the nervous system can process and use it. Thus, it appears that well-learned, rapid, and automatic tasks may not require feedback for their execution. If people adopt the movement strategies that they use for lifting light objects, they may wrongly apply these adopted strategies in lifting objects that are neither light nor heavy. The risk that an improper strategy is used to lift a relatively heavy object will be increased.

It is unlikely that the loads induced to the lumbar spine for lifting objects of medium weight using improper movement strategies are the single cause of LBP injury. However, if improper movement strategies are frequently used, the cumulative effects may lead to injury or pain. In a clinical or laboratory setting, it is difficult to document people's lifting habits as they will be assessed under a conscious situation. It is advised that people should keep conscious in performing manual handling tasks. The movements should be performed slowly enough to receive internal sensory feedback that will control and modify the postural requirement. Repeated and rapid lifting should be avoided.

6. SUDDEN UNEXPECTED EFFECTS

Epidemiological data suggests that sudden unexpected maximal efforts are particularly harmful to the back. However, there have been very few investigations into the effects of unexpected loads on the back. Unexpected loads can take many forms; they can be a sudden load application or a release condition. It has been proposed that a sudden release of load, perhaps when a load slips, can generate an unexpected acceleration and a very large muscle force which may damage the tissues of the spine. The back muscles tend to overrespond to regain equilibrium under unexpected events.

Slips or falls in the workplace are relatively common and account for about 50% of accidental back injuries. In an experimental investigation, a subject held a box in a standing position and a designated weight was suddenly dropped into the box. The mean muscle force for the unexpected cases exceeded the mean for the expected cases by nearly 2.5 times, and the peak muscle forces were on average 70% greater in the unexpected cases. The lumbar muscle activities were not fast enough to cope with the sudden change of loading and this produced an increased swaying of the trunk, which in turn induced a load on the lumbar spine.

7. RECOMMENDATIONS

- Minimize loads on the lumbar spine to reduce the risk of low back pain.
- A simplified free body technique can be used to make rough estimates of the loads on the lumbar spine in symmetric lifting.
- In general, holding the object close to the body instead of away from it can reduce the loads on the lumbar spine.
- Leg lift is not always better than back lift.

- Avoid lifting with an asymmetric posture.
- Be conscious in performing manual material handling.
- Avoid repeated and rapid lifting.
- The loads on the lumbar spine can be dramatically increased under sudden unexpected conditions.

REFERENCES

RIIHIMÄKI, H., 1991, Low-back pain, its origin and risk indicators. *Scandinavian Journal of Work and Environmental Health*, 17, 81–90.

SCHMIDT, P., 1989, *Motor Control and Learning* (2nd ed.) (Champaign IL: Human Kinetics Publishers).

SCHOLZ, J.P., 1993, The effect of load scaling on the coordination of manual squat lifting. *Human Movement Science*, 12, 427–59.

TROUP, J.D.G. and Edwards, F.C., 1985, *Manual Handling and Lifting: An Information and Literature with Special Reference to the Back* (London: HMSO).

WHITE, A.A. and PANJABI, M.M., 1990, *Clinical Biomechanics of the Spine* (2nd ed.) (Philadelphia: Lippincott).

Maximum Holding Times of Static Standing Postures

M.C. Miedema and M. Douwes

TNO Work and Employment, PO Box 718, 2130 AS Hoofddorp, The Netherlands

1. INTRODUCTION

Many work situations require standing postures which have to be maintained for a long period of time (e.g. machine operation, assembly work, VDU work). Depending on the posture and the duration of holding the posture, there is a risk of acute discomfort and long-term health effects (musculoskeletal disorders). Static work load can be diminished by improving the working posture (by optimization of the workplace and the equipment), by reducing the holding time of postures and by supplying sufficient and properly distributed rest pauses.

A posture can be maintained for a limited period of time. The Maximum Holding Time (MHT) is the maximum time that a posture, with or without external force exertion, can be maintained continuously until maximum discomfort, from a rested state. Acute discomfort can be considered as an independent evaluation criteria for static postures (Miedema 1992; Dul *et al.* 1994; Miedema *et al.* 1996). Several studies have shown a linear relationship (at group level) between the duration of a continuous static task, and the level of discomfort as measured with the Borg rating scale (Manenica 1986; Bishu and Wei 1992). To avoid uncomfortable durations of static postures discomfort do not exceed score 2 or 3 on a 10 point rating scale (Hagerup and Time 1992; Douwes *et al.* 1993)

EVALUATION METHOD FOR HAND POSITIONS

Due to the increment of discomfort in time, holding time (% MHT) can be taken as a measure for making recommendations concerning the maximum duration of static postures. In seven studies information was gathered about the MHT of 19 different standing postures which were maintained without rest pauses and without external load (Corlett and Manenica 1980; Hagberg 1981; Boussenna *et al.* 1982; Milner 1985; Taksic 1986; Manenica 1986; Meijst *et al.* 1995). All postures were defined by two parameters, i.e. the horizontal distance (% shoulder height) and vertical distance (% arm reach) of the position of the hands with respect to the feet in upright standing posture. Shoulder Height (SH) is defined as the distance from acromion to the floor in the upright position. The Arm Reach (AR) is defined as the maximum distance from the knuckles to the wall when standing upright with the back against the wall and the shoulder in 90 degrees anteflexion. The 19 postures differ in the combination of 25, 50, 75, 100, 125 or 150% SH and 25, 50, 75 or 100% AR and are shown in figure 1. In all studies the participants were asked to maintain the posture as long as they could. In almost all studies the subjects had to perform a task while holding the posture. These tasks implied television games, spot-tracking or tapping tasks. During maintenance of the posture, location and amount of perceived discomfort was registered. The experiments ended when maximum discomfort was reached (score 10 on a 10 point rating scale; Borg 1990).
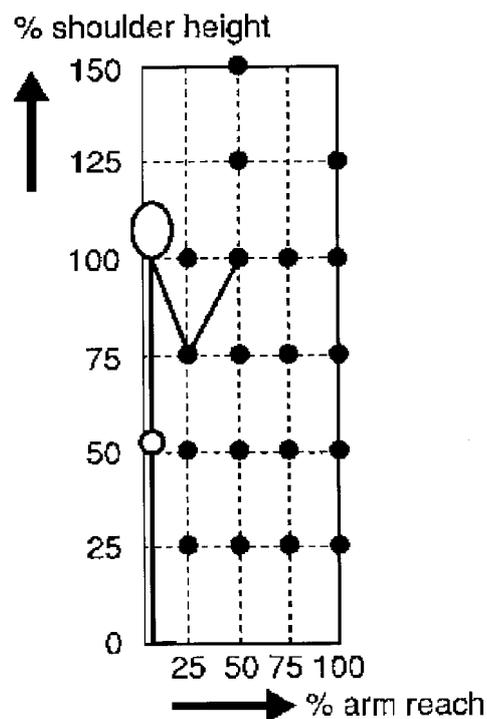
Figure 1. The 19 postures found in literature are indicated with ●. Posture is defined by the horizontal and vertical hand position. The vertical hand position is expressed as a percentage of the shoulder height and the horizontal hand position as a percentage of the arm reach in the upright standing posture.

There is much variation in MHTs of similar postures within and between studies. In spite of this variation a ranking of the 19 postures was made, based on the mean MHT from all available data. These ranked postures are shown in figure 2. The posture 75% SH/50% AR has the highest MHT (35.7 min.) and the posture 25% SH/100% AR has the lowest MHT (2.7 min.).

The ranked postures can be arbitrarily classified into three groups: "comfortable", "moderate", and "uncomfortable" postures, with relatively large, medium, and small MHTs, respectively. Uncomfortable postures are defined as postures with a MHT smaller than 5 minutes, which implies that maintaining an uncomfortable postures will lead to a relatively quickly increased feeling of discomfort. All postures with an extremely low or high working height (25% and 150% SH) appear to be uncomfortable postures. According to the classification the postures with a combination of a moderate working height (50%, 75%, 100%, and 125% SH) and small working distance (25% and 50% AR) are comfortable postures (MHT longer than 10 minutes). Postures with a moderate working height (50%, 75%, 100% and 125% SH) and a large working distance (75% and 100% AR) appear to be moderate postures (MHT $ 5 and #10 minutes).

Holding a comfortable, moderate, or uncomfortable posture for the maximum period of time causes extremely strong (maximal) discomfort in (a part of) the body. By limiting the actual holding time of a posture, discomfort can be limited, even in uncomfortable postures. As mentioned, we propose that for a group of individuals the maximum acceptable holding time is 20% of the MHT, corresponding to a discomfort score of 2 of the
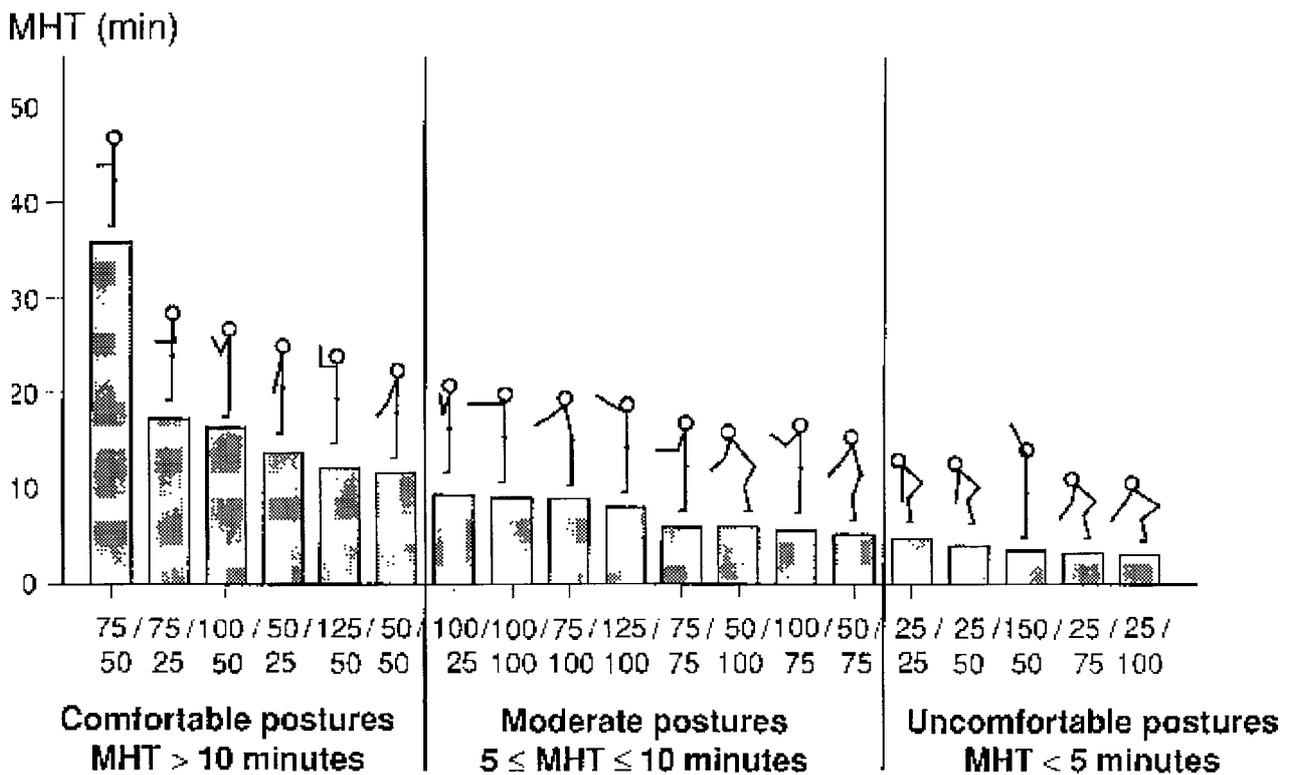
Figure 2. Ranking of the postures on the basis of the mean values of Maximum Holding Time (MHT) of available data. The recommended holding time is 20% of MHT.

Borg scale (weak discomfort). To calculate the maximum acceptable holding time of a posture the MHT has to be divided by 5. To make the recommendations safe for all postures in the three classes, the maximum acceptable holding time valid for each class of hand position corresponds with the lowest maximum acceptable holding time of that class (most uncomfortable posture of that class). Thus, comfortable postures with a MHT of more than 10 minutes are recommended to be maintained 2 minutes maximally. Following the same procedure for a moderate posture the maximum acceptable holding time is 1 minute and an uncomfortable posture is not acceptable. In figure 3 the possible hand positions are divided into three areas corresponding to these recommendations (area 1: 2 minutes; area 2: 1 minute; area 3: 0 minutes). It is estimated that for a mean discomfort of 2, 95% of the population will have less than "strong discomfort" (score 5 on the 10 points scale).

The body part(s) in which discomfort is felt, depends on the posture. All healthy subjects who adopt the same posture (independent of the study) perceive discomfort in approximately the same body part(s). Postures with hand positions at or below 50% SH are terminated by discomfort in the lower back and legs. In postures with hand positions at or above 100% SH the shoulders and arms are the critical body parts. Also, it appears that a larger work distance results in a higher discomfort in shoulders and arms.

APPLICATIONS AND LIMITATIONS OF THE METHOD

The evaluation method for hand positions relates to:
- standing postures
- postures without external force
- postures that are symmetric in the sagittal plane
- postures that are maintained without rest pauses (static work)
- healthy, young adults.

The classification can be a guidance in practical situations for occupational health officers, designers, labour inspectors and ergonomists to match the working time to the working posture.

The method has been developed for pure static postures without body motions. In most working postures minor changes in posture and loading may occur. This may result in partial recovery due to changing the critical muscle group or variations of the muscle effort. Another point of attention is that in many work situations body parts are supported by a table, an arm rest, or a machine. This support unloads the muscles and the joint. It can be assumed that the MHT of a "dynamic" or supported posture is longer than the MHT of a static posture. For these kind of working situations the recommendations are expected to be relatively safe.

REFERENCES

BISHU, R.R. and WEI, W., 1992. Evaluation of handle position: Comparison of psychophysical force/endurance and biomechanical criteria. *Journal of Applied Physiology* 9: 221–34.

BOUSSENNA, M., CORLETT, E.N., and PHEASANT, S.T., 1982, The relation between discomfort and postural loading at the joints. *Ergonomics* 25: 315–22.

CORLETT, E.N. and BISHOP, R.P., 1976, A technique for assessing postural discomfort. *Ergonomics 19*: 175–82.

CORLETT, E.N. and MANENICA, I., 1980, The effects and measurement of working postures. *Applied Ergonomics* 11: 7–16.
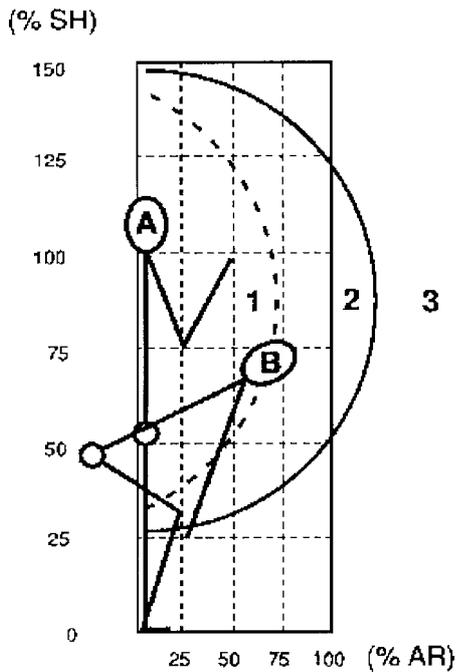
Figure 3. Recommendations concerning the maximum holding times of static postures: — = demarcation line between uncomfortable and moderate postures; – – – = demarcation line between moderate and comfortable postures; 1 = hand positions of comfortable postures with MHT longer than 10 minutes and recommended maximum holding time of 2 minutes; 2 = hand positions of moderate postures with MHT between 5 and 10 minutes and recommended maximum holding time of 1 minute; 3 = hand positions of uncomfortable postures with MHT less than 5 minutes which are advised against. Person A adopts a comfortable posture and person B adopts an uncomfortable posture.

DOUWES, M. and DUL, J., 1993, Studies on the validity of a work-rest model for static postures. In *Proceedings of the international association world congress 1993. Ergonomics of materials handling*, Warsaw, Poland.

DUL, J., DOUWES, M., and SMITT, P.,1994, Ergonomic guidelines for the prevention of discomfort of static postures can be based on endurance data. *Ergonomics* 37: 807–15.

HAGBERG, M., 1981, Electromyographic signs of shoulder muscular fatigue in two elevated arm positions. *Am. J. of Phys. Med.*, 60: 111–21.

HAGERUP, A.B. and TIME, K., 1992, Felt load on shoulder in the handling of 3 milking units with one- and two-handgrips in various heights. In Proceedings of International Scientific Conference on Prevention of work-related musculoskeletal disorders (PREMUS), pp. 105–7, Sweden.

MANENICA, I., 1986, The Ergonomics of Working Postures: A Technique for Postural Load Assessment, (London: Taylor & Francis) pp. 270–7.

MEIJST, W., DUL, J., and HASLEGRAVE, C., 1995, Maximum holding times of static standing postures. Thesis of extended essay. TNO Institute of Preventive Health Care, Leiden, The Netherlands.

MIEDEMA, M.C., 1992, Static working postures. Part 1: Classification of static working postures on the basis of maximum holding time. Part 2: Secondary analysis on endurance data. Thesis of extended essay. TNO Institute of Preventive Health Care, Leiden, The Netherlands.

MIEDEMA, M.C., DOUWES, M., and DUL, J., 1996, Recommended holding times for prevention of discomfort of static standing postures. *Industrial Ergonomics*, 19: 9–18.

MILNER, N., 1985, Modelling fatigue and recovery in static postural exercise, University of Nottingham, Nottingham. PhD thesis.

TAKSIC, V., 1986, The ergonomics of working postures: comparison of some indices of postural load assessment. (London: Taylor & Francis) pp. 278–82.

Models of Attention

R. W. Proctor and L. E. Read

Department of Psychological Sciences, Purdue University, West Lafayette, IN 47907-1364, USA

1. INTRODUCTION

Most research on attention is concerned with either the selectivity of processing or the capacity to perform multiple tasks concurrently. Experimental research on attention has been conducted since the earliest days of psychology. However, contemporary research is usually dated from the early 1950s. Interest in attention arose in part because of the increasing performance demands of technology growing out of World War II. Also, around this same time, the shift from behavioral to information processing approaches began to occur.

The information-processing approach provided a descriptive language for conceptualizing the nature of attention. Using the computer as a metaphor, early models of attention viewed processing as serial. Hence, theoretical issues were concerned with the point at which attention intervened in the processing sequence. Research on attention continues to be guided by metaphors, including attention as an allocatable resource supply, attention as a spotlight, and attention as "glue." Although the dominant metaphors tend to capture some aspects of attention and to direct research toward solving particular issues, there is still debate as to what metaphor, if any, best captures the nature of attention. Though many issues regarding the nature of attention remain unresolved, much has been learned about attention that is relevant to applied system design through a better understanding of the contexts and conditions conducive to efficient task performance.

2. THE FIRST FORMAL MODELS DEFINING ATTENTION

2.1. Bottleneck Models

Bottleneck models propose a structural limitation on processing at some point in the human–information-processing system. Donald Broadbent developed the first major bottleneck model, filter theory, in the 1950s. Filter theory was developed to explain results of studies examining performance limitations in listening tasks implying that only one message at a time could be processed. The primary assumption of this model is that the nervous system acts as a single communication channel. Therefore, for the sequential nature of processing to be most efficient, Broadbent argued that the system must perform a selection operation at an early stage. He postulated that information is first received in a pre-attentive temporary store of unlimited capacity. This information is then selectively filtered to allow only one input to enter the limited-capacity channel. Selection is based on physical features such as the intensity, pitch, and spatial location of sounds. Although Broadbent's original filter theory is no longer accepted as providing an adequate characterization of attention, it is a useful model for the design of human–machine systems because it is simple and captures many of the aspects of attention that are of consequence for human performance.

The primary reason for rejection of Broadbent's filter theory was numerous findings demonstrating that the meaning of an unattended message can influence performance. To accommodate these findings, Ann Treisman reformulated filter theory in the 1960s, developing what is called filter-attenuation theory. According to attenuation theory, early selection still precedes stimulus identification. The filter does not entirely block out the unattended information, but attenuates the information on unattended channels. If expectancies or familiarity cause recognition units to have low thresholds, the attenuated signal may still be sufficient to trigger recognition.

An extreme opposite of Broadbent's early-selection filter theory is late-selection theory, proposed initially by Deutsch and Deutsch, also in the 1960s. According to late-selection theory, within sensory limitations, all stimulus information is identified fully regardless of whether it is attended. The bottleneck is presumed to occur subsequent to stimulus identification. The central issue differentiating filter-attenuation theory and late-selection theory is whether meaning is always analyzed fully. This issue has been debated to the present without resolution. Pashler (1998b) recently concluded from a review of the literature that early attenuation does occur, but that it is an optional strategy rather than a structural limitation.

2.2. Resource Models

Bottleneck models are based to a large extent on selective attention studies in which one message is to be attended and others ignored. Divided attention tasks require concurrent processing of multiple sources of information and are more amenable to models that view attention in terms of resource limitations. Unitary resource models, of which Kahneman's (1973) is the most cited, view attention as a single pool of resources that can be distributed in various amounts to different tasks depending on task demands and voluntary allocation strategies. This is operationalized in performance operating characteristic curves representing a smooth tradeoff for concurrent performance of two tasks as a function of emphasis changes. Unitary resource models provided the impetus for dual-task methodologies and mental workload analyses that are used widely in Human Factors. Accordingly, multiple tasks should produce interference and mental overload when their combined resource demands exceed the available capacity.

Unitary resource models imply that the difficulty of tasks that are to be performed concurrently should be the primary determinant of performance. However, numerous studies have demonstrated that two tasks are easier to perform together when the stimulus or response modalities differ than when they do not. Furthermore, performance is better when one task requires verbal codes and the other visuo-spatial codes than when the same type of code is required for both tasks. These modality-specific patterns of interference have provided the basis for multiple resource models of attention, of which Wickens's (1984) is the most cited in Human Factors. Multiple resource models presume that there are distinct pools of attentional resources for different sensory-motor modalities and coding domains. Multiple resource theory provides a useful heuristic for predicting which display and control combinations can be performed efficiently together and which cannot. The primary criticism of multiple resource theory is that new resources can be proposed in an *ad hoc* manner to accommodate any finding of specificity of

interference. The fact that distinctions beyond broad modality and code differences can produce similar patterns of results has led some to conclude that the multiple resource view is insufficient.

3. CONTEMPORARY MODELS

Whether a structural bottleneck or limited capacity processor, a central issue in the study of attention was, and still is, the degree of processing that occurs with or without attention. Because of a shift in emphasis from auditory tasks to visual tasks beginning in the 1970s, the primary phenomena that many of these models attempt to explain involve processing of visual stimuli.

3.1. Models of Attention in Perception

One metaphor that has been widely used is to characterize attention as a spotlight. The main idea behind this metaphor is that attention serves to focus processing on a location within the visual field. The spotlight is presumed to have a spatial extent such that anything within the spotlight is attended, but attention cannot be directed to different, non-contiguous locations in the visual field. Although the direction of attention is often the same as the direction of gaze, evidence indicates that it is possible to direct attention independently of fixation. A distinction is made between exogenous and endogenous cueing of attention. An exogenous cue is the onset of a visual event in the field that involuntarily attracts the attentional spotlight to that location. An endogenous cue is a central signal, such as an arrow pointing to the right, that requires a voluntary shift of attention to the cued location.

A spatially based theory that has been particularly influential is Treisman's Feature Integration Theory (FIT). It was developed primarily to explain results from visual search studies in which a subject must indicate whether a target stimulus is present in an array. For feature search, in which the target is distinguished from distractors by a basic attribute such as color, the number of distractors has little effect on response time. However, for conjunctive search, in which two or more attributes must be conjoined to distinguish the target from distractors, response time increases as an approximately linear function of the number of distractors.

As with the early selection models, FIT assumes that basic features of stimuli are coded automatically and in parallel at the perceptual level. Each feature is then mapped onto specific "modules" in the brain. Detection of single features is automatic. However, attention is required for conjunction of features, proper selection of feature combinations, and matching them to previously stored representations. Treisman coined the focal attention needed in the conjoining of features as "glue."

Whereas FIT, in its original formulation, is an entirely space-based theory of attention, numerous authors, most prominently, John Duncan, have developed object-based models of attention that do not treat location as unique. Proponents of object-based theories argue that attention is drawn to perceptual groupings of stimulus elements, rather than to spatial locations. Among the findings cited as evidence for object-based models are the following: (1) identifying several targets in a complex visual display is facilitated if the targets form a perceptual group; (2) search for a conjunctive target is much more efficient when the distractors are homogeneous rather than heterogeneous; and (3)

it is easier to make judgments about two attributes from a single object in a display than about attributes from two objects, even when spatial separation of the attributes is controlled.

Treisman has modified FIT considerably to accommodate findings such as those described above. In a recent formulation (see her chapter in Baddeley and Weiskrantz 1993), a variant of the attentional spotlight, termed the attentional window, has been introduced. In the original formulation, though the theory was meant to explain object recognition, spatial location took precedence. Feature locations were implicitly coded, but attentional focus made feature location explicit. In the current formulation, though location takes precedence, there is parallel processing of both "what" and "where." Hence, dependent on prior specification, either objects or locations can direct control of attention. However, without prior specification, the system adjusts the attentional window to the task and searches serially.

3.2. Models of Attention in Response Selection and Initiation

The issue of whether response selection or initiation requires attention has been investigated extensively using the psychological refractory period (PRP) effect. Two discrete choice–reaction tasks must be performed by the subject, with the stimulus for the second task presented at varying stimulus onset asynchronies (SOA) relative to the stimulus for the first task. Second-task responses typically are slowed at short SOA, and it is this phenomenon that is called the PRP effect. Various types of models of attention have been evaluated with the PRP task, but most of the research has been conducted from the perspective of bottleneck models. The experimental results have been interpreted with what is called locus of slack logic. The basic idea is that if a Task 2 variable has its effect prior to the bottleneck, there will be an underadditive interaction with SOA. That is, at short SOA, the "slack" period during which post-bottleneck processing cannot begin can be used for continued processing for the more difficult condition. If a Task 2 variable has its effect subsequent to the bottleneck, the effect will be additive with SOA.

The most widely accepted account of the PRP effect since the 1950s has been to place the bottleneck at response selection. Pashler (1998b) has been the most vocal advocate of this position in recent years. The primary reason for accepting this locus for the bottleneck is that perceptual variables tend to show underadditive interactions with SOA, whereas post-perceptual variables tend to show additive effects. One exception to this pattern was obtained in a study by Karlin and Kestenbaum that manipulated whether Task 2 involved simple or choice reaction time. The effect of this manipulation showed an underadditive interaction with SOA; assuming that the variable has its effect on response selection, these data can be interpreted as implying that response selection occurs prior to the bottleneck. Hence, an alternative model was offered by Keele, who placed the bottleneck at response initiation. De Jong has recently proposed that bottlenecks exist at both response selection and response initiation, and Meyer and Kieras (1997) have argued that there is no structural bottleneck and that the locus of the bottleneck can vary depending on the strategies adopted to perform the assigned tasks.

3.3. Selection for Action

Recent theorizing takes the view that cognitive psychology cannot be complete without an understanding of both mind and brain. This has lead to an increased melding of the disciplines of neuroscience and cognitive psychology (for example, Pashler 1998a, Styles 1998). The bridge between neuroscience and information processing approaches has been aided by linking studies of animal visual systems with the findings from human behavioral studies of attention, including changes during development or as a consequence of brain damage. Localization of intact brain function has also been aided by brain imaging techniques such as positron emission tomography and functional resonance imaging. The temporal properties, as well as implied spatial localization of brain function, are studied by conjoining behavioral techniques with measures of brain activity, event-related electrical or magnetic potentials.

The selection for action approach, described in detail by Styles (1998), is consistent with the emerging use of the brain as metaphor for models of attention, and in the advances in computer models that more closely mimic brain function. Selection for action is a direct contrast to theories postulating structural bottlenecks or limited capacity. Apart from the constraints set by the sensory systems, the brain is considered to be of unlimited capacity. The processing of information is viewed as parallel and distributed. Information regarding attributes as well as possible actions is coded in separate dissociable systems throughout the brain. Attention is needed for perceptual integration, as well as for control of appropriate actions. In this sense, selectivity of attention prevents "crosstalk" or interference among competing activation from multiple streams of informational input. The coupling of relevant input and the blocking of irrelevant competing actions accomplish this. The only limitations are in the actual translation of input into action, in that usually only one action can be performed at a time with any effector.

3.3.1. A computational model

Recently, selection for action has been elaborated and translated into a more concrete analysis of attentional processes within a detailed computational model developed by Meyer and Kieras (1997). The model is implemented within the EPIC (Executive-Process Interactive Control) architecture, a comprehensive framework that includes perceptual, cognitive and motor components. This model, called the Strategic Response-Deferment, was formulated specifically to account for attentional processing in multiple-task performance, specifically PRP tasks. It includes a step-by-step analysis of the processes involved in the performance of each individual task, as well as of the executive control processes that coordinate joint performance. Attention begins at the perceptual level, orienting focus (i.e. moving the eyes) on sensory input. Limits in the systems are attributed to the sensory and motor effectors, as well as individuals' strategies for satisfying task demands (e.g. making sure that the responses for the two tasks are made in the instructed order). The model allows for flexible scheduling in the processing chain during dual-task performance, thus accounting for the effects of response–selection difficulty and SOA on observed reaction times without need of a central, structural bottleneck.

The EPIC architecture resembles the Model Human Processor

architecture developed in the early 1980s by Card, Moran and Newell. However, it has the advantages of being implemented in a computational model and describing in greater detail the operation of the perceptual and motor processes. Also, it has a larger processing capacity and, therefore, can model multiple-task performance without cognitive resource limitations. These features, along with the emphasis placed on sensory-motor limitations and the control strategies that a person can use to accomplish task goals, should make the model attractive to Human Factors specialists.

3.3.2. A neural network model

Neurophysiological evidence has demonstrated attentional inhibition in both intra-cellular recordings of primate cortex and human electrophysiological measurements. Behaviorally, the significance of inhibitory processes of attention is demonstrated with the phenomena of inhibition of return and negative priming (see Milliken and Tipper's chapter in Pashler 1998a). Inhibition of return refers to the finding that although an exogenous cue facilitates processing of a target stimulus at the cued location when the cue–target interval is short, it inhibits processing when the interval is longer. This phenomenon reflects a bias toward novel events and may be a consequence of automatic activation of the oculomotor system. Negative priming refers to the finding that, in selective attention tasks, if the distractor on one trial becomes the target on the following trial, responses are slowed. This effect has been demonstrated with objects and locus of attention. It can also occur with semantically related items and when the stimuli are presented in different perceptual modalities. Negative priming is used as evidence that irrelevant information is processed beyond the perceptual level. Inhibitory effects may not be modality specific but necessary for the coordination of perception and action by assisting enhancement of the selected stimulus and reducing interference.

Houghton and Tipper (1994) have successfully demonstrated both negative priming and inhibition of return in a neural network model that treats the excitatory and inhibitory properties of selective attention as an opponent process mechanism. Inhibition is placed in the flow of activation between perceptual systems and motor planning and execution. Attention requires the matching of a target stimulus with its internally maintained representation. Situated in the frontal lobes, inhibition is a self-regulating mechanism allowing non-targets to achieve an equilibrium activation below that of targets, thus allowing enhanced activation of the target. Negative priming and inhibition of return are viewed as the result of inhibitory rebound and suggest that this effect of the opponent mechanism is not confined to ignored inputs but can happen to selected inputs as well.

4. AUTOMATICITY

It has been customary to distinguish processing that requires attention from processing that does not, which is usually called "automatic" processing (see Proctor and Dutta 1995 for a review). Starting as early as the 1890s with Bryan and Harter, automatic processes were thought of as habits, or spontaneous chains of behavior initiated by environmental stimulus events and not subject to conscious control. The concept of automaticity received renewed interest in the 1970s due to the popularity of limited-capacity models of attention. The general reason for this was

numerous findings suggesting that under many circumstances there is no limitation on information processing for highly overlearned tasks and activities, as well as little if any control over them. One example from laboratory tasks is the well-known Stroop color-naming effect, in which the time to name the color in which a stimulus is printed is much longer if the stimulus spells a conflicting color word than if it is neutral elements or a congruent color word. Phenomena such as these led to the view that automatic processing is qualitatively different from controlled, attentional processing. Its defining characteristics include occurrence without intention, lack of conscious awareness, and lack of interference with other mental processes.

Shiffin and Schneider developed the most prominent dual-process theory of controlled and automatic processing. Their theory was developed primarily to explain results from search tasks in which the number of items in visual displays and the size of the target set held in memory were varied. A central distinction made was between tasks for which the mapping of stimuli as targets or distractors remained consistent (i.e. a member of the target set on one trial could not occur as a distractor on another) versus tasks for which the mapping varied. Their major finding was that the set-size effects were very small in consistent mapping tasks compared to varied mapping tasks. However, once a person became skilled at a consistent mapping task, considerable negative transfer occurred when the target and distractors were switched. Shiffrin and Schneider proposed that performance with highly practiced consistent mappings occurs in an automatic mode that operates in parallel across the display, whereas performance with a varied mapping occurs in a controlled mode that operates serially across the display.

The major implication of the dual-process view is that some tasks require controlled attentional processing, whereas others do not. Not too surprisingly, individuals who have questioned the value of the limited-capacity resource assumption have tended to question whether there really is a qualitative distinction between automatic and controlled processing. The general nature of the critiques has been to argue that "automatic" processes are not independent of intentions, that they may show signs of interference if responses are similar, and that they can show evidence of control when the task demands change. In short, these individuals see the automatic-controlled distinction as excess baggage in the same way as the concept of limited capacity.

Rather than assuming that attention is discarded as a person becomes practiced at a task, an alternative view is that it is more important to consider the change in attentional control strategies and the organization of behavior. For example, in keeping with the selection for action view, practice is seen as constraining the parameters of action. Hence, less attention is needed for the processing of a response. Automatic processing is seen as a quantitative grading of attentional control through the acquisition of attentional skills or procedures for action stored as schemata in long-term memory.

A third viewpoint of automaticity, Logan's instance theory, does not concentrate on the response but the overall processing of stimuli. Attention is seen as the interface between encoding and retrieval. Attention to a stimulus obliges encoding of that instance as well as retrieval of past instances associated with the stimulus. Further, each encounter with a stimulus creates an independent trace. With experience, the accumulation of these traces leads to a transition from algorithmic-based to memory-based processing. Logan's theory relies on individual memory traces rather than on a change in strength of associative links. Performance variability and task-load effects will diminish with practice not because of a qualitative change but because direct memory retrieval can eliminate the slower, step-by-step algorithmic process. This account both predicts the power function speed-up seen in skill acquisition as well as the poor transfer effects to novel situations.

5. IMPLICATIONS FOR HUMAN FACTORS

1. As captured by Filter Theory, information intake is often serial and little is remembered regarding the meaning of unattended messages. Selection between streams of information can be performed efficiently on the basis of physical distinctions such as location.

2. As implied by filter-attenuation and late-selection models, meaning of unattended information can be processed and influence performance, under at least some circumstances.

3. Unitary resource models provide an algorithm for depicting the trade-off between two tasks, as represented by performance operating characteristic curves.

4. Multiple resource models provide a useful heuristic for determining which combinations of stimulus and response modalities and codes will likely lead to optimal performance in multiple-task contexts. They also provide a rationale for varying the tasks used for assessment of mental workload with secondary-task procedures to provide a complete profile of task demands.

5. Models of visual search make clear that selection of information from visual displays is more efficient if done on the basis of distinctive features, perceptual groupings, or object properties.

6. Response competition is a major factor influencing performance in many situations, and models emphasizing limitations in response selection and initiation are useful in deciding the timing structure between multiple tasks.

7. The selection for action approach emphasizes the importance of control strategies adopted by operators in order to coordinate action in complex task environments. The procedural and connectionist models developing out of this approach are promising tools for describing and simulating task performance.

8. Attention often has inhibitory as well as facilitatory effects on performance that must be taken into consideration in any assessment of performance limitations.

9. Models of automaticity present guidelines for the conditions of efficient time-sharing. Most important is that automaticity develops when a consistent mapping of task elements is maintained.

REFERENCES

BADDELEY, A. D. and WEISKRANTZ, L. (eds), 1993, *Attention: Awareness, Selection, and Control* (Oxford: Oxford University Press).

BROADBENT, D. E. (1958), *Perception and Communication* (Oxford: Pergamon Press).

BRYAN, W. L. and HARTER, N. (1999), Studies on the telegraphic language: The acquisition of a hierarchy of habits. *Psychological Reiew*, 6, 345–75.

CARD, S. K., MORAN, T. P. and NEWELL, A. (1983), *The Psychology of Human–computer Interaction* (Hillsdale, NJ: Lawrence Erlbaum).

DE JONG, R. (1993), Multiple bottlenecks in overlapping task performance. *Journal of Experimental Psychology: Human Perception and Performance,* 19, 965–80.

DEUTSCH, J. A. and DEUTSCH, D. (1963), Attention: Some theoretical considerations. *Psychological Review*, 70, 80–90.

DUNCAN, J. (1984), Selective attention and the organization of visual information. *Journal of Experimental Psychology: General*, 113, 501–17.

HOUGHTON, G. and TIPPER, S.P., 1994, A model of inhibitory mechanisms in selective attention. In D. Dagenbach and T. Carr (eds), *Inhibitory Processes in Attention, Memory, and Language* (Orlando: Academic Press), 53–112.

KAHNEMAN, D., 1973, *Attention and Effort* (Englewood Cliffs: Prentice-Hall).

KARLIN, L and KESTENBAUM, R. (1968), Effects of a number of alternatives on the psychological refractory period. *Quarterly Journal of Experimental Psychology*, 20, 167–78.

KEELE, S. W. (1973), *Attention and Human Performance* (Pacific Palisades, CA: Goodyear).

LOGAN, G. D. (1988), Toward an instance theory of automatization. *Psychological Review*, 95, 492–527.

MEYER, D. E. and KIERAS, D. E., 1997, A computational theory of executive cognitive processes and multiple-task performance: Part 1. Basic mechanisms. *Psychological Review*, 104, 3–66.

MILLIKEN, B. and TIPPER, S. P. (1998), Attention and inhibition. In H. Pashler (ed.), *Attention* (191–221) (Hove, East Sussex: Psychology Press Ltd).

PASHLER, H. E. (ed.), 1998a, *Attention* (Lewes: Psychology Press).

PASHLER, H. E., 1998b, *The Psychology of Attention* (Cambridge, MA: MIT Press).

PROCTOR, R. W. and DUTTA, A., 1995, *Skill Acquisition and Human Performance* (Thousand Oaks: Sage).

SHIFFRIN, R. M. and SCHNEIDER, W. (1977), Controlled and automatic human information processing: II. Perceptual learning, automatic attending, and a general theory. *Psychological Review* 84, 127–90.

STYLES, E. A., 1998, *The Psychology of Attention* (Lewes: Psychology Press).

TREISMAN, A. M. (1960), Contextual cues in selective listening. *Quarterly Journal of Experimental Psychology*. 12, 242–8.

TREISMAN, A. (1993), The perception of features and objects. In A. Baddeley and L. Weiskrantz (eds.), *Attention: Selection, awareness, and control* (5–35) (Oxford: Clarendon Press).

WICKENS, C. D., 1984, Processing resources in attention. In R. Parasuraman and D. R. Davies (eds), *Varieties of Attention* (New York: Academic Press), 63–102.

Multiple Resources

D. B. Boles

Department of Philosophy, Psychology, and Cognitive Science, Rensselaer Polytechnic Institute, Troy, NY 12180, USA

1. INTRODUCTION

Multiple resources are mental entities that permit the processing of information in more than one channel simultaneously. Resources were originally conceived as specifically attentional in nature. Most recently they have been identified with mental processes more generally. The exploitation of multiple resources can allow more than one task to be performed together with minimal interference, a matter of great importance in the design of complex human–machine interfaces.

Comprehending the nature of multiple resources requires understanding how the resource construct evolved in the fields of cognitive psychology and human factors. To some extent earlier versions of resource theory are still viable, in the form of possible "layers" of resources (Wickens 1984: 305) that vary from a generalized pool usable by any task, down to numerous specialized resources associated with specific mental processes.

2. RESOURCES AS ATTENTION
2.1. Single Channel Theory

Early attentional theory held that the human could be modeled as a single channel when first attending and subsequently processing information through to response (Broadbent 1958). On this view every attended task would require completion before any other task could be initiated beyond anything but the most rudimentary pre-attentional processing. Thus, attention was conceived as a limited-capacity resource required in every conscious action, and the simultaneous performance of multiple tasks was stated, in effect, to be impossible.

The single-resource viewpoint, however, soon proved problematic when instances were found of parallel processing. For example, it was discovered that people could usually recognize different digits presented simultaneously at the two ears, and they could detect their name when it was unexpectedly presented at the ear opposite to a continuous message requiring close monitoring. Such results implied that multiple sources of information could be processed simultaneously, a seeming violation of single-channel theory.

Attentional theorists proved adept at amending single-channel theory to account for the new findings, variously proposing that attention can switch rapidly back and forth between inputs; that all events can be processed to some extent for meaning, although only one is actually attended at a time; or that attention operates subsequent to processing for meaning. Nevertheless, all such alterations abandoned the notion that complex human performance could be sufficiently modeled using a single channel resource called attention, whose operation was limited in real time.

Because it had been demonstrated that multiple channels could sometimes be processed together, the question became how best to characterize the nature, capabilities and limitations of those channels. This question was addressed first by positing the existence of a generalized resource pool that could be drawn upon by multiple tasks, and later by proposing the existence of multiple resources.

2.2. Resource Pool Theory

Kahneman (1973) among others proposed that attention was best conceived as a pool of undifferentiated resources that could be parceled out to one or more tasks as required. Multiple-task performance was, therefore, possible using multiple channels without substantial interference up to the point where aggregate demand on the pool exceeded the supply.

Although resource pool theory proved useful in accounting for breakdowns in human performance as multi-tasking became more demanding, problems developed as it became clear that certain pairings of tasks or task components resulted in less mutual interference than other pairings. Brooks (1968), for example, showed that a task embedding visuo-spatial and auditory–verbal components produced much less interference than a task embedding either two visuo-spatial or two auditory–verbal components. Accordingly it appeared that interference could not always be predicted by aggregate demand on a single pool of resources.

2.3. Multiple Resource Theory

Wickens (1984) drew on an increasing empirical base of studies investigating dual-task performance, to derive an explicit resource model. His multiple resource theory (MRT) incorporated three orthogonal dimensions: (1) stages of processing (encoding/central processing and response stages); (2) modality (visual and auditory on the encoding end of processing, and manual and vocal on the response end); and (3) processing code (verbal and spatial). Each cell of the model was proposed to have resources dedicated to it, allowing relative predictions of interference between pairs of tasks by examining overlapping characteristics. Little interference was predicted if the two tasks used completely different cells, for instance with one task requiring visual, spatial and manual processing, and with the second task requiring auditory, verbal and vocal processing, especially if the execution of each task was most demanding at a different stage of processing (e.g. central processing for one versus responding for the other). Experimental tests using dual tasks showed that the model faired well: interference generally increased as resource overlap increased.

Although MRT was framed in terms of type of processing, it is important to note that it was developed within an attentional context, and made the implicit assumption that resources underlying the different types of processing were attentional in nature. This formulation essentially preserved the "attentional pool" of Kahneman by fractionating it, and then predicted interference in cases where two tasks drew on the same pool or pools out of multiple possible pools.

Wickens (1992) made minor alterations to MRT, for example by suggesting that interference between two visual tasks might be due more to physical scanning demands than resource competition. However, the bulk of the model remained implicitly attentional in nature.

3. RESOURCES AS PROCESS

Boles and Law (1998) recently proposed a reinterpretation and expansion of MRT. They noted that while the theory had been

described in an attentional context, it may correctly predict dual task performance due to interference in resources either attentional or structural (e.g. by competing for structural areas of the brain devoted to particular mental processes, or for peripheral effectors). They suggested that a more general viewpoint might be to identify resources with particular processes and to leave aside, for the moment at least, the question of whether resources are better viewed in attentional versus structural terms. Finally, they proposed that orthogonal mental processes have orthogonal resources, a principle that if true could potentially require a substantial expansion of MRT.

Putting the orthogonality principle to an empirical test, Boles and Law conducted a series of experiments in which dual tasks

Table 1. Some Probable Process-Specific Mental Resources.

A. Encoding/central processing resources

1. **Auditory emotional** (Boles and Law 1994): resources associated with recognizing emotional tones of voice, using auditory input

2. **Auditory-spatial** (Wickens 1984): resources associated with generalized spatial processing, using auditory input

3. **Auditory-verbal-linguistic** (Wickens 1984, Boles and Law 1998): resources associated with generalized verbal processing, using auditory input

4. **Tactile figural** (Boles and Law 1998): resources associated with recognizing shapes, using tactile input

5. **Visual-spatial** (Wickens 1984): resources associated with generalized spatial processing, using visual input

   a. **Facial figural** (Boles and Law 1998): resources associated with the processing of faces or facial emotions, using visual input

   b. **Planar categorical** (Boles 1998): resources associated with simple left-versus-right or above-versus-below relationships, using visual input

   c. **Spatial attentive** (Boles and Law 1998): resources associated with the deployment of attention in space, using visual input

   d. **Spatial concentrative** (Boles 1998): resources associated with recognition of the density of clustering of numerous objects, using visual input

   e. **Spatial emergent** (Boles 1998): resources associated with the separation of figure and ground, using visual input

   f. **Spatial positional** (Boles and Law 1998): resources associated with the identification of precise locations, using visual input

   g. **Spatial quantitative** (Boles and Law 1998): resources associated with the recognition of analog numerical quantities, using visual input

6. **Visual temporal** (Boles and Law 1998): resources associated with the recognition of the timing of events, using visual input

7. **Visual-verbal** (Wickens 1984): resources associated with generalized verbal processing, using visual input

   a. **Visual lexical** (Boles and Law 1998): resources associated with recognizing words, letters, or digits, using visual input

   b. **Visual phonetic** (Boles and Law 1998): resources associated with evoking verbal sounds, using visual input

B. Responding Resources

1. **Facial motive** (Boles and Law 1998): resources associated with responding with facial movement unrelated to speech or emotion

2. **Manual** (Wickens 1984): resources associated with responding with the hands

3. **Vocal** (Wickens 1984): resources associated with responding with the voice

either did or did not use orthogonal perceptual processes as deduced from prior factor analytic research. In general they found that competition for the same process resulted in substantially degraded performance compared with the use of orthogonal processes. Furthermore, multiple visuo-spatial processes found to be orthogonal in factor analytic work yet not differentiated by MRT, were among those that did not selectively interfere. Evidence was likewise found for resources associated with temporal processing that were not encompassed by MRT. Using inductive logic, the authors concluded that orthogonal processes do have orthogonal resources, and that any complete MRT requires a full accounting of them.

Because the processes investigated by Boles and Law were drawn from a larger set determined to be orthogonal from factor analysis, the inductive conclusion is that each of the larger set most likely has resources associated with it. Combining these resources with those of Wickens' MRT, and with more recent factor analytic work (Boles 1998), it is possible to list a fairly large set of process-specific resources, with the often more specific resources of Boles and Law, and of Boles, typically nested under the more general resources of Wickens. It should be understood that any such tabulation can only be as complete as our current taxonomy of orthogonal processes (Table 1).

It should also be noted that besides process-specific resources, some type of general resource may have influenced performance in the experiments of Boles and Law. Dual-task pairings resulted in substantially lengthened reaction time (RT) compared with single-task RT as measured in equivalent experiments. This outcome may indicate competition for generalized resources, resulting in interference even when two tasks use dissimilar processes. In short, there is good evidence for multiple resources that correspond to multiple processes, yet a full accounting of multiple-task performance will need to consider the possible existence of "layers" of resources that are less process-specific (Wickens 1984: 305). Candidates for such layers include Kahneman's general pool and Wickens' "stage of processing" resources, which were proposed to be available to all processes within a given stage.

4. PRIVILEGED LINKAGES BETWEEN RESOURCES

Beyond describing a specific model of multiple resources, Wickens *et al.* (1983) proposed that privileged linkages exist between certain of the model's cells. Specifically, verbal processing was proposed to be best served by auditory input and vocal output, and spatial processing by visual input and manual output. Such stimulus-central processing-response (S-C-R) compatibility was viewed as a natural expansion of earlier S-R compatibility concepts, when applied to complex systems requiring that the operator maintain a mental model using a substantial amount of central processing resources.

Empirical support for the privileged access proposal is based on studies examining both single- and dual-task performance under varying S-C-R arrangements. Of these, single-task performance provides the less ambiguous test. Wickens *et al.* (1983) found that spatial acquisition of a target stimulus was performed best when its identity was cued visually and its acquisition was confirmed manually, and performed worst when identity was cued auditorily and acquisition was confirmed vocally. The opposite result was obtained in a verbal command

task. Further confirmatory results were found in a dual-task study by Wickens *et al.* (1984), who investigated performance over a continuum of tasks in which predicted resource competition and predicted S-C-R incompatibility were negatively correlated (i.e. as one increased, the other decreased).

While these and similar results support the existence of privileged linkages between resources, it is presently unclear how expanded conceptions of multiple resources might affect their expression. As noted, Boles and Law (1998) found empirical evidence for multiple spatial resources, and they cited factor analytic evidence arguing for multiple verbal resources as well. It remains to be seen whether all spatial processes have privileged access to visual input and manual output, and whether all verbal processes have privileged access to auditory input and vocal output. It is possible that S-C-R compatibilities will be found to differ depending on the exact spatial and verbal resources involved.

5. APPLICABILITY OF MULTIPLE RESOURCES

From a human factors perspective, the value of the multiple resources concept is determined by its applicability to real-world design. Generally speaking, most situations in which the concept should be of value are those in which humans perform multiple tasks simultaneously. However, the concept also has implications for the measurement of workload and for training.

5.1. Multiple-task Application

One of the major areas for the application of multiple resource concepts is that of flight control, due to the extensive multi-tasking required of pilots. Wickens and Liu (1988) examined a simplified simulation of flight under threat conditions. Participants were asked to manually track a moving target while either categorizing the threat represented by spatial (vector) depictions of enemy aircraft, or selecting a weapon based on verbal (digital) depictions of the probability of its reaching its target. Responses to the categorization and selection tasks were made either manually or vocally, providing a manipulation of predicted response-level interference with the tracking task. In general, performance was found to be lower when the spatial tracking task was paired with the spatial threat task, and when responses to both the tracking and the other tasks were made manually, compared with conditions in which spatial and verbal processes or manual and vocal processes were paired. Thus, MRT was supported.

Generally speaking, any dual-task setting requiring high levels of performance and separate responses for the two tasks, could be expected to benefit from design influenced by multiple resource considerations.

5.2. Workload Application

Wickens *et al.* (1988) pursued the intriguing possibility that multiple resource concepts could be extended to the measurement of workload. In their study, a single expert rated the degree of overlap in resource demands incurred by multiple flight tasks. Predictions of interference based on the overlap in demand on multiple resources were then compared with the predictions of workload models variously based on the total number of tasks performed, the number of tasks performed per unit time, and the total resource demand. The overlap (multiple resources) model performed remarkably better than the others at predicting

flight performance. However, the other models performed much better than the overlap model at predicting overall subjective workload as rated by the participants.

What the Wickens *et al.* (1988) results appear to indicate is that a multiple-resource-based approach holds promise if one wishes to measure objective workload as it impacts on actual performance. On the other hand, if one is interested in subjective workload as it is perceived by participants, a more global measure may be a better choice.

5.3. Training Application

Recently Boles (1997) suggested that training programs can also benefit from multiple resource concepts, if resources are thought of as separately trainable components that can be used flexibly in a variety of tasks. The training of a complex task might therefore benefit from targeting critical resources, even if a task very different from the complex task is used during training. Empirical support for this idea came from an experiment in which participants made quantitative judgments based on either bargraph or numerosity (dot cluster) representations, both types of judgment using the same spatial quantitative resource according to prior research. It was found that training in numerosity judgments benefited bargraph judgments nearly as much as training in bargraph judgments themselves, and much more than training using verbal representations of number. Thus, a training task may not require high fidelity with respect to a more complex end task, as long as the same resources are used.

5.4. Limitations on Application

Of equal importance to the applicability of mental resources are indications that its applicability has limits. Attempts to implement the resource construct outside the limits may prove unsuccessful.

Although as indicated, dual-task design benefits from the use of separate resources in the two tasks, this may be true only when they require distinctly separate responses. Boles and Wickens (1987) found that when subjects responded separately to two numerical indicators, performance was better when the indicators were formatted to use separate perceptual resources than when they used the same resource, a replication of the multiple resource advantage in dual-task design. However, if a single response was required, the multiple resource advantage was eliminated. Indeed, in an integration task requiring a single judgment of whether either indicator exceeded a criterion value, there was actually an advantage when they used the same resource. Most likely the outcome derived from a translation cost incurred when different-format indicators had to be compared with the same criterion and the results combined prior to the single response.

Boles and Ruffles (1987) explored a further limit to applicability when comparing experimental conditions in which two numerical indicators had values that were either uncorrelated or highly correlated ($r = 0.9$). Subjects making separate responses to the indicators performed better when they were formatted to use separate resources, congruent with multiple resource concepts, but only when the indicators had values that were uncorrelated over trials. When a high correlation existed, the separate-resource advantage was eliminated. However, performance became better overall, in accordance with the authors' prediction that correlated indicators would encourage subjects to integrate them, generally improving performance but

eliminating the multiple-resource advantage because of translation costs.

Both these results and those of Boles and Wickens illustrate the risks involved in applying multiple resource concepts outside of true multiple-task situations. If multiple tasks become less so, either by reducing response requirements to a single response or by using correlated information sources, translation costs appear to overwhelm any advantage that might otherwise accrue from the use of separate resources. In a true multiple-task situation where information sources and responses are distinct, translation is a moot issue and multiple resource considerations apply. Wickens (1992: 379) drew similar conclusions from research on tracking performance, suggesting that multiple resource considerations are less important when dual tasks require the same mental set, processing routine or timing mechanism.

6. RECOMMENDATIONS

The multiple resource construct holds considerable promise in the prediction of multi-task performance, in workload measurement, and in training. The following recommendations may help fulfill that promise.

Ascertain whether the multiple resource construct is appropriate before applying it in a given context. In the prediction of multi-task performance, the multiple resource construct is most likely to be appropriate if the tasks are intended to be processed independently. Caution should be exercised in situations that might by design or circumstance reward the integration of information across tasks. Examples include the integration of multiple information sources into a single response; the use of correlated information sources; and convergence of multiple information sources onto a common processing routine. In such cases, the savings from multiple resource use may be overwhelmed by the costs of translating information sources into a common format prior to their integration.

In appropriate multi-task situations, consider designing tasks using a dissimilarity heuristic. Because the only complete multiple resource model is one that includes resources associated with all possible mental processes, it may be some time before a complete model is approximated. Nevertheless, in multi-task situations it should be possible to take advantage of multiple resources simply by designing tasks to be as dissimilar as possible. Very often a given task can be achieved through alternate input and output modalities, and sometimes it can be achieved through alternate perceptual and cognitive processes. A judicious selection of these to make multiple tasks as dissimilar as possible should reduce interference between them.

Consider implementing possible S-C-R compatibilities in both single- and multi-task design. Verbal processing may best be served by auditory input and vocal output, and spatial processing by visual input and manual output. Other factors being equal, it may be beneficial to employ these linkages in both single- and multi-task design. However, all such designs should be thoroughly tested since it is presently unclear whether all spatial (or all verbal) processes interact similarly with input and output modalities.

In workload assessment, consider using or developing a multiple-resource-based measure if performance is to be predicted. There is some reason to believe that multiple-resource-based measures of workload more accurately predict performance than global measures. Among MRT-influenced approaches are

W/INDEX and the Workload Profile (North and Riley 1989, Tsang and Velazquez 1996). Alternatively, more complete measures could presumably be developed to include workload associated with task-critical processes not in the original MRT.

Consider supplementing target task training with resource training if economic or risk considerations favor it. A target task may benefit from the training of resources used in performing it. This becomes a particularly desirable strategy if training on the target task itself is expensive or risky. The training of resources may be found to show considerable transfer to the target task, and thus can be used to supplement a reduced program of target-task training.

Finally, it should be noted that further empirical work is needed to more fully identify the set of process-related resources to be included in a multiple resources model. Such work should include empirical assessments in applied settings to identify, if possible, those resources that might be discarded for the sake of parsimony, versus those most important in multi-tasking, workload assessment, and training.

REFERENCES

BOLES, D.B., 1997, Multiple resource contributions to training. *Proceedings of the Human Factors Society*, 41, 1176–9.

BOLES, D.B., 1998, Multiple lateralized spatial processes. Presented at the 39th Annual Meeting of the Psychonomic Society, 19–22 November, Dallas, TX.

BOLES, D.B. and LAW, M.B., 1998, A simultaneous task comparison of differentiated and undifferentiated hemispheric resource theories. *Journal of Experimental Psychology: Human Perception and Performance*, 24, 204–15.

BOLES, D.B. and RUFFLES, J.W., 1987, Multiple resources theory and correlated displays. *Proceedings of the Annual Conference on User System Interfaces*, 27–35.

BOLES, D.B. and WICKENS, C.D., 1987, Display formatting in information integration and nonintegration tasks. *Human Factors*, 29, 395–406.

BROADBENT, D.E., 1958, *Perception and Communication* (London: Pergamon).

BROOKS, L.R., 1968, Spatial and verbal components of the act of recall. *Canadian Journal of Psychology*, 22, 349–68.

GOETTL, B.P. and WICKENS, C.D., 1989, Multiple resources versus information integration. *Proceedings of the Human Factors Society*, 33, 1454–8.

KAHNEMAN, D., 1973, *Attention and Effort* (Englewood Cliffs: Prentice Hall).

NORTH, R.A. and RILEY, V.A., 1989, W/INDEX: A predictive model of operator workload. In G.R. McMillan, D. Beevis, E. Salas, M.H. Strub, R. Sutton and L. van Breda (eds), *Applications of Human Performance Models to System Design* (New York: Plenum), 81–9.

TSANG, P.S. and VELAZQUEZ, V.L., 1996, Diagnosticity and multidimensional subjective workload ratings. *Ergonomics*, 39, 358–81.

WICKENS, C.D., 1984, *Engineering Psychology and Human Performance* (Columbus: Charles E. Merrill).

WICKENS, C.D., 1992, *Engineering Psychology and Human Performance* (2nd ed.) (New York: HarperCollins).

WICKENS, C.D., HARWOOD, K., SEGAL, L., TKALCEVIC, I. and SHERMAN, B., 1988, TASKILLAN: a simulation to predict the validity of multiple resource models of aviation workload. *Proceedings of the Human Factors Society*, 32, 168–72.

WICKENS, C.D. and LIU, Y., 1988, Codes and modalities in multiple resources: a success and a qualification. *Human Factors*, 30, 599–616.

WICKENS, C.D., SANDRY, D.L. and VIDULICH, M., 1983, Compatibility and resource competition between modalities of input, central processing, and output. *Human Factors*, 25, 227–48.

WICKENS, C.D., VIDULICH, M. and SANDRY-GARZA, D., 1984, Principles of S-C-R compatibility with spatial and verbal tasks: the role of display-control location and voice-interactive display–control interfacing. *Human Factors*, 26, 533–43.

Muscle Strength

K. H. E. Kroemer

Department of Industrial and Systems Engineering, Virginia Tech, Blacksburg, VA, and Director, Eri, Inc., A Consulting Company, PO Box 3019, Radford, VA 24143-3019, USA

Muscle strength, the capability to generate lengthwise tension (force) between its origin and insertion, depends on the cross-sectional thickness of the muscle (i.e. its number of fibers), on the types of fibers, and the geometric arrangement of its fibrils and fibers.

The frequency of contractions and their strength are controlled by the frequency of exciting nervous signals. This is called rate coding. Recruitment coding determines how many motor units of a muscle contract simultaneously; the more, the higher the contractile strength exerted by the total muscle. Thus, nervous control of muscle strength involves a complex pattern of rate and recruitment coding.

1. MUSCLE FATIGUE

If, after contraction, not enough time is provided for relaxation and recovery before the next contraction, the muscle becomes 'fatigued'. (It experiences an oxygen deficit, lactic acid is not removed sufficiently, potassium ions accumulate while sodium is depleted in extracellular fluid, phosphate accumulates in intracellular fluid, and ATP rebuilding is hindered — see the references for details.) Fatigue reduces muscle force and endurance capabilities. Muscular fatigue is overcome by rest, during which accumulated metabolic byproducts are removed.

Muscle fatigue depends on the frequency and intensity of muscular contraction, and on the period of time over which it is maintained. The larger the portion of its strength that a muscle must exert, the shorter the period through which this exertion can be maintained. As a practical rule, maximal muscle strength can be maintained for only a few seconds; 50% of strength is available for ~1 min; < 20% can be applied continuously for long times.

2. MUSCLE LENGTH AND TENSION

Under a 'no-load' condition, in which no external force applies and no internal contraction occurs, the muscle is at its resting length. Sufficient nervous stimulation causes the muscle to contract and, in doing so, exert force against any resistance. In its smallest possible length, the actin proteins are completely curled around the myosin rods; this is at ~60% of resting length. In this position, the muscle cannot shorten any farther nor develop any further active contraction force.

Stretching the muscle beyond resting length by external forces between insertion and origin lengthens it passively. When the muscle is stretched to 120–130% of resting length, the cross-bridges between the actin and myosin rods are in optimal position to generate a contractile force. But elongating the fibrils further reduces the overlap in the cross-bridges between the protein rods. At ~160% of resting length, so little overlap remains that no contraction force can be developed internally. Thus, the curve *active contractile force* developed within a muscle is zero at ~60% resting length, ~0.9 at resting length, at unit value at ~120–130% of resting length, and then falls back to zero at ~160% resting length. (These values apply to an isometric, or static, contraction — see the chapter on Static and Dynamic Strength.)

Passive reaction force to external stretch also occurs, as in a rubber band. This passive stretch resistance increases strongly from resting length to the point of muscle or tendon (attachment) breakage. Thus, above resting length, the total tension in the muscle is the sum of active and passive strain. This explains why we "preload" muscles by stretching them for a strong force exertion, as in bringing the arm behind the shoulder before throwing.

3. VISCOELASTIC BEHAVIOR

In engineering terms, skeletal muscles exhibit viscoelastic properties. They are viscous because their reaction depends both on the amount by which they are deformed, and on the rate of deformation. They are elastic in that, after deformation, they return to their original length and shape. These behaviors, however, are not pure in the muscle, because it is non-homogeneous, anisotropic, and discontinuous in its mass. Nevertheless, viscosity and nonlinear elastic theory can be used to explain major features of muscular performance. This includes the reasons why the muscle tension that we can develop statically by holding a stretch is the highest possible, while in active shortening (a dynamic concentric movement) muscle tension is decidedly lower. The higher the velocity of muscle contraction, the faster actin and myosin filaments slide by each other and the less time is available for the cross-bridges to develop and hold (Schneck 1992). This reduction in force capability of the muscle holds true for both concentric and eccentric activities. In eccentric activities, however, where the muscle becomes increasingly lengthened beyond resting length, the total force resisting the stretch increases with larger length, owing to the — just discussed — summing of active and passive tension within the muscle.

REFERENCES

KROEMER, K.H.E., KROEMER, H.J. and KROEMER-ELBERT, K.E., 1997, *Engineering Physiology* (3rd ed.) (New York: van Nostrand Reinhold–Wiley).

KROEMER, K.H.E., KROEMER, H.B. and KROEMER-ELBERT, K.E., 2000, *Ergonomics: How to Design for Ease and Efficiency* (2nd ed.) (Englewood Cliffs: Prentice-Hall).

SCHNECK, D.J., 1992, *Mechanics of Muscle* (2nd ed.) (New York: New York University Press).

WINTER, D.A., 1990, *Biomechanics and Motor Control of Human Movement* (2nd ed.) (New York: Wiley).

Muscle Terms – Glossary

K. H. E. Kroemer

Industrial And Systems Engineering, Virginia Tech, Blacksburg, VA, and Eri, Inc., A Consulting Company, PO Box 3019, Radford, VA 24143-3019, USA

- Activation of muscle — see contraction.
- Co-contraction — simultaneous contraction of two or more muscles.
- Concentric (muscle effort) — shortening of a muscle against a resistance.
- Contraction — literally, "pulling together" the *Z* lines delineating the length of a sarcomere, caused by the sliding action of actin and myosin filaments. Contraction develops muscle tension only if the shortening is resisted. Note that during an isometric "contraction" no change in sarcomere length occurs and that in an eccentric "contraction" the sarcome is actually lengthened. To avoid such contradiction in terms, it is often better to use the terms activation, effort or exertion.
- Distal — away from the center of the body.
- Dynamics — subdivision of mechanics that deals with forces and bodies in motion.
- Eccentric (muscle effort) — lengthening of a resisting muscle by external force.
- Fiber — see muscle.
- Fibril — see muscle fibers
- Filament — see muscle fibers.
- \n Force — as per Isaac Newton's Third Law, the product of mass and acceleration; the proper unit is the Newton, with $1\,N = 1\,kg\,m\,s^{-2}.$ On Earth, 1 kg applies a (weight) force of 9.81 N (1 lb exerts 4.44 N) to its support. Muscular force is defined as muscle tension multiplied by transmitting cross-sectional area.
- Free dynamic — in this context an experimental condition in which neither displacement and its time derivatives, nor force are manipulated as independent variables.
- Iso — prefix meaning constant or the same.
- Iso-acceleration — condition in which the acceleration is kept constant.
- Isoforce — condition in which the muscular force (tension) is constant, i.e. isokinetic. This term is equivalent to isotonic.
- Iso-inertial — condition in which muscle moves a constant mass.
- Isojerk — condition in which the time derivative of acceleration, jerk, is kept constant.
- Isokinetic — condition in which muscle tension (force) is kept constant. See isoforce and isotonic; compare with isokinematic.
- Isokinematic — condition in which the velocity of muscle shortening (or lengthening) is constant. (Depending on the given biomechanical conditions, this may or may not coincide with a constant angular speed of a body segment about its articulation.) Compare with isokinetic.
- Isometric — condition in which the length of the muscle remains constant.
- Isotonic — condition in which muscle tension (force) is kept constant — see isoforce. (In the past, this term was occasionally falsely applied to any condition other than isometric.)
- Kinematics — subdivision of dynamics that deals with the motions of bodies, but not the causing forces.
- Kinetics — subdivision of dynamics that deals with forces applied to masses.
- Mechanical advantage — in the context of muscle terms, the lever arm (moment arm, leverage) at which a muscle pulls about a bony articulation.
- Mechanics — branch of Physics that deals with forces applied to bodies and their ensuing motions.
- Moment — product of force and the length of the (perpendicular) lever arm at which it acts. Mechanically equivalent to torque.
- Motor unit — all muscle filaments under the control of one efferent nerve axon.
- Muscle — bundle of fibers that can contract or be lengthened. Striated (skeletal) muscle moves body segments about each other under voluntary control.
- Muscle contraction — result of contractions of motor units distributed through a muscle so that the muscle length is shortened. See contraction.
- Muscle fibers — elements of muscle, containing fibrils which consist of filaments.
- Muscle fibrils — elements of muscle fibers, containing filaments.
- Muscle filaments — muscle fibril elements, especially actin and myosin (polymerized protein molecules) capable of sliding along each other, thus shortening the muscle and, if doing so against resistance, generating tension.
- Muscle force — product of tension within a muscle multiplied with the transmitting muscle cross-section.
- Muscle strength — ability of a muscle to generate and transmit tension in the direction of its fibers. See also body strength.
- Muscle tension — pull within a muscle expressed as force divided by transmitting cross-section.
- Myo — prefix referring to muscle (Greek *mys*, muscle).
- Mys — prefix referring to muscle (Greek *mys*, muscle).
- Proximal — towards the center of the body.
- Repetition — performing the same activity more than once. (One repetition indicates two exertions.)
- Statics — subdivision of mechanics that deals with bodies at rest.
- Strength — see body strength and muscle strength.
- Tension — force divided by the cross-sectional area through which it is transmitted.
- Torque — product of force and the length of the (perpendicular) lever arm at which it acts. Mechanically equivalent to moment.

Definitions are adapted from Kroemer (1999) by permission of the publisher.

REFERENCE

KROEMER, K.H.E., 1999, Human strength evaluation. In W. Karwowski and W.S. Marras (eds) *The Occupational Ergonomics Handbook* (Boca Raton: CRC Press), 205–27.

Musculo-skeletal System

K. H. E. Kroemer

Industrial And Systems Engineering, Virginia Tech, Blacksburg, VA, and Eri, Inc., A Consulting Company, PO Box 3019, Radford, VA 24143-3019, USA

Following Giovanni Alfonso Borelli (1608–79), the human body is modeled as volumes (masses) built on a structure of solid links (the bones) connecting in joints (the articulations with various degrees of freedom); the body segments are powered by muscles that cross joints.

Many body components may be well considered in terms of engineering analogies, for example:

- Articulations — joints and bearing surfaces.
- Blood vessels — tubular transport network.
- Body contours — surfaces of geometric bodies.
- Bones — lever arms, central axes, structural members.
- Flesh — volumes, masses.
- Muscles — motors, dampers or locks.
- Nerves — control and feedback circuits.
- Organs — generators or consumers of energy.
- Tendons — cables transmitting muscle forces.

The human skeleton is normally composed of 206 bones, with associated connective tissues and articulations. The main function of human skeletal bone is to provide an internal framework for the whole body. The long, more or less cylindrical bones that run between body joints are the lever arms at which muscles pull.

Several types of tissues connect parts of the body:

- Muscles are the organs generating force and movement between bone linkages.
- Tendons are strong yet elastic extensions of muscle, connecting it with bones.
- Ligaments connect bones and provide capsules around joints.

Cartilage is visco-elastic flexible material at the ends of the ribs, as discs between the vertebrae, and in general as articulation surfaces at the joints.

The mobility of body joints depends on the shape of the bones at their articulations, the encapsulation of the joint by ligaments, the supply of cartilaginous membranes, and the provision of discs or volar plates, together with the action of muscles.

Some bony joints have no mobility left, such as the seams in the skull of an adult; others have very limited mobility, such as the connections of the ribs to the sternum. Joints with "one degree of freedom" are simple hinge joints, like the elbow or the distal joints of the fingers. Other joints have two degrees of freedom, such as the loosely defined wrist joint where the hand may be bent in flexion and extension, and pivot laterally (deviate). The capability to twist (pronate and supinate) is located in the forearm, not in the wrist. Shoulder and hip joints have three degrees of freedom.

Synovial fluid in a joint facilitates movement of the adjoining bones by providing lubrication. For example, while a person is running, the cartilage in the knee joint can show an increase in thickness of ~10%, brought about in a short time by synovial fluid seeping into it from the underlying bone marrow cavity. Similarly, fluid seeps into the spinal discs (which are composed of fibrous cartilage) when they are not compressed, e.g. while lying down to sleep. This makes them more pliable directly after getting up than during the day, when they are "squeezed out" by the load of body masses and their accelerations. Thus, immediately after a bed rest, one stands taller than after a day's effort.

The term mobility, or flexibility, indicates the extent of angular displacement that can be achieved voluntarily in a body articulation. The actual range depends on training (use), age and gender. It is usually measured as the enclosed angle between the smallest and largest excursions achieved by adjacent body segments about their common joint.

For more information see any book on physiology or kinesiology, or Kroemer *et al.* (1997).

REFERENCE

KROEMER, K.H.E., KROEMER, H.J. and KROEMER-ELBERT, K.E., 1997, *Engineering Physiology* (3rd ed.) (New York: van Nostrand Reinhold–Wiley).

Physical Ability Testing

C. K. Anderson

Advanced Ergonomics, Inc., 5550 LBJ Freeway, #350, Dallas, TX 75240, USA

1. INTRODUCTION

Physical ability testing allows the identification of individuals who will have the physical ability to perform a particular job. This approach is a complement to the traditional ergonomic approach of designing the job to fit the worker. Occasions arise where the traditional ergonomic approach is technically or economically infeasible, at least in the short-term. In those instances, testing may be one of the few options available for achieving the benefits of worker–job matching. As an example, it is possible to reduce the risk of injury by 20–40% and to improve performance through the appropriate use of physical ability testing (Anderson 1998).

Physical abilities that might be considered could include ability to reach a particular distance, manually to handle a specific weight, to maintain a certain energy expenditure level for a given period, to tolerate a particular posture for a prolonged period or repetitively to perform a specific movement, etc. Typically, the focus of a testing program would be on those physical abilities for which a considerable number of individuals have difficulty demonstrating the level required to perform the task. Obviously, there could be a significant difference in capability if one is considering military recruits as opposed to individuals returning to work after back surgery. Hence, when designing a physical ability testing program, it is important to consider the nature of the tasks at issue and the population from which individuals will be drawn. Two basic categories of testing will be considered. The first is the type of testing used to evaluate individuals in the general population for placement on a specific job. This will be of primary focus. The second is associated with testing a specific individual who is temporarily or permanently disabled, which often means that the individual has capabilities that are distinctly different from the overall population.

2. USES

2.1. Placement

From an ergonomics perspective, the most common use for physical ability testing would be in the assignment or placement of individuals relative to alternative tasks. The basic idea would be to compare an individual's specific physical abilities to the specific physical job requirements.

2.1.1. Validity

Typically, one or both of two questions are being addressed: "does this individual have the physical ability to perform this task?" and "does this individual have a higher risk of injury for themselves or others if performing this task than others would?" In the past, some employers have made decisions about people's ability to perform the job or risk of injury simply on the basis the employer's beliefs or stereotypes about the physical abilities of women or older individuals, for instance. In general, it is true that, for instance, females have ~60% of the strength of males and 70% of the aerobic capacity (a measure of ability to sustain particular levels of energy expenditure for extended periods). It is also true that the average aerobic capacity of a 40-year-old male is ~70% of the aerobic capacity of a 20-year-old male. These facts do not warrant the blanket exclusion of women or individuals > 40 years of age from physically demanding jobs. As a side note, a number of the pieces of the employment-related legislation in the USA (e.g. the Americans With Disabilities Act, the Age Discrimination in Employment Act, and Title VII of the Civil Rights Act) have come about in reaction to employers who were making placement decisions on the basis of unfounded stereotypes regarding the capabilities of females, older individuals or those with disabilities. The more appropriate approach, which is the approach advocated in these pieces of legislation, is individually to assess each candidate and base a decision on the physical ability s/he has demonstrated relative to the job's physical requirements. It is important to rely on tests that are accurate indicators of the individual's ability to perform the job and/or risk of injury to self or others. The accuracy of a test in predicting these outcomes is an indication of its validity.

In many countries it is illegal to make employment decisions on the basis of results from test batteries that have not been demonstrated to be valid, particularly if groups protected from employment discrimination, such as females, older individuals or individuals with disabilities are more adversely impacted than their counterparts. Adverse impact is generally assessed by determining whether the pass rate for the protected group, such as females, is lower than the rate for their unprotected counterpart, which in this case would be males. Given the facts stated above regarding the relative strength and endurance of males and females, for instance, it is common to find adverse impact with physical ability test batteries. Said another way, if the task demands are at least moderate, it can be expected that the pass rate for females, and perhaps older individuals, will be lower. Hence, it becomes critical in many countries, for at least legal reasons, to take the necessary steps to validate the test battery before it is implemented. Even if there is not a legal imperative for validating a test battery, the information is valuable to the employer in assessing the cost-effectiveness of the test battery. Obviously, the employer would want to invest in a battery for which the cost to implement can be demonstrated to be heavily outweighed by the benefits received in terms of reduced turnover, injuries, etc.

The evidence needed to demonstrate the validity of the test battery differs substantially depending on the question being considered (Miner and Miner 1979). The strongest form of validation evidence is a statistical study that demonstrates a significant relationship between test scores and indicators of job performance such as productivity, turnover, supervisor evaluation or injury experience. Studies of this sort can be difficult to perform for technical or economic reasons. An alternative is to evaluate "content validity" of a battery. This basically involves assessing whether the content of the job is reflected by the content of the tests, and whether the test cutoffs represent the job requirements. Any evidence from statistical studies of a battery's validity for jobs with similar demands can be useful in determining the content validity of a test battery for a particular job.

2.1.2. Job analysis

There are two key elements essential when developing a battery that will have a high degree of validity. The first is a thorough ergonomic job analysis; the second is the utilization of tests that have a high degree of relation to the tasks that are most physically demanding. One would expect that the most demanding tasks would be the ones that limit the ability of some people to successfully perform the job. Success in identifying physical ability tests that are highly job-related is contingent, in large part, on the accuracy of the ergonomic job analysis.

The first step in the job analysis is to identify those tasks that are physically difficult to perform for a substantial portion of the population from which one would draw. It is important to underscore the point that the labor pool may be larger than what is reflected by the composition of the current workforce, particularly if there has been a hiring bias regarding the characteristics of individuals, such as size, gender, age or disability. For instance, police departments had traditionally hired large males. An analysis of tasks that would be difficult for a population of large males may overlook tasks that would be difficult for a population of smaller females, which is a group these departments are being encouraged additionally to consider for employment.

Tasks that would be of particular concern are heavy lifts, pushes, pulls and carries. Particular postures or movements can also be of concern, such as going through small openings or areas with tight clearances. In some cases, the performance-limiting factor is the overall stamina or endurance required to perform all of the activities involved over the course of a shift. Any one task, such as lifting 1 kg, may not be particularly demanding, but the combined effect of high repetition or large amounts of body movement such as walking or climbing, may yield a high-energy expenditure requirement over the course of a shift. A second factor to consider in the assessment of the energy expenditure requirement is the period over which the energy expenditure must be maintained (Rodgers and Yates 1990).

The second step in the job analysis is to evaluate the degree to which the physically demanding tasks identified in the first step are essential to satisfactory performance of the job. Considerations in this assessment might be the frequency with which the task is performed, the degree to which assistance might be available (such as with a heavy lift), the opportunity to assign the task to someone in another job classification or whether the physical demands of the task can be reduced or eliminated through redesign of the work organization or process.

2.1.3. Test battery design

Once the essential functions that have high physical demand have been identified, it is possible to move into the phase of designing the test battery. There are three key issues to consider in identifying the tests that will be used in the battery. The first issue is the degree to which the test reflects the task in question. The ideal test is the job task itself, though this may be difficult to incorporate into a test battery. For instance, a test of ability to lift 20 kg from floor level to table height could actually be the demonstration of lifting a case from floor level to table height. The safety of the test could be enhanced by starting with a negligible weight and increasing the weight up to 20 kg on later repetitions. In contrast, for some jobs the main question is whether the individual can maintain a given energy expenditure level for the entire shift day after day. One way to assess an individual's ability to do this would be to have them perform the job for a number of days, but this could be rather impractical. An alternative is assess the individual's aerobic capacity, which can be accomplished in < 15 min, and use that to determine the individual's likelihood of being able to perform the job on an ongoing basis. The greater the similarity between the task and the test, the easier it is to validify the utility of the test in assessing an individual's ability to perform the job.

The second consideration is the safety of the test itself. Clearly, it is not desirable to have individuals perform a test that may put them at increased risk of injury. For instance, in the lift-test example above, there would be no reason to have test participants demonstrate the ability to lift > 20 kg on the test, since the job itself would not require the ability to lift > 20 kg. As a general statement, individuals should only be required to demonstrate what the job requires, as opposed to the maximum ability that they have, since the risk of injury typically increases with the degree of exertion.

A third consideration is the objectivity of the test. It is useful to incorporate tests that yield quantified results that are repeatable from test administrator to test administrator as well as stable from day to day. Tests that involve subjective evaluations on the part of the test administrator, such as an assessment of lifting technique, can be unduly influenced by the test administrator's personal judgment of proper behavior, which may or may not have any real bearing on ability to perform the job.

2.2. Norm–based Comparisons

It needs to be briefly mentioned that physical ability tests are sometimes used to compare an individual to particular populations, such as "healthy" individuals or individuals with similar health conditions. This application of physical ability testing can yield very different types of tests since the emphasis is not on determining whether an individual can perform particular job-related tasks. Tests, such as range-of-motion in the low back, may fall in this category. The test can be useful as a measure of impairment, but its use for placement purposes can be questionable since one's low-back range-of-motion may not have a lot of bearing on the ability to perform a particular task. For instance, it may be possible to perform the task even with a limited range-of-motion if an alternative posture is used. The critical point to be made here is that care must be taken when considering the application of tests designed for one purpose, such as impairment rating, for another purpose, such as job placement.

3. SUMMARY

Physical ability testing can be a useful injury control strategy in conjunction with other ergonomic interventions. Research indicates that testing has the potential to reduce injuries to new-hires by 20–40%. The effectiveness of a testing program is a function of the physical difficulty of the job and the degree to which the tests are able accurately to assess an individual's ability to meet or exceed those physical requirements. Critical steps in the implementation process are the job analysis and test battery design.

REFERENCES

ANDERSON, C.K., 1999, Physical ability testing for employment decision purposes. In Karwowski, W. and Marras, W. S. (eds) *The Occupational Ergonomics Handbook* (Boca Raton: CRC Press). pp. 1291–1298.

MINER, M.G. and MINER, J.B., 1979, *Employee Selection Within The Law* (Washington, DC: Bureau of National Affairs).

RODGERS, S.H. and YATES, J.W., 1990, Physiological basis for manual lifting guidelines. In *Scientific Support Documentation for the Revised 1991 NIOSH Lifting Equation: Technical Contract Reports* (Cincinnati: National Institute for Occupational Safety and Health).

Physical Strength in the Older Population

S. N. Imrhan

Engineering Department, University of Texas at Arlington, USA

1. INTRODUCTION

The ratio and number of people over 60 years of age (older people) in the technologically developed world has been steadily increasing. In the USA, the population over 60 years of age increased by 20.5% from 1980 to 1989 to reach a total number of 31 million (US Bureau of Census 1991), and the trend is expected to continue. The favorable conditions producing such a trend — enhanced medical care and consequent longer lifespans — are expected to propel the percentage further. A concomitant overall population growth is not expected and so older people are expected to work longer into their lifespans to offset a relative decrease in the population of younger workers. There is, therefore, an increasing need to design work, equipment, consumer products, and the environment, in general, to accommodate older people.

2. FUNCTIONAL CAPACITIES IN OLDER PEOPLE

As people become older, beyond adulthood, many physical and mental functions tend to decline. Among the most important are sensory perception, psychomotor skills, memory and learning, balance, cardio-respiratory strength and endurance, and muscular strength and endurance. While the exact causes and rates of decline of these functions are debatable, both objective scientific measurements and personal observations indicate clearly that the older population, in general, does not perform many types of tasks as well as younger cohorts. Tasks that require great physical resources, heavy sensory load, a high level of manual dexterity, or rapid responses tend to decline with increasing age in older workers. In the industrial workplace, for example, they require more intense illumination levels than young adults to read and cannot lift loads as heavy as those lifted by their younger cohorts. However, some types of performances tend to improve with age — especially those that require a high level of skill, experience or decision-making.

Few people would doubt that, except for special designs targeted at special sub-populations, many tasks and equipment for work or daily living were designed to benefit younger adults more than any other age group. The arguable point is whether it was done consciously or not. A widely held belief among ergonomists and other work design professionals is that the older population would be able to perform tasks at work and at home significantly better if, at the design stage of those tasks and associated equipment, the reduced functional capacities of the older population were taken into account. This chapter discusses muscular strength in older populations and how task design and performance are affected by strength.

3. MUSCULAR STRENGTH AND AGING

Scientific data confirms and quantifies what is known from personal observations and anecdotal evidence — that muscular strength in people declines gradually after adulthood and not suddenly at any particular age. People seem to be strongest at about 30–35 years of age and maintain most of that strength into the late forties as long as they are physically active and free of diseases and other debilitating ailments. The decline in strength seems to be accelerated by lack of physical activity and is common soon after people retire from their jobs and assume a relatively inactive lifestyle. The main physiological cause of strength decline is (gradual) atrophication and loss of some of the fibers that make up the muscles. Research has shown that between ages of 30 and 80 years the loss may be as much as 30% with greater losses in the less active muscle groups. Loss in muscle size is also accompanied by a decline in muscular movements and coordination which affects body balance and tone, and speed of contraction.

Muscle strength in older people is used for the same purpose as in any other age group. A certain minimum strength is necessary for moving the whole body or a body segment from one place to another, maintaining body posture, or effecting small movements for changing posture. Greater muscular strength is required for moving or stabilizing objects during manual materials handling. Manual materials handling tasks comprise mostly lifting, lowering, pushing, pulling, carrying and holding objects. Pressing down on objects, gripping and squeezing with the hands (handgrip) or the fingers (pinching), and turning or twisting objects with the hands may also require considerable strength, especially for job tasks. A comprehensive database and profound knowledge of strengths in older people are, therefore, required to assess how well this segment of the population may be able to perform existing physical tasks, design new tasks, or redesign existing tasks to conform to their strength levels. Unfortunately, because the working population has relatively few people beyond 60 years of age, research related to work tasks has traditionally been concentrated in adults below 50 years of age. Research in strength capacities in older people has also been avoided because of their greater susceptibility to physical injuries. The result is that few data sets and empirical models of strength exist for this segment of the population for use in the design of work tasks, equipment, living environments, products for activities of daily living, or products for leisure activities.

Many of the available strength data sets (mostly small ones) on older people were gathered from clinical studies designed to explore basic theories of aging, rather than from ergonomic studies on work capabilities. They have been concentrated mainly on strengths of single joints or basic easy-to-measure strengths, such as elbow flexion/extension, knee flexion/extension, pinch, handgrip, etc., and do not relate to work task variables. The data are, therefore, almost useless to ergonomists and other work designers.

4. HANDGRIP STRENGTH

Hand gripping is a precursor to many tasks, activities of daily living, or leisure activities. In the workplace, great or sustained gripping forces may be required to operate hand tools and the older population may be at a disadvantage in those tasks. In home and leisure activities, tasks requiring grip strength, which may not normally present problems for the adult population, may prove to be stressful for older people. Opening medicine bottle caps and food jars and tearing plastic wrappers are examples of such tasks. Unfortunately, there are no published data for grip

strength in the older population on objects (e.g. jar lids) associated with these tasks, but grip strength from standardized tests using dynamometers are available. These tests have shown mean strengths (per sample of subjects) of 294–352 Newtons (N) in males and 113–204 N in females among 60–80 years olds (9.81 Newtons (N) = 1 kilogram force); in this age group, grip strength decreases steadily with age. These strengths are 60–80% as great as the corresponding strengths of younger adults 20–40 years old.

Grip strength in both older people and adults follows an increasing–decreasing trend over a (sufficiently wide) range of grip widths — weaker strength for small or large grip widths and the greatest strength for intermediate widths. However, there is some evidence that the trend is more pronounced in the older people than in younger adults. The reason is not clear but it seems that the decrease in biomechanical leverage in the closed or widely opened hand (small or large width) has a greater detrimental effect on hand muscle contraction in older people.

When gripping, while turning or twisting an object with maximal strength (as in exerting torque on a circular handle), the grip force may not necessarily be maximal. Data for such forces are rare but one set of published data (Imrhan and Loo 1989) indicate that the grip force varies little at small to medium size grips (31–74 mm diameter) and then falls rapidly beyond 74 mm.

5. PINCHING AND PULLING WITH THE FINGERS

As in other human age groups (adults and children) pinch strength in older people depends on the type of pinch. The order of magnitude for the different types is same as for adults and children — that is, (i) lateral, (ii) chuck, (iii) pulp-2 and pulp-3, (iv) pulp-4 and (v) pulp-5 in decreasing order of magnitude, as measured on a standard pinch gauge. The difference between the strengths of the lateral pinch (thumb pressing on the lateral aspect of the index finger, as when using a flat key) and chuck pinch (thumb versus index and middle fingers) is small. The weakest pinch (pulp-5, thumb versus little finger) is about one-quarter as strong as the strongest pinch (lateral) which is about one-quarter as strong as handgrip; the strongest pulp pinch (pulp-2 or pulp-3: thumb versus index or middle finger) is two and a half times as strong as the weakest pulp pinch but only two-thirds as strong as the lateral pinch; and ring finger (pulp-3) pinch is about two-thirds as strong as lateral pinch. The relative magnitudes among the various pinch and handgrip strengths in older people are about the same as in any other age group. It is important to note, however, that these relative magnitudes were derived from standardized tests in which pinching is executed in a restrictive manner unlike the way in which it is executed in many work tasks. Recent research has shown that the chuck pinch is significantly stronger than the lateral pinch under pinching conditions that are different from those in standardized tests, and that the lateral pinch becomes weak for great pinch widths (for thick objects) — beyond 5.6 cm. Standardized pinch tests have been performed at pinch widths of 1.6–1.8 cm.

6. PINCHING AND PULLING

Pulling while pinching with the fingers is necessary for opening some types of packages such as plastic wrappers and glass bottles with flexible metal caps. While pinch-pull strength data would

be useful for the design of such products, there is no empirical data for older people to guide designers. However, data for adults are available and if it is assumed that older people are only about 0.63 times as strong as adults, on average, they (older people) would be expected to pull maximally with a force of about 63 N with the lateral grip, 41 N with the chuck pinch and 28 N with the pulp-2 pinch.

7. PUSHING AND PULLING WITH THE HANDS

On average, older people can be expected to pull toward the body with one hand with a force of about 71 N with the fingers in a hook grip (as when holding a briefcase) on a handle. When using the power grip (fingers wrapped around the handle) the increase in strength is insignificant. However, research has shown that strength depends on the mode of living. People living independently have been found to be 1.4 times as strong as those who have assistance with daily tasks (91.6 N versus 64.9 N). These forces, however, are not much greater than the forces required to perform certain common activities of daily living, such as opening refrigerator doors (32.5 N) and oven doors (36.9 N). This implies that these two tasks would be difficult for some older people, especially when performed frequently.

Tests by this author have shown that the pull forces required to open doors in public buildings average about 64 N. This requires an enormous effort for most older people.

8. GRIPPING, TURNING AND TWISTING

Torquing strength of the hands for opening or closing and tightening circular lids (discs) have been investigated in older people. This kind of strength is important for designing containers (food jars and medicine bottle caps, for example) which are difficult for small children to open yet easy enough for older people. Test results have shown that older males can produce maximal torques on circular jar lids in the range 1.5–7.9 Nm depending on lid diameter, and females 0.9–4.4 Nm. Also, older people are significantly stronger than children in respect of jar lids with diameters above 64 mm than below 48 mm. This implies that child-resistant lids should preferably be above 64 mm in diameter rather than below 48 mm. Other important findings about torquing strength in older people are: (i) torque strength on circular (jar) lids increases with diameter up to a point; (ii) the trend is linear up to about 74 mm diameter for both rough and smooth-surface lids; but (iii), above 74 mm the trend is monotonic increasing (slower increase) for rough lids, and increasing—decreasing with a maximum at about 83 mm for smooth lids; and (iv) roughness on the grip surface of commercial jar lids does not enhance torque generation with the hands.

There is also some torque strength data for opening faucets. Properly designs faucet handles can produce three times as much manual torque as poorly designed ones. For the best handles the strengths fall in the range 5.7–10.0 Nm, and for the worst ones 2.7–3.7 Nm, the exact amount depending on hand laterality, direction of torque application and body posture. The best commercial handles for torque exertion are the paddle types with a significant lever arm for the fingers, and the worst are spheroid-shaped ones with ridges. By comparison, torques required to close a sample of faucets were found to be in the range 2.0–2.2 Nm, indicating that certain handle designs were biomechanically stressful on the hands.

9. LIFTING, LOWERING AND CARRYING

Research indicates that lifting, lowering, and carrying tasks in industry are among the most strenuous physical tasks. Injuries and cumulative strains to the back are common, and are costly and debilitating to the industrial population. Older people are less likely to be assigned to these stressful tasks. However, if tasks and equipment are designed to accommodate their (reduced) strength capacities they would be more useful and employable. Activities of daily living also require lifting, lowering, and carrying, and can be made safe and achievable only if the objects to be lifted are designed within their strength capabilities. Unfortunately, empirical data to guide designers of these jobs are almost non-existent for people beyond 60 years of age. The only practical solution is to use subjective estimates of lifting, lowering or carrying strength with extreme caution. Kovar and La Croix (1987), in a questionnaire study for the National Center for Health Statistics, found that 36% of women and 18% of men aged 60–74 years had difficulty in lifting or carrying 25 pounds; and 11% of women and 6% of men had difficulties with 10 pounds. This information underlies the importance of exercising care in designing objects that must be manually moved or tasks requiring the application of muscular forces, and the urgent need for empirical task-related data on muscular strengths.

For manual material handling applications (lift, lower, hold, carry, push and pull) maximal single effort (MVC) strength is not typically used for task or equipment design. Instead, the maximal acceptable loads (or force) people are willing to exert during a work shift is used. This is determined psychophysically while people perform a simulated task. It is much less than MVC strength — typically 25–35% MVC. The psychophysical method for determining maximal acceptable work loads is strenuous and it is unlikely that such strength data will be available in large sets from the US population in the near future.

10. CONCLUSIONS

Strength data in older people are needed for designing tasks and equipment, placing workers in tasks, pre-employment screening, predicting or assessing task performance, and evaluating functional impairment. It is doubtful whether there is any biomechanical or statistical model to predict strengths in the older population accurately enough for ergonomic applications. Also, very little empirical (tabulated) data exists for these purposes. Almost all of the data on task-related strengths in older people are for static exertions — that is, the object or body part does not move from the muscular contractions — and are for manual performance only. Many work activities are, however, dynamic in nature and involve the whole body. More research is therefore needed to generate dynamic strength data simulating task performance.

REFERENCES

FALETTI, M.V., CLARKE, M.C., GREENE, R. and BASTACKY, S., 1983, *Human Factors analysis of elderly impairment*. NIA (NIH) grant #RD1-AG-02727.

IMRHAN, S.N., 1994, Muscular strength in the elderly: implications for ergonomic design. International Journal of Industrial Ergonomics, 13, 125–38.

IMRHAN, S.N. and LOO, C.H., 1989, Trends in finger pinch strength in children, adults, and the elderly. *Human Factors*, 31(6), 689–701.

KOVAR, M.G. and La CROIX, A.Z., 1987, *Aging in the eighties, ability to perform work-related activities*. Vital Health Statistics, US Department of Public Health Services, Washington, DC.

ROHLES, F.H., MOLDRUP, K.L. and LAVIANA, J.E., 1983, *Opening Jars: An Anthrompometric Study of the Wrist-twisting Strength of Children and the Elderly*. Report no. 83-03, Institute for Environmental Research, Kansas State University.

Physical Work Capacity (PWC)

J. L. Smith

Department of Industrial Engineering, Texas Tech University, Lubbock, TX 79409-3061, USA

Physical work capacity (PWC) implies the determination of a physical limit of performance. However, the term is used in ergonomics to refer to a physiological limit of human capabilities (as opposed to biomechanical limitations). Thus, PWC is defined as the maximum level of physiological exertion that can be achieved by an individual. The measurement of PWC is generally given in terms of oxygen consumption (e.g. liters of oxygen consumed per min). PWC is used synonymously with aerobic capacity, or VO_{2max}.

PWC depends on several factors (figure 1). The three examples of each of the factors identified are not meant to be exhaustive, but rather are shown to indicate the potential level of complexity in determining an individual's PWC. In the assessment of PWC for an individual, one must determine which of the factors shown in figure 1 are appropriate and should be accounted for in arriving at a PWC value for that individual. The work environment will determine many of the factors that must be considered.

PWC can be assessed in a number of different ways, depending on the accuracy required and the level of risk willing to be assumed. The most common definition of PWC is the point at which an increase in workload does not produce a corresponding increase in oxygen consumption or heart rate. Such a definition implies that the physical task demands must be continually increased until the worker is no longer showing an increased physiological response to work, which indicates that he or she has reached his/her capacity or VO_{2max}. Taking an individual to their physiological capacity is normally done during a stress test in a clinical environment under the direction of a physician. However, in the industrial environment, it is generally not felt that the benefits of obtaining more accurate capacity data outweigh the risks of subjecting an older, less well-trained population of workers to maximal testing. Therefore, a variety of submaximal techniques have been developed over the years to predict PWC.

Most submaximal techniques to assess PWC involve the worker performing a standardized test (e.g. riding a cycle ergometer, or walking on a treadmill) at a constant workload
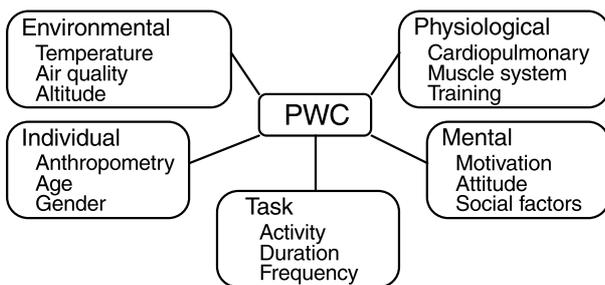
until steady-state physiological conditions have been established. As shown in figure 2, physiological responses to changes in work demand are not instantaneous. As the workload is increased, a new set of steady-state physiological responses, notably heart rate (HR) and oxygen consumption (VO_2), will be reached. In most cases, a level of steady-state physiological response will be reached in 3–4 min after changing a workload. This process could be continued until the individual's capacity (HR_{max} and VO_{2max}) is reached. However, if one assumes a linear relationship between HR and VO_2, one can use linear regression to estimate VO_{2max} from a set of submaximal workloads. A common technique is to find three steady-state HR and VO_2 combinations and determine the least-squares linear relationship for the data. Then, based on an estimate of maximum heart rate, VO_{2max} (PWC) can be estimated for that individual. A commonly used, simple estimate for maximum heart rate is:

$$HR_{max} = 220 - Age. \tag{1}$$

For example, assume that a 37-year-old female worker is tested on a cycle ergometer at three workloads of 75, 100 and 125 watts. Her heart rate and oxygen consumption are monitored during the test and recorded for the three steady-state levels of work:

Workload (watts)	Heart rate (bpm)	VO_2 (l min^{-1})
75	121	1.2
100	136	1.5
125	151	1.7

Calculating a linear regression line for the three steady-state data pairs above (HR, VO_2), the regression equation

$$VO_2 = 0.01667 \cdot HR - 0.8$$

is obtained. The worker's maximum heart rate is estimated as $220 - 37 = 183$. Substituting 183 into the regression equation for the HR, VO_{2max} is estimated as 2.25 liters min^{-1}.

Using the submaximal regression technique described above requires that oxygen consumption and heart rate be measured. Heart rate is easily measured through palpation or by using a relatively inexpensive commercially available heart rate monitor. However, oxygen consumption measurement requires more sophisticated and costly equipment. Over the years, a number of researchers have established models to predict oxygen consumption from measurements that do not require oxygen consumption data. Perhaps the most famous of the VO_2 or VO_{2max} prediction models is the nomogram developed by Astrand and Rodahl (1986). The nomogram allows the user to predict VO_{2max} from a variety of input data. VO_{2max} can be predicted from a single steady-state condition using HR and oxygen consumption.

Figure 1. Factors affecting physical work capacity
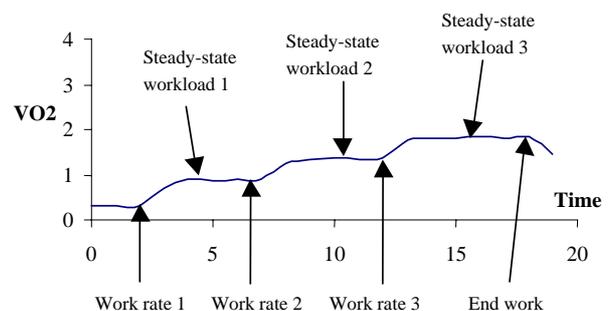
Figure 2. Physiological response to work

However, VO_{2max} can also be predicted without requiring the use of oxygen consumption data. The Astrand nomogram does require HR and one other parameter to estimate VO_{2max}. The other parameter could be work rate in watts from a standard source such as a cycle ergometer, or body weight if using a step test as the work source. The VO_{2max} obtained from the Astrand nomogram must be adjusted for age. Astrand and Rodahl (1986) states that a 15% standard error exists for using the nomogram for moderately trained individuals of different ages.

Another alternative to the Astrand nomogram is to utilize a step test that is less strenuous than that recommended by Astrand and Rodahl (1986). Such step tests can be based on a model developed by the American College of Sports Medicine (ACSM) (1991). The ACSM model is based on step height and stepping frequency and can be expressed as:

$$VO_2 = 0.35 \cdot f + 2.395 \cdot f \cdot h, \tag{2}$$

where f is the frequency of stepping in steps·min^{-1}, h is height of the step (m) and VO_2 is steady-state oxygen consumption (ml·kg^{-1}·min^{-1})

Thus, for a step height of 0.25 m and a stepping frequency of 20 steps·min^{-1}, the predicted oxygen consumption for that task is 12.675 ml·kg^{-1}·min^{-1}. Multiplying that by the individual's body weight (kg) and dividing the result by 1000 ml·l^{-1} gives the predicted oxygen consumption (liters·min^{-1}). That VO_2 along with the steady-state HR at the given workload can be utilized in the Astrand nomogram to predict the individual's VO_{2max}. Again, an age correction adjustment must be made.

Once an individual's PWC is determined, it can be compared with the physiological demands of the job to determine whether the worker is being overworked and needs additional rest breaks. The most widely accepted standard is that the work should require no more than 33% of an individual's VO_{2max} without additional rest breaks.

REFERENCES

AMERICAN COLLEGE OF SPORTS MEDICINE, 1991, *Guidelines for Exercise Testing and Prescription* (London: Lea & Fèbiger).

ASTRAND, P. and RODAHL, K., 1986, *Textbook of Work Physiology: Physiological Bases of Exercise* (New York: McGraw-Hill).

KROEMER, K.H.E. and GRANDJEAN, E., 1997, *Fitting the Task to the Human* (5th ed.) (London: Taylor & Francis).

TAYYARI, F. and SMITH, J., 1997, *Occupational Ergonomics: Principles and Applications* (London: Chapman & Hall).

Postural Adaptation

R. S. Bridger and R. S. Whistance

Department of Biomedical Engineering, University of Cape Town Medical School, and Groote Schuur Hospital, Observatory 7925, South Africa

1. INTRODUCTION

Human beings are the only remaining bipedal hominids to walk the surface of the Earth. The upright posture and bipedal gait that characterize modern man have a long evolutionary history. In Laetoli, East Africa there is a set of fossilized footprints that it is thought to have been made 3.7 million years ago by early hominids known as *Australopithecus afarensis*. Reconstruction of the walking patterns from the footprints, using data on modern human gait, indicates that *Australopithecus* was fully bipedal. According to anthropologists such as Lovejoy (1988), *Australopithecus* had a cranial capacity similar to that of a modern chimpanzee and weighed ~30 kg, but its pelvis, femur and lumbar spine resembled those of modern man. Both the human and the australopithecine bones are very different from those of a chimpanzee. The former are adapted to enable the animal to walk on two legs rather than four and to stand upright with the weight of the upper body balanced on top of the bones of the spine, pelvis and femur.

Clearly, humans are well adapted to an upright, bipedal posture. If this has enabled them to develop for > 3.5 million years, why is it that postural faults and posturally related musculoskeletal pain are so common, particularly in the workplace?

2. FUNDAMENTALS OF THE STANDING POSTURE

2.1 Bones and Joints

The upright posture of bipedal hominids is illustrated in Figure 1. It is a general solution to the problem of rotating the trunk with respect to the ground from a parallel (quadrupedal) position to a perpendicular position in an energy-efficient way.

In upright standing, the pelvis is held in an anteriorly tilted position by the iliopsoas muscles and the hip joint is free to extend, as happens during the stance phase of gait when the superincumbent body parts move ahead of the planted foot. The trunk and head are rotated until they are vertically above the legs. This is achieved by extension of the lumbar and cervical spines. This is why the vertically held spine is s-shaped in humans whereas in quadrupeds it is c-shaped and held horizontally to the ground.

In the erect posture, the line of gravity of superincumbent body parts passes through the lumbar, sacral and hip joints and in front of the knee and ankle joints. This places an extension torque around the knee joint, which is resisted because the joint is already fully extended. The flexion torque around the ankle is resisted by the plantar flexors.

2.2 Muscles and Ligaments

A person standing erect under the influence of gravity is never in a state of passive equilibrium. The body can be conceived of as a pillar of segments stacked one on top of the other and linked by
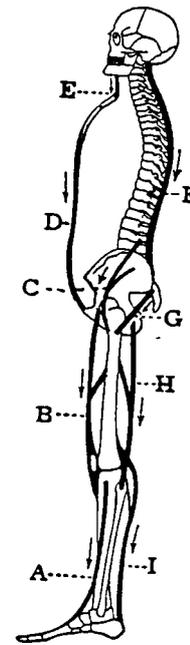
Figure 1. Upright posture in man. A, Tibilalis anterior; B, quadriceps femoris; C, iliopsoas; D, abdominals; E, neck flexors; f, spinal extensors; G, gluteus maximus; H, hamstrings; I, soleus gastrocnemius. (From Roaf 1997. Reproduced woth kind permission.)

joints. The pillar is momentarily balanced when the resultant effect of all forces acting on it is zero. The system is designed to minimize any displacement of the line of action beyond the base of support described by the position of the feet. Compensatory mechanisms come into play to maintain balance immediately when this happens.

Muscles and ligaments play a stabilizing role by means of the active and passive torques they exert around joints to correct small, fleeting displacements of the lines of action away from the joints. A "good" posture may be defined as one in which destabilizing moments are minimized and the posture is maintained by the resistance of the relatively incompressible bones (as well as interleaved soft tissues such as the intervertebral discs).

When the body is pulled "off-balance" by the requirements of badly designed jobs or workspaces, the anti-gravity muscles come into play and a new equilibrium position is established, but with the associated cost of isometric muscle activity.

2.2.1 The erector spinae muscles

These are the main extensors of the trunk; they also control flexion. In relaxed standing very little muscle activity occurs since the lumbar lordosis minimizes the trunk flexion moment. When the trunk is flexed even slightly forwards or when a weight is held in front of the body, the erector spinae muscles come into play. Work situations that set up static loading of these muscles include:

1. Working with the hands and arms held away from the body.
2. Holding a tool or a weight.
3. Standing with the trunk flexed to reach for work objects placed too far away or made inaccessible due to a lack of foot space.

Fatigue of the back muscles may be experienced as low back pain. Recent electromyographic research suggests that people with chronic pain in the low back have back muscles that fatigue more quickly than those of people without pain.

2.2.2 The leg muscles

The soleus and gastrocnemius muscles are true postural muscles in the sense that they are always "switched on" when standing. When leaning forward, the activity of the gastrocnemius muscle increases. Prolonged standing causes significant localized leg muscle fatigue and is one of the causes of leg discomfort. Research indicates that "anti-fatigue" mats for standing on do not delay the onset of leg muscle fatigue, although they do reduce discomfort in the lower legs, feet and back. This is consistent with the view that the mats work by cushioning the feet and by reducing the impact forces that are passed up the kinetic chain when workers move. Mats would not be expected to lower the postural load on the gastrocnemius muscle.

If workers can place one foot on a footrest, this reduces pressure on the tissues of the foot and leg and allows the leg muscles to relax. It releases the ipsilateral iliopsoas muscle and helps prevent excessive lordosis (by up to 15°). Essentially, footrests for standing workers encourage them to stand on one leg and to increase the load on the supporting leg. Such devices should be designed to allow workers to alternate their posture and in this way to change the static leg load into a periodic regime of load and rest.

2.2.3 The abdominal muscles

There is very little abdominal muscle activity in standing. These muscles may help to maintain a proper relationship between the thorax and pelvis by preventing excessive anterior pelvic tilt and hyperlordosis. The abdominals certainly play a role in manual handling by contracting to increase the intra-abdominal pressure which is thought to exert an extension moment (although this is controversial) and may help stiffen the joints of the lumbar spine. The abdominals can prevent trunk extension, caused, for example, by loads placed high on the back (or when putting on a backpack, for example) or when walking down steep hills.

2.2.4 The hamstring and gluteal muscles

The hamstring and gluteal muscles are hip extensors. The gluteal muscles exhibit hypertrophy in man and their function is to stop the trunk from "jack-knifing" forwards over the legs — unlike in quadrupeds where the trunk is already jack-knifed and the gluteals are used for locomotion. The gluteals are, however, used for locomotion in ladder or stair climbing. Activity in the hamstrings is slight in the standing position but increases when the stander leans forward, holds a weight or pulls.

2.2.5 The iliopsoas muscles

Psoas major and iliacus are hip flexors and are constantly active in normal standing as they prevent extension of the hip joint (the trunk jack-knifing backwards over the legs or loss of lumbar lordosis if the head position is maintained). The iliopsoas muscles act against the hip extensors. Factors that upset the balance between these two muscle groups cause abnormal lumbar curvatures. If the iliopsoas weaken relative to the hip extensors, the pelvis rotates posteriorly over the hip joint and the lumbar curve is flattened; if the iliopsoas strengthen with respect to the

hip extensors, the lumbar curve is exaggerated. This is discussed further in the section on sitting posture.

2.2.6 The adductors and abductors of the hip

When standing on two feet, these muscles provide lateral stability, preventing translation of the pelvis in the frontal plane. When standing on one foot (and also during the stance phase of gait), the pelvis tends to tilt in the direction of the unsupported side. The hip abductors on the side of the supporting leg contract to keep the pelvis level.

3. PHYSIOLOGY OF STANDING

The increase in energy expenditure when a person changes from a supine to a standing position is only ~8%. However, erect standing imposes a hydrostatic handicap that makes people liable to peripheral circulatory collapse. Peak plantar (foot) pressures of 137 kPa exceed the normal systolic pressure of 17 kPa, resulting in occlusion of blood flow through the foot. Venous and circulatory insufficiencies in the lower limbs also contribute to the discomfort that results from prolonged standing. It has been shown that venous reflux is more common in symptom-free surgeons (who stand for long periods) than in a comparison group who experience discomfort in standing.

Prolonged standing causes physiological changes including peripheral pooling of blood due to the influence of gravity, a decreased stroke volume and an increased heart rate, diastolic and mean arterial pressure, peripheral resistance and thoracic impedance. Lying supine brings an increase in whole-body electrical impedance, blood and plasma volume, and lower serum potassium levels. Standing-up from a supine position is accompanied by an increase in the dimensions of the nasal passages.

Constrained standing for long periods causes venous pooling and is particularly troublesome for older workers or for those with peripheral vascular disease or varicose veins. Poor health, obesity, tobacco smoking and any other factors that increase circulatory inefficiency magnify these effects. The "venous muscle pump", which returns blood to the heart, is activated by moving and by postural sway. Fidgeting is a preconscious defense against the postural stresses of constrained standing or sitting. Its purpose is to redistribute and relieve loading on bones and soft tissues and to rest muscles.

4. POSTURAL BEHAVIOR

Much postural behavior and many of our attitudes towards posture are culturally determined. Cross-cultural surveys do enable a few fundamental conclusions to be drawn. First, humans are capable of adopting > 1000 comfortable resting positions but, when unconstrained, they never adopt any posture for any length of time. Postural behavior is, therefore, a process of continual adaptation to changing internal and external circumstances. Second, although humans are adapted to stand on two legs, they are not adapted to stand still. In many cultures, people who have to stand for long periods (such as shepherds, sentries, etc.) use standing aids such as poles, "shooting sticks" or chinstraps, or prop themselves up against external objects. "Natural" standing tends to be tripedal rather than bipedal with the third leg providing extra stability and enabling the weight to be periodically taken off the other legs.

Constrained standing without the use of standing aids or sufficient rest squashes the intervertebral discs, reducing the vertical distance between adjacent vertebrae and increasing the load on the facet joints. It may cause inflammation of the tendon sheaths and chronic degeneration of joints in the form of arthritis.

Standing may be symmetrical or asymmetrical. Asymmetrical standing occurs four times as often as symmetrical standing. Body weight is supported almost entirely on one leg while the other helps maintain balance with its knee slightly flexed and its foot anterolateral to the supporting leg. The pelvis tilts towards the resting leg and the lumbar spine flexes laterally and is concave to the supporting limb. Asymmetrical standing seems a postural strategy aimed at reducing static loading of the spine, pelvis and lower limbs.

The body can be thought of as an open chain of linked segments that, in the absence of external stabilization, is maintained in position by tonic or postural muscle activity. People constantly seek to turn the open chain into a closed one by means of postural strategies such as folding the arms, crossing the legs or cradling the chin in the hand. These behaviors, far from being maladaptive, increase the energy efficiency of posture and facilitate muscle relaxation. It has been demonstrated that leg crossing in the seated position lowers electromyographic activity in the oblique abdominal muscles. Workstations should be designed to encourage postural behavior of this sort.

5. FUNDAMENTALS OF THE SEATED POSITION

When a person flexes their hip and knee joints to sit down, the iliopsoas muscles immediately shorten and the hip extensors lengthen, resulting in an abnormal lumbar curvature as described above. The balance of antagonistic muscle forces that kept the pelvis in its anteriorly tilted position is changed and the pelvis tilts posteriorly almost immediately and continues in proportion to the flexion at the hip. To keep the head erect, the lumbar spine flexes to compensate for the tilting pelvis and the lumbar lordosis diminishes and eventually disappears. The spine of a person seated erect now resembles that of a quadrupedal animal attempting to stand upright, there is a single c-shaped curve in the thoraco-lumbar region and an increased flexion moment about the lumbar spine. This puts the posterior spinal ligaments under tension and causes the intervertebral discs to be "wedged" anteriorly and to protrude into the intervertebral foramen.

6. SEATED POSTURES FOR RESTING AND WORKING

In most cultures and throughout history people either sit on the floor, cross-legged or in other ways, or squat. Squatting is the habitual resting position of the chimpanzee and of young humans of all cultures.

In squatting, the line of action of the body center of gravity passes through the foot support base and some of the weight of the upper body is born on the legs. Humans exhibit great developmental plasticity and can continue to sit comfortably in this and other ways into old age. However, when deprived of practice, these postural options are lost, as happens to modern-day Western children when they start school and sit all day on seats.

Sitting on chairs with the trunk erect and with 90° of flexion at the trunk–thigh and knee joints "takes the weight off the legs"

and puts it onto the back in the form of an increased flexion moment. Modern thinking suggests that conventional chairs should be designed to enable seated workers to recline with a trunk–thigh angle of 105°. Backrests should be fitted with a convexity in the lumbar region to prevent the lumbar spine flexing further. Some researchers have suggested that seats should slope forward (with or without the knee rest found on "kneeling chairs") so that sitters have an upright trunk and a trunk–thigh angle > 105°. This design philosophy is essentially a compromise, or half-way measure, that attempts to take the weight off the legs while preserving a sufficient amount of lumbar extension to allow the trunk to be held erect without a backrest. There is evidence that the concept is valid but its widespread use would depend on it being able to overcome user habits, preferences and usability issues, which are beyond the scope of the present discussion.

7. WORKING POSTURE

7.1 Postural Adaptation

As far as the lumbar spine is concerned, a good working posture is one in which the spine is towards the mid-point of its range of movement and the trunk is unconstrained (i.e. free to move anteriorly and posteriorly). Figure 2 presents a range of postures varying in extremity and constraint.

In practice, postural adaptation in the workplace is a result of the interaction of three classes of variables:
1. Task requirements (visual, postural and temporal).
2. Workspace design.
3. User characteristics.

The position of the head is a major determinant of posture and is determined by the visual requirements of the tasks. If the main visual area is a flat desktop or workbench > 30° below the "straight ahead" line of sight, people will fixate it by flexing the neck. This causes seated workers to slump forwards and it increases neck flexion, particularly in standing workers. Solutions are to tilt the worksurface ~15° towards the user or to provide document holders to lessen the visual angle. Check also that illumination levels and the quality of lighting are adequate and that workers with visual defects wear appropriate corrective lenses.

7.2 Visual Cues and Standing Balance

In the healthy young person balance is achieved by means of feedback via stretch receptors in the muscle spindles and from the vestibular system. Vision plays a secondary role. Restoration of lost balance is achieved by means of control strategies at the ankle and hip joints. With age (in those > 50 years), proprioceptive feedback degrades and visual feedback becomes more important. When investigating cases of slipping, tripping and falling, one area of investigation is the analysis of the foot–floor interface and the task and organizational factors at the time. Another is the analysis of the visual environment. It has been stated that people are quite capable of walking on ice if they know about it. A more thorough analysis might also look for appropriate visual cues in the work environment and to check whether prevailing light conditions were adequate for revealing them.

7.3. Manual Factors

The position of the hands and arms also has a major influence on the posture of the body. It is strongly determined by the
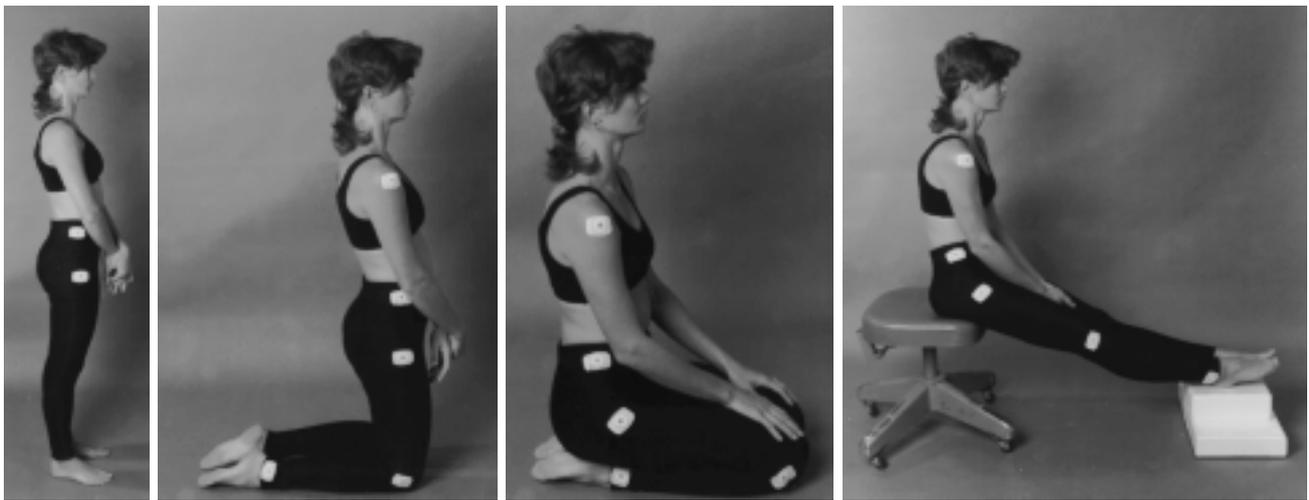
Figure 2. Common postures according to lumbar position and constraint: 1, standing – lumbar spine extended, constrained posteriorly; 2, "praying" posture – lumbar spine highly extended; very constrained in posteriorly and anteriorly (due to lack of anterior base of support); 3, kneeling – lumbar spine in mid-range; unconstrained; 4, "long" sitting (stool) – lumbar spine flexed, constrained anteriorly; 5, "kneeling" chair – limbar spine in mid range (slightly extended); unconstrained; 6, supported kneeling – lumbar spine in mid-range; unconstrained; 7, "long" sitting (floor) – lumbar spine highly flexed; very constrained anteriorly; 8, "90 degree" sitting – lumbar spine unextended or slightly flexed and unstable due to tendency to slump forwards; 9, squatting – lumbar spine highly flexed; highly constrained.

location of work objects with respect to the user, and the forces required to carry out the task. The optimal location for work objects is ~40 cm in front of the worker, between the level of the shoulders and elbows. Objects should not be placed to the worker's side as this encourages axial rotations of the trunk which in turn increase stress on the facet joints and intervertebral discs. Objects should not be placed above shoulder height or so low as to demand stooping. The need to exert large forces may cause workers to adopt flexed or awkward postures in an attempt to close the postural chain by "locking" a limb against a part of the workstation. Seated workers should be able to work close to the front of the desk so that they can work while reclining against a backrest. Standing workers should be able to get their knees and feet underneath the worksurface and rest one foot on a footrail.

7.4 Temporal Factors

Temporarily constrained workers are particularly at risk and must be given the opportunity to alternate between sitting and standing wherever possible. Split-level worksurfaces should be considered so that some part of the job, even superficial or secondary tasks like the ordering of supplies or the sending of a fax, can be done in the alternate posture.

A detailed discussion of workstation design is beyond the present scope. Actual dimensions depend on the anthropometry of the user population, which is often unknown. A list of general recommendations given below will enable workers to adapt appropriately to the constraints of the task and workstation and to alternate between several comfortable working and resting postures.

8. RECOMMENDATIONS

1. Provide clearance under desks and benches so that foot and knee position are not constrained.

2. The feet must never be confined to a small area as this degrades balance and shifts the load of postural adaptation to vulnerable structures higher up the kinetic chain.
3. For every task, find an optimal visual and manual distance that will minimize forward flexion of the trunk and flexion of the neck.
4. Provide footrails for standing workers and footrests for seated workers.
5. Enrich or enlarge the job to increase postural variety. Differentiate new tasks by introducing configurations demanding alternate postures.
6. Activate the venous muscle pump. Require walking, wherever possible. A period of 5 min h^{-1} may be optimal.
7. Design sit–stand workstations.
8. If multi-user worksurfaces are of a fixed height, provide height-adjustable platforms for standing workers. For seated workers, choose a height that will fit taller workers and provide footrests for shorter workers. The height of floors is always adjustable in the upward direction.
9. Alternating between asymmetric postures is a strategy for minimizing static loading. Design workstations and jobs to encourage this strategy.

REFERENCES

BRIDGER, R. S., 1995, *Introduction to Ergonomics* (New York: McGraw-Hill).

KUORINKA, I. and FORCIER, L., 1995, *Work-Related Musculoskeletal Disorders (WMSD's): a reference book for prevention* (London: Taylor & Francis).

LOVEJOY, C.O., 1988, Evolution of human walking. *Scientific American*, November, 82–9.

ROAF, R., 1997, *Posture* (London: Academic Press).

ZACHARKOW, D., 1988, *Posture: sitting, standing, chair design and exercise* (Springfield: Thomas).

Principles of Simple Movements

R.A. Schmidt†, T.D. Lee‡ and D. E. Young†

†Exponent Failure Analysis Associates, Inc., 5401 McConnell Avenue, Los Angeles, CA 90066, USA

‡Department of Kinesiology, McMaster University, Hamilton, Ontario, Canada L8S 4K1

This article is concerned with the principles that describe the behavior in some of the human's simplest movements, discovered and refined over the past 50 years or so. Such actions might involve one-directional hand movements to a target, where the action is brief and usually rapidly done. The principles discussed here are those that define the relationships among various manipulable or measurable features in these actions, such as the movement's duration, the distance moved, the loads applied, the type of load (friction, mass, etc.), and the accuracy demands in achieving a target. Understanding these simple actions is probably fundamental for understanding more complex actions, assuming that more complex actions are in some sense "built" out of simpler ones. We can conceptualize these principles of simple movement as analogous to the laws of Newtonian physics (e.g. the relations among force, velocity, time, etc.) in movements of a mass, which form the cornerstone for understanding more complex mechanical devices (e.g. automobiles).

1. MOTOR PROGRAMMING PRINCIPLES

The rationale for the research in this area, and the basis for understanding many of the results, is based on the concept of motor programming discussed by us elsewhere (see Control of Rapid Actions). Briefly, the idea is that, in rapid movements, the central nervous system (CNS) sets up patterns of activity that control the musculature for several hundred milliseconds or more. In this sense, this activity is "pre-programmed" and, when run off, results in patterns of activity in the muscles that define the movement. In such brief actions there is not sufficient time for sensory information to be transduced, delivered to the CNS, translated and delivered back to the responding limb. In many cases, the actions are completed before any such sensory-based compensations can begin to have an effect. For this reason, primarily, the system organizes the actions in advance. Our most rapid actions are thus relatively immune to sensory information that would alter the movement in some way, and the action tends to be run off as planned.

One major hypothesis concerning how motor programs are thought to operate is by the control of two essential features of muscle activity: (1) the amount of force produced in a muscle and (2) the temporal features of the contraction (when it occurs, and for what duration). In this sense, the "language" of the motor program is thought to be *impulses*, i.e. the integral of force over time in physics (the area "under" the force–time curve of a muscle contraction). A fundamental assumption is that muscular impulses are timed by the CNS so that the muscles generate the desired pattern; hence, the term *impulse timing*.

2. DETERMINANTS OF MOVEMENT ACCURACY

One of the important questions for human factors and ergonomics concerns what features determine the movement's accuracy. Accuracy can be evaluated in many ways, but an important one in this regard concerns the *variability* of the movement, usually measured as the standard deviation of the trajectory about some target value to which the movement is aimed. Based on the notion that the CNS controls impulses, we turn next to the fundamental ways that movements can become inaccurate — variability in force and variability in time.

2.1. Force Variability Principles

It has been known for most of the twentieth century that when muscles contract to produce force on bones, they do so somewhat unreliably. This force variability tends to be distributed approximately normally about the mean target-force, unless the contractions are very small or are near the operator's maximum capabilities. This variability, usually measured as the standard deviation of the force produced about the subject's own mean force, increases with the amount of force produced. From very small contractions up to forces of ~65% of the subject's force capability, the positive relationship between force and force variability is essentially linear, or perhaps slightly concave downward. The slope of this function varies with the muscle tested, with larger muscles having steeper slopes. For the elbow flexor muscles, this variability is ~8% of the mean.

When the forces become larger, such as are found in many activities where violent, rapid actions are required (e.g. manual materials handling, sports, etc.), the relationship between force and force variability adopts an inverted-U shape. In some studies, the maximum force variability occurs at ~65% of the subject's maximum force capability, with the force variability actually becoming smaller as the maximum force is approached. In other studies, the curve merely becomes increasingly concave downward in this range. In any case, the linearity between force and force variability disappears at lower forces.

An important implication is that, when an action is to be produced more forcefully — by decreasing the duration of the movement (with distance constant), increasing the distance (with time constant), or increasing the inertial load — the necessary increases in muscular force brings a change (usually an increase) in the variability of the contractions (except for near-maximal contractions, as discussed above). This locus of this variability is almost certainly at the neuromuscular level related to the recruitment of motor neurons and/or the contraction of muscle itself. As discuss below, these sources of force variability influence variability of the movement's trajectory — especially so in a brief action that cannot be modified en-route.

2.2. Temporal Variability Principles

The second component of the impulse is its duration. It has been long known from studies of timing and time-estimation tasks (such as rhythmical tapping) that the variability of a temporal interval is directly and strongly related to the duration of that interval, even for intervals as long as several seconds. More recent work has shown that other intervals (such as the duration of an impulse in a rapid action) also follow this general rule. When the duration of impulses ranging from 200 to 500 ms have been examined in reciprocal (and other) movements, the variability of
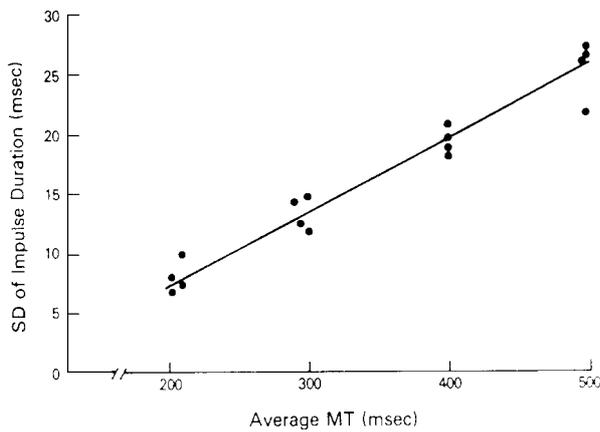
Figure 1. Variability in impulse duration as a function of the movement time in reciprocal actions (from Schmidt and Lee 1999).

Figure 2. Effects of movement time and movement amplitude on variability in the movement's endpoint (effective target width, W_e) (redrawn from Schmidt *et al.* 1979).

the impulse duration is almost directly proportional to the duration of the impulse (Figure 1). The amount of temporal variability is ~5% of the mean duration under these conditions. Various studies have revealed that the source of this increased variability is in the central timing processes, and is not caused by increased variability at the muscle level (unlike force variability, see Section 2.1). It is as if these various intervals are produced by "noisy" central timing processes, with longer intervals having more of the variability included. This general trend holds for much longer intervals as well, although the relationship is not always linear.

These principles are important in several ways. For rapid movements, the variability in the duration of pre-programmed impulses provides another source of error that is related not only to the movement trajectory but also to its intended duration. As we will see, this provides a basis for understanding why increasing the action's duration increases the temporal variability of the action (Schmidt *et al.* 1979). In addition, in many tasks an operator must produce an action at a precise time with respect to some other event; if the action is programmed, then increased variability in the duration of that action (by, for example, increasing its movement time) increases the variability in the timing error.

3. SPEED–ACCURACY TRADE-OFFS

We have long recognized that when an action is to be produced more quickly, it is usually done with less accuracy. Furthermore, we seem to have the capability of "trading" speed for accuracy in many tasks, so that one can maximize accuracy by producing actions more slowly. Recent work on this so-called "speed–accuracy trade-off" has formalized this idea somewhat, and various statements about this kind of trade-off can now be made.

3.1. The Linear Speed–Accuracy Trade-off

When actions are very brief, such as rapid aiming of a hand-held stylus to a target, a clear relationship emerges among the movement's amplitude, the movement's duration and the movement's accuracy (measured as the variability at the movement's endpoint). It is important to note that variability is measured in *spatial* units, and not in terms of time. As shown in
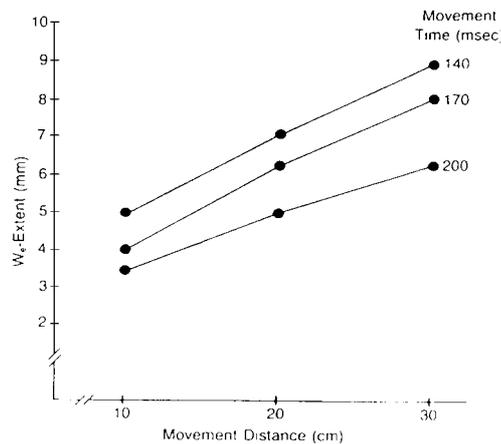
Figure 2, when the distance is varied (holding the time constant via instructions), the variability in the movement's endpoint (effective target width, W_e) increases almost linearly with distance. And, when the movement time is varied (changing the time instruction, with distance constant), W_e decreases almost linearly as the movement time increases.

Both of the manipulations in Figure 2 (movement distance and movement time) have the effect of altering the movement's mean velocity. In fact, when W_e is plotted against the movement's mean velocity (that is, movement amplitude divided by movement time, in cm/s), the relationship has been found to be linear; movement endpoint variability (W_e) is linearly related to average velocity as shown in equation 1:

$$W_e = a + b \ [A/MT], \tag{1}$$

where A is the movement amplitude, MT is the movement time, and a and b are arbitrary constants (Schmidt *et al.* 1979). The slope constant b varies with the particular limb used; smaller limbs with musculature that is more oriented toward precision (hands) have smaller slope constants than larger limbs (arms). With hand movements to targets, the slope is ~3.3 mm/100 ms. This general trade-off holds when the variability is measured in the direction parallel to the overall movement direction (errors in distance) as well as in the direction perpendicular to the movement (errors in direction); direction errors are about half the size of distance errors, at least in hand movements. This general finding has been produced in a variety of movement tasks, and even holds with saccadic eye movements.

3.2. The Temporal Speed–Accuracy Trade-off

Another type of trade-off emerges for rapid tasks in which the accuracy is evaluated in terms of *time* rather than space as it was above. In this trade-off, the operator makes a movement of a given distance, but the *time* of the action is measured. In other tasks, the operator moves a given distance, with a follow-through beyond the point of contact (e.g. batting a baseball), where the time from initiation until the target point is measured. With distance held constant, numerous studies have shown that shortening the movement time produces decreases in the

variability in movement timing, with the variability in timing being almost proportional to the movement time (Schmidt *et al.* 1979). While the major determinant of movement time variability is movement time itself, there is a small effect of movement velocity (when time is held constant and distance is varied), especially if the movement velocities are very low. Here, low-velocity movements are especially "noisy" in terms of temporal variability. Many rapid tasks (such as batting) require precise timing of the action to the movement of an external object (e.g. a ball). Here, shortening the movement time via instructions to move more quickly reduces not only the variability in the overall movement duration, but also the variability of both the movement initiation time (relative to the ball's flight) and the time of the limb's arrival at the target-point (where the ball is to be struck). This makes the entire action more temporally stable and increases the capability for accurate temporal anticipation.

This temporal effect seems related to the idea, presented above, that the central production of time intervals is a "noisy" process, and that shortening the interval reduces these processes to make the action more temporally stable. It seems to make little difference whether the intervals are brief (e.g. 50 ms) and are produced non-consciously via the output of a motor program (such as the duration of a muscle contraction), or are longer and consciously produced (as in estimating 5 s). Note also that this effect is quite different from the movement time effect found when spatial errors are measured. At the same time that decreases in movement time reduce temporal variability, the decreased movement time produces *increases* in spatial variability (see Section 3.1). In tasks like batting, where both temporal and spatial accuracy, and the speed of the bat and its effects on the baseball, are important, the trade-off of these two sources of variability can be quite complex. These various findings are described well by the so-called *impulse-variability* view of movement accuracy (Schmidt *et al.* 1979, Schmidt and Lee 1999).

3.3. Speed–Accuracy Trade-offs and Violent Actions

The principles of rapid actions presented above have an exception when the movements are very rapid and violent. When the level of force production is < 65% of the operator's maximum, decreased movement time results in an almost linear increase in spatial error (see above). However, as the movement is made more violent by reducing the movement time further (and/or by adding inertial load), a decrease in the spatial error is found, with a maximum spatial error at ~60% of the subject's maximum force. This value agrees rather well with the finding that force variability peaks at ~65% of the subject's force capabilities (see above). It appears that, in some of the most violent and forceful actions (e.g. a kick in football), performers seem to enjoy the benefits of larger forces and higher velocities, but without the detriment of increased movement variability as a result. Much more needs to be done in this area to determine, for example, whether such principles also work in actions with multiple joints and/or high levels of required precision (e.g. a golf swing), as well as the role of practice in these relationships.

3.4. Logarithmic Speed–Accuracy Trade-off: Fitts' Law

Certainly one of the most famous and well-established principles of rapid actions was developed by Fitts (1954). In what has been termed the Fitts Paradigm, the operator taps a hand-held stylus alternately between two target plates, attempting to maximize the number of taps in a trial of fixed duration without missing the target more than ~5% of the time. These targets are varied in width (*W*, along the direction of the movement) and amplitude (*A*, or the separation between target centers). A typical trial might be 20 s in duration, and the experimenter measures the number of successful taps in that period. These data are converted to number of seconds/tap, as an estimate of the average movement time. Note that this paradigm is somewhat different from that used with the linear speed–accuracy trade-off. The independent variables in the Fitts work are *A* and *W*, with the movement time being the dependent variable; with the linear trade-off, the independent variables are *A* and MT, and the dependent variable is the effective target width, or W_e.

The typical finding is that the movement time increases (i.e. slows) as the distance to be covered (*A*) increases or the size of the target (*W*) to be contacted decreases. This is not particularly surprising. But Fitts' main discovery was the finding that these variables relate to each other in a relatively simple way. In the expression of Fitts' Law:

$$MT = a + b \, [\log_2 (2A/W)], \tag{2}$$

where MT is the average movement time, *A* is the separation between targets, *W* is the width of the targets, and *a* and *b* are empirical constants. In this work, the expression $\log_2 (2A/W)$ has been termed the index of difficulty (ID), as "difficulty" seems to be expressed both in terms of the separation of the targets and their widths. Thus, Fitts' Law states that the average movement time is linearly related to the ID. The nature of this relationship is shown in Figure 3.

An important implication from the equation is that the movement time will be constant whenever the ratio of *A* and *W* are constant; that is, doubling both *A* and *W* together produce no effects on movement time. As well, the Fitts principle embodies
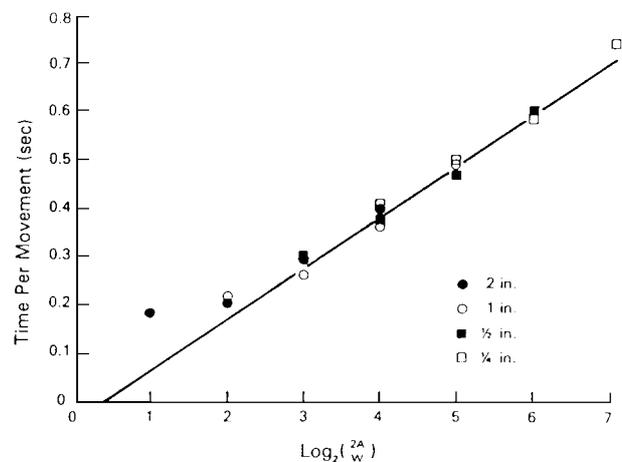
Figure 3. Average movement time as a function of the index of difficulty [(ID = log₂ (2A/W)], where *A* is movement distance and *W* is target width (adapted from Fitts 1954).

the general finding that the movement time is logarithmically related to movement distance; doubling movement distance (*W* constant) increases movement time by only a factor of 1.3. The Fitts principle also accounts well for discrete, single-aiming movements, in which the subject makes a single movement as quickly as possible from a starting location to a target, where the distance to the target (*A*) and the target's width (*W*) are varied.

The Fitts principle appears to hold in a remarkably wide variety of situations. It has been studied in a variety of subject populations, with different limbs and manipulanda, in various body orientations, with microscopic movements, and even under water. In fact, recent findings with movements of a computer "mouse" and joystick from a starting point to a target area on the screen followed very well the expectations of Fitts' Law. Target width has been operationalized in various ways, such as the "play" between a rod and a washer that was to be fit over it, and the results agree well with the general findings. There is a change in the slope constant (*b*) with different limbs, with slopes systematically increasing when the movements are done with the finger, wrist and arm. Various attempts have been made to refine the mathematical statement of Fitts' Law to account for greater variance in movement times, such as a power law presented by Kvalseth [$MT = a(A/W)^b$]; but little or no gains in precision have been produced, and the original statement seems to be preferred. It is certainly one of the most robust principles in the area of movement control. See Schmidt and Lee (1999), Meyer *et al.* (1990) and Plamondon et al. (1997) for a more complete discussion.

Besides the strong generality to many different kinds of movement situations, Fitts' Law has many implications for design issues in human factors. In human–computer interaction, the size of the target and the distance to which the cursor needs to be moved should be considerations in software design. In designing workstations or assembly lines, the time for an action is limited strongly by the target width or "fit" required at the endpoint. Also, the law allows auto engineers to compute the effect of design changes in the size of the brake pedal and/or the distance away from the accelerator on the resulting time to move the foot to the brake (Schmidt 1989).

3.5. Logarithmic versus Linear Trade–off

On the surface, the Fitts principle appears to contradict the linear speed–accuracy trade-off described earlier here (see above). Several features seem to distinguish them, though, and the current understanding is that these actions are descriptors of different phenomena. An important consideration seems to be that the linear trade off is especially relevant for very rapid actions, where the movements are almost certainly programmed; the movement times in this work are usually < 200 ms (Figure 1). The data which fit the Fitts equation most effectively are typically slower (Figure 3), where the movement times often range from 200 to ~1 s. However, it is probably not correct to say that the linear

model is for quick actions and the Fitts' model is for slow ones; one would expect that, even if the movement were slow, if it were programmed without corrections, it would follow the predictions from the linear model, just as if it were faster. In terms of mechanisms, the present view is that the linear trade-off is descriptive of programmed actions, whereas the logarithmic trade-off is descriptive of actions with programmed initial portions and feedback corrections later. One important modeling effort has used the linear trade-off principles to describe the programmed initial part of slower movements; in these views, visually based corrections may or may not later be used to bring the hand to the target depending on various external constraints on the action.

4. APPLICATIONS

The basic principles of human movement discussed here represent a component human performance that resides at the "lowest" level — the functioning of the effector system itself. Often, these components are ignored in analyses of human errors, as they seem to lie "downstream" from errors in perception, in decision-making and judgement, and other more cognitive aspects of performance that are so critical in some performances. Even so, these low-level processes are important for predicting human performance in tasks for which the perceptual and decision-making aspects are minimized, such as in free-throw shooting in basketball or striking a nail with a hammer. Often, in such cases, these lower-level processes are a primary determinant of success. As well, these processes are useful in the understanding of more complex actions, in which variability in components of actions "sum" to give an indication of the proficiency of the larger performance.

REFERENCES

FITTS, P.M., 1954, The information capacity of the human motor system in controlling the amplitude of movement. *Journal of Experimental Psychology*, 47, 381–91.

MEYER, D.E., SMITH, J.E.K., KORNBLUM, S., ABRAMS, R.A. and WRIGHT, C.E., 1990, Speed–accuracy trade-offs in aimed movements: toward a theory of rapid voluntary action. In M. Jeannerod (ed.) *Attention and Performance XIII* (Hillsdale: Erlbaum), 173–226.

PLAMONDON, R. and ALIMI, A.M., 1997, Speed/accuracy trade-offs in target-directed movements. *Behavioral and Brain Sciences*, 20, 279–349.

SCHMIDT, R.A., 1989, Unintended acceleration: A review of human factors considerations. *Human Factors*, 31, 345–64.

SCHMIDT, R.A. and LEE, T.D., 1999, *Motor Control and Learning: A Behavioral Emphasis* (3rd ed.) (Champaign: Human Kinetics).

SCHMIDT, R.A., ZELAZNIK, H.N., HAWKINS, B., FRANK, J.S. and QUINN, J.T., 1979, Motor output variability: a theory for the control of rapid motor acts. *Psychological Review*, 86, 415–51.

Psychophysiological Fitness for Work

A.Y. Burov
Centre of Ergonomics and Safety, ELAN Ltd, Kyiv, Ukraine

1. INTRODUCTION

The analysis of accidents in power plants shows that one of the main causes of increase of their number is a discrepancy in operators' psychophysiological possibilities and professional demands. The economic losses are stipulated not only for accidents, but also for the uneffient operation of equipment. Besides, the operators who do not have appropriate professional requirements become ill and experience industrial trauma more often. There is a necessity to monitor an operator's psychophysiological professional important qualities (PPIQ) during all stages of his professional life (initial professional selection, periodical and daily check). Such an approach allows not only a reduction of economic damage because of uneffient and mistaken human action, but also prevents the premature aggravation of his health and prematurely leaving the profession.

Ergonomic approaches to the evaluation of operator's fitness to work have typically assumed that the functional characteristics of the operator (e.g., knowledge about task, skill level, information processing capacity, etc.) are invariant. They are assumed to not be influenced by variables such as operator mood state, sleep history, time on task, etc. Research efforts aimed at enhancing the safety and reliability of complex process plant operations have, as a result, focused on the improvement of the technical skills of the operator, as well as by reducing operator requirements by automating processes. In complex technological systems, human involvement in the production of accidents is estimated at between 40 and 60% (power plants) and up to 80% (aircraft accidents). This testifies to the inadequate evaluation of operator efficiency by existing techniques. The monitoring of an operator's physiological parameters during his professional life known to be related to effective skill utilization can be used to enhance system performance. While the requirements for professional knowledge and skills may be defined for various kinds of physical and mental work, the evaluation of their functional state over time suffers from a failure to take into consideration not only levels of psychophysiological parameters, but also changes of state of the operator, especially those occurring over different periods of time (year to year, day to day variation, within work schedule variation, variation produced by changing shifts, etc.). The direct or indirect measurement of psychophysiological changes allows for the evaluation of the structure (or moment to moment state of the operator) of the functional state of the operator, allowing us to predict individual fitness and reliability of the operator for effective work.

The objective is the development of the methodology and applied systems of psychophysiological maintenance of operator's fitness to work during all stages of his professional life in order to reduce human errors and to raise his efficiency.

2. PSYCHOPHYSIOLOGICAL RESEARCH

To make psychophysiological research methods valid in industry it is necessary to convert scientific researches into industrial systems when using new developments in the field of psychophysiological initial professional selection, for example, of operators, and periodical and day-by-day checks of human fitness to work as applications for the industry.

To solve this problem a computer system was developed for psychophysiological researches which could be simply transformed to the industry systems for psychophysiological initial professional selection and periodical and day-by-day checks of operators' fitness to work (Burov 1996). The system is intended for psychophysiological researches and provides:

- the opportunity of running a sequence of the presentation, and dimensioned characteristics of test tasks;
- registration of characteristics of each test task performance;
- registration of physiological parameters;
- synchronization between test tasks and physiological parameters;
- database control; and
- time series analysis and statistical modeling.

To record physiological parameters, a hardware–software complex is used which consists of a set of sensors, multi-channel analog–digital converter, personal computer and software. The software realizes functions of control process of registration, processing and database control of physiological signals realization. In dependence on tasks, the analysis of data received during experimental researches can be conducted in different directions, for example

- assessment of the level of anindividual's psychophysiological parameters in relation to a surveyed group;
- dynamics of psychophysiological parameters during psychological tests performance;
- connection of physiological parameters and parameters of mental activity; and
- study of psychophysiological reaction of a human organism affected by external factors (industrial, household, medical).

The researcher has some flexibility to control the experiment.

The developed interactive display systems (IDS) of psychophysiological security allow the solving of research and assessment problems on the basis of the following principles:

- suitability of using tests to the functions which are investigated;
- ergonomic information display on the screen;
- minimal touch;
- taking into account the time factor;
- taking into account the dynamic of operator's work;
- specialized database management system (DBSM) as a heart of the system; and
- using of external criterion.

3. SYSTEMS FOR INDUSTRIAL USE

The observance of these principles allows the creation of psychodiagnostical systems on the basis of the system of psychophysiological researches as being a specialized variant for decision of concrete psychodiagnostical tasks – a system for initial professional selection, daily and periodic psychophysiological checks. Thus the architecture of the system allows for:

- exception of subsystem for registration of physiological parameters,
- transmission of test control from researcher to the DBSM; and
- storage in the database of only the resulting data and, as the consequence, a more economical use of computer resources.

Such a structure provides the availability of managing the program (monitor); subsystems, ensuring of functions of training, testing, data processing, construction of valuations and of predictive models. The carrying out of all these functions, as well as the functions of database control is executed by specialized DBSM which provides as the integrity of internal database an exchange by data of the system with the external program environment.

The IDS allow one to unify the effect on method and to retain information gained all control stages of functional state (FS) and operator fitness to work (OFW):

- psychophysiological professional selection (with recommendations on teaching individualization);
- periodic psychophysiological control FS and operator OFW including monitoring of professional aging;
- daily (preshift) check of professional OFW of operator in "regular" routines;
- daily (intra-shift) check of professional OFW of operator in extreme routines;
- means of functional rehabilitation (psychophysiological correction of functional operator's state);
- training process support included in teaching means; and
- PPIQ's training with specialized psychological playing simulators.

All IDS are intended to work in real-life conditions and can be combined with automated systems and do not need such as supplementary personnel specialists from medical and psychological services.

The systems of initial psychophysiological selection are aimed to estimate unspecific professionally important qualities which make possible (or prevent) good work in the profession. The means and estimation methods must be sufficient to make a decision concerned with potential professional fitness.

The periodical check systems must make possible the estimation of characteristics and personality, and are used periodically (once a year). Time of inspection can be considerable (> 1 h), but the inferences of status must be sufficient to make a decision concerning renewal options of the professional activity.

Systems for initial psychophysiological selection and periodic checks can use objective medical-biological and psychological information concerning the human state in psychophysiological parameters registered by special apparatus in laboratory conditions.

Daily pre-shift and intra-shift check. The pre-shift psychological check must be directed at revealing unfitness for professional activity affected by functional (inconstant) changes in the physiological or psychological status of individual.

An intra-shift psychophysiological check must be directed at the clearing up of any psychophysiological activity dysfunction. The means and methods of verification of this must be effected without reduced professional activity, but be sufficient for the possibility of its continuation on the right levels.

The means and methods of functional rehabilitation must restore the functional abilities which were decreased as result of profes-

sional activity. Such methods can include diverse exercises and measures on renewal of health, and also the full arsenal of physio- and psychotherapeutic means.

At present, to realize this approach were developed and used systems of industrial purpose:

- for a psychophysiological initial professional selection;
- for a periodical check; and
- for a daily check.

The systems permit the evaluation of the reliability and efficiency of the operator's work, as well as to construct the prognosis of changes of these parameters during his professional activity. The results of valuation and prognosis allow either an unreliable operator not to work or to increase his reliability and fitness to work and make preventive measures for accident precautions.

Psychophysiological initial professional selection (Burov 1995). The efficiency of an operator's work depends not only on his professional knowledge and working conditions, but also on the conformity of psychophysiological features of a person to the requirements of his profesion. As the results of preliminary researches have shown, the following structure of tests for determination of the professional fitness group of the operator is optimum: structure of personality, structure of intelligence, an individual's psychodynamic features, and his bent for either kinds of mental activity. The most informative parameters of efficiency of tests performance are included in the model of a "standard operator" enabling the psychophysiological prediction of the group of the professional fitness of a candidate.

Periodical check. The periodical check is intended for a valuation of changes of PPIQ which vary evaluation of slowly, that permits the moment when the operator needs to do some rehabilitation steps or when it is necessary for him to be prepared to leave his profession, if irreversible age-related changes occur. With this purpose the system of valuation of operators' professional aging rate is used (Burov and Chetvernya 1996a, b). The results of valuation of staff's professional age are presented as an integrated age value, which is calculated by the chosen model, as well as an "age profile", which is a vectorial diagram of main parameters determining biological and professional human age.

The daily check (Burov 1997). This is used by the system of valuation of an operator's functional state and for the prediction of professional fitness to work in the workplace which is a method of providing feedback to an operator, enabling him to make necessary changes in case of deviation from his individual "norm." The following tasks are solved in the system:

- development of objective methods of valuation of a human's psychophysiological condition in a relation to norm;
- individual approach to valuation of human's serviceability that is a construction of the individual psychophysiological norm, instead of the group's one; and
- development of organization-technical aspects of valuation's use.

The system allows:

- economical running of the technological object (power industry, aviation);
- economy of energy resources;
- to increase of operators' professional "longevity";
- to do the system as a self-adjustment. It forms automatically the individual "norm" of operator's working ability; and
- to achieve a high precision of prognosis (85–90%).

The use of the systems allows the monitoring of the operator's psychophysiological fitness to work on all stages of his professional life to secure the necessary level of the human–technical system's effectiveness.

REFERENCES

BUROV, A. and CHETVERNYA, YU.V., 1995a, The reliability and productivity of operators of various age on enterprises of energetics. In Bittner, A.C. and Champney, P.C. (eds) *Advances in Industrial Ergonomics and Safety VII* (London: Taylor & Francis).

BUROV, A.YU. and CHETVERNYA, YU.V., 1995b, Structure of psychophysiological indexes of operators' reliability. In *ErgoCon'95 Conference Proceedings. SVEI Conference & Exposition*, San Jose, Hilton & Towers, 21–24 May, 1995.

BUROV, A. and CHETVERNYA, YU., 1996, Methodological principles of psychophysiological forecasting of operators' reliability. In *Advances in Occupational Ergonomics and Safety I. Proceedings of the XIth Annual International Occupational Ergonomics and Safety Conference.* 8–11 July 1996, Zurich, Switzerland. International Society for Occupational Ergonomics and Safety, OH, USA, vol. 2.

BUROV, A. and CHETVERNYA, YU., 1997, Development and implementation of information technology for the psychophysiological valuation and forecasting of the efficiency and safety of operators' work. In *Proceedings of the XIIth Annual International Occupational Ergonomics and Safety Conference*, Washington.

Push and Pull Data

B. J. Daams

Delft University of Technology, Department of Industrial Design Engineering, Jaffalaan 9, 2628 BX Delft, The Netherlands

1. INTRODUCTION

The amount of force that must be exerted on or with a product is important to the perception of the consumer and to the product satisfaction (or especially to the dissatisfaction if the product does not fulfil its task properly). The importance of force exertion in product design is discussed here and in Chapter XX.

When designing for force exertion, a few steps must be taken to get a good result. First, the problem has to be defined the right way during the information stage. Second, an ergonomically sound concept should be developed in the concept stage with the help of global information (rules of thumb) on force exertion. Third, the actual size of maximal and/or minimal forces should be determined in the detailing stage.

Thus, the information needed for the design consists of two parts:
- general information (rules of thumb) for the early stage of design; and
- detailed information (exact forces) for the detailing stage of design.

These rules of thumb are important, because they force the designer to think about a good design concept. If no effort is put in at this first stage, it is no use detailing a basically bad ergonomical design. Rules of thumb can be found in Chapter XX, together with examples of product design. Also, some tips on measuring force exertion for design purposes are given.

After the concept phase of the design, the details should be looked into. How much force can the envisaged user group exert? Or better, how much force can the weakest users exert? This question must preferably be answered by research, but usually there is lack of time, money and/or interest to do so. Therefore, the question about the force that can be exerted can only be answered by literature. Adequate information on force exertion in literature is very scarce, because a situation in which force exertion is investigated seldom corresponds with the situation in which a product is expected to be used, and even less with the real situation. For an attempt to gather design-relevant information and to present it in a way that makes it useful for the designer, see Daams (1994). A choice from these tables are given in this article and in Chapter XX. This chapter comprises the pushing and pulling forces of adults, children, the elderly and the disabled.

2. OVERVIEW OF THE DATA

2.1. Maximal and Comfortable Force

Although the object was to gather a variety of data for this collection, only information on static force exertion could be found. Nearly all investigations concerned maximal force. Only Kanis (1989), Schoorlemmer and Kanis (1992) and Daams (1994) measured comfortable force exertion. Daams (1994) also measured maximal endurance times of submaximal forces.

2.2. Posture

Data on pushing and pulling in free posture can be obtained from Daams (1993) and Steenbekkers and van Beijsterveld (1998). The other investigations concern force exerted in standardized postures.

2.3. Support

Van de Kerk and Voorbij (1993) measured standing force exertion with feet support. Caldwell (1962) investigated the influence of back supports at various heights.

2.4. Fingers

Most push and pull forces are exerted using the whole hand. Kanis (1989) and Steenbekkers (1993) measured push force with the forefinger. Push force with the thumb was measured by Kanis (1989) and Schoorlemmer and Kanis (1992). Pull force with forefinger and thumb was measured by Fothergill *et al.* (1992) and Steenbekkers (1993). Maximal finger forces are in general smaller than maximal forces exerted with the whole hand.

2.5. Laterality

In most cases, the strength of one hand was measured. Force exertion with two hands was measured by Frank *et al.* (1985) in sitting position, and by van de Kerk and Voorbij (1993) in both standing and sitting position.

2.6. Sex

Information on women's force exertion is in Daams (1993 and 1994), Thompson *et al.* (1975), Fothergill *et al.* (1992), Kanis (1989), Schoorlemmer and Kanis (1992) and Steenbekkers and van Beijsterveld (1998), who measured both sexes. Frank *et al.* (1985), Steenbekkers (1993) and van de Kerk and Voorbij (1993) investigated both boys and girls.

2.7. Children

Frank *et al.* (1985) investigated push and pull forces exerted by two groups of children, at 5.5–8.5 and 10–13 years old respectively. Push and pull exerted with hand and with fingers by Dutch girls and boys aged from 4 to 13 years are measured by Steenbekkers (1993). Push and pull in sitting and standing posture by girls and boys aged 4–12 years was investigated by van de Kerk and Voorbij (1993). Bovend'eerdt *et al.* (1980) measured the pull force of children from 12 to 18 years old.

2.8. Healthy Adults

Forces exerted by healthy adults were measured by Daams (1993, 1994), Kroemer (1968), Rohmert (1988), Caldwell (1962), Fothergill *et al.* (1992) and Schoorlemmer and Kanis (1992). In these investigations, relatively young adults are used with an average age of 23–30 years. An exception is Schoorlemmer and Kanis (1992) whose subjects have an average age of 49 years.

2.9. Elderly

Forces in various directions exerted by the elderly aged 60–75 were measured by Thompson (1975). Steenbekkers and van Beijsterveld (1998) measured maximal grip force, push, pull and torque for various age groups from 50 to > 80 as well as for a group of young people.

2.10. Disabled

Kanis (1989) investigated forces exerted by arthritic persons and those suffering from muscle diseases. Schoorlemmer and Kanis

(1992) looked into push force with the thumb, exerted by healthy people and persons suffering from various diseases. They both measured female and male subjects.

2.11. Summary

The information gathered should preferably be summarized to facilitate comparison. However, this is not possible because all experiments describe and prescribe posture in different ways. This makes plotting of various results in one figure, or even a comparison of some of the values, virtually impossible.

For push and pull in standing posture, maximal forces exerted in free postures are higher than maximal forces exerted in standardized postures. Maximal forces exerted while standing with the feet close together are lower than those with one foot in front of the other. Maximal forces exerted with two hands are larger than those exerted with one hand, although usually less than twice as much.

3. PUSH AND PULL DATA

3.1. Bovend'eerdt (1980)

3.1.1. Subjects
- Number: 6000.
- Sex: girls and boys.
- Age: 12–18 years.
- Other characteristics: –

3.1.2. Method
- Direction of force: pull.
- Posture: standing. The instruction for the subject was the following: "stand with legs apart, with the shoulder of the strongest arm furthest removed from the wall bars. Lean with your 'weak' hand horizontally and with the arm stretched against the wall bars. Grasp the dynamometer with your strongest hand, thumb up. At my sign, start pulling. Do not use your weight. Your weak hand should stay on the bar during pulling".
- Sort of force: maximal static force.
- Laterality: the strongest arm.
- Measurement: on a given sign, subjects should pull as hard as possible. The score is the best of two trials.
- Sort of handle: a dynamometer, see figure.
- Size of handle: ?
- Position of handle: at shoulder height, see figure for orientation.
- Other characteristics: the measurements were carried out in a gymnasium with wall bars.
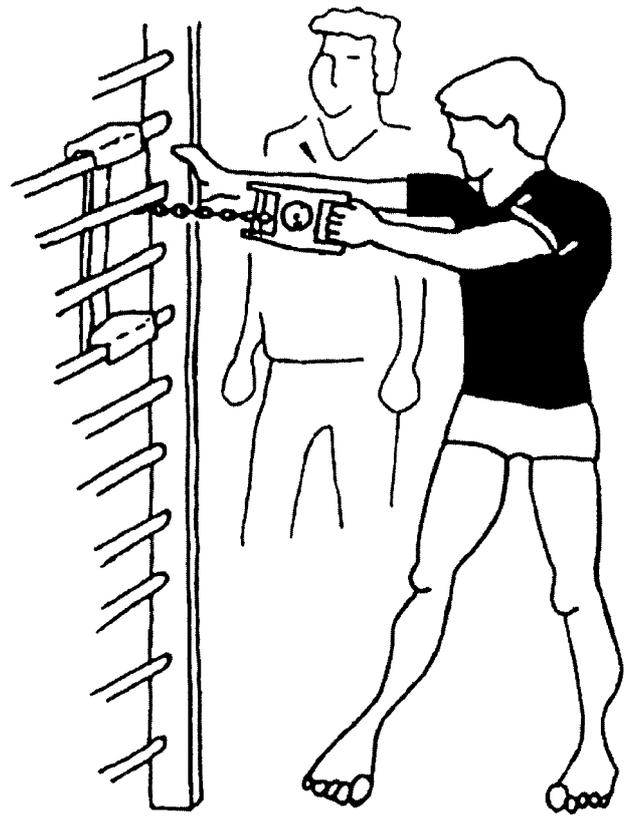
Figure 1. Posture during the experiment (Bovend'eerdt 1980).

For results; see Table 1.

Table 1. Pull (N), categories of percentiles of girls and boys of various ages (Bovend'eerdt 1980).

Results age [years]	P20	P21 – P40	P41 – P60	P61 – P80	P81
girls					
12	249	250 – 279	280 – 318	319 – 357	358
13	279	280 – 318	319 – 347	348 – 387	388
14	298	299 – 338	339 – 377	378 – 426	427
15	318	319 – 357	358 – 396	397 – 445	446
16	328	329 – 367	368 – 416	417 – 465	466
17	347	348 – 387	388 – 416	417 – 475	476
18	347	348 – 396	397 – 436	437 – 494	495
boys					
12	269	270 – 308	309 – 338	339 – 377	378
13	298	299 – 338	339 – 377	378 – 445	446
14	357	358 – 416	417 – 465	466 – 543	544
15	416	417 – 485	486 – 553	554 – 632	633
16	475	476 – 534	535 – 592	593 – 661	662
17	514	515 – 583	584 – 651	652 – 710	711
18	573	574 – 622	623 – 690	691 – 779	780

3.2. ⚡ Daams (1993)

3.2.1. Subjects

- Number: 20.
- Sex: 10 female, 10 male.
- Age: average 23 years (SD 4).
- Other characteristics: mainly healthy students of Delft University.

3.2.2. Method

- Direction of force: push and pull (only horizontal components measured).
- Posture: see figure, standing in: free posture; functional posture: position of feet determined; standard posture: one foot 30 cm in front of the other, elbow in 90° flexion.
- Sort of force: maximal static force exertion.
- Laterality: the preferred hand.
- Measurement: duration 6 s, the score is the average of the last four, the average of two trials.
- Sort and size of handle: push: doorknob, round model, Ø 59 mm, slightly convex; pull: bar Ø 32 mm.
- Position of handle: at shoulder height, elbow height, 0.70, 1.30 and 1.70 m.
- Other characteristics: non-skid flooring was used.
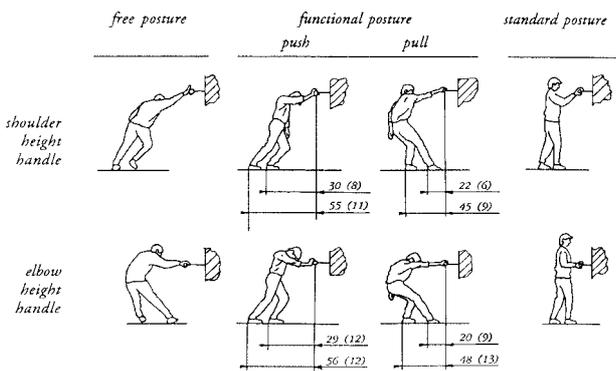
Figure 2. Postures during the experiment, three types of posture and two handle heights (Daams 1993). Functional posture: horizontal distance from heel to vertically projected position of handle is expressed in percentage of body height.

For results; see Table 2.

Table 2. Push and pull (N), averages and SD of 10 women and 10 men (Daams 1993).

Results

handle height	force direction	posture	females x	s	males x	s	both x	s
Shoulder height	pull	free	201	56	301	65	251	79
		functional	198	51	253	48	225	56
		standard	122	29	145	25	134	29
	push	free	198	72	304	53	251	82
		functional	186	42	236	34	211	45
		standard	107	16	136	26	122	26
elbow	pull	free	235	106	400	56	327	111
		functional	249	86	351	54	300	87
		standard	128	27	168	33	148	36
	push	free	225	93	349	96	287	112
		functional	194	61	285	93	240	89
		standard	108	18	147	25	126	28
0.70 m	pull	free	292	97	541	81	416	154
	push	free	185	57	393	134	289	147
1.30 m	pull	free	223	80	347	55	285	92
	push	free	221	103	337	83	279	109
1.70 m	pull	free	196	56	263	60	229	66
	push	free	181	75	300	50	241	87

3.3. ⚡ Kroemer (1968)

3.3.1. Subjects

- Number: 30–43, varying per condition.
- Sex: male.
- Age: adults.
- Other characteristics: –

3.3.2. Method

- Direction of force: push.
- Posture: see figures.
- Sort of force: maximal static force.
- Laterality: see figures and tables.
- Measurement: ?
- Sort of handle: force plate, 20 cm high _ 25 cm long.
- Size of handle: –
- Position of handle: vertically oriented, height of center of force plate: see tables.
- Other characteristics: all conditions include a structural support to 'push against'.

For results; see Table 3.

3.4. ⚡ Rohmert et al. (1988)

3.4.1. Subjects

- Number: 21.
- Sex: male.
- Age: average 23.6 (SD 3.1).
- Other characteristics: body height: average 180.2 cm (SD 8.1); body weight: average 70.6 kg (SD 9.6).

3.4.2. Method

- Direction of force: push, pull, lift, press, and forces in various other directions in the sagittal plane.
- Posture: standing, with the handle at shoulder height at a

Table 3. Push (N), average and SD of males. *Not indicated in literature. However, the sizes of the exerted forces indicate that both hands exert them (Kroemer 1968).

Results

posture		height of centre of force plate	horizontal distance between force plate and vertical support	exerted force	
				x	s
	both hands	acromial height	% of thumb–tip reach		
			50	594	145
			60	680	163
			70	1002	277
			80	1311	408
			90	998	308
			100	658	259
	preferred hand	acromial height	% of thumb–tip reach		
			50	268	68
			60	304	73
			70	367	100
			80	531	145
			90	503	172
			100	435	177
	both hands*	40% of acromial height	% of thumb–tip reach		
			80	1941	753
			90	1651	549
			100	1687	513
			110	1969	576
			120	2000	608
			130	1801	526
		acromial height	% of span		
			50	376	141
			60	354	127
			70	531	168
			80	721	195
			90	331	136
	both hands	% of acromial height	% of acromial height		
		50	80	676	181
		50	100	789	218
		50	120	794	168
		70	80	730	163
		70	100	744	236
		70	120	835	141
		90	80	640	150
		90	100	689	200
		90	120	880	145
		% of acromial height	% of acromial height		
		60	70	776	172
		60	80	871	181
		60	90	807	145
		70	60	590	113
		70	70	712	127
		70	80	739	145
		80	60	531	132
		80	70	630	132
		80	80	649	143
	both hands	% of acromial height	% of acromial height		
		70	70	635	150
		70	80	703	159
		70	90	599	136
		80	70	558	127
		80	80	553	127
		80	90	544	82
		90	70	440	95
		90	80	458	95
		90	90	494	82

distance of 80% of arm reach, and the feet in a fixed position (see figures, a = 0 and 80%).

- Sort of force: maximal static force.
- Laterality: one or two hands? (the exerted forces are limited by the posture, so this may not affect the results).
- Measurement: ?
- Sort of handle: ?
- Size of handle: ?
- Position of handle: at shoulder height, at a distance of 80% of arm reach.
- Other characteristics: –
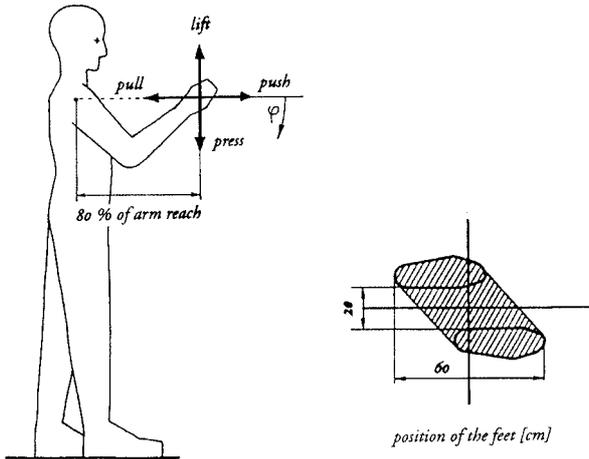
Figure 3. Posture during the experiment (Rohmert *et al.* 1988).

For results, see Figure 4.
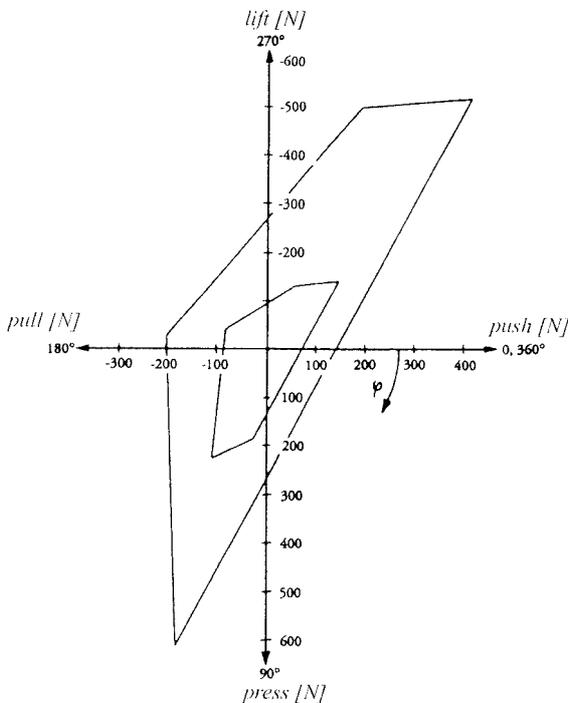
Figure 4. Push, pull, lift, press and forces in various other directions in the sagittal plane (N). Range of 21 men. The inner contour represents the minimum maximal force, the outer contour represents the maximum maximal force (Rohmert *et al.* 1988).

3.5. Thompson (1975)

3.5.1. Subjects
- Number: 100.
- Sex: 62 female, 38 male.
- Age: 60–65 years: 22 female, 13 male; 66–70 years: 23 female, 14 male; 71–75 years: 17 female, 11 male.
- Other characteristics: subjects were selected locally at random from elderly persons within this age range who were living in unassisted occupation of a dwelling. However, preliminary screening eliminated from selection those with any indication of coronary disease.

3.5.2. Method
- Direction of force: push, pull, to the left, to the right, lift and press.
- Posture: a line was drawn on the floor in line with, and vertically beneath, the handle of the apparatus. The subject's leading foot was positioned on this line to ensure maximum force application. Apart from this positioning, subjects were allowed to adopt whatever stance was natural to them.
- Sort of force: maximal static force.
- Laterality: the dominant hand.
- Measurement: subjects were told to exert force as hard as they could until they felt they had reached their maximum, and they were then to release the handle. The highest value of three trials is the value used for computation of the strength data.
- Sort of handle: for push, pull, to the left and to the right: handle in vertical position.
- for lift and press: handle in horizontal position.
- Size of handle: handle Ø 3.2 cm.
- Position of handle: 0.83, 1.0, 1.3 and 1.6 m above the floor.
- Other characteristics: –
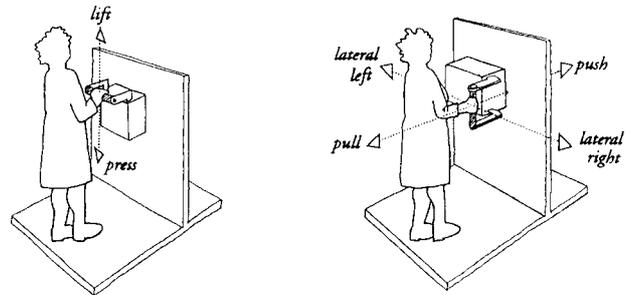
Figure 5. Postures during the experiment (Thompson 1975).

For results; see Table 4

Table 4. Various forces (N) exerted on a handle at various heights. Averages, SD, ranges and 5th percentiles of 62 elderly women and 38 elderly men.

force direction	height [cm]	age [years]	females				males			
			x	s	P5	range	x	s	P5	range
push	83	60–65	149.3	54.9	58.9		282.7	122.1	81.9	
		66–70	168.9	47.8	90.4		220.3	88.0	75.5	
		71–75	147.0	68.4	34.4		204.8	43.0	134.1	
		all	155.9	56.0	63.8	75.7 – 327.7	237.2	94.7	81.4	87.6 – 574.9
	100	60–65	159.3	40.7	92.4		282.7	103.6	112.3	
		66–70	165.4	35.9	106.3		225.0	76.9	98.5	
		71–75	160.6	57.0	66.8		199.3	34.8	142.2	
		all	162.0	43.2	90.9	83.1 – 309.3	237.3	83.1	100.7	92.2 – 540.2
	130	60–65	143.0	53.4	55.3		235.8	81.9	101.0	
		66–70	142.1	35.4	87.8		202.9	63.8	97.9	
		71–75	119.1	32.8	65.2		183.6	34.3	127.2	
		all	136.1	42.4	66.4	61.1 – 299.3	208.6	65.4	101.1	83.1 – 424.4
	160	60–65	134.1	44.2	61.3		219.7	75.1	96.1	
		66–70	126.9	42.7	56.7		182.3	67.7	70.9	
		71–75	117.8	35.2	59.8		170.9	54.1	82.0	
		all	126.9	40.9	59.7	52.0 – 267.4	191.8	79.8	60.5	84.9 – 479.1
pull	83	60–65	175.2	62.5	72.4		387.9	144.5	150.2	
		66–70	223.8	72.1	105.2		281.4	102.0	113.5	
		71–75	185.4	82.6	49.4		284.5	116.1	93.6	
		all	196.0	73.5	75.1	83.0 – 402.8	318.7	127.2	109.4	98.1 – 664.0
	100	60–65	197.4	66.0	88.8		378.6	164.0	108.7	
		66–70	207.3	56.1	115.0		277.8	92.6	125.4	
		71–75	195.3	76.3	69.9		257.8	68.0	146.1	
		all	200.5	64.3	94.9	100.0 – 406.5	306.5	124.1	102.3	102.8 – 785.8
	130	60–65	175.5	66.5	66.1		315.0	128.6	103.4	
		66–70	181.5	45.2	107.2		253.3	81.9	118.6	
		71–75	157.8	54.5	68.0		217.0	33.1	162.6	
		all	172.9	55.5	81.5	80.2 – 389.6	263.9	96.8	104.6	105.7 – 577.3
	160	60–65	159.1	54.2	70.0		299.6	132.6	81.4	
		66–70	150.2	34.5	93.4		227.1	85.0	87.2	
		71–75	145.7	46.5	69.3		210.8	68.3	98.5	
		all	152.2	49.5	70.6	48.1 – 305.6	247.2	103.6	76.8	87.7 – 582.9
lateral, to the left	83	60–65	97.9	35.0	40.2		230.6	79.8	99.5	
		66–70	101.2	49.8	19.2		158.2	60.7	58.4	
		71–75	101.6	44.3	28.7		144.5	58.6	48.2	
		all	100.2	42.6	30.2	27.3 – 225.3	179.0	74.8	56.1	63.8 – 320.0
	100	60–65	105.3	35.9	46.3		221.0	85.6	80.1	
		66–70	97.9	40.9	30.7		163.3	54.5	73.9	
		71–75	98.6	39.0	34.6		143.3	44.7	69.8	
		all	100.8	37.9	38.6	34.0 – 215.4	177.3	70.1	61.9	61.9 – 381.2
	130	60–65	105.8	54.5	16.3		159.0	56.3	66.5	
		66–70	84.5	22.9	46.9		142.3	45.7	67.1	
		71–75	86.6	37.1	25.5		142.1	43.9	70.0	
		all	92.6	40.5	26.0	23.5 – 149.8	147.9	47.8	69.4	55.1 – 245.1
	160	60–65	86.5	30.9	35.6		156.8	66.3	47.6	
		66–70	80.3	27.7	34.8		125.0	46.2	48.9	
		71–75	75.1	33.3	20.2		121.2	54.5	31.6	
		all	81.0	35.2	23.2	24.1 – 234.0	134.8	56.2	42.3	44.0 – 264.3
lateral, to the right	83	60–65	60.9	24.3	20.8		156.2	80.3	24.2	
		66–70	59.1	24.0	19.5		104.9	29.4	56.5	
		71–75	60.6	27.7	15.1		75.8	35.0	18.2	
		all	60.1	24.5	19.7	19.2 – 142.3	114.0	55.9	22.0	43.2 – 293.4
	100	60–65	60.5	28.3	14.1		147.5	63.6	43.0	
		66–70	57.2	26.9	13.1		106.2	51.4	21.8	
		71–75	68.0	31.0	17.0		85.0	39.0	20.7	
		all	61.3	28.2	15.1	15.0 – 159.6	114.2	56.8	20.8	23.5 – 255.0
	130	60–65	61.9	26.1	18.9		129.9	45.9	54.4	
		66–70	66.3	23.9	26.9		101.0	34.1	44.8	
		71–75	60.1	21.4	24.9		79.0	35.1	21.2	
		all	63.1	23.6	24.1	21.0 – 120.6	104.6	42.5	34.6	19.8 – 196.8
	160	60–65	62.5	24.3	22.6		133.0	50.5	49.9	
		66–70	61.6	16.4	34.6		98.8	36.5	38.8	
		71–75	54.4	17.3	25.9		76.7	32.9	22.9	
		all	59.9	19.6	27.7	21.6 – 147.6	104.1	45.3	29.5	26.4 – 208.8

force direction	height [cm]	age [years]	females x	s	P5	range		males x	s	P5	range	
press	83	60–65	208.6	81.8	73.9			464.6	143.2	229.1		
		66–70	256.6	90.1	108.4			343.2	142.1	109.8		
		71–75	182.4	54.9	92.0			363.7	129.9	150.0		
		all	219.2	82.8	83.0	80.4 –	531.0	390.6	144.0	153.8	100.9 –	650.5
	100	60–65	140.9	55.9	49.1			379.5	138.9	151.0		
		66–70	172.4	64.4	66.5			287.1	91.2	137.1		
		71–75	121.2	48.4	41.5			261.5	90.3	113.0		
		all	147.2	59.7	49.0	46.4 –	290.9	311.3	116.6	119.4	93.5 –	578.7
	130	60–65	192.5	55.6	101.0			391.2	129.3	178.5		
		66–70	213.3	73.2	92.8			298.6	129.9	84.9		
		71–75	158.7	51.7	73.7			250.3	58.0	154.8		
		all	190.9	64.1	85.5	69.4 –	406.0	316.3	123.4	113.2	133.7 –	596.6
	160	60–65	230.3	58.1	134.8			426.5	132.3	208.9		
		66–70	254.1	69.9	139.1			360.3	109.1	180.9		
		71–75	200.6	48.8	120.4			338.5	80.3	206.3		
		all	231.0	62.8	127.7	108.3 –	448.1	376.6	112.5	119.5	174.5 –	591.1
lift	83	60–65	138.8	62.3	36.3			402.1	201.3	70.9		
		66–70	192.1	87.4	48.4			289.4	161.1	24.5		
		71–75	120.4	47.4	42.5			294.3	138.3	66.7		
		all	153.5	74.4	31.1	47.4 –	405.0	329.4	171.7	47.0	78.6 –	864.0
	100	60–65	107.9	42.9	37.4			320.6	163.3	51.9		
		66–70	134.9	67.1	24.6			244.4	97.4	84.1		
		71–75	100.9	34.6	44.1			194.1	75.1	70.6		
		all	116.1	52.4	29.8	28.2 –	375.6	255.9	117.0	63.5	85.8 –	667.8
	130	60–65	166.7	67.5	55.5			489.3	245.6	85.3		
		66–70	206.4	99.5	42.9			348.5	184.7	44.6		
		71–75	136.7	57.6	41.9			233.3	105.2	60.3		
		all	173.2	81.8	38.6	56.4 –	510.6	363.3	186.9	55.9	99.6 –	937.3
	160	60–65	157.9	35.9	98.9			283.0	107.9	105.6		
		66–70	163.0	46.6	86.4			212.6	68.2	100.5		
		71–75	142.0	40.8	75.0			221.0	64.0	115.8		
		all	155.5	41.3	87.4	71.4 –	291.0	239.2	85.7	98.0	119.4 –	476.4

3.6. van de Kerk & Voorbij (1993)

3.6.1. Subjects
- Number: total 203.
- Sex: 95 girls, 108 boys.
- Age: 4–12 years.
- Other characteristics: –

3.6.2. Method
- Direction of force: push and pull.
- Posture: standing and sitting, see figures.
- Sort of force: maximal static peak force.
- Laterality: two-handed.
- Measurement: sudden peak force was measured. The children were encouraged during the measurements.

pull, standing

push, standing

push with legs

push, sitting
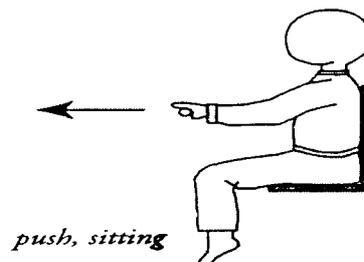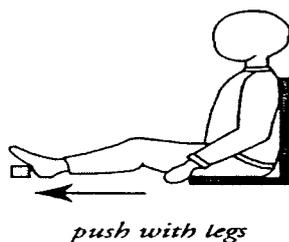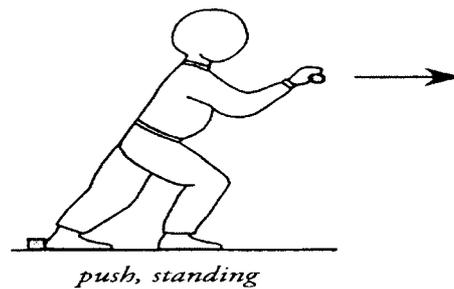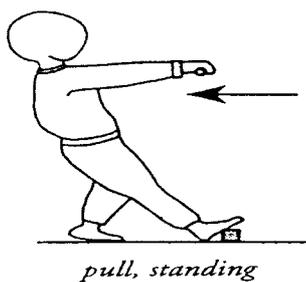
Figure 6. Postures during the experiment (van de Kerk and Voorbij 1993).

- Sort of handle: a cylindrical bar.
- Size of handle: Ø 3 cm.
- Position of handle: see figures.
- Other characteristics: –
  For results; see Table 5.

3.7. ⌐→ Caldwell (1962)

3.7.1. Subjects

- Number: 9.
- Sex: probably men.
- Age: average 24 years, range 22–26.
- Other characteristics: body height: average 174.6 cm, range 1647.6–188.6 cm; body weight: average 71.7 kg, range 57.6–88.4 kg.

3.7.2. Method

- Direction of force: push.
- Posture: sitting, knee angle 140°, with five different elbow angles: 60, 85, 110, 135 and 160°.
- Sort of force: maximal static force.
- Laterality: probably one (the right) hand.
- Measurement: each trial lasted 7 s. The subject was told to push as hard as he could on the handle and to reach maximum output in ~3 s.

- The subject was instructed to note his peak output on the meter and to do his best to exceed this on the next trial. The first trial of each pair was used as a 'motivator'. The results presented are based only on the second trials.
- Sort of handle: ? see figure.
- Size of handle: ?
- Position of handle: at shoulder height, at a horizontal distance changing with the elbow angle (see posture).
- Other characteristics: five backrest conditions were included:
1. No backrest.
2. Backrest height 20% of shoulder height.
3. Backrest height 40% of shoulder height.
4. Backrest height 60% of shoulder height.
5. Backrest height 80% of shoulder height.

A footrest was provided.

Table 5. Various forces (N), averages of girls and boys between 4 and 12 years old (van de Kerk and Voorbij 1993).

force and posture	age [years]	girls				boys			
		n	x	s	range	n	x	s	range
a) push, standing	4	9	119.5	27.5	85.0 – 161.5	11	148.4	42.3	63.5 – 223.5
	5	16	151.1	37.8	79.0 – 216.5	19	191.3	37.4	130.5 – 258.0
	6	8	183.9	36.3	117.5 – 237.5	5	195.2	35.8	151.0 – 243.0
	7	13	274.5	52.5	218.5 – 361.0	12	273.8	35.6	226.0 – 331.0
	8	9	265.3	51.1	181.0 – 337.0	15	282.1	47.3	194.5 – 359.5
	9	14	319.0	71.1	197.0 – 500.5	14	365.8	80.9	247.0 – 529.0
	10	5	479.0	67.8	396.0 – 568.5	8	390.7	90.6	262.5 – 576.5
	11	13	386.4	56.1	293.5 – 468.0	10	432.2	154.9	288.0 – 727.0
	12	8	420.4	99.8	321.5 – 615.5	14	539.8	153.2	333.5 – 958.5
b) pull, standing	4	9	144.3	43.2	84.0 – 200.5	11	172.0	43.5	87.0 – 240.5
	5	16	182.2	69.4	99.5 – 371.0	19	217.6	66.1	122.5 – 345.0
	6	8	246.2	96.4	122.5 – 410.5	5	246.8	28.0	220.5 – 290.0
	7	13	351.2	109.5	227.0 – 627.10	12	402.1	101.1	254.0 – 540.0
	8	9	399.7	127.6	221.5 – 598.0	15	399.9	118.4	160.5 – 604.0
	9	14	433.2	118.2	280.5 – 677.5	14	509.8	127.6	285.5 – 788.0
	10	5	538.8	106.1	453.0 – 713.5	8	596.5	173.7	232.5 – 766.0
	11	13	582.0	152.5	315.5 – 820.5	10	622.5	146.0	407.5 – 864.5
	12	8	565.7	160.5	264.0 – 804.5	14	737.7	243.6	356.5 – 1182.0
c) push, sitting	4	9	214.6	78.9	107.5 – 358.0	11	193.1	55.2	110.0 – 263.5
	5	16	220.8	109.8	54.0 – 420.5	19	245.0	86.5	77.0 – 376.5
	6	8	295.8	57.1	180.5 – 350.5	5	299.4	46.7	242.0 – 366.5
	7	13	331.4	99.0	177.5 – 566.5	12	379.0	58.8	263.5 – 465.5
	8	9	416.1	89.7	238.0 – 532.5	15	378.7	72.7	252.0 – 492.5
	9	14	476.9	104.5	288.5 – 671.5	14	502.1	115.3	332.5 – 767.0
	10	5	602.6	84.4	518.5 – 729.5	8	532.0	78.1	401.0 – 625.5
	11	13	490.4	93.9	341.5 – 618.0	10	518.4	119.9	286.5 – 671.0
	12	8	497.9	95.5	363.0 – 651.5	14	673.9	174.6	354.5 – 978.0
d) push with legs	4	9	309.1	145.3	122.5 – 640.5	11	369.9	214.6	151.5 – 925.0
	5	16	408.0	228.5	89.0 – 819.5	19	470.0	287.9	98.0 – 1027.5
	6	8	516.2	257.0	180.0 – 933.0	5	372.2	39.2	325.5 – 425.5
	7	13	600.3	332.5	179.5 – 1234.0	12	718.3	285.8	320.0 – 1098.0
	8	9	793.3	378.2	370.0 – 1527.5	15	660.4	364.6	308.0 – 1828.5
	9	14	997.0	435.7	350.5 – 1843.0	14	1024.0	427.1	461.5 – 1924.5
	10	5	1321.3	383.4	821.1 – 1829.0	8	1206.7	455.9	448.5 – 1745.5
	11	13	808.8	353.7	424.5 – 1420.5	10	1006.8	572.2	360.5 – 1949.5
	12	8	739.6	222.8	391.5 – 1134.0	14	1291.3	515.9	552.0 – 1959.0
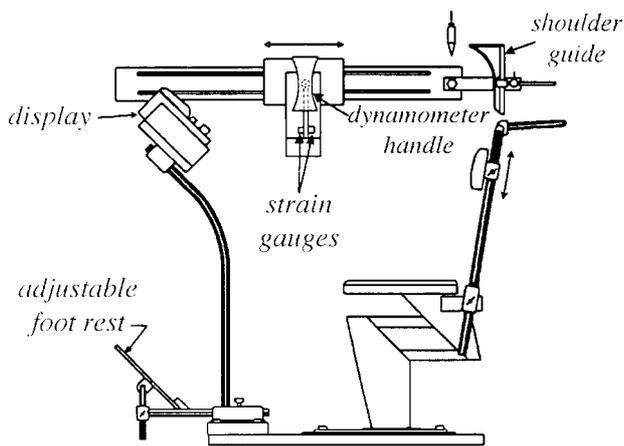
Figure 7. Posture during the experiment (Caldwell 1962).

For result; see Table 6 and Figure 8.

Table 6. Arm extension force (N) at five elbow angles with different backrest positions. Averages of nine subjects (Caldwell 1962).

backrest height [% of shoulder height]	elbow angle					
	60° x	85° x	110° x	135° x	160° x	all angles x
0%	17.4	20.1	22.4	23.2	22.2	21.0
20%	20.5	24.8	35.8	47.5	45.7	34.9
40%	19.9	25.5	35.0	51.3	51.9	36.7
60%	20.0	25.7	37.1	58.6	67.6	41.8
80%	20.9	27.5	37.9	64.4	64.8	43.1
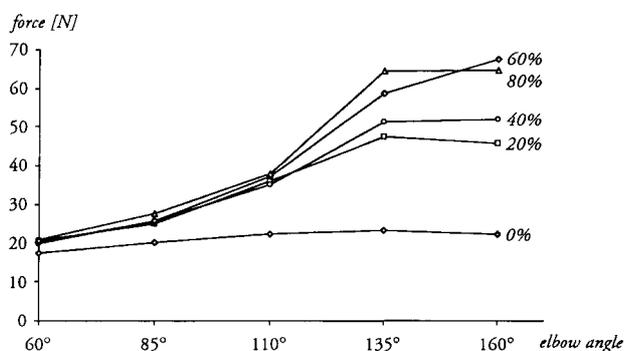
Figure 8. Arm extension force (N) at five elbow angles with different backrest positions. Averages of nine subjects (Caldwell 1962).

3.8. Daams (1994)

3.8.1. Subjects
- Number: 24 for maximal force and endurance measurements, 17 for discomfort measurements.
- Sex: maximal force and endurance: 12 female, 12 male; discomfort: eight female, nine male.
- Age: average 28.2 years (SD 11.2).
- Other characteristics: university students and staff.

3.8.2. Method
- Direction of force: push.
- Posture: maximal force and endurance: sitting and standing, see figures;
- discomfort: sitting.
- Sort of force: maximal static force, endurance and discomfort of submaximal static force. 'discomfort' is measured as the time to the first change of hand when maintaining a submaximal push force with one hand.
- Laterality: maximal force and endurance: non-preferred hand sitting, preferred hand standing;
- discomfort: alternating, starting with the preferred hand.
- Measurement: duration of maximal force: 2 s build up, 4 s maximal force.
- Duration of endurance: measurement limit 30 min. Duration of discomfort measurements:
1. At 80% of maximal force 69 s.
2. At 60%: 144 s.
3. At 40%: 282 s.
4. At 30%: 474 s.
5. At 20%: 1170 s.
6. At 15%: 1800 s (30 min).
- Sort and size of handle: sitting: a slightly convex doorknob, round, Ø 59 mm. Force is exerted with the hand; standing: a metal plate covered with soft plastic, size 13.5 _ 4 cm. Force is exerted with the distal part of the lower arm.
- Position of handle: see figure.
- Other characteristics: –
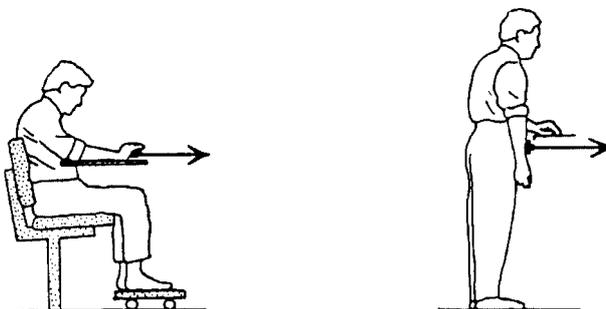
Figure 9. Postures during the experiment (Daams 1994).

For results; see Tables 7–9 and Figure 10.

Table 7. Push (N), maximal forces exerted sitting and standing, averages and SD and numbers of measurements of 12 women and 12 men (Daams 1994).

	females			males			both		
	x	s	n	x	s	n	x	s	n
sitting, push	176.9	48.5	138	311.1	86.6	161	249.0	98.0	299
standing, push	63.8	13.8	99	114.6	29.7	100	89.3	34.4	199

Table 8. Maximal endurance time (s) per force level. Median, percentiles, range, average, SD, coefficient of variation and number of measurements of 24 subjects (Daams 1994).

force level	P25	P50	P75	range			x	s	coeff. of variarion	n
women (n=12)										
80%	58.8	75.0	107.0	27	–	166	86.3	37.8	0.44	23
60%	95.0	133.0	172.8	54	–	290	144.7	63.9	0.44	23
40%	150.3	230.0	381.5	95	–	457	258.7	119.9	0.46	23
30%	225.0	353.0	494.0	123	–	1389	426.6	293.8	0.69	22
20%	385.5	485.0	732.3	182	–	1800	652.5	450.3	0.69	19
men (n=12)										
80%	52.0	58.5	82.5	28	–	76	65.2	23.7	0.36	24
60%	76.5	101.0	135.0	44	–	254	111.2	48.5	0.44	24
40%	162.0	237.5	334.5	93	–	574	254.8	119.4	0.47	24
30%	267.5	389.0	640.8	139	–	1800	557.7	460.0	0.82	23
20%	418.0	729.5	1187.0	253	–	1800	855.1	525.4	0.61	22
15%	785.0	1088.0	1800.0	409	–	1800	1229.0	535.9	0.44	23
women and men (n=24)										
80%	52.0	64.0	90.0	27	–	166	75.5	32.8	0.43	47
60%	85.0	116.0	156.5	44	–	290	127.6	58.4	0.46	47
40%	160.0	230.0	353.0	93	–	574	256.7	118.3	0.46	47
30%	252.8	378.0	609.3	123	–	1800	493.6	389.1	0.79	47
20%	401.3	621.5	915.8	182	–	1800	761.2	496.6	0.65	42
15%	603.3	946.0	1800.0	313	–	1800	1056.4	536.2	0.51	41

Table 9. Time to the first change of hand (s) per force level. Median, percentiles, range, average, SD and coefficient of variation of 17 subjects (eight female, nine male) (Daams 1994).

	force level	time[s]	P25	P50	P75	range			x	s	coeff. of variation
time to the first change of hands[s]	80%	69	22.0	27	38.5	17	–	>69	31.8	14.1	0.44
	60%	144	29.0	41	56.3	14	–	87	45.0	21.4	0.48
	40%	282	58.0	92	107.0	28	–	182	89.1	39.3	0.44
	30%	474	66.8	95	125.3	44	–	344	112.3	72.2	0.64
	20%	1170	109.3	134	212.5	88	–	848	211.2	208.3	0.99
	15%	1800	143.8	212	329.0	90	–	1295	321.9	323.3	1.00
total number of changes(n)	80%	69	1.0	2	2.0	0	–	4	1.71	0.92	0.54
	60%	144	2.0	3	3.3	1	–	5	2.88	1.11	0.39
	40%	282	3.0	4	6.0	1	–	9	4.35	2.09	0.48
	30%	474	3.0	4	7.0	1	–	8	4.82	2.13	0.44
	20%	1170	4.8	9	12.0	1	–	15	7.94	4.28	0.54
	15%	1800	4.8	9	12.3	1	–	19	8.82	4.94	0.56
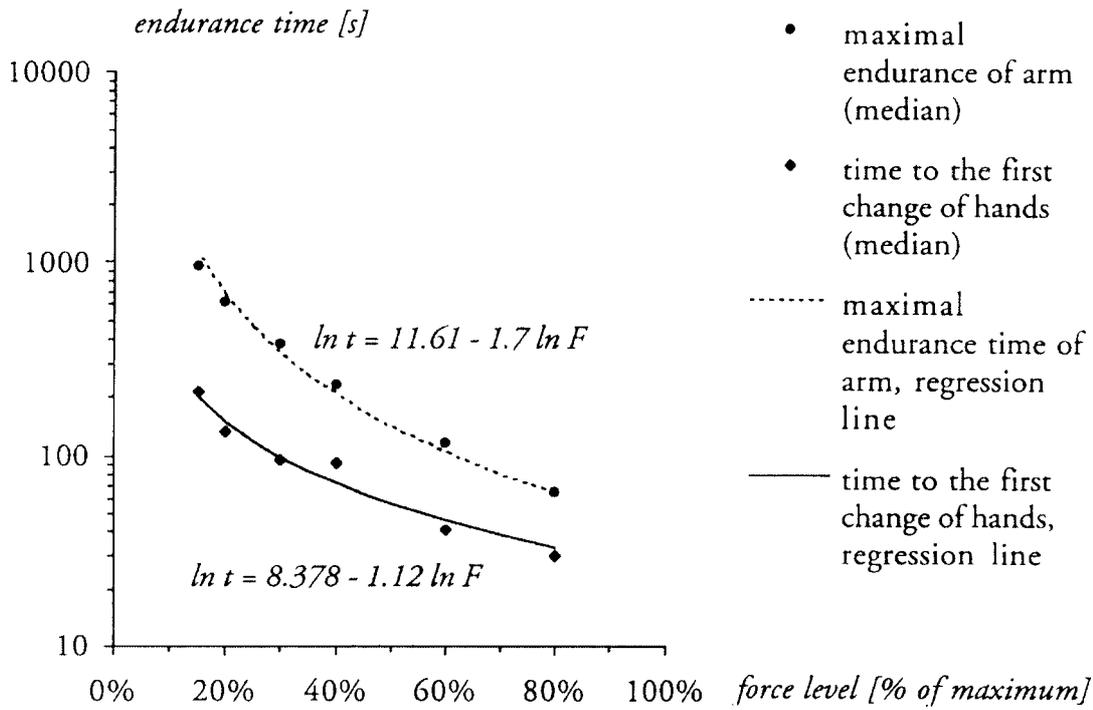
Figure 10. Medians and interpolations of maximal endurance time and time to the first change of hand (s), for arms only (Daams 1994).

3.9. 🏃↔ Frank et al. (1985)

3.9.1. Subjects

- Number: 59.
- Sex: youngest group: 13 girls, 18 boys; oldest group: 14 girls, 14 boys.
- Age: youngest group: 5.5–8.5 years; oldest group: 10–13 years.
- Other characteristics: length and length/weight ratio are sufficiently representative for Dutch children.
- Body height, youngest group: average 125.6 cm (SD 6.7); oldest group: average 149.1 cm (SD 8.5); body weight, youngest group: average 24.6 kg (SD 3.9); oldest group: average 39.1 kg (SD 8.1).
- Distance of back to feet when the knees are at an angle of 110°: youngest group: average 59.1 cm (SD 6.1); oldest group: average 76.9 cm (SD 7.4).

3.9.2. Method

- Direction of force: push and pull.
- Posture: sitting in an experimental situation similar to a 'Vliegende Hollander' (Flying Dutchman, children's vehicle propelled by pushing and pulling with arms and upper body), see figures.
- Two different postures: with knees at an angle of 110° and at an angle of 160°. Apart from this, the posture was free.
- Sort of force: maximal static force.
- Laterality: both hands.
- Measurement: ?
- Sort of handle: a cylindrical bar.
- Size of handle: ?
- Position of handle: see figure, height not known.

- Other characteristics: a foot support was provided, to push against.

posture during push   *posture during pull*
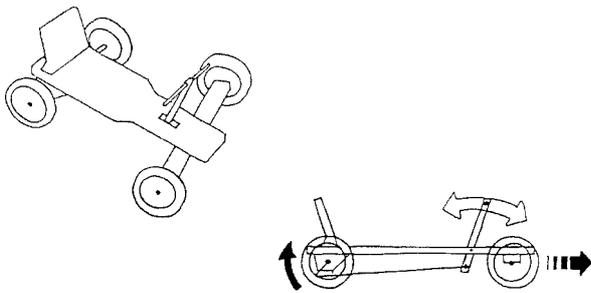
Figure 11. Postures during the experiment (Frank *et al.* 1985).

For results; see Table 10.

Table 10. Push and pull (N), average, SD and percentiles of children (Frank *et al.* 1985).

Results

age	n	force	x	s	P5	P95
5.5–8.5 years	31	push	151.9	33.7	96.3	207.6
		pull (knee 110°)	333.9	103.3	163.4	504.4
		pull (knee 160°)	373.4	82.1	237.9	508.9
10–13 years	28	push	230.5	57.2	136.1	324.9
		pull (knee 110°)	573.9	159.1	311.3	836.4
		pull (knee 160°)	603.6	133.8	382.8	824.4

3.10. Steenbekkers (1993)

3.10.1. Subjects
- Number: 782.
- Sex: 392 girls, 390 boys.
- Age: 4–13 years.
- Other characteristics: Dutch children, selected to be representative for the Netherlands.

3.10.2. Method
- Direction of force: push and pull with hand and fingers.
- Posture: sitting before the measuring device, which is in front of the forearm. The forearm is horizontal and in a sagittal plane. Upper arm and forearm are at an angle of about 150°. The legs hang down freely or are positioned forward: (a) pulling with the hand: the bar is placed vertically, its axis between middle finger and ring finger; (b) pulling with thumb and forefinger: the finger is placed on top of a round knob, the thumb is on the bottom of the knob. The other fingers are flexed into a fist; (c) pushing with the hand: force is exerted with the palm of the hand; (d) pushing with the forefinger: the finger is placed in the middle of a round concave knob. The other fingers and the thumb are flexed into a fist.

- Sort of force: maximal static force.
- Laterality: preferred and non-preferred hand (they are not significantly different and therefore represented in one table).
- Measurement: duration 3 s.
- Sort and size of handle: (a) vertical bar, Ø 1.1 cm; (b) round concave knob, Ø 2 cm; (c) round convex knob, Ø 4 cm; (d) round concave knob, Ø 2 cm.
- Position of handle: mounted on a table.
- Other characteristics: –

For results, see Figures 12–15 and Tables 11–14.
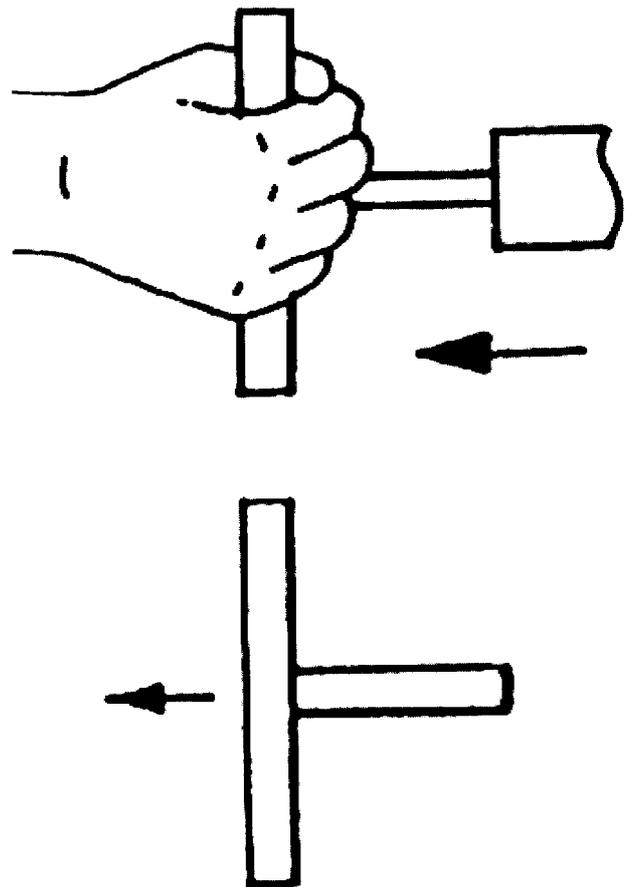
Figure 12. Posture during experiment A (Steenbekkers 1993).
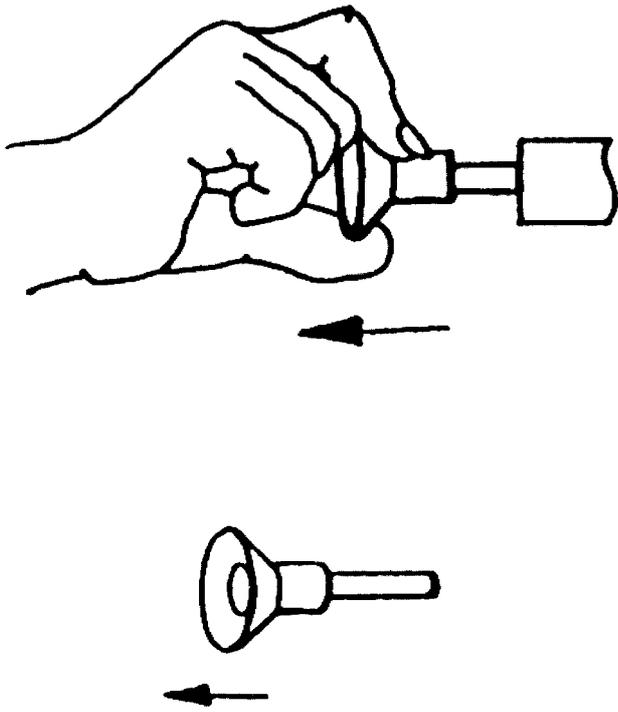
Figure 13. Posture during experiment B (Steenbekkers 1993).
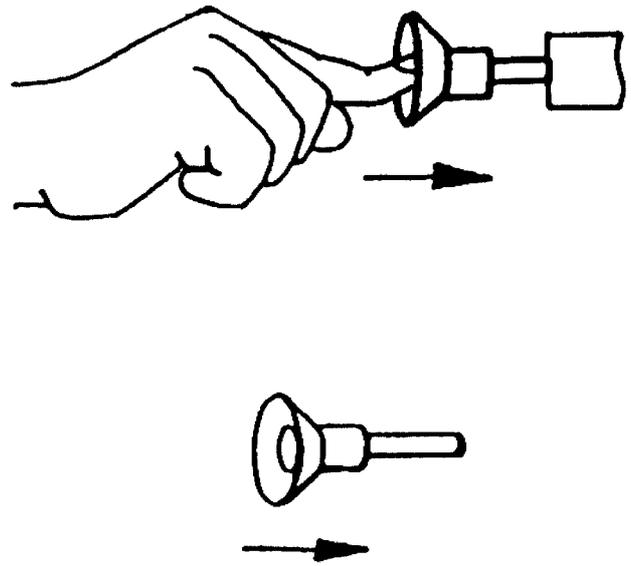
Figure 15. Posture during experiment D (Steenbekkers 1993).
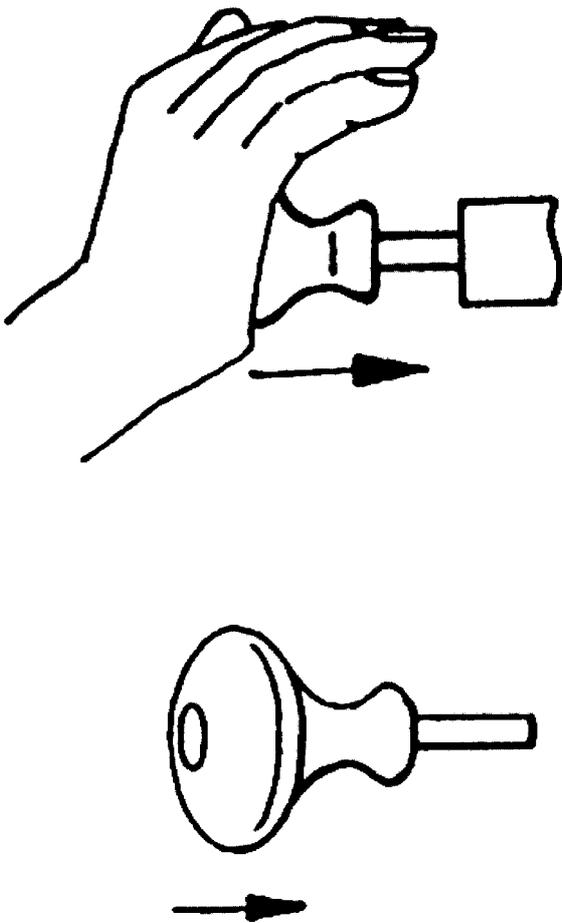
Figure 14. Posture during experiment C (Steenbekkers 1993).

Table 11. Experiment A: pull (N) with one hand. Averages, SD and percentiles of girls and boys (Steenbekkers 1993).

age [years]	girls					boys					all	
	n	x	s	P3	P97	n	x	s	P3	P97	x	s
4.0– 4.9	46	31.2	12.5	15.9	60.6	41	36.9	15.1	16.9	71.4	33.9	14.0
5.0– 5.9	59	36.3	12.1	13.7	63.7	53	43.3	17.7	16.9	80.2	39.6	15.3
6.0– 6.9	57	44.0	16.4	18.6	88.3	59	48.1	21.7	21.0	110.3	46.1	19.3
7.0– 7.9	42	55.0	28.5	17.6	135.7	50	59.0	24.5	27.0	122.9	57.2	26.3
8.0– 8.9	40	53.8	17.7	25.9	90.1	41	66.3	28.6	30.6	145.1	60.1	24.5
9.0– 9.9	42	72.2	29.8	31.4	150.6	41	81.3	32.9	37.8	145.6	76.7	31.5
10.0–10.9	39	80.6	28.5	38.9	138.6	35	99.7	49.7	34.3	210.2	89.7	40.8
11.0–11.9	44	93.5	36.8	42.9	170.3	40	100.6	37.2	57.0	181.5	96.9	36.9
12.0–12.9	23	97.8	40.0	39.2	191.2	30	127.0	45.9	50.2	244.9	114.3	45.5

Table 12. Experiment B: pull (N) with thumb and forefinger. Averages, SD and percentiles of girls and boys (Steenbekkers 1993).

age [years]	girls					boys					all	
	n	x	s	P3	P97	n	x	s	P3	P97	x	s
4.0– 4.9	46	15.3	5.7	4.8	25.9	41	18.3	5.5	11.4	29.3	16.7	5.8
5.0– 5.9	59	20.0	5.3	10.9	32.6	53	21.6	6.0	10.1	34.7	20.7	5.7
6.0– 6.9	57	23.7	6.4	8.8	36.3	59	25.3	8.5	13.5	43.6	24.5	7.5
7.0– 7.9	42	27.7	7.6	15.9	42.1	50	30.5	11.2	16.8	61.0	29.2	9.8
8.0– 8.9	40	28.9	8.7	16.6	45.3	41	34.7	10.9	18.5	64.9	31.8	10.2
9.0– 9.9	42	38.7	10.8	20.9	57.4	41	39.0	9.6	22.8	52.3	38.9	10.2
10.0–10.9	39	42.0	11.8	25.9	69.2	35	45.7	13.9	25.4	70.8	43.7	12.9
11.0–11.9	44	48.1	14.9	23.0	81.2	40	51.3	14.5	31.8	95.5	49.6	14.7
12.0–12.9	23	51.4	15.2	24.8	75.9	30	60.9	17.2	31.7	108.2	56.8	16.9

Table 13. Experiment C: push (N) with one hand. Averages, SD and percentiles of girls and boys (Steenbekkers 1993).

age [years]	girls					boys					all	
	n	x	s	P3	P97	n	x	s	P3	P97	x	s
4.0– 4.9	46	38.7	15.0	16.1	60.0	41	45.9	15.0	25.7	81.7	42.1	15.3
5.0– 5.9	59	44.4	13.8	23.8	73.9	53	54.2	16.6	28.8	90.4	49.0	15.9
6.0– 6.9	57	55.0	16.9	26.3	90.1	59	61.9	15.6	36.5	95.7	58.5	16.6
7.0– 7.9	42	60.9	18.9	34.3	96.2	50	69.7	19.6	40.4	108.0	65.6	19.7
8.0– 8.9	40	62.4	18.8	28.6	93.2	41	79.4	22.5	44.3	136.9	71.0	22.3
9.0– 9.9	42	74.5	19.3	46.6	114.9	41	88.6	25.6	46.2	133.6	81.5	23.6
10.0–10.9	39	86.4	23.6	42.8	136.7	35	108.9	40.6	59.9	195.5	97.0	34.4
11.0–11.9	44	94.3	24.8	59.9	137.0	40	108.5	23.3	70.6	144.9	101.1	25.0
12.0–12.9	23	104.0	32.5	27.0	165.5	30	118.9	31.7	74.8	186.9	112.5	32.6

Table 14. Experiment D: push (N) with forefinger. Averages, SD and percentiles of girls and boys (Steenbekkers 1993).

age [years]	girls					boys					all	
	n	x	s	P3	P97	n	x	s	P3	P97	x	s
4.0– 4.9	46	18.3	5.5	7.8	29.9	41	20.0	5.3	12.3	29.0	19.1	5.4
5.0– 5.9	59	22.8	6.7	14.2	34.7	53	25.2	6.1	16.0	38.6	24.0	6.5
6.0– 6.9	57	26.3	6.0	15.0	39.4	59	27.5	6.8	14.8	45.2	26.9	6.5
7.0– 7.9	42	30.8	8.4	17.9	44.2	50	33.1	9.9	17.6	59.0	32.1	9.3
8.0– 8.9	40	32.3	8.4	18.4	47.6	41	39.1	9.8	23.2	59.7	35.7	9.7
9.0– 9.9	42	38.1	9.7	24.6	53.9	41	41.2	9.6	24.3	60.4	39.6	9.7
10.0–10.9	39	44.7	11.0	29.7	66.7	35	49.4	14.0	32.1	76.8	46.9	12.7
11.0–11.9	44	45.7	11.3	29.5	69.1	40	52.8	12.7	31.9	73.5	49.1	12.5
12.0–12.9	23	50.1	16.4	16.9	81.6	30	62.3	15.1	31.9	88.4	57.0	16.7

3.11. ✏ ⫞ ert Icon K> Fothergill et al. (1992)

3.11.1. Subjects

- Number: 30.
- Sex: 16 female, 14 male.
- Age: average 30.0 (SD 8.2).
- Other characteristics: volunteer staff and students: body height: average 170.4 cm (SD 9.6); body weight: average 67.5 kg (SD 9.4).

3.11.2. Method

- Direction of force: pull.
- Posture: standing. The type of grip on each handle and the posture adopted were freely chosen, provided that (1) only the dominant hand was used on the handle/bar; (2) only the feet made contact with the floor; and (3) the leading foot was not placed in front of the handle.
- Sort of force: maximal static force.
- Laterality: the dominant hand.

- Measurement: subjects were instructed to exert a steady maximal pull on each handle for 5 s, in a direction as close to the horizontal plane as possible. For each measurement, peak force was determined and steady pull strength was calculated from a 3-s average.
- Sort and size of handle: three handles and one bar were used (see figure).
- Position of handle: two handle heights: 1.0 and 1.75 m.
- Other characteristics: the floor was covered with a coarse emery paper to prevent slipping. Subjects wore their normal everyday footwear. The majority wore rubber soled shoes or trainers.
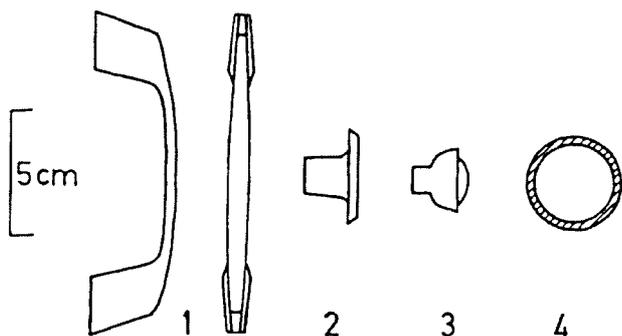
Figure 16. Side views of the four handles used in this study (the front view for handle 1 is also shown). Handles 1 and 2 were cast alloy, handle 3 was of Bakelite, and handle 4 was a mild steel bar (Fothergill *et al.* 1992).

For results; see Table 15.

3.12. Kanis (1989)

3.12.1. Subjects
- Number: 34.
- Sex: 28 female and six male.
- Age: average 54 years (SD 14.2).
- Other characteristics: 28 arthritic subjects (24 women, four men) and six subjects (four women, two men) with a muscle disease.

3.12.2. Method
- Direction of force: push with thumb and push with forefinger.
- Posture: sitting, with the elbow 90° flexed; see figures.
- Sort of force: maximal and comfortable static force. "Comfortable" refers to a level of pain or fatigue found acceptable by the subject.
- Laterality: preferred hand.
- Measurement: subjects were instructed to build up their force gradually and to hold the maximum force for at least a few seconds.
- Sort of handle: plastic cylinder.
- Size of handle: height 15 mm, Ø 20 mm.
- Position of handle: mounted on a table.
- Other characteristics: –
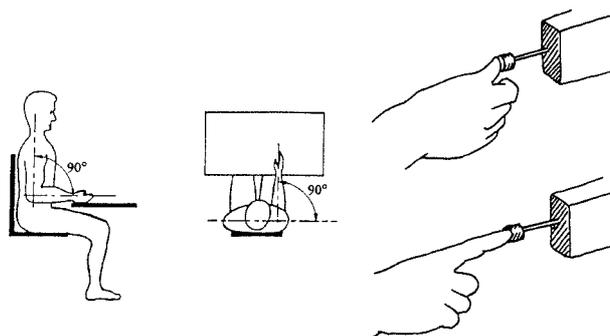
Figure 17. Posture during the experiment (Kanis 1989).

For results; see Table 16.

Table 15. Pull (N) on four different handles at two different heights, defined as "peak strength" and "steady maximal strength". Averages and SD of 30 subjects (Fothergill *et al.* 1992).

	bar height	handle type	female x	female s	male x	male s	both x	both s
peak pull	1 m	1	276	58	400	125	334	113
		2	186	36	256	86	219	72
		3	112	18	127	27	119	23
		4	282	76	403	114	338	112
	1.75 m	1	164	30	232	58	196	56
		2	122	20	191	46	154	49
		3	100	16	121	28	110	24
		4	178	40	241	61	207	59
steady max, pull	1 m	1	247	51	364	125	302	109
		2	171	45	225	87	197	72
		3	90	21	105	22	97	23
		4	244	62	361	109	299	104
	1.75 m	1	139	28	197	50	166	49
		2	101	21	168	44	133	47
		3	82	17	104	31	93	26
		4	144	25	201	53	171	49

Table 16. Push (N) with thumb and forefinger. Averages and SD of 34 impaired subjects (Kanis 1989).

grip	force	females x	s	males x	s	arthiritis x	s	muscle diseases x	s	all x	s
thumb	maximal	25.0	16.3	35.8	29.5	28.8	19.2	18.2	18.1	26.9	19.2
forefinger	maximal	17.2	9.9	19.3	15.9	18.8	10.6	12.0	11.9	17.6	11.0
	comfortable	7.6	4.7	9.1	9.5	8.1	5.0	7.1	7.9	7.9	5.6

3.13. ☞ Schoorlemmer and Kanis (1992)

3.13.1. Subjects
- Number: 27.
- Sex: 21 female, six male.
- Age: average 49.4 years (SD 17.4)
- Other characteristics: average body height 165 cm (SD 12), average body weight 67 kg (SD 12). Twelve healthy subjects, five spastics, four visually impaired, three blind subjects, three subjects with Parkinson's disease.

3.13.2. Method
- Direction of force: push with the thumb.
- Posture: sitting, with the elbow 90° flexed (see figures) and in free posture.
- Sort of force: maximal and comfortable static force. 'Comfortable' refers to a level of pain or fatigue found acceptable by the subject.
- Laterality: left and right hand (averaged in the table).
- Measurement: subjects were instructed to build up their force gradually and to hold the maximum for at least a few seconds.
- Sort of handle: plastic cylinder.
- Size of handle: height 15 mm, Ø 20 mm.
- Position of handle: mounted on a table.
- Other characteristics: –
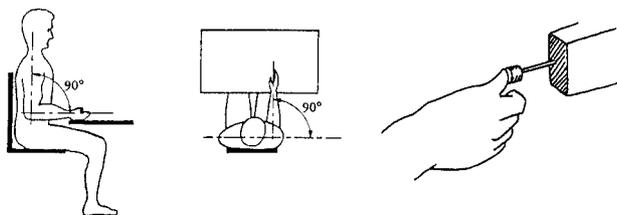
Figure 18. Posture during the experiment (Schoorlemmer and Kanis 1992).

For results; see Table 17.

3.14. ⊢→ ⊢ Steenbekkers and van Beijsterveld (1998)

3.14.1. Subjects
- Number: several hundreds; see tables.
- Sex: women and men; see tables.
- Age: 20 to > 80, divided in nine different age groups; see tables.
- Other characteristics: Dutch adults and elderly, selected to be representative for the Netherlands.

3.14.2. Method
- Direction of force: push (with two hands); pull (with one hand).
- Posture: the subject stands before the measuring device and adopts a free posture, see figure.
- Sort of force: maximal static force.
- Laterality: pushing with both hands; pulling with one hand (own choice).
- Measurement: duration 3 s.
- Sort and size of handle: for pushing: a horizontal bar, Ø 3.0 cm; for pulling a horizontal bar, Ø 3.2 cm (see figure).
- Position of handle: at elbow height standing.
- Other characteristics: –

Table 17. Push (N) with the thumb. Averages and SD of 12 healthy and 15 impaired subjects (Schoorlemmer and Kanis 1992).

Results

force	posture	females n = 21 x	s	males n = 6 x	s	healthy, blind & visually impaired n = 19 x	s	spastics n = 5 x	s	Parkinson's desease n = 3 x	s	all n = 27 x	s
max.	standard	64.1	32.4	81.3	26.7	76.2	31.3	54.3	25.9	38.1	18.5	67.9	35.6
	free	68.5	36.1	86.9	43.0	79.1	39.0	67.2	34.7	40.8	19.1	72.6	37.7
comf.	standard	31.8	15.4	22.4	16.6	33.0	16.5	19.0	9.1	26.9	15.1	29.7	15.8

Figure 19. Posture during push with two hands (Steenbekkers and van Beijsterveldt 1998).
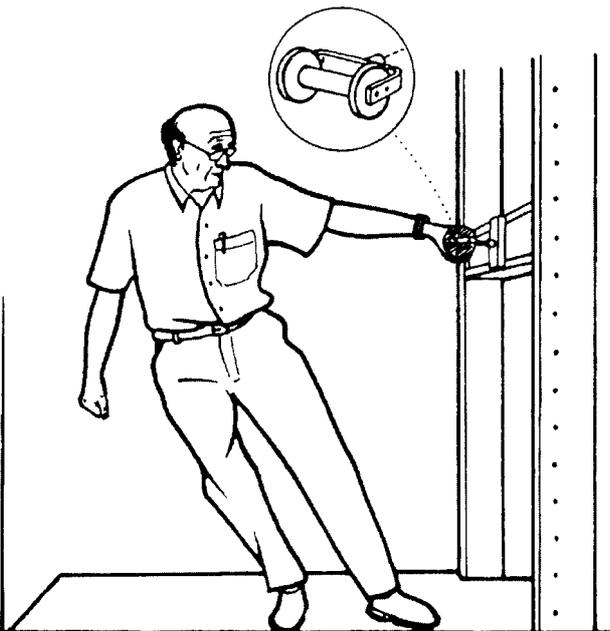
Figure 20. Posture during pull with one hand (Steenbekkers and van Beijsterveldt 1998).

For results, see Figures 19 and 20, and Tables 18 and 19.

REFERENCES

BOVEND'EERDT, J., KEMPER, H. and VERSCHUUR, R., 1980, *De Moper Fitness Test* (Haarlem: De Vrieseborch).

CALDWELL, L.S., 1962, Body stabilization and the strength of arm extension.

DAAMS, B.J., 1993, Static force exertion in postures with different degrees of freedom. *Ergonomics*, 36, 397–406.

DAAMS, B.J., 1994, *Human Force Exertion in User–Product Interaction*. Physical Ergonomics Series (Delft: Delft University Press).

FOTHERGILL, D.M., GRIEVE, D.W. and PHEASANT, S.T., 1992, The influence of some handle designs and handle height on the strength of the horizontal pulling action. *Ergonomics*, 35, 203–12.

FRANK, P., HAN, F. and SPANGENBERG, S., 1985, *Krachtuitoefening bij kinderen in een Vliegende Hollander*. Internal report no. 66 (Delft: Delft University of Technology, Faculty of Industrial Design Engineering).

KANIS, H., 1989, *Bediening & Handikap* (Delft: Delft University of Technology, Faculty of Industrial Design Engineering).

KROEMER, K.H.E., 1968, *Push Forces Exerted in Sixty-five Common Working Postures*. Arml-tr-68-143 (Aerospace Medical Research Laboratory, Wright-Patterson Air Force Base, Ohio).

ROHMERT, W., MAINZER, J. and KANZ, U., 1988, Individuelle Unterschiede der Vektogramme von isometrischen Stellungskräften. *Zeitschrift für Arbeitswissenschaft*, 42, 102–5.

SCHOORLEMMER, W. and KANIS, H., 1992, Operation of controls on everyday products. In *Proceedings of the Human Factors Society 36th Annual Meeting — 1992* (Santa Monica: Human Factors Society), 509–13.

STEENBEKKERS, L.P.A, 1993, *Child Development, Design Implications and Accident Prevention*. Physical Ergonomics Series (Delft: Delft University Press).

STEENBEKKERS, L.P.A. and Van Beijsterveldt, C.E.M. (eds), 1998, *Design-Relevant Characteristics of Ageing Users* (Delft: Delft University Press)

THOMPSON, D., 1975, *Ergonomics Data for Evaluation and Specification of Operating Devices on Components for Use by the Elderly* (Loughborough: Institute for Consumer Ergonomics).

VAN DE KERK, B. and VOORBIJ, L. 1993, *Hoe sterk zijn kinderen? Onderzoek naar krachten van kinderen*. Internal report (Delft: Faculty of Industrial Design Engineering).

Table 18. Push (N) with two hands. Averages, SD and number of subjects per age category (Steenbekkers and van Beijsterveldt 1998).

age [years]	women					men					all		
	n	s	P5	x	P95	n	s	P5	x	P95	n	s	x
20–30	68	90	235	333	520	54	123	309	508	746	122	137	411
31–49		93		317			121		490				
50–54	35	70	177	307	407	35	131	196	459	687	70	129	383
55–59	49	122	177	294	407	46	107	299	485	638	95	149	386
60–64	53	76	196	288	432	44	88	280	424	584	97	106	350
65–69	51	66	137	250	363	50	111	265	409	569	101	121	329
70–74	62	71	132	230	348	59	74	211	361	466	121	98	294
75–80	37	57	113	219	314	36	103	150	349	525	73	105	283
80+	34	44	88	165	245	33	87	186	295	456	67	95	229
total							115		468				

Table 19. Pull (N) with two hands. Averages, SD and number of subjects per age category (Steenbekkers and van Beijsterveldt 1998).

age [years]	women					men					all		
	n	s	P5	x	P95	n	s	P5	x	P95	n	s	x
20–30	68	56	151	240	336	55	73	246	349	504	123	84	289
31–49		71		247			80		354				
50–54	35	52	152	247	315	35	88	191	342	522	70	86	294
55–59	50	120	159	262	341	46	86	243	374	517	96	118	315
60–64	53	64	146	238	349	44	80	205	330	497	97	85	280
65–69	51	60	114	212	307	50	75	229	321	473	101	87	266
70–74	62	55	114	194	275	59	60	184	288	401	121	75	240
75–80	37	46	112	184	264	36	71	139	287	399	73	79	234
80+	35	40	76	155	211	33	66	161	258	374	68	75	205
total		65		233			78		343				

Pushing and Pulling Strengths

S. N. Imrhan

Industrial and Manufacturing Systems, Engineering Department, University of Texas, Arlington, TX 76019, USA

1. INTRODUCTION

Pushing is the act of applying a muscular force in a direction away from the body; pulling is the opposite. Directions oblique to the trunk may be regarded as push or pull depending on the direction of the resultant force. In some cases the distinction between a pull and a lift may not be clear, as when extending the arm in front the body and gripping and raising an object from below the shoulder. A pull is almost always characterized by grasping the object, except in rare cases as when using a hook grip. A push may be applied with or without a grasp, as when the palmar surface of the open hand and fingers makes contact with the wall of a large box. The hand-handle contacts during pushing or pulling vary considerably but, for convenience, they may be classified into three main types: (i) full-hand power grasping, (ii) finger pinch grasping, and (iii) flat hand contact.

Pushing and pulling are executed for either moving an object from a state of rest or accelerating it from a state of constant motion (dynamic exertion), or stabilizing it (static exertion). Push forces may be exerted by any part of the body but only hand pushing/pulling have been identified as workplace activities of enough frequency or importance for ergonomic investigation.

Depending on the workspace design and muscular strength requirements, pushing/pulling tasks in the workplace may be performed in a variety of postures, among the most common are standing erect, stooping, squatting, kneeling, sitting, and lying down, and they may be performed with one hand or both hands. Thus any discussion of push/pull strengths must specify the posture and hand-object coupling characteristics clearly.

Push/pull activities occur in a wide variety of tasks; they assume importance when push/pull muscular strengths may be inadequate for a task and may result in (i) failure to accomplish the task properly, (ii) impairment to health (such as cumulative strains), or (iii) unsafe work conditions. Push/pull strengths have been investigated to determine the strength capabilities of people while pushing or pulling under a variety of conditions. In some of these studies the variables were experimentally controlled whereas others have simulated in industrial tasks. The essence of their results is briefly discussed in this article. More details may be obtained from my literature review (Imrhan 1999) or from the individual studies listed there.

2. FACTORS INFLUENCING PUSH/PULL STRENGTHS

Push or pull strength may be defined as the maximum force that a person can generate at the interface of the body (usually the hand) and the object being pushed or pulled. It is usually measured at this interface as the peak or mean force from a maximal voluntary push/pull (MVC) exertion lasting about 4–6 s under specified conditions. (An MVC exertion implies the maximum effort a person is willing to produce without pain or significant discomfort.) The mean is usually taken as the average force within the middle 3 s. Push/pull strengths, like almost all types of muscular strength, are influenced significantly by a number of factors relating to (i) the type of task, (ii) the method of performing the task, (iii) work equipment, and (iv) personal characteristics (e.g., body size and weight).

The degree of influence has been investigated for many factors, including type of exertion (static or dynamic), number of hands performing the exertion, general body posture, handle characteristics, hand–handle interface characteristics, use of gloves, height of application of the forces by the hands, distance between the object or handle and the body (or reach distance), direction of force application, frequency of force application, endurance, body support characteristics, foot–floor or shoe–floor traction, speed of motion of the body or hands, environmental stressors, and characteristics of the person. An understanding of how these factors influence push/pull strength can help practitioners in reducing, minimizing, or eliminating strains; this may help to improve work and workplace design, or equipment design. These influences are discussed below. To clarify the conditions of pulling, "strength" in these discussions will refer to static push/pull exertions while standing, unless otherwise stated.

2.1. Body Posture

A comparison of all the major body postures in similar push/pull conditions has never been investigated. Hence comparisons are related only to conditions that have been tested (Table 1). These comparisons indicate significant differences among different working postures and in variations within each posture, such as standing and leaning backward, standing and leaning forward, and sitting. Results indicate that (i) in general, two-handed hor-

Table 1. Average push/pull force ranges from experimental studies

Type of strength	Posture	Male force (N)	Female force (N)
Two-handed static push	Standing	400–620	180–335
Two-handed static pull	Standing	310–370	180–70
Two-handed dynamic pull	Standing	200–290	170–430
Two-handed dynamic push	Standing	225–500	160–180
One-handed push/pull	Standing	70–327	70–327
One-handed push/pull	Sitting with restraints	350–540	335–673
One-handed pull	Lying down prone	270–283	270–283
One-handed push	Lying down prone	285–330	285–330

izontal push and pull strengths were slightly greater when standing, but kneeling strengths may be greater for some positions of the arm; and (ii) for one-handed pull strength, standing strength was 37% greater than sitting strength.

The postures associated with the greatest strength are those which allow the body (i) to capitalize on the leverages about various joints, (ii) to use the body weight for enhancing push/pull force, and (iii) to maintain stability. Spreading the legs, one in front of the other, will allow greater pull or push strength than keeping them together; and kneeling on one knee is also likely to produce greater strength than standing with the feet together. More specifically, research has shown that pushing is strongest when the person leans forward with the pivoting (rear) foot behind the body's center of mass, and pulling is strongest when the pivoting (front) foot is in front of the center of mass. Also, push strength seems to be as great as or greater than pull strength when the feet are spread one in front of the other but are about the same when the feet are together.

2.2. Height of Pushing and Pulling

The height of the hand above the feet that produces the greatest strength depends strongly on body posture and degree of leaning forward or backward. In general, however, the best height for two-handed pushing while standing is between the elbow and hip; this is true whether pushing horizontally or at an angle to the horizontal. The best height for pulling is between hip and knee. Pushing upward with the hand above the shoulder or pulling upward with the hand below the elbow generates greater forces than with the hand in other positions because these forces are enhanced by the reactive forces of the feet on the floor (while standing) or buttocks on a seat (while sitting). More specific information on these heights depends on other factors, such as type of exertion (static or dynamic), elbow angle, foot–floor traction, etc.

2.3. Foot and Hand Reach Distances

The degree of reaching forward to push or pull while maintaining good balance is influenced by the positions of the feet, and determines the body's overall posture. Reaching should therefore influence push/pull strengths. Again, the nature of the effect is complex because of the interplay of other task variables. Among other muscular actions, pushing involves elbow extension and pulling involves elbow flexion. For many postures, push/pull strengths should therefore be greatest in the reaching range where the elbow is strongest for extension or flexion — i.e., with the elbow angle at about 90 to 110° — and where the additional effect of body weight is significant. The strength–reach trend is therefore an increasing/decreasing trend for both static and dynamic exertions.

When standing, the foot distance is defined as the distance from the hand–handle contact to the ankle of the pivot foot; and when sitting or standing, the reach distance is defined as the distance from the hand–handle contact to the shoulder (acromion). It may be expressed as a percentage of a person's shoulder height. Reach distance is influenced by the amount of space available for the arms, the height of hand–handle or hand–object contact, the nature of foot–floor contact, shoe–floor contact, or buttocks–seat contact, and it is also influenced by other factors (Kroemer 1974). Experiments have shown that the strongest standing pushing strengths occur when foot distance is about 90–100% of shoulder height, behind the hand; and the strongest pulling strengths occur when reach distance is 10%, in front of the hand.

2.4. Direction of Force Exertion

The direction which produces the greatest strength is where the body's reaction force is maximized at the contact with the floor. Pulling with the hands below the horizontal and pushing with the hands above the horizontal are among the best directions for maximizing strength when standing. Also, the force is greatest when directed along the arm axis for the fully extended arm. But when sitting and pulling with a harness around the chest, the horizontal direction (in the sagittal plane) seems to be the strongest.

For one-handed or two-handed exertions, pull/push strengths in the horizontal direction are strongest when the hands are directly in front of the body at about chest to shoulder height; and when the exertion is only with the right hand, the forces are greater to the right of the body than to the left. Pulling toward the body is about 10% stronger than pulling across the body; and for finger pulling at least, strengths obliquely across the body are about the same as strengths horizontally.

2.5. Body Supports and Traction

There are body supports and mechanisms for enhancing traction at the shoe-floor interface when standing, and for enhancing the buttocks–seat interface when sitting. They enable greater push/pull strengths by enhancing the reaction to the applied manual forces. Body supports are most influential when the applied force is perpendicular to them. Suitable supports for standing exertions include panels, walls, and fixed footrests; and for sitting exertions, a stable chair backrest or footrest may be used for pushing and a harness may be used for pulling. Research suggests the enhancement may be as much as 50 N when the shoe-floor coefficient of friction increases from 0.3 (poor traction) to 0.6 (good traction). Postural effects seem to be more influential for lower traction than for higher traction.

2.6. Speed of Walking

Research by Resnick and Chaffin (1995) has shown that, regardless of their (MVC) muscular strengths, people push (dynamically) with greater forces at greater speeds of walking; and the maximum speed of walking while pushing or pulling a load depends on a number of factors, such as shoe–floor coefficient of friction, personal strength, load weight, handle type, and handle height. Within the shoulder to knee range, people also seem to push faster when the load is lighter. For one-handed pulls, peak static strength is achieved later in the range of pull for higher pull speeds, and earlier for lower pull speeds. In other words, greater space must be allowed for a person to achieve maximum pull strength at higher pull speeds.

2.7. Frequency of Pushing and Pulling

Dynamic strength decreases as the number of repetitions of a muscular exertion increases. Experimental data confirms that, in simulated work tasks, people choose lighter loads to push or pull as the frequency of task performance increases.

2.8. Personal Characteristics and Gender

The available empirical data indicates that females are 0.5–0.9 times as strong as males for pushing or pulling. Some of this variation is due to differences in the sample characteristics and test procedures among the various experimental studies. It seems also that

the greater ratios occur where the legs play a more important role in strength exertion, such as pulling at low to medium height, and in situations where it is difficult to exert one's maximal effort. Body weight is the most influential anthropometric variable on push/pull strength. It adds to strength when leaning forward during pushing or leaning backward during pulling. However, no single anthropometric variable or combination of anthropometric variables can predict push/pull strengths accurately enough for ergonomic applications. Practitioners should therefore consult tables (Snook and Ciriello 1991, Ayoub and Mital 1989) to determine push/pull strength capabilities under specific work conditions.

3. MAXIMAL (MVC) STRENGTH VALUES

Measured average MVC push/pull strength values can be found in the ergonomics literature for several different studies conducted by different researchers. These values vary widely because of the different conditions in which the tests were conducted and because of the different physical characteristics among subject samples in the different studies. In view of that, only generalized results will be given here. More details, including test conditions, can be found in the ergonomics literature. Be careful when applying any data to task design or evaluation.

- Two-handed standing static push   Male   400–620 N
                                     Female 180–335 N
- Two-handed standing static pull   Male   310–370 N
                                     Female 180–270 N
- Two-handed standing dynamic push   Male   225–500 N
                                     Female 160–180 N
- Two-handed standing dynamic pull   Male   310–370 N
                                     Female 180–270 N
- One-handed static push or pull,   Standing   70–327 N
  combined male and female sets   Sitting
                                     restrained 350–540 N
                                     Lying down 270–330N

4. AT WORK

The previous sections provide information on the various factors that affect maximal muscular efforts. They may be used for rough estimates of how stressful or helpful a change in work conditions would be. For example, we know that suitable shoe and floor material (kept clean and dry) can allow a person to push 12 times as hard as on poor shoe–floor contact, and having handles on carts at elbow to hip height minimizes muscular strain when push forces are applied horizontally but not when they are applied at a sharp angle to the horizontal, and so on. However,

another kind of application of push/pull strengths is in determining how much load a worker is expected to work with under various conditions of work. These loads or forces are actually submaximal values and should be called "maximal acceptable forces" instead of strengths.

There are no definite and extremely accurate values for maximal acceptable forces under all conditions, but useful practical guidelines exist for the industrial workplace. The best guidelines seem to be the tables of Snook and Ciriello (1991); they are based on simulated industrial tasks and each of their force values represents the "maximum initial or sustained push or pull force which a given percentage of the population can exert over a specified distance, at a specified frequency, and at a specified height without a significant chance of being injured or developing cumulative trauma disorders." Remember these forces are much less than the maximum forces people can exert voluntarily (MVC forces) in a single (maximal) effort, and sustained force is about 20% less than an initial force.

5. CONCLUSIONS

Push or pull strength depends on a variety of factors; among the most important are posture and factors associated with posture. The condition or level of each factor must be specified for any meaningful strength estimate. Push/pull forces that people are willing to exert for a typical industrial task are known as maximum acceptable forces, not strengths (MVC forces). Maximum acceptable forces are significantly less than MVC forces. There are no accurate quantitative models that can predict MVC forces or maximum acceptable forces for pushing or pulling, and the best estimate should be derived from published tables which typically include means, standard deviations, and sometimes percentiles.

REFERENCES

AYOUB, M.M. and MITAL, A., 1989, *Manual Materials Handling* (New York: Taylor & Francis).

IMRHAN, S.N., 1999, Push-pull force limits. In Marras , W. S. and Karwowski, W. (eds) *The Occupational Ergonomics Handbook* (Boca Raton FL: CRC Press).

KROEMER, K.H.E., 1974, Horizontal push and pull forces. *Applied Ergonomics*, 5(2), 94–102.

RESNICK, M.L. and CHAFFIN, D.B., 1995, An ergonomic evaluation of handle height and load in maximal and submaximal cart pushing. *Applied Ergonomics*, 26(3), 173–78.

SNOOK, S.H. and CIRIELLO, V.M., 1991, The design of manual handling tasks: revised tables of maximum acceptable weights and forces. *Ergonomics*, 34(9), 1197–1213.

Recumbents

B. J. Daams

Daams Ergonomie, Mauvezand 17, 1251 JC Laren, The Netherlands

1. INTRODUCTION

Recumbent bicycles are relatively popular in The Netherlands, but they are still few in number. There are > 15,000 recumbents to a population of 16 million. Recumbent cyclists are generally convinced of the advantages of their vehicle. Their opinion is investigated by Daams (1999a, b), who gives an overview of the ergonomic aspects of recumbents and compares them with upright bicycles.

2. ADVANTAGES

The results indicate that recumbents are generally more comfortable, more efficient and safer than upright bicycles. Advantages are more comfortable sitting and looking around, more efficient and thus faster cycling, and better safety (stability is at least equal, better overview, better brake properties because of bicycle geometry and the location of the center of gravity, and less damage when falling because it is not the head that hits the ground first). Disadvantages are the price, getting on and off takes

Figure 1. Typical recumbent for city use: the Basic from Flevobike (courtesy: Flevobike).

Figure 2. Typical recumbent for touring: a M$_5$, fully loaded (courtesy: Bram Moens, M$_5$).

slightly more time, looking over the shoulder may be more difficult and adjusting for leg length takes some effort.

3. SUITABLE FOR VARIOUS PURPOSES

The usefulness of a recumbent depends very much on the type of recumbent (Figures 1–4) and the purpose for which it is used.

Recumbents are most advantageous on long distance rides (for commuting and cycling tours), in races and as a post delivery bike for long distances. Furthermore, recumbents are ideal for cyclists with certain physical problems, like back pain, saddle pain, and pain in wrist and arm.

Recumbents are less suited than conventional bikes for city use (short distances shopping), for all terrain use and as post delivery bike in the city (short distance).

4. ERGONOMIC ASPECTS

Some ergonomic aspects of recumbents are discussed in the following.

4.1. Recumbent Seat

The seat supports three parts of the human body (Figure 5): shoulders, back and bottom. The right size is very important in a recumbent seat. An oversized bottom support interferes with the movement of the legs while cycling; an undersized bottom support makes the cyclist slide down the seat. If the back support is not the right size, the chair does not fit and is very uncomfortable. A lumbar support is important.

(a)

(b)

Figure 3. Two typical recumbents for racing: a carbon low racer with tail fairing (a) and a fully faired carbon racing recumbent (b) (both courtesy: Bram Moens, M$_5$).
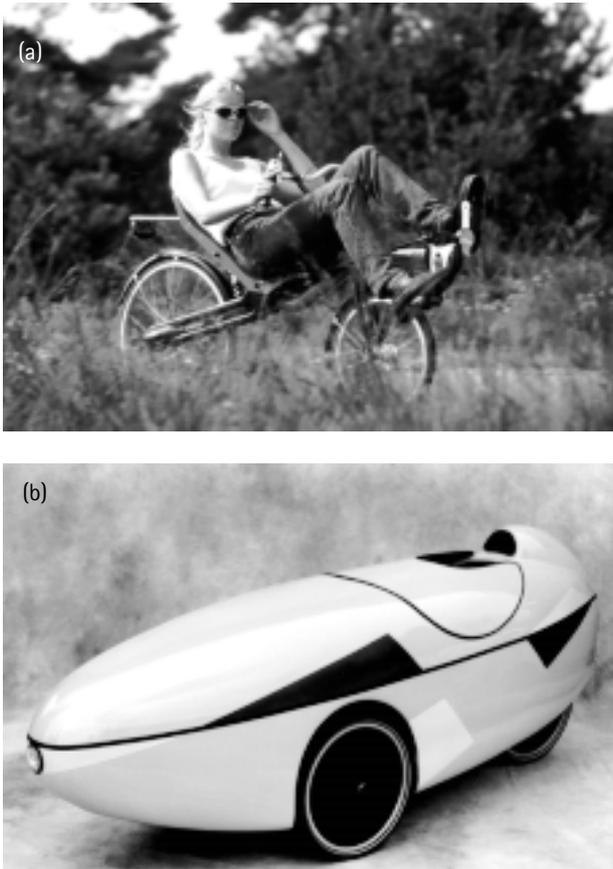
Figure 4. Two typical commuting recumbents: in summer the Focus from Challenge (a; courtesy: Challenge) and in winter the faired C-Alleweder from Flevobike (b; courtesy: Flevobike).

Most seats are available only in one or two sizes. Ideally, a recumbent seat should be made to measure because no two backs are alike (Staarink 1995).

Women should be taken into account when designing recumbent seats. They tend to have more of a bottom, for which the seat should provide sufficient space. Until now no such recumbent seat could be found. Insufficient bottom space leads to insufficient lumbar support, and thus to discomfort during long cycling tours.

4.2. Speed

Recumbents are much faster than conventional bikes, especially when they are faired. The World Speed Record in 1999 was 110 km/h in a 200-m sprint with a flying start, the Hour Record is 81.158 km, both in a fully faired recumbent by amateur cyclists. The Hour Record on a conventional bike is 56.357 km — by Chris Boardman, a professional, though he used an "arms-forward" aerodynamic position that has now been banned by the world governing body for cycling, the UCI. For consumer use, the speed of recumbents is slightly higher than that of conventional bikes (depending on the types) due to better aerodynamics.

4.3. Posture

The posture of the rider on a recumbent bicycle should be such that the feet should not be located higher than the heart, to prevent circulation problems. Cold and "sleeping" feet will be the consequence of a bottom bracket that is too high relative to the heart. When the feet are too low, relative to the seat, the rider will push himself out of the seat when cycling.

The posture of a recumbent rider is defined (Figure 6) by: χ, the shoulder angle; ψ, the back angle; θ, the pedaling angle. The resulting angle between body and upper leg (ω) determines to a large extent the maximal power output. An angle between 105
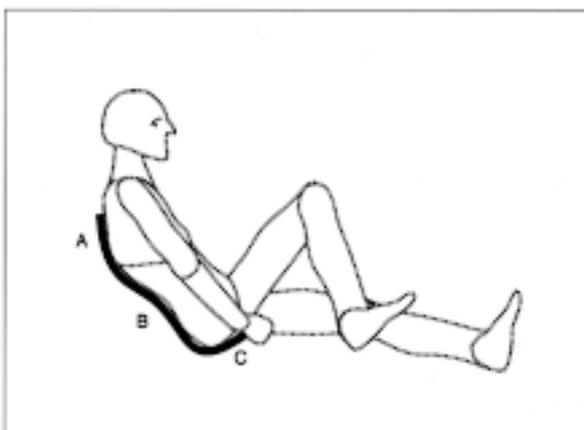
Figure 5. Recumbent seat consisting of three supports (Daams 1999a): support of the shoulders; support of the back; and support of the bottom.
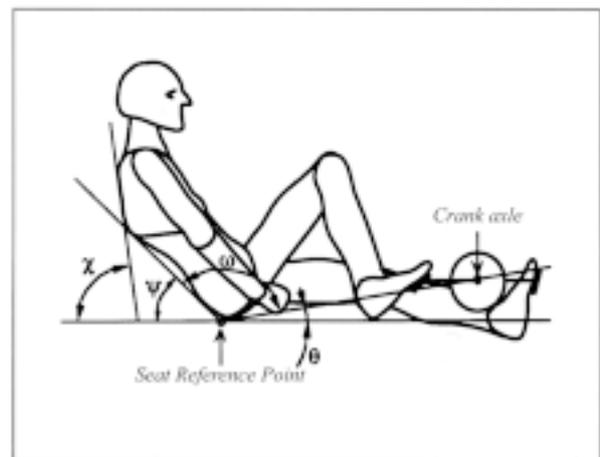
Figure 6. Posture of a recumbent rider is defined by: χ, the shoulder angle; ψ, the back angle; θ, the pedaling angle. The resulting angle between body and upper leg (ω) determines to a large extent the maximal power output (Daams 1999a).

and 130° is advised (Reiser and Peterson 1998). The posture also influences the aerodynamics of the recumbent (which influences the speed), the feelings of comfort and the view on the road. Investigation is needed into the relation between posture, power output, aerodynamics and endurance.

4.5. Muscle Use

There is some controversy whether the muscles used for recumbent cycling are the same as those used for conventional cycling. Received wisdom says the leg muscles need about as much as a year to adapt to recumbent cycling, imply that muscle use is different. However, no research on this subject could be found.

5. CONCLUSION

The conclusion is that recumbents are more ergonomic than conventional bicycles for commuting, for cycling tours, in races and for cyclists with certain physical problems. The popularity of recumbents is, however, not proportional to the ergonomical advantages. Why recumbents are not more popular may have a number of reasons, mainly concerning unfamiliarity with recumbents, the image of the bicycle, the attitude of other people, and the price. These factors are slowly changing in favor of the recumbent. In The Netherlands, recumbents are growing more popular and less expensive, and it is expected that this will continue over the coming years.

REFERENCES

DAAMS, B.J. 1999a, Een inventarisatie van de ergonomische aspecten van ligfietsen (I). *Tijdschrift voor Ergonomie*, 24, 42–50.

DAAMS, B.J. 1999b, Een inventarisatie van de ergonomische aspecten van ligfietsen (II). *Tijdschrift voor Ergonomie*, 24, 68–74.

REISER, R.F. and PETERSON, M.L. 1998, Lower-extremity power output in recumbent cycling: a literature review. *Human Power*, 13, 6–13.

STAARINK, H.A.M. 1995, *Sitting Posture, Comfort and Pressure* (Delft: Delft University Press).

Sleeping Postures

B. Haex and T. Huymans

Biomechanics and Engineering Design, Celestijnenlaan 200 A, B-3001
Leuven, Belgium

1. INTRODUCTION

Both psychological factors and physical conditions during bed rest affect the quality of sleep. Concentrating on the latter component, most authors consider a resting place as being ergonomically justified when the entire musculoskeletal system can recuperate well: muscles have to relax while the unloaded intervertebral disks are rehydrating. Owing to an insufficiently adapted sleeping system or an incorrect sleeping posture the human body — especially the vertebral column — is often supported unsatisfactorily, leading to low back pain or sleep disorders in general. This text will concentrate mainly on the prevention of back pain and on pressure distribution since other components (e.g. moisture, heat transportation) are only affected marginally by sleeping posture.

During sleep a local ischemy will arise in body zones in contact with the sleeping system. This ischemy generates metabolic substances that stimulate the sensible nerve extremities, which will cause the person to change their posture before it gets painful. Posture changes are necessary to avoid a pressure overloading of soft tissues and to prevent muscle stiffness. Most people have a preferred sleeping posture and change posture ~20 times a night. These facts make it almost impossible to define an overall ideal sleeping posture. Section 2, therefore, discusses different sleeping postures and how they influence back support. Section 3 discusses how back support can be measured and modeled to evaluate correctly an individual's posture on a sleeping system.

2. POSTURES

Most people have a preferred posture to fall asleep in, it often being the fetal position. Furthermore, the sleep locomotion system starts working, resulting in 65% of people sleeping in lateral recumbency, 30% in posterior recumbency and the remaining 5% in a prone position. Sleeping postures are also age-related: older people mostly have a preferred side to sleep on (especially on the right side), they rarely sleep in a prone position, they sleep for a shorter time and change their position less frequently (De Koninck *et al.* 1992).

2.1. Prone Recumbency

In spite of a good distribution of body weight over a large contact surface, a prone sleeping position is the most unfavorable posture in relation to back support, even on a well-qualified sleeping system. Gravity causes the most heavy middle part of the human body to sink deeper into the mattress while the legs remain stretched. Consequently, the lumbar lordosis will increase significantly and will augment the pressure on the facet joints at the posterior side, while soft tissues (e.g. ligaments) will be under tension at the anterior side. This effect may result in a hyperlordosis on soft mattresses and some pressure-distributing sleeping systems due to an increased mattress indentation in the pelvic zone.

When sleeping in a prone position, the combined effect of (1) the body weight resting on the rib cage and (2) the intestines being
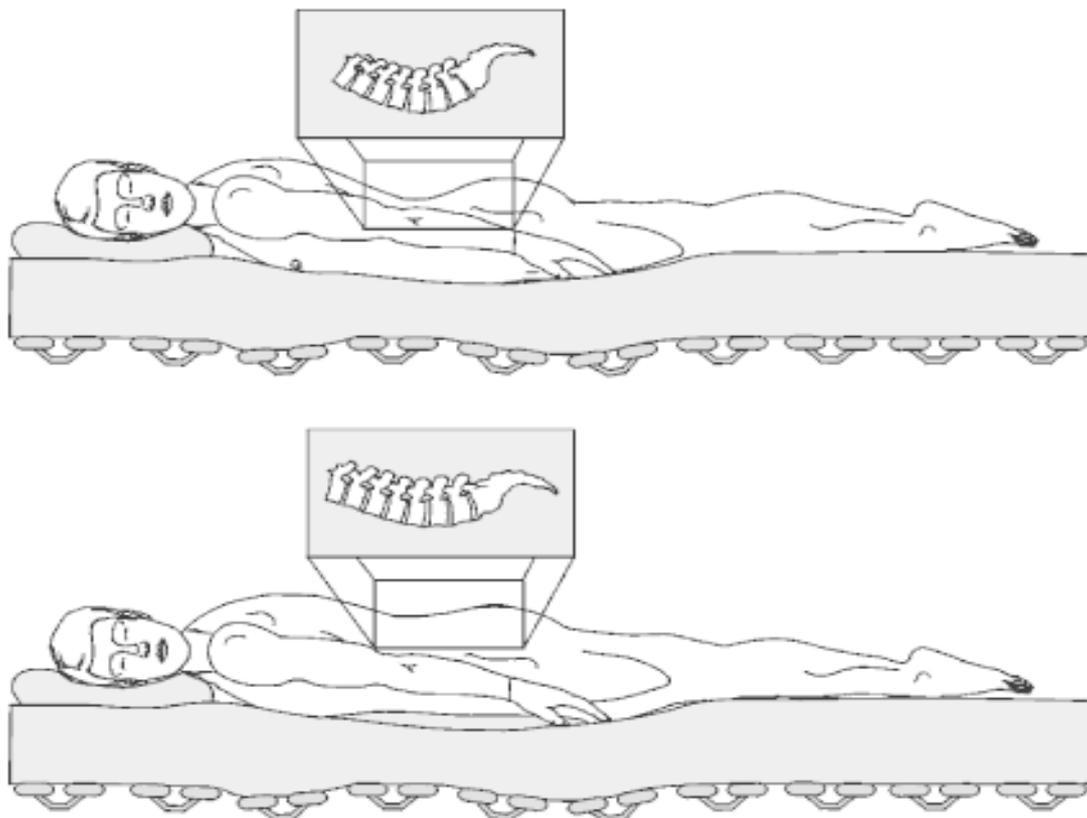
Figure 1. Increased lumbar lordosis when sleeping in a prone position (top) and improved the back support (bottom).
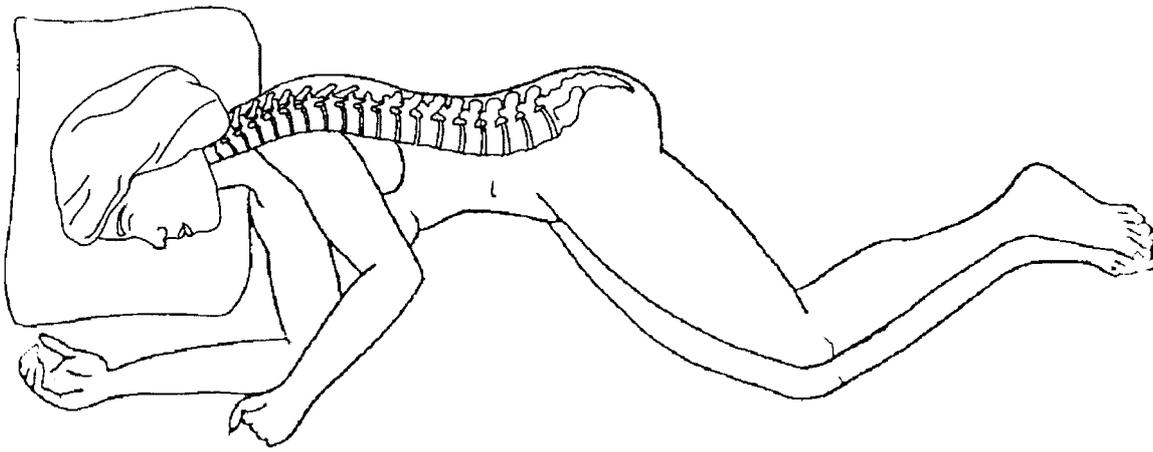
Figure 2. Correct lateral position.

pressed against the diaphragm enlarges the pressure on the lungs, which may cause respiratory problems. Further, the head is mostly turned sideward to improve breathing, which increases neck rotation. Consequently, an increased spinal loading occurs because several facet joints of most cranial vertebra are compressed at the side to which the head is turned, while ligaments at the other side are under tension. Also blood vessels may be compressed causing headaches, dizziness and other disorders.

In spite of all enumerated disadvantages — especially the increased lumbar and cervical spinal loading — many people prefer a prone position. It is, therefore, advisable to improve spine support by a simple means: sleeping without a head cushion will reduce neck rotation, which can even be improved by putting a pillow under the shoulder and the rib cage at the side to which the head is turned, or by lifting the arm at that side. Putting a pillow under the belly easily restores the natural lumbar lordosis. The same smoothening of an eventual hyperlordosis can be reached by elevating the knee and the hip at the side to which the head is turned, which can be enhanced by placing a pillow under the hip and the knee at this side. In fact, most presented corrections implement a slight position shift to lateral recumbency.

2.2. Lateral Recumbency

The lateral position is the most adopted sleeping posture, and it can support the human spine correctly in case both sleeping system and head cushion are well conceived: the spinal column is a straight line when projected in a frontal plane, while natural curves (cervical lordosis, thoracic kyfosis and lumbar lordosis) are maintained. There are no differences between sleeping on the left or on the right side, except from the weight of the liver working on the stomach and the lungs when sleeping on the left side.

Owing to the decreased contact surface and the center of gravity being more elevated, lateral recumbency is an unstable sleeping position, which can be altered by a correct position of the extremities. Bending arms and legs enlarges the support area and thus improves stability, but care should be taken not to apply torsion on the spine, to which intervertebral disks are especially vulnerable: when turned about a longitudinal axis, both pelvic and shoulder girdle should adopt the same angle.

Pillows can be an easy solution for the prevention of torsion of both pelvic and shoulder girdle: putting a head cushion or blanket between the knees will stabilize an elevated leg in a horizontal position. This position also avoids an asymmetric loading of the spine, which might induce a scoliosis. A similar effect can be reached by putting a pillow between the elbows. Because the human body will strive for minimal loading while sleeping, most people will automatically adjust their position and use a blanket to obtain the described minimal loading of the spine.

People — and muscles — are relaxed when sleeping in a fetal position; upper legs and trunk should form an angle of 135° to avoid a hyperlordosis (when stretching legs) or a smoothened lordosis (when elevating knees to close to the trunk). Further, it is clear that people with major spinal disorders (e.g. people with
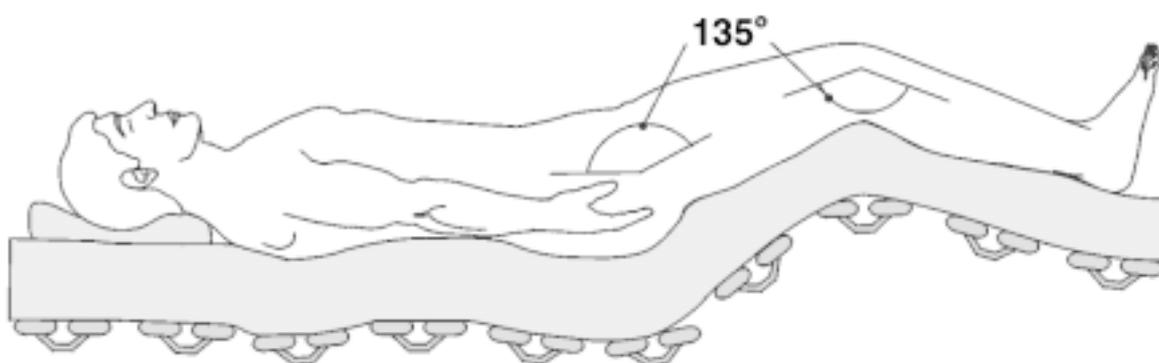
Figure 3. Semi-Fowler's position on an adjustable bed.

a spinal injury) need a special treatment: a moderate spine torsion or flexion may yield a stress relief at the level of injury, implying only a temporary solution because the spinal column will suffer an increased loading at other places.

2.3. Posterior Recumbency

The main advantage of sleeping in posterior recumbency is the fact that body weight is distributed over a large surface, resulting in pressure distribution and stability being optimized. The lumbar part of the vertebral column will mostly be positioned between a smoothened lordosis and a slight kyphosis, depending on (1) the kind of sleeping system, (2) the natural curves of the spine and (3) muscle tension while sleeping. When a sleeping system is too soft, places where body weight is concentrated (e.g. the hip zone) will sink deeply into the mattress. Some muscles may be well relaxed in this position, but the spine certainly will not: the pelvis will cant backward resulting in a complete and unnatural lumbar kyphosis. At the anterior side intervertebral disks will be compressed while soft tissues (e.g. ligaments) will be under tension at the posterior side. To support the natural cervical lordosis it is best to use a head cushion, although it is less obligatory compared with lateral recumbency. Further, people with respiration problems had better avoid sleeping in posterior recumbency: apnea and snoring frequency are much higher in this posture.

To discuss how posture should be adjusted in case of low back pain, neck problems or muscle stiffness is out of the focus of this chapter. Only the semi-Fowler's position will be mentioned here, it being a common relaxing posture for many patients. It can be adopted in case of an adjustable bed: both hip and knee joints have an angle of 45°, resulting in a relaxed iliopsoas muscle and a slightly smoothened lumbar lordosis. It is remarkable that astronauts experience more or less the same posture (Drerup and Hierholzer 1994).

2.4. Posture Evaluation

A reliable sleeping posture diagnosis can be made based only on invariant shape properties of the spine, e.g. preservation of the lumbar lordosis. Mattresses and mattress supports with different stiffness and different comfort zones (softer shoulder zone, harder pelvic zone, etc.) can be evaluated by two different techniques

measuring spinal deformations during bed rest. First, a 2D camera system is suited for both posterior and lateral recumbency; second, 3D video-raster-stereography (VRS) is appropriate to measure 3D postures in case the back surface are visible, i.e. all postures are between lateral and prone recumbency.

3. EVALUATION

3.1. 2D Evaluation

A first experimental set-up designed to measure lumbar and thoracic spinous process positions during bed rest is a camera system. For lateral recumbency 17 reflecting markers are glued to the skin to cover the spinous processes and are detected by a Qualisys MAC-Reflex, camera system. A 2D approximation of the vertebral curvature is obtained by projecting the markers in a frontal plane and by calculating their position relative to two reference markers. For posterior recumbency a system with pins piercing the mattress measures the spinous process positions. A conducting strip glued to the spinous processes and an electrical circuit permit to ensure permanent contact between the pins and the spinous processes. Two reference markers allow the recalculation of the positions of the markers mounted on the lower side of the pins to the actual spinous process positions.

When concentrating on the optimal sleeping system for normal healthy people — as it is suggested by most ergonomic specialists — it has to support the human spine such that it adopts its natural position, which is assumed to be the same as it takes in the upright position. If this thesis is presupposed then an optimal body support for lateral recumbency gives rise to the spinal column being a straight line when projected in a frontal plane. For posterior recumbency an optimal support gives the spinal column the same thoracic kyfosis and lumbar lordosis as in the upright position, yet slightly smoothened by the loss of body weight working in longitudinal direction on the spinal column. Consequently a small prolongation of the spine occurs, as during weightlessness.

3.2. 3D Evaluation

The fact that people do not limit themselves to 2D sleeping positions justifies the need for more reliable and realistic
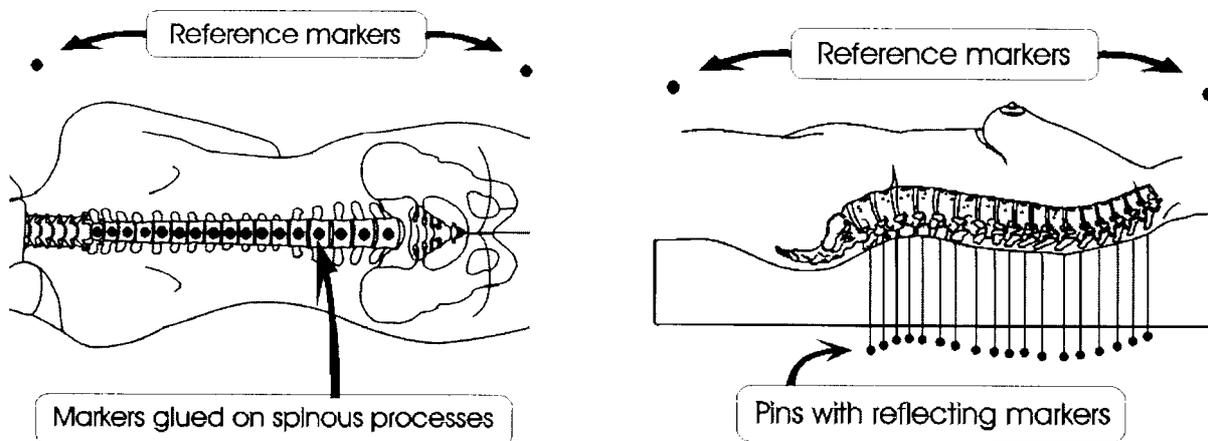
Figure 4. Camera measurements for lateral (left) and posterior recumbency (right).

evaluations, by measuring both mattresses and supporting structures in 3D. As a result of comparing different measurement techniques, VRS is now used to locate the processi spinosi of a person on a sleeping system.

VRS is an objective analysis method for the assessment of 3D objects using a line raster diapositive and a camera. It is applied to measure the back shape of test persons and, therefore, acts as a powerful tool for the evaluation of different sleeping postures. The measurement technique can scan any 3D surface by projecting raster lines on the surface and captures these lines under a known and fixed angle. The measured surface can be reconstructed mathematically by fitting a dense point cloud, which is created based on triangulation algorithms (Krag *et al.* 1990).

Frobin and Hierholzer developed hardware and image processing techniques for the reconstruction of the 3D back surface to detect a lateral scoliosis. These techniques could be adapted to lateral recumbency applications. The camera and raster diapositive were optimally chosen: the image processing reconstructs the back surface with a mean error of 0.75 mm. A raster with more raster lines or a camera with a higher resolution is, therefore, not necessary.

As a result of the image processing a cloud of randomly distributed points, indicating the raster lines, describes the back surface. Using 1D linear interpolation the randomly distributed data points are transformed to a regular grid, which will simplify further calculations.

Owing to major differences between the upright and lying position, e.g. soft tissue influences, new dedicated software had to be developed starting from the regular point grid. We, therefore, developed Matlab, procedures (1) to model a back surface accurately and (2) to detect processi spinosi and other anatomical landmarks — sacrum point, vertebra prominens — based on asymmetry functions and convex/concave characteristics.

A reliable diagnosis only can be made based on invariant shape properties of the back. Frobin and Hierholzer proposed surface curvatures to analyze the shape of a human back. Owing to this, the landmark localization is independent of a patient's position and insensitive to moderate asymmetry and posture changes of the patient. The landmarks are, therefore, well suited for the objective definition of a body-fixed coordinate system.

In case of healthy people the medial sagittal plane is a symmetry plane that divides the back in a symmetrical left and right part. In case of lateral asymmetry of the back this is no longer true: a deformation of the spine — due to scoliosis or an incorrect mattress support — can cause lateral asymmetry. Software routines were developed to create a profile for these cases, dividing the back in a left and right part with minimal asymmetry. The entire profile of minimal asymmetry is found by connecting symmetry points, which are calculated in each horizontal cross-section. This line is called the symmetry line and is a rough approximation of the line through the processi spinosi. Drerup and Hierholzer (1994) described how the symmetry line is used to locate anatomical landmarks on the back, including the vertebra prominens C_7 (VP), the sacrum point (SP) and the left and right dimple of the spina iliaca posterior superior (DL en DR). All landmarks are determined from local surface curvature extrema. Further, the 3D line through the vertebral bodies is constructed out of (1) the symmetry line and (2) the positions of the anatomical landmarks, based on an empirical formula assuming that the vertebral rotation is equal to the surface rotation (Refshauge 1994).

4. RECOMMENDATIONS

- In a correct sleeping posture the human spine adopts its natural position, which is assumed to be the same as it takes in the upright position, yet slightly smoothened by the loss of weightlessness working in longitudinal direction.
- It is almost possible to obtain a correct back support in a prone sleeping position. In case a person has no complaints (as to respiratory or back problems), lateral and posterior positions are equivalent.
- Owing to the fact that the human body will strive for minimal loading while sleeping, most people will automatically adjust their posture to their sleeping system.
- When changing a sleeping system it usually takes 2 weeks to get used to it. A temporarily worse back support on a new and better sleeping system is possible due to these adaptation symptoms.

REFERENCES

DE KONINCK, J., LORRAIN, D. and GAGNON, P., 1992, Sleep positions and position shifts in five age groups: an onthogenic picture. *Sleep*, 15, 143–9.

DRERUP, B. and HIERHOLZER, E., 1994, Back shape measurement using video rasterstereography and three dimensional reconstruction of spinal shape. *Clinical Biomechanics*, 9, 28–36.

KRAG, M.A., COHEN, M.C., HAUGH, L.D. and POPE, M.H, 1990, Body height change during upright and recumbent posture. *Spine*, 15, 202–7.

REFSHAUGE, K.M., 1994, The relationship between surface contour and vertebral body measures of upper spine curvature. *Spine*, 19, 2180–5.

Static and Dynamic Strength

K. H. E. Kroemer

Industrial And Systems Engineering, Virginia Tech, Blacksburg, VA, and Eri, Inc., A Consulting Company, PO Box 3019, Radford, VA 24143-3019, USA

Generation of muscle strength is a complex procedure of nervous activation and control of myofilaments. It may involve substantial shortening or lengthening of muscle, i.e. a concentric or an eccentric effort, respectively; or there may be no perceptible change in length, i.e. the effort is isometric. Mechanically, the main distinction between muscle actions is whether they are "dynamic" or "static".

Static strength. Per definition, muscle length is constant during an isometric effort. Thus, involved body segments do not move; in physics terms, all forces acting within the system are in equilibrium, as Isaac Newton's First Law requires. Therefore, the physiological "isometric" case is equivalent to the 'static' condition in physics.

The static condition is theoretically simple and experimentally well controllable. It allows rather easy measurement of muscular effort. Therefore most of the information currently available on human strength describes outcomes of static (isometric) testing.

Measurements of isometric strength yield reasonable estimates of maximally possible exertion for most slow body link movements, if they are eccentric. However, the data do not estimate fast exertions well, especially if they are concentric and of the ballistic-impulse type, such as throwing or hammering.

Dynamic strength. Dynamic muscular efforts are more difficult to describe than static contractions. In dynamic activities, muscle length changes, and therefore involved body segments move. This results in displacement. The amount of travel is relatively small at the muscle but usually amplified along the links of the internal transmission path to the point of application to the outside, for example at the hand or foot.

Displacement and its time derivatives, velocity and acceleration, are of importance both for the muscular effort (as discussed earlier) and the external effect: for example, change in velocity determines impact and force, as per Newton's Second Law.

Definition and experimental control of dynamic muscle exertions are much more complex tasks than in static testing. Various new classification schemes for independent and dependent experimental variables can be developed. One such system for dynamic and static efforts includes the traditional tests as well as new ones (Marras *et al.* 1993). See the chapter on Muscle Terms — Glossary.

Setting the displacement (muscle length change) to zero — the *isometric* condition — one may either measure the magnitude of force (or torque) generated, its duration, or the number of repetitions that can be performed. Of course, since there is no displacement, its time derivatives are also zero.

Alternatively, one may choose to control velocity as an independent variable, i.e. the rate at which muscle length changes. Keeping velocity to a constant value, one generates an *isokinetic* muscle strength measurement. (Note that this *isovelocity* condition is often mislabeled "isokinetic".) Time derivatives of constant velocity, acceleration and jerk, are zero. Mass properties are usually controlled in isokinetic tests. The variables displacement, force (torque), time and repetition can either become dependent variables or they may be controlled independent variables. Most likely, force and/or repetition are chosen as the dependent test variables.

In some tests the amount of muscle tension (force) is set to a constant value. In such an *isoforce* test mass properties and displacement (and its time derivatives) are likely to become controlled independent variables, and time or repetition a dependent variable. This *isotonic* condition is, for practical reasons, often combined with an isometric condition, such as in holding a load motionless.

(Note that the term "isotonic" has often been applied falsely. Some older textbooks described the examples of lifting or lowering of a constant mass (weight) as typical for isotonics. This is false for two physical reasons. The first is that according to Newton's Laws the change from acceleration to deceleration of a mass requires application of changing (not: constant) forces. The second fault lies in overlooking the changes that occur in the mechanical conditions (length, pull angles and lever arms) at the muscle. Hence, even if there were a constant force to be applied to the external object (which is not the case), the changes in geometry would result in changes in muscle tonus.)

In the *iso-inertial* test, the external mass is set to a constant value. In this case, load (force, torque), time and repetition of moving the constant mass (as in lifting) and its displacement may be selected as independent or dependent variables.

The most common and general case of motor performance is "free dynamic": the person freely chooses all the independent variables of muscle use. Force, torque or some other performance measure serve as dependent output.

The complexity of dynamic testing explains why, in the past, dynamic measurements (other than isokinematic and iso-inertial) have been rarely performed in the laboratory. On the other hand, if one is free to perform as one pleases, such as in the "free dynamic" tests common in sports, very little experimental control can be executed.

For more information see the references.

REFERENCES

KROEMER, K.H.E., 1999, Human strength evaluation. In W. Karwowski and W. S. Marras (eds) *The Occupational Ergonomics Handbook* (Boca Raton: CRC Press), 205–27.

KROEMER, K.H.E., KROEMER, H.J. and KROEMER-ELBERT, K.E., 1997, *Engineering Physiology* (3rd ed.) (New York: van Nostrand Reinhold–Wiley).

KROEMER, K.H.E., KROEMER, H.B. and KROEMER-ELBERT, K.E., 2000, *Ergonomics: How to Design for Ease and Efficiency* (2nd ed.) (Englewood Cliffs: Prentice-Hall).

MARRAS, W.S., McGLOTHLIN, J.D., McINTYRE, D.R., NORDIN, M. and KROEMER, K.H.E., 1993, *Dynamic Measures of Low Back Performance* (Fairfax: American Industrial Hygiene Association).

SCHNECK, D.J., 1992, *Mechanics of Muscle* (2nd ed.) (New York: New York University Press).

WINTER, D.A., 1990, *Biomechanics and Motor Control of Human Movement* (2nd ed.) (New York: Wiley).

Static Muscle Strength

J. L. Smith

Department of Industrial Engineering, Texas Tech University, Lubbock, TX 79409-3061, USA

Static, or isometric, muscle strength is defined as a muscle contraction that results in no movement of the object that the muscles are resisting. Common examples of static muscle exertions include pushing against a solid wall, or pushing your hands together. When relating muscular activity to work, many ergonomists are concerned with only the dynamic aspects of the work activity. However, there are many examples of static muscle activity in industry. One example involves the carrying of loads, while the legs are involved in the dynamic muscle activity of walking, the arms are often in a fixed (static) posture, holding an object as the worker walks. Other static activities are associated with postures and maintenance of postures for long periods of time without moving. Typically, such activities are referred to as static muscle loading. Upper extremity problems can occur when workers have to hold their arms up while working. The actual hand work may be very dynamic, but the shoulders and upper arms may be under a great deal of static load to maintain a required posture.

To "match the capabilities and limitations of workers to the demands of their jobs," static strength capabilities must be established. Generally, static strength is isolated to one (or a few) muscle groups performing a simple activity. Most commonly reported static strength measures are for arms, legs, shoulder and back.

Assessment of static strength is popular due to its relative simplicity to measure, low equipment costs, short testing time and test–re-test reliability. The basic equipment needed for static strength testing includes a load cell with a display unit, a platform or testing station, and a variety of handles, connectors and braces. A variety of load cells are commercially available ranging in cost from a few hundred to > $1000. A typical setup would involve a load cell connected to a handle on one end and anchored to a fixed point on the other end. A chain or cable that is easy to adjust in length is used as a connector.

Instructions to subjects typically ask them to build up to a maximal exertion (within 2–5 s), and then hold that maximal exertion for 3–5 s. The steady-state maximal exertion is then recorded. It should be noted that the data being collected are maximal voluntary exertions, which may vary under differing testing conditions. The ergonomist should attempt to minimize variability through: (1) standardized instructions, (2) isolation of testing from other subjects and/or observers and (3) not providing feedback to subjects regarding their strength performance. Adherence to these guidelines can help reduce competition among subjects, and reduce the chances of someone overexerting himself/herself. Adequate rest time (1–3 min) should be allowed between exertions. Some test protocols require repeated testing until three values are obtained that are within 15% of each other. A frequent problem encountered in isometric strength testing is a false reading due to the subject jerking on the load cell and adding an acceleration component to the reading. This problem is most prevalent in systems that have a maximum hold option for the display. Thus a momentary false reading could be "held in memory" and recorded as the maximum value for that trial. By watching, a continuously changing display, the test administrator can observe a "steady-state" value and record that data point.

Static strength data obtained from the ergonomics laboratory at Texas Tech University are summarized in Table 1. The data were obtained from a population of relatively young (mostly 18–25-year-old) male subjects who participated in a variety of research projects. Table 2 presents static strength data for 551 industrial workers (Chaffin *et al.* 1977) ranging in age from 18 to 62 years (average age 28.9 years). The two data sets agree closely for male arm strength, but differ on leg strength values. The difference in leg strength values is likely due to slight differences in the posture assumed for the strength tests, which points out the importance of documenting the strength test protocol, so that when it is replicated in the future, the values will be comparable. In general, average female strength values would be expected to be 40–70% of the average male values. The strength values presented in Tables 1 and 2 are provided for reference only and may not reflect the strength capabilities of specific industrial populations.

Table 1. Average static strengths for male subjects (n=129) in Texas Tech ergonomics studies.

Muscle group	Average static strength (N)
Arms	378
Legs	1410
Stoop back	905
Standing back	722
Composite	1162

Table 2. Average static strength values from Chaffin, *et. al.*(1977) for male and female industrial workers (n=443 males and n=108 females).

Muscle group	Average static strength (N)	
	Males	Females
Arms	381	200
Legs	941	417
Torso	544	266

There is a good deal of variability in static strength data, especially if male and female data are pooled and as wider age ranges of subjects are incorporated. The coefficients of variation (the ratio of standard deviation to the mean) for the data in tables 1 and 2 range from < 15% to > 50%.

Static strength testing has some appeal as a screening tool to match the capabilities and limitations of workers to the demands of their jobs. In such cases, job demands are often examined as a "worst case" or "most typical case" in which the appropriate posture of the worker is assumed for purposes of static strength testing. The worker's capabilities are then assessed as the worker performs a static exertion. The primary argument against using static strength testing is that dynamic activities can not be adequately represented as a series of static tasks.

However, if a screening test is desired, NIOSH (1981) recommends that following questions be asked before accepting the screening test:

- Is it safe to administer?
- Does it give reliable, quantitative values?
- Is it related to specific job requirements?
- Is it practical?
- Does it predict the risk of future injury or illness?

The ergonomist should carefully consider the above questions before recommending a screening test. If satisfactory answers are not available for the above questions, legal issues over job discrimination might arise.

REFERENCES

CHAFFIN, D.B., 1975, Ergonomics guide for the assessment of human strength. *American Industrial Hygiene Journal*, 36, 505–10.

CHAFFIN, D.B. and ANDERSSON, G.B.J., 1991, *Occupational Biomechanics* (2nd ed.) (New York: John Wiley).

CHAFFIN, D.B., HERRIN, G., KEYSERLING, W.M. and FOULKE, J., 1977, *Pre-employment Strength Testing.* CDC-99-74-62 (Cincinnati: Department of Health, Education, and Welfare, NIOSH).

NIOSH, 1981, *Work Practices Guide for Manual Lifting.* DHHS (NIOSH) Publication no. 81-122 (Cincinnati: US Department of Health and Human Services).

WATERS, T.R., PUTZ-ANDERSON, V. and GARG, A., 1994, *Applications Manual for the Revised NIOSH Lifting Equation* (Cincinnati: US Department of Health and Human Services, Public Health Service).

Strength Testing

S. Gallagher

National Institute for Occupational Safety and Health, Pittsburgh Research Laboratory, PO Box 18070, Pittsburgh, PA 15236-0070, USA

1. INTRODUCTION

If we consider ergonomics to be an exercise in matching job demands to worker capabilities, one of the principal capabilities we must be concerned with is that of *human strength*. Our ability to evaluate different characteristics of muscular strength has increased dramatically over the past couple of decades with the development of new and increasingly sophisticated instrumentation. One would think that armed with such advanced techniques, we might be able to develop methods to conclusively identify workers at risk of injury in physically demanding jobs. Unfortunately, this has not yet proven to be the case. Instead, what these instruments have continued to point out is how intricate a function muscular strength really is, and how complicated and ambiguous its relationship is to musculoskeletal injury.

While we cannot just use isolated tests of strength to specify precisely who may be at risk of injury, studies have indicated that strength testing can be a useful tool for job design and, under certain circumstances, selection of workers for demanding jobs. However, because strength is such a complex phenomenon, there has often been some confusion regarding the proper application and interpretation of strength tests in ergonomics, especially among persons not thoroughly familiar with the limitations and caveats associated with the available procedures. The purpose of this chapter is to discuss some of the fundamental principles of strength assessment in ergonomics, so that these procedures can be better applied to control the risk of musculoskeletal disorders in the workplace.

1.1. What is Strength? (And what are we Measuring?)

Many of the complications associated with strength assessment arise from the simple fact that even our most sophisticated machinery *does not directly measure* the force or tension developed by a muscle in a living person. Instead, we can only observe the consequences of force development by a contracting muscle, or more likely, by a *combination* of muscles. There are many ways in which we can measure the effects of muscular contraction, and the techniques we use can have a dramatic impact on the strength readings we will obtain. Consider the situation illustrated in Figure 1. In this example, the muscle exerts a constant force of 1000 Newtons (N). However, the forces we measure can vary quite dramatically depending on where we place the force cuff — from 167 N if we place it near the wrist to 500 N if we place it near the elbow. Which value should we select as properly representing the muscular strength for this elbow flexion exertion?

The preceding example illustrates some important points with regard to strength assessment. Perhaps the most important is that "muscular strength is what is measured by an instrument" (Kroemer *et al.* 1990). It should also be clear from this example
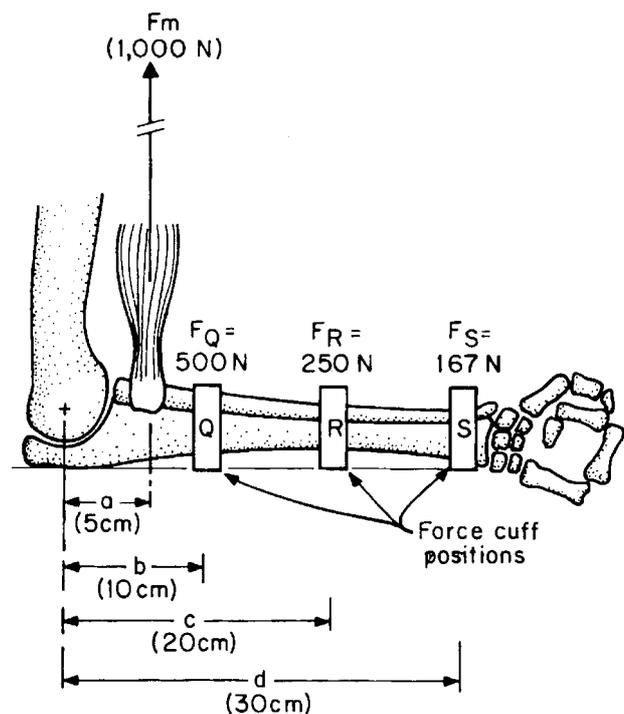
Figure 1. Given a constant muscle force (F_m), forces measured at various distances from the elbow will result in different force readings (F_Q, F_R or F_S).

that two researchers could perform an elbow flexion strength experiment on the same group of subjects, but if each selected different force cuff positions, they might end up with wildly differing estimates of strength. Differences in the strengths of various muscle groups in the published literature may be the results of differences in the procedures and measurement methods used by the experimenters. Thus, it is critical that any strength data presented be accompanied by a detailed account of the manner in which the data were obtained.

A few additional points need to be made with regard to the testing of human strength. We must be clear that what we are obtaining in such tests are not a person's maximal strength capability, but their maximal *voluntary* strength. The voluntary nature of the exertion introduces an unknown, but surely substantial, amount of variability in our measurements of strength. One can imagine two subjects with identical muscular strength capabilities, but with varying levels of motivation or discomfort tolerance, for example. We are likely to observe considerable differences in the voluntary force exerted by the two, but it should be understood that such results may be largely the result of psychological factors, and not differences in muscular strength *per se*. The important point to be made here is that not only are we unable to directly measure muscular force, what we *are* able to measure is modified by an invisible filter — a filter subject to a wide variety of influences and which will differ considerably for every person we test. To make matters worse, this filter would be expected to change even within a given individual on a given day. From the foregoing discussion, one can perhaps better appreciate some of the difficulties with establishing a definitive relationship between an individual's measured strength and that individual's risk of injury.

2. PURPOSES OF STRENGTH ASSESSMENT IN ERGONOMICS

There are a number of reasons people may want to collect human strength data. This article will discuss two of the most common uses of physical strength assessment in ergonomics: *job design* and *worker selection.*

2.1. Job Design

Probably the most effective use of worker strength evaluations is in the area of job design. Job design has been a primary focus of the psychophysical method of determining acceptable weights and forces. The psychophysical method attempts to determine workloads that are "acceptable" (a submaximal strength assessment) for populations of workers. Once the acceptable workloads for a population are determined, the job or task is designed to accommodate the vast majority of that population. It has been estimated that this approach to the design of lifting tasks might reduce the risk of back injuries by up to 33%.

2.2. Worker Selection

The purpose of worker selection and placement programs is to ensure that jobs which involve heavy physical demands are not performed by those lacking the necessary strength capabilities. It should be noted that this method is not the preferred strategy of the ergonomist, but is a provisional measure for the control of work-related musculoskeletal disorders (WMSD) where job design cannot be used to alleviate task demands. Nonetheless, this method can be effective in reducing the harmful physical effects caused by the mismatch of worker and job, given adherence to two fundamental principles. These principles are: (1) ensuring that the strength measures are related to the demands of the job, and (2) that strength assessment is performed only under circumstances where they can predict who may be at risk of WMSD. The literature has shown that worker selection is *only effective when a worker's strength capacity is equated with the demands of the job.* All too often, emphasis is placed on collecting data on the former attribute, while the latter receives little or no attention. The second issue that must be considered when worker selection is to be implemented is that of the test's *predictive value.* The predictive value of a test is a measure of its ability to determine who is at risk of future WMSD. In the case of job-related strength testing, the predictive value appears to hold only when testing individuals for jobs *where high risk is known.* Strength testing does not appear to predict the risk of injury or disease to an individual when job demands are low or moderate.

3. TYPES OF MUSCULAR STRENGTH ASSESSMENT AND THEIR USE IN ERGONOMICS

Muscular exertions can be divided into those which produce motion about a joint (dynamic exertions), and those which do not (isometric or static exertions). The vast majority of occupational tasks involve dynamic motions. Unfortunately, the complexity of such motion makes it more difficult to quantify. Static exertions, on the other hand, are easier to control and measure, but may be inappropriate to apply in situations where dynamic activity is present. Neither mode of strength testing is inherently better than the other — the key is to make sure that the test that is used relates to the application being studied. The following sections briefly describe the most common strength analysis techniques used in ergonomics. The first deals with isometric strength testing, the remaining sections describe various dynamic tests of strength. Greater detail on these strength assessment procedures can be found elsewhere (Gallagher *et al.* 1998).

3.1. Analysis of Isometric Strength

When a worker is called upon to perform a physically demanding lifting task, moments (or torques) are produced about various joints of the body by the external load. Often these moments are augmented by the force of gravity acting on the mass of various body segments. For example, in a biceps curl exercise, the moment produced by the forearm flexors must counteract the moment of the weight held in the hands, as well as the moment caused by gravity acting on the center of mass of the forearm. In order to successfully perform the task, the muscles responsible for moving the joint must develop a greater moment than that imposed by the combined moment of the external load and body segment. It should be clear that for each joint of the body, there exists a limit to the strength that can be produced by the muscle to move ever increasing external loads. This concept has formed the basis of isometric muscle strength prediction modeling.

The following procedures are generally used in this biomechanical analysis technique. First, workers are observed (and usually photographed or videotaped) during the performance of physically demanding tasks. For each task, the posture of the torso and the extremities are documented at the time of peak exertion. The postures are then re-created using a computerized software package, which calculates the load moments produced at various joints of the body during the performance of the task. The values obtained during this analysis are then compared to population norms for isometric strength obtained from a population of industrial workers. In this manner,
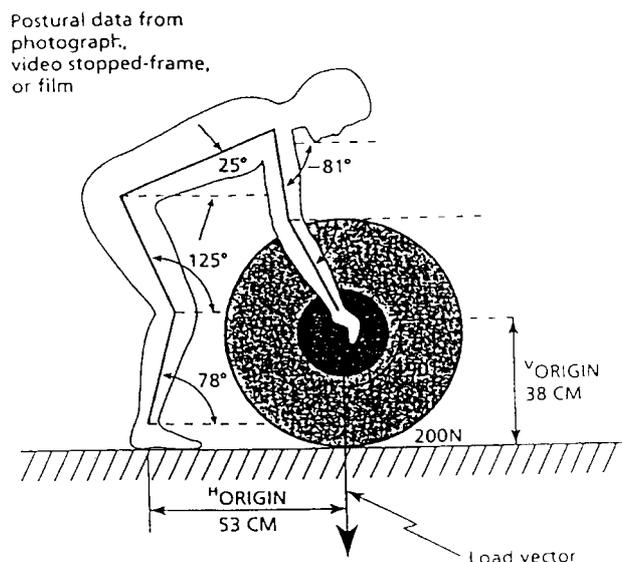
Figure 2. Postural data required for analysis of joint moment strengths using the isometric technique.

the model can estimate the proportion of the population capable of performing the exertion, as well as the predicted compression forces acting on the lumbar discs resulting from the task (Chaffin and Andersson 1991).

3.2. Iso-inertial Methods

3.2.1. The strength aptitude test

The Strength Aptitude Test (SAT) is a classification tool for matching the physical strength abilities of individuals with the physical strength requirements of jobs in the Air Force (McDaniel *et al*. 1983). The SAT is given to all Air Force recruits as part of their pre-induction examinations. Results of the SAT are used to determine whether an individual possesses the minimum strength criterion which is a prerequisite for admission to various Air Force Specialties (AFS). The physical demands of each AFS are objectively computed from an average physical demand weighted by the frequency of performance and the percent of the AFS members performing the task. Objects weighing less than 10 pounds are not considered physically demanding and are not considered in the job analysis. Prior to averaging the physical demands of the AFS, the actual weights of objects handled are

Figure 3. Incremental Weight Lift Machine. The barrier has been removed to expose the hidden stack of weights.

converted into equivalent performance on the incremental weight lift test using statistical procedures developed over years of testing. These relationships consider the type of task (lifting, carrying, pushing, etc.), the size and weight of the object handled, as well as the type and height of the lift. Thus, the physical job demands are related to, but are not identical to, the ability to lift an object to a certain height. Job demands for various AFS are re-analyzed periodically for purposes of updating the SAT.

In this technique, a preselected mass, constant in each test, is lifted by the subject using a device such as that shown in Figure 3. The amount of weight to be lifted is relatively light at first, but the amount of mass is continually increased in succeeding tests until it reaches the maximal amount that the subject voluntarily indicates s/he can handle. A unique aspect of this technique is that it is the only strength measurement procedure discussed in this document where results are based on the success or failure to perform a prescribed criterion task. The criterion tasks studied have typically included lifting to shoulder height, elbow height or knuckle height.

3.2.2. Psychophysical strength assessment

As mentioned previously, job design has been a primary focus of the psychophysical method of determining acceptable weights and forces. In this technique, subjects are typically asked to adjust the weight or force associated with a task in accordance with their own perception of what is an *acceptable workload* under specified test conditions. It can be seen from this description that this technique does not attempt to evaluate the maximum forces a subject is capable of producing. Instead, this procedure evaluates a type of "submaximal," endurance-based estimate of acceptable weights or forces.

In the context of lifting tasks, the following procedure is usually used in psychophysical strength assessments. The subject is given control of one variable, typically the amount of weight contained in a lifting box. There will usually be two 20-min periods of lifting for each specified task: one starting with a light box (to which the subject will add weight), the other starting with a heavy box (from which the subject will extract weight). The box will have a hidden compartment containing an unknown (to the subject) amount of weight, varied before each test, to prevent visual cues to the subject regarding how much weight is being lifted. The amount of weight selected during these two sessions is averaged and is taken as the maximum acceptable weight of lift for the specified conditions. In psychophysical assessments, the subject is instructed to work consistently according to the concept of "a fair day's pay for a fair day's work": working as hard as s/he can without straining himself, or becoming unusually tired, weakened, overheated, or out of breath. As psychophysical strength data is collected on large numbers of subjects, it becomes possible to design jobs so that they are well within the strength capabilities of the vast majority of workers. One criterion that is often used is to design the job so that 75% of workers rate the load as acceptable. Studies have indicated that if workers lift more than this amount, they may be three times more likely to experience a low back injury. On the other hand, designing jobs in accordance with this criterion has the potential to reduce the occurrence of low back injuries by up to 33% (Snook and Ciriello 1991).
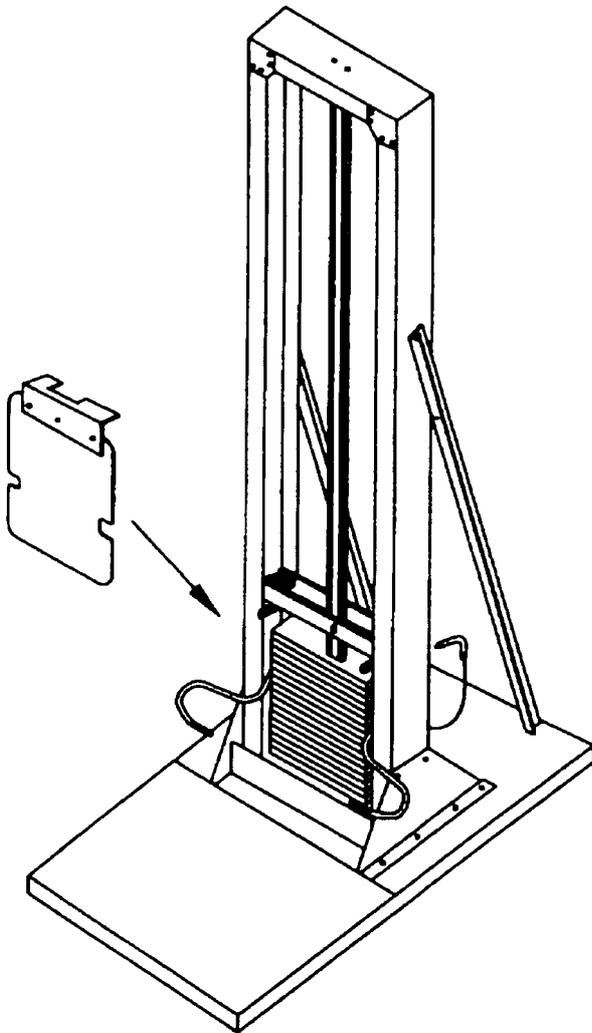
3.3. Isokinetic Strength

A technique of dynamic testing that has been growing in popularity is that dealing with the measurement of isokinetic strength. As defined previously, this technique evaluates muscular strength throughout a range of motion and at a constant velocity. It is important to realize that people do not normally move at a constant velocity. Instead, human movement is usually associated with significant acceleration and deceleration of body segments. Thus, there is a perceptible difference between isokinetic strength and free dynamic lifting. In the latter instance, subjects may use rapid acceleration to gain a weight lifting advantage. Acceleration is not permitted in isokinetic tests of strength.

The majority of isokinetic devices available on the market focus on quantifying strength about isolated joints or body segments, for example, trunk extension and flexion. This may be useful for rehabilitation or clinical use, but isolated joint testing is generally not appropriate for evaluating an individual ability to perform occupational lifting tasks. One should not make the mistake of assuming, for instance, that isolated trunk extension strength is representative of an individuals ability to perform a lift. In fact, lifting strength for a task may be almost entirely unrelated to trunk muscle strength. Strength of the arms or legs (and not the trunk) may be the limiting factor in an individuals lifting strength. For this reason, machines that measure isokinetic strengths of isolated joints or body segments should not be used as a method of evaluating worker capabilities related to job demands in most instances.

Many investigators have used dynamic isokinetic lifting devices specifically designed to measure whole-body lifting strength. These devices typically have a handle connected by a rope to a winch, which rotates at a specified isokinetic velocity when the handle is pulled (Figure 4). Studies using this type of
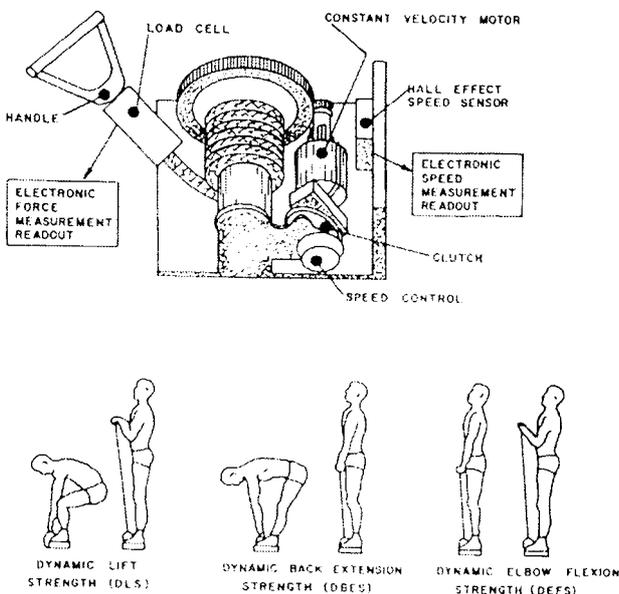
device have demonstrated good correlations between isokinetic Dynamic Lift Strength (i.e. a lift from floor to chest height) and the maximum weights individuals were willing to lift for infrequent tasks using the psychophysical approach (Pytel and Kamon 1981). Thus, under certain circumstances, this device appears to possess some validity for assessment of job related dynamic lifting strength capabilities of individuals. Some investigators have attempted to modify this type of instrument by providing a means to mount it so that isokinetic strength can be measured in vertical, horizontal, and transverse planes. However, while advances have been made in the use of isokinetic devices for worker strength evaluation, this procedure cannot be thought to be fully developed in the context of worker selection procedures.

4. CONCLUSIONS

In spite of advances in measurement techniques and an explosive increase in the volume of research, our understanding of human strength remains in its introductory stages. It is clear that muscle strength is a highly complex and variable function dependent on a large number of factors. It is not surprising, therefore, that there are not only substantial differences in strength between individuals, or that strength measurements for a single individual can vary a great deal even during the course of a single day. Strength is not a fixed attribute — strength training regimens can increase an individual's capability by 30–40% or more. Disuse can lead to muscle atrophy.

The use of physical strength assessment in ergonomics has focused on both job design and worker selection techniques. Of these, the former has a much greater potential to significantly reduce WMSD. Worker selection techniques must be considered a method of last resort — where engineering changes or administrative controls cannot be used to reduce worker exposure to WMSD risk factors. This technique has only shown a moderate effect in truly high-risk environments, and only in short-term studies. It is not known whether worker selection procedures have a protective effect over the long-term.

Figure 4. An isokinetic device allowing assessment of various muscular strengths (such as those shown) at a constant velocity.

REFERENCES

CHAFFIN, D.B. and ANDERSSON, G.B.J., 1991, *Occupational Biomechanics* (2nd ed.) (New York: Wiley).

GALLAGHER, S., MOORE, J.S. and STOBBE, T.J., 1998, *Physical Strength Assessment in Ergonomics* (Farifax: American Industrial Hygiene Association Ergonomics Press).

KROEMER, K.H.E., MARRAS, W.S., MCGLOTHLIN, J.D., MCINTYRE, D.R. and NORDIN, M., 1990, On the measurement of human strength. *International Journal of Industrial Ergonomics*, 6, 199–210.

MCDANIEL, J.W., SHANDIS, R.J. and MADOLE, S.W., 1983, *Weight Lifting Capabilities of Air Force Basic Trainees*. AFAMRL-TR-83-0001 (Wright-Patterson AFBDH, Air Force Aerospace Medical Research Laboratory).

PYTEL, J.L. and KAMON, E., 1981, Dynamic strength test as a predictor for maximal and acceptable lift. *Ergonomics*, 24, 663–72.

SNOOK, S.H. and CIRIELLO, V.M., 1991, The design of manual handling tasks: revised tables of maximum acceptable weights and forces. *Ergonomics*, 34, 1197–1213.

Torque Data

B. J. Daams

Delft University of Technology, Department of Industrial Design Engineering, Jaffalaan 9, 2628 BX Delft, The Netherlands

Presently: Daams Ergonomie, Mauvezand 17, 1251 JC Laren, The Netherlands

1. INTRODUCTION

The amount of force that must be exerted on or with a product is important to the perception of the consumer and to the product satisfaction (or especially to the dissatisfaction if the product does not fulfil its task properly). The importance of force exertion in product design is discussed in the article 'Force Exertion for (Consumer) Product Design: Problem Detinition'.

When designing for force exertion, a few steps must be taken to get a good result. First, the problem has to be defined the right way during the information stage. Second, an ergonomically sound concept should be developed in the concept stage with the help of global information (rules of thumb) on force exertion. Third, the actual size of maximal and/or minimal forces should be determined in the detailing stage.

Thus, the information needed for the design consists of two parts:

- general information (rules of thumb) for the early stage of design; and
- detailed information (exact forces) for the detailing stage of design.

These rules of thumb are important because they force the designer to think about a good design concept. If no effort is put in this first stage, it is no use detailing a basically bad ergonomical design. Rules of thumb are given in the article 'Force Exertion for (Consumer) Product Design: Information for the design Process', together with examples of product design. Also, some tips on measuring force exertion for design purposes are given.

After the concept phase of the design, the details should be looked into. How much force can the envisaged user group exert? Or better, how much force can the weakest users exert? This question must preferably be answered by research, but usually there is lack of time, money and/or interest to do so. Therefore, the question about the force that can be exerted can only be answered by literature. Adequate information on force exertion in literature is very scarce because a situation in which force exertion is investigated seldom corresponds with the situation in which a product is expected to be used and even less with the real situation. For an attempt to gather design relevant information and to present it in a way that makes it useful for the designer, see Daams (1994). A choice from these tables is given here. This chapter comprises the pushing and pulling forces of adults, children, the elderly and the disabled.

2. OVERVIEW OF THE DATA

2.1. Various Sorts of Torque

Torque can be exerted in various ways. Not only is the orientation of the lid/knob/handle is important, but also its shape. With a circular knob, the torque force cannot exceed the tangential force caused by the friction between hand and knob (which increases with increasing grip force). A small grip force thus limits the maximal exertion of torque. It is easier to operate a torque device that enables the hand to exert torque without the exertion of much grip force, like a T-bar or a lever with paddle. Torque knobs with levers or handles are recommended therefore for products that require much force and in designs for weak persons.

2.2. Products

Some investigators used "real products" to measure the torque exerted, and Daams (1994) used aluminum replicas of jam jars. Putto (1988) measured torque on light bulbs.

Not included in the data, but containing useful information on forces exerted on products: Berns (1981) investigated forces exerted on replicas of jam jars and bottles. Imrhan and Loo (1988) investigated glass jars with circular screw-type container lids. Bordett et al. (1988) used various faucet handles. Adams and Peterson (1988) used electrical connectors. Mital and Sanghavi (1986) and Mital (1986) used various hand tools, including wrenches and screwdrivers. Pheasant (1983) used screwdriver-like handles.

2.3. Maximal and Comfortable Torque

All measured torques are static. Most are maximal, but Kanis (1989) and Schoorlemmer and Kanis (1992) measured comfortable torque. Putto (1988) investigated the actual torque exerted when screwing in light bulbs, the maximal torque exerted when unscrewing bulbs and the maximal torque exerted on dummy bulbs.

2.4. Clockwise and Counter-clockwise

Kanis (1989), Schoorlemmer and Kanis (1992) and Steenbekkers (1993) investigated both torque directions and came to the conclusion that the exerted moments were not significantly different. They presented their results in one table. Mital (1986) measured clockwise torque only. Berns (1981), Daams (1993) and Imrhan and Loo (1988) measured counter-clockwise torque only. Bordett et al. (1988), Adams and Peterson (1988), Thompson (1975) and Rohmert and Hettinger (1963) give both values in a table. From the reports of Swain et al. (1977), Pheasant (1983) and Mital and Sanghavi (1986), it is not clear what the measured torque direction is.

2.5. Fingers

In general, torque is exerted with the whole hand, certainly when large lids or knobs are involved. With smaller knobs, either the whole hand can be used, or just the fingers. Adams and Peterson (1988) define and measure both fingertip grasp and full wraparound grasp. Kanis (1989) and Schoorlemmer and Kanis (1992) use a lateral grip. The subjects of Steenbekkers (1993) exert torque both with the whole hand and with forefinger and thumb only.

For laterality, nearly all forces are exerted with one hand, although in the case of the jam jar of Daams (1994), in free posture the subjects could use both hands. Probably the same applies to the jam jar of Berns (1981).

2.6. Sex

The torque exerted by women was investigated by Bordett et al.

(1988). The torque exerted by men was investigated by Swain *et al.* (1970). All other researchers measured both sexes.

2.7. Children

For children, Steenbekkers (1993) measured torque exerted by Dutch girls and boys between 4 and 13 years of age.

2.8. Elderly

The following investigators measured torque exerted by elderly subjects. Berns (1981) had subjects aged from 20 to > 71 years. Imrhan and Loo (1988) investigated elderly people aged between 60 and 97 years. Bordett *et al.* (1988) looked into the capacities of females aged between 60 and 90. The age of the subjects of Thompson (1975) ranged from 60 to 75 years.

2.9. Disabled

Berns (1981) and Schoorlemmer and Kanis (1992) measured torque of both healthy persons and various disabled subjects. Kanis (1989) investigated torque exerted by arthritic persons and persons with muscle diseases. All measured female and male subjects.

2.10. Summary

Part of the collected data is summarized, as far as possible, in the following tables. Torques around a horizontal axis and a vertical axis are summarized in different tables. Only torque on circular knobs or lids is included, because the moment arm of torque on handles cannot be determined exactly. Averages and SD are included. A selection is made of the values of some authors. Of Steenbekkers' values, for example, the results of three age groups were selected. For more detailed information, the relevant tables should be studied.

3. TORQUE DATA

3.1. 🖼 Daams (1994)

3.1.1. Subjects
- Number: 22
- Sex: 10 female, 12 male.
- Age: eight young women and nine young men, average 22 years. Two women and three men were older.
- Other characteristics: healthy students and staff of Delft University.

3.1.2. Method
- Direction of force: torque, counter-clockwise.
- Posture: standing in free posture (figure 1), (a) torque on a fixed jar; (b) torque on a freely movable jar.
- Sort of force: maximal static force.
- Laterality: one-handed on the fixed jar, two-handed on the freely movable jar.
- Measurement: duration 6 s, measure is the average of the last 4 s. The score is the average of two trials.
- Sort of handle: jam jar shaped with built-in force transducer, material aluminum, weight 650 g.
- Size of the handle: lid Ø 66 mm, jar Ø 75 mm, jar height 113.5 mm.
- Position of handle: fixed jar: lid fixed at 0.95 m from the floor.
- freely movable jar: not fixed.

Table 1. Torque (Nm) on a jam jar, exerted in two postures. Averages and SD of 10 women and 10 men (Daams 1994).

posture	females x	females s	males x	males s	both x	both s
free, jar fixed at 0.95m	6.36	1.61	11.70	3.45	9.27	3.84
free, jar not fixed	5.91	1.26	9.67	2.20	7.96	2.62
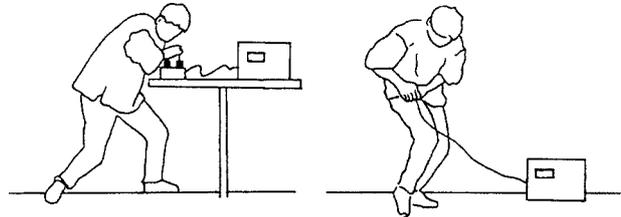
Figure 1. (a) Torque on a fixed jar; (b) torque on a freely movable jar (Daams 1994).

- Other characteristics: non-slip plastic material was placed between hand and jar if maximal torque could otherwise not be obtained.

3.2. 🖼 Steenbekkers and van Beijsterveld (1998)

3.2.1. Subjects
- Number: several hundreds; see tables.
- Sex: women and men; see tables.
- Age: 20–80+, divided in nine different age groups; see tables.
- Other characteristics: Dutch adults and elderly, selected to be representative for the Netherlands.

3.2.2. Method
- Direction of force: torque with two hands.
- Posture: the subject holds the measuring device (which resembles a jam jar) in his or her hands and adopts a free posture, whether holding the jar or leaving it on the table.
- Sort of force: maximal static force.
- Laterality: exerting torque with both hands.
- Measurement: duration 3 s.

Figure 2. Example of hand posture during the experiment (Steenbekkers and van Beijsterveldt 1998).

Table 2. Torque (Nm) with two hands. Averages, SD and number of subjects per age category (Steenbekkers and van Beijsterveldt 1998).

age	women					men					all		
[years]	n	s	P5	x	P95	n	s	P5	x	P95	n	s	x
20 – 30	68	1.4	3.6	5.6	8.2	55	2.2	5.6	8.7	12.3	123	2.4	11.7
31 – 49		1.4		5.2			2.1		8.2				
50 – 54	35	1.5	2.6	4.8	7.8	35	1.9	4.5	7.7	10.8	70	2.2	9.8
55 – 59	50	1.4	2.5	4.7	7.2	46	2.2	4.7	7.5	10.6	96	2.3	10.4
60 – 64	53	1.4	2.8	4.8	7.5	44	1.8	4.3	6.4	9.6	97	1.7	9.1
65 – 69	51	1.2	2.1	4.0	6.2	50	2.1	4.1	6.5	10.4	101	2.1	9.0
70 – 74	62	1.1	1.8	3.7	5.1	59	2.1	2.5	5.3	8.9	121	1.8	7.6
75 – 80	38	1.3	1.7	3.5	6.2	35	1.7	2.5	5.1	7.8	73	1.7	7.1
80+	35	0.9	0.9	3.4	4.5	33	1.7	2.5	4.9	7.9	68	1.5	6.8
total		1.4		4.9			2.1		7.7				

- Sort and size of handle: a measuring device with the weight, shape and surface texture of a full jam jar.
- Position of handle: free.
- Other characteristics: –
  For results, see Figure 2 and Table 2.

3.3. Swain et al. (1970)

3.3.1. Subjects

- Number: 120.
- Sex: male.
- Age: civilians: average 31.4 range 22–40 years; military, army: average 27.2 range 21–38 years; military, navy: average 27.4 range 23–32 years.
- Other characteristics: 96 civilians, 12 army, 12 navy.

3.3.2. Method

- Direction of force: torque.
- Posture: standing. The steadying hand was always placed on top of the measuring apparatus. The subjects were required to stand approximately in front of the device, to keep both feet on the floor and not to move the device.
- Sort of force: maximal static force.
- Laterality: preferred hand.
- Measurement: the subjects were instructed to get a firm grip and give the knob a single, hard twist and then stop.
- Sort of handle: small knob.
- Size of handle: knob Ø 9.5 mm, 12.7 mm and Ø 19.1 mm.
- Position of handle: about 81 cm above the floor.
  Two orientations were used:
- control panel facing the subject; and
- the edge of the control panel toward the subject so that the knob was on his right.

Table 3. Torque (Nmm) exerted on knobs, with bare hand and gloves, averages and SD of males (Swain *et al.* 1970).

knob		mil. under 29		mil. over 29		civ. under 29		civ. over 29		all	
		n = 16		n = 8		n = 31		n = 65			
		x	s	x	s	x	s	x	s	x	s
	age [years]	24.50	2.16	32.62	2.88	25.26	1.97	34.32	2.88		
9.5 mm	with gloves										
	front	44.66	10.39	47.94	7.03	41.98	8.28	46.82	7.88	45.35	8.48
	side	46.08	11.89	49.54	11.68	46.19	11.06	47.82	8.48	47.28	9.82
	bare hand										
	front	49.23	8.94	47.06	13.06	51.42	13.64	54.94	12.92	52.75	12.80
	side	52.56	11.98	58.87	11.02	52.36	12.21	56.78	12.87	55.21	12.54
12.7 mm	with gloves										
	front	57.00	10.82	60.02	11.42	56.00	10.28	58.02	10.48	57.50	10.46
	side	58.73	9.52	62.15	9.49	57.26	13.64	61.06	11.39	59.84	11.68
	bare hand										
	front	59.66	14.54	62.15	14.77	62.48	13.38	68.41	15.60	65.29	15.09
	side	63.00	15.21	64.46	18.33	61.66	14.00	69.74	16.35	66.40	16.00
19.1 mm	with gloves										
	front	92.69	23.94	97.75	15.85	88.97	14.89	93.57	16.39	92.54	17.13
	side	95.22	19.85	84.96	14.21	92.08	18.32	95.06	17.63	93.64	17.90
	bare hand										
	front	105.17	21.81	104.33	22.00	103.82	25.05	117.89	27.04	111.65	26.20
	side	112.31	27.37	105.75	32.01	109.41	26.72	118.27	25.88	114.35	26.74

Other characteristics: bare hands and gloves were used. To avoid the effects of sweaty hands, paper towels were provided and subjects were instructed to wipe their fingertips before each trial.

For results, see Table 3.

3.4. Thompson (1975)

3.4.1. Subjects
- Number: 38.
- Sex: 23 female, 15 male.
- Age: 60–75 years.
- Other characteristics: subjects were selected locally at random from elderly persons within this age range, who were living in unassisted occupation of a dwelling. However, preliminary screening eliminated from selection those with any indication of coronary disease.

3.4.2. Method
- Direction of force: (a) torque, clockwise and counter-clockwise; (b) torque on door/window handles.
- Posture: a line was drawn on the floor in line with and vertically beneath, the handle of the apparatus. The subject's leading foot was positioned on this line to ensure maximum force application. Apart from this positioning, subjects were allowed to adopt whatever stance was natural to them.
- Sort of force: maximal static force.
- Laterality: the dominant hand.
- Measurement: subjects were told to exert force as hard as they could until they felt they had reached their maximum and they were then to release the handle. The highest value of three trials is the value used for computation of the strength data.
- Sort and size of handle: (a1) circular knob, Ø 2.8 cm; (a2) square knob, 3.1 x 2.5 cm; (b1) handle, oval diameter; length 11 cm; (b2) handle, square diameter, length 12 cm.
- Position of handle: 1 m above the floor.
- Other characteristics: –

For results, see Table 4.

3.5. Kanis (1989)

3.5.1. Subjects
- Number: 34.

- Sex: 28 female; six male.
- Age: average 54 years (SD 14.2).
- Other characteristics: 28 arthritic subjects (24 women, four men) and six subjects (four women, two men) with a muscle disease.

3.5.2. Method
- Direction of force: torque, averaged for clockwise and counter-clockwise together.
- Posture: sitting, with the elbow 90° flexed. Lateral grip. See figures.
- Sort of force: maximal and comfortable static force. 'Comfortable' refers to a level of pain or fatigue found acceptable by the subject.
- Laterality: left hand, right hand and preferred hand.
- Measurement: subjects were instructed to build up their force gradually and to hold the maximum for at least a few seconds.
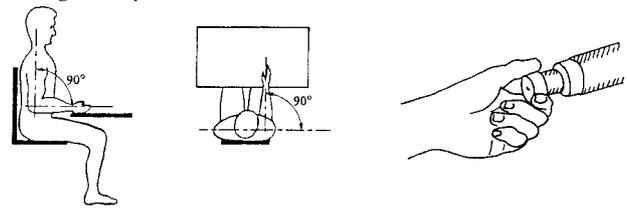
Figure 3. Posture during the experiment (Kanis 1989).
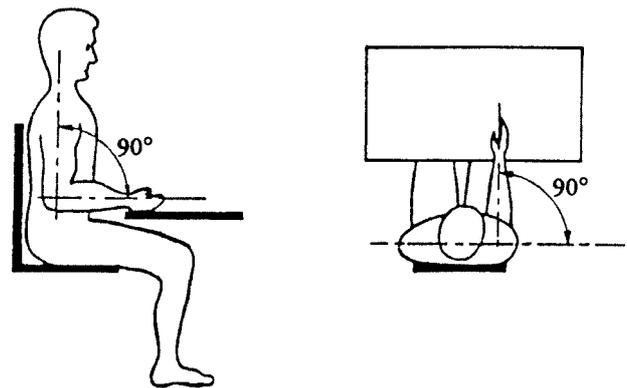
Figure 3a.

Table 4. Torque (Nm) exerted on two different knobs and two different handles. Averages, SD, ranges and 5th percentiles of 38 elderly people (Thompson 1975).

handle	force direction*	female x	s	range		P5	male x	s	range		P95
circular knob											
circular knob	cw	0.58	0.22	1.27	- 0.23	0.21	0.83	0.39	1.66	- 0.33	0.19
	ccw	0.56	0.23	1.60	- 0.28	0.18	0.94	0.46	2.02	- 0.31	0.19
square knob	cw	0.69	0.21	1.04	- 0.19	0.35	1.04	0.36	1.70	- 0.57	0.44
	ccw	0.78	0.20	1.24	- 0.35	0.45	1.12	0.46	2.01	- 0.40	0.36
handle											
oval diameter	push	16.84	5.74	28.83	- 10.12	7.40	24.35	4.70	28.93	- 10.81	16.61
	pull	11.29	3.60	18.42	- 6.94	5.36	15.98	4.67	21.55	- 7.34	8.29
square diameter	push	10.73	4.13	28.83	- 6.47	3.94	16.67	5.92	28.72	- 9.04	6.94
	pull	9.51	3.30	18.78	- 6.12	4.08	15.74	3.77	20.76	- 7.79	9.53

* clockwise and counter-clockwise

Table 5. Torque (Nm) exerted with lateral grip. Averages and SD of 34 impaired subjects (Kanis 1989).

handle diameter [mm]	force	hand	females		males		arthritis		muscle desease		all	
			x	s	x	s	x	s	x	s	x	s
40	maximal	left	0.35	0.18	0.45	0.40	0.40	0.23	0.24	0.18	0.37	0.23
		right	0.36	0.19	0.59	0.54	0.44	0.29	0.24	0.20	0.40	0.29
	comfortable	preferred	0.17	0.10	0.26	0.28	0.19	0.15	0.14	0.11	0.18	0.15
13	maximal	preferred	0.10	0.04	0.15	0.12	0.12	0.07	0.07	0.03	0.11	0.07

- Sort of handle: two aluminum cylinders with anti-slip texture.
- Size of handle: height 15 mm, Ø 13 mm and Ø 40 mm.
- Position of handle: mounted on a table.
- Other characteristics: –

For results, see Table 5.

3.6. ⌖ Schoorlemmer and Kanis (1992)

3.6.1. Subjects
- Number: 27.
- Sex: 21 female, six male.
- Age: average 49.4 years (SD 17.4).
- Other characteristics: average body height 165 cm (SD 12), average body weight 67 kg (SD 12).

- Twelve healthy subjects, five spastics, four visually impaired, three blind subjects, three subjects with Parkinson's disease.

3.6.2. Method
- Direction of force: torque, clockwise and counter-clockwise (they are not significantly different and thus represented in one table).
- Posture: sitting, with the elbow 90° flexed (see figures) and in a free posture. Lateral grip.
- Sort of force: maximal and comfortable static force. 'Comfortable' refers to a level of pain or fatigue found acceptable by the subject.
- Laterality: left hand and right hand (averaged in the table).
- Measurement: subjects were instructed to build up their force gradually and to hold the maximum for at least a few seconds.

ridged knob   *circular knob*

Torque with lateral grip on a ridged knob   *Torque with lateral grip on a circular knob*
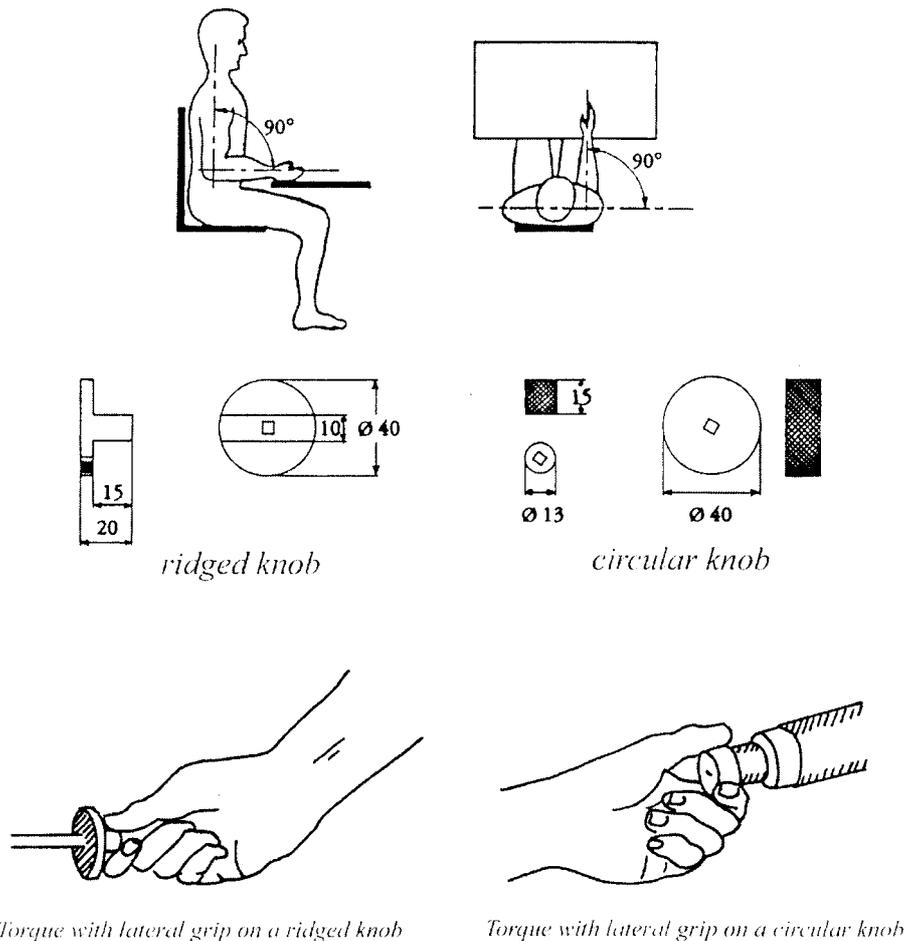
Figure 4. Posture during the experiment (Schoorlemmer and Kanis 1992).

Table 6. Torque (Nm) exerted with lateral grip. Averages and SD of 12 healthy and 15 impaired subjects (Schoorlemmer and Kanis 1992).

force	posture	males n = 6		females n = 21		healthy, blind & visually impaired n = 19		spastics n = 5		Parkinson's disease n = 3		all n = 27	
		x	s	x	s	x	s	x	s	x	s	x	s
ridged knob													
max.	standard	1.46	0.51	1.14	0.45	1.35	0.42	0.89	0.48	0.85	0.48	1.21	0.47
	free	1.42	0.64	0.94	0.35	1.16	0.45	0.83	0.49	0.69	0.30	1.05	0.46
round knob Ø 13 mm													
max.	standard	0.32	0.15	0.26	0.11	0.31	0.10	0.19	0.12	0.15	0.13	0.27	0.12
round knob Ø 40 mm													
max.	standard	1.12	0.41	0.85	0.29	1.03	0.29	0.68	0.29	0.51	0.25	0.91	0.34
	free	1.12	0.54	0.92	0.40	1.03	0.40	0.86	0.58	0.71	0.35	0.96	0.43
comf.	standard	0.54	0.38	0.50	0.34	0.59	0.34	0.25	0.16	0.43	0.42	0.51	0.34

- Sort of handle: three handles: two aluminum cylinders with anti-slip texture, one with a rim (see figure).
- Size of handle: height 15 mm for all, Ø 13 mm and Ø 40 mm (see figure)
- Position of handle: mounted on a table.
- Other characteristics: –

For results, see Table 6

3.7. Steenbekkers (1993)

3.7.1. Subjects
- Number: 782.
- Sex: 392 girls, 390 boys.
- Age: 4–13 years.
- Other characteristics: Dutch children, selected to be representative for the Netherlands.

3.7.2. Method
- Direction of force: torque, clockwise and counter-clockwise (they are not significantly different and thus represented in one table).
- Posture: the child sits upright before the measuring device, which is in front of the forearm. The forearm is horizontal and in a sagittal plane. Upper arm and forearm are at ~150°. The legs hang down freely or are directed forward. (Torque of forefinger and thumb: the thumb and fingers are positioned on the knob opposite one another. The other fingers are flexed into a fist.
- Sort of force: maximal static force.
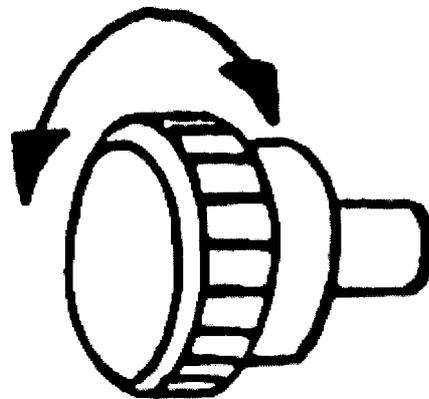
Figure 5. Posture during experiment (a) (Steenbekkers 1993).
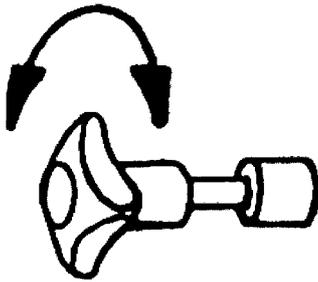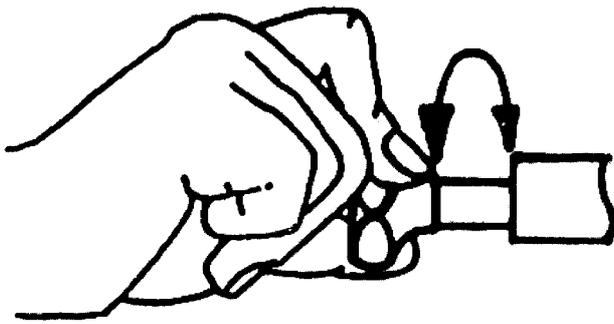
Figure 6. Posture during experiment (b) (Steenbekkers 1993).

• Laterality: preferred and non-preferred hand (they are not significantly different and thus represented in one table).
• Measurement: duration 3 s.
• Sort of handle: (a) cylindrical knob; (b) star knob (see figure).
• Size of handle: (a) Ø 4 cm; (b) maximum Ø 2.5 cm.
• Position of handle: mounted on a table.
• Other characteristics: –
For results, see Tables 7 and 8

3.8. Putto (1988)

3.8.1. Subjects
• Number: 16.
• Sex: two female, 14 male.
• Age: average 41.9 years (SD 10.2), range 23–59 years.
• Other characteristics: Philips staff. Some had more than average knowledge of light bulbs, but their results did not differ from the others.
• Body height average 1.79 m (SD 0.10); body weight average 77.1 kg (SD 1.3); hand length average 19.2 cm (SD 1.3).

3.8.2. Method
• Direction of force: torque, clockwise and counter-clockwise.
• Posture: free posture when screwing and unscrewing a light bulb.
• When exerting maximal force, the subject's own posture adopted when unscrewing a light bulb was reproduced (i.e. the grip, the orientation relative to the bulb and the arm posture were the same).

Table 7. Torque (Nm) exerted with one hand. Averages, SD and percentiles for girls and boys (Steenbekkers 1993)

| age | girls | | | | | boys | | | | | all | |
[years]	n	x	s	P3	P97	n	x	s	P3	P97	x	s
4.0–4.9	46	0.83	0.32	0.33	1.50	41	0.88	0.30	0.52	1.54	0.85	0.32
5.0–5.9	59	1.02	0.35	0.34	1.61	53	1.16	0.39	0.53	1.93	1.09	0.38
6.0–6.9	57	1.23	0.43	0.49	2.08	59	1.43	0.42	0.80	2.21	1.33	0.43
7.0–7.9	42	1.20	0.46	0.46	2.01	50	1.50	0.53	0.66	2.81	1.36	0.52
8.0–8.9	40	1.33	0.52	0.55	2.32	41	1.56	0.57	0.86	2.68	1.45	0.56
9.0–9.9	42	1.45	0.55	0.60	2.64	41	1.74	0.56	0.88	2.75	1.59	0.57
10.0–10.9	39	1.54	0.56	0.78	2.72	35	2.07	0.70	0.96	3.18	1.80	0.68
11.0–11.9	44	1.74	0.66	0.74	3.18	40	1.98	0.62	1.07	3.14	1.85	0.65
12.0–12.9	23	1.94	0.94	0.82	5.08	30	2.18	0.83	0.96	4.33	2.08	0.88

Table 8. Torque (Nm) exerted with thumb and forefinger. Averages, SD and percentiles of girls and boys (Steenbekkers 1993).

| age | girls | | | | | boys | | | | | all | |
[years]	n	x	s	P3	P97	n	x	s	P3	P97	x	s
4.0–4.9	46	0.27	0.07	0.17	0.42	41	0.27	0.08	0.12	0.41	0.27	0.07
5.0–5.9	59	0.33	0.08	0.18	0.48	53	0.35	0.09	0.17	0.50	0.34	0.08
6.0–6.9	57	0.38	0.09	0.23	0.56	59	0.41	0.10	0.22	0.57	0.40	0.09
7.0–7.9	42	0.43	0.10	0.25	0.64	50	0.44	0.09	0.30	0.62	0.44	0.09
8.0–8.9	40	0.44	0.10	0.28	0.62	41	0.50	0.09	0.35	0.64	0.47	0.10
9.0–9.9	42	0.49	0.10	0.32	0.64	41	0.53	0.10	0.35	0.73	0.51	0.10
10.0–10.9	39	0.54	0.12	0.32	0.75	35	0.61	0.11	0.41	0.78	0.57	0.12
11.0–11.9	44	0.59	0.11	0.39	0.79	40	0.66	0.14	0.46	0.94	0.63	0.13
12.0–12.9	23	0.64	0.15	0.38	0.88	30	0.75	0.14	0.44	1.09	0.70	0.16

- Sort of force: (a) screwing in a light bulb (actual force, not maximal); (b) unscrewing a real (although fixed) bulb. The exerted force is static and as high as the subjects dare exert; (c) real maximal static torque, unscrewing a dummy bulb.
- Laterality: the preferred hand.
- Measurement: (a) and (b): exerted as in real bulb screwing; (c) for the maximal force, the subjects were instructed to slowly build up their force to a maximum and to maintain it for a few seconds.
- Sort and size of handle: five different types of bulbs and three different types of dummies were used (see figure):
- Bulbs (real): 1, small bulb, Ø 45 mm; 2, standard bulb, Ø 60 mm; 3, PLC-E small energy-saving bulb (square), diagonal cross section 45 mm; 4, fancy bulb (square), diagonal cross-section 55 mm; 5, SL large energy-saving bulb, Ø 70 mm.
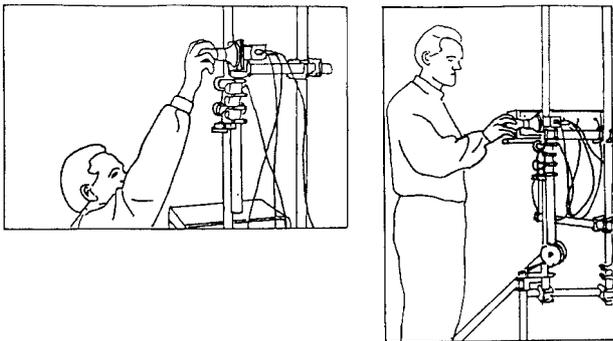
Figure 7. Posture with bulb at reach height and at heart height (Putto 1993).

- Dummies, made of polished PC (Polycarbonate): similar to the small bulb, Ø 45 mm (no. 1); similar to the standard bulb, Ø 60 mm (no. 2); similar to the large energy-saving bulb, Ø 70 mm (no. 5).
- Position of handle: orientation of the bulbs: with the axis around which torque is exerted, horizontal.
- Two different heights: chest (heart) height and maximum reach height.
- Other characteristics: –

3.8.3. Results

Comparing the maximal forces exerted on real bulbs and dummy bulbs, it can be seen that the characteristics of the real bulb limit the maximal force that is exerted (table 9).

REFERENCES

ADAMS, S.K. and PETERSON, P.J., 1988, Maximum voluntary hand grip torque for circular electrical connectors. *Human Factors*, 30, 733–45.

BERNS, T., 1981, The handling of consumer packaging. *Applied Ergonomics*, 12, 153–161.

BORDETT, H.M., KOPPA, R.J. and CONGELTON, J.J., 1988, Torque required from elderly females to operate faucet handles of various shapes. *Human Factors*, 30, 339–46.

DAAMS, B.J., 1994, *Human Force Exertion in User–Product Interaction.* Physical Ergonomics Series (Delft: Delft University Press).

KANIS, H., 1989, *Bediening & Handikap* (Delft: Faculty of Industrial Design Engineering).

Table 9. Torque (Nm) exerted on real and dummy bulbs. Averages, SD and ranges of 16 subjects (Putto 1988).

sort bulb		direction of force	force	bulb height	x	s	range	
standard	real	screw in	actual	heart	1	0.84	0.30 –	1.20
		unscrew	maximal	heart	3	3.76	1.20 –	5.40
				reach	3	3.76	1.30 –	5.40
	dummy	unscrew	maximal	heart	6	7.76	3.20 –	9.70
				reach	7	9.05	2.60 –	11.10
small	real	screw in	actual	heart	0.60	0.85	0.30 –	1.10
		unscrew	maximal	heart	1.88	2.72	0.83 –	3.40
				reach	1.89	2.73	0.63 –	3.40
	dummy	unscrew	maximal	heart	4.70	6.30	2.20 –	8.80
				reach	5.11	6.74	2.50 –	8.50
SL	real	screw in	actual	heart	2.00	3.32	0.46 –	6.20
		unscrew	maximal	heart	6.03	7.81	3.40 –	10.30
				reach	5.59	6.87	3.90 –	7.70
	dummy	unscrew	maximal	heart	17.47	22.73	5.30 –	24.20
				reach	11.86	17.41	5.50 –	28.80
PLC-E	real	screw in	actual	heart	1.36	2.07	0.53 –	3.50
		unscrew	maximal	heart	4.19	5.91	1.50 –	7.20
	dummy	unscrew	maximal	reach	4.28	6.04	1.00 –	7.60
fancy	real	screw in	actual	heart	1.40	1.92	0.80 –	2.30
		unscrew	maximal	heart	4.53	6.35	2.10 –	8.40
	dummy	unscrew	maximal	reach	4.84	6.55	2.60 –	8.50

MITAL, A., 1986, Effect of body posture and common hand tools on peak torque exertion capabilities. *Applied Ergonomics*, 17, 87–96.

MITAL, A. and SANGHAVI, N., 1986, Comparison of maximum volitional torque exertion capabilities of males and females using common hand tools. *Human Factors*, 28, 283–94.

PHEASANT, S.T., 1983, Sex differences in strength — some observations on their variability. *Applied Ergonomics*, 14, 205–11.

PUTTO, R., 1988, Hanteren en kracht. Thesis, Faculty of Industrial Design, Delft University.

SCHOORLEMMER, W., 1993, *Bediening van Knoppen*. Internal report, Faculty of Industrial Design Engineering, Delft University.

SCHOORLEMMER, W. and KANIS, H., 1992, Operation of controls on everyday products. In *Proceedings of the Human Factors 36th Annual Meeting — 1992* (Santa Monica: Human Factors Society), 509–13.

STEENBEKKERS, L.P.A., 1993, *Child Development, Design Implications and Accident Prevention*. Physical Ergonomics Series (Delft: Delft University Press).

STEENBEKKERS, L.P.A. and VAN BEIJSTERVELDT, C.E.M. (eds), 1998, *Design-relevant Characteristics of Ageing Users* (Delft: Delft University Press).

SWAIN, A.D., SHELTON, G.C. and RIGBY, L.V., 1970, Maximum torque for small knobs operated with and without gloves. *Ergonomics*, 13, 201–8.

THOMPSON, D., 1975, *Ergonomics Data for Evaluation and Specification of Operating Devices on Components for Use by the Elderly* (Loughborough: Institute for Consumer Ergonomics)

Trunk Muscle Force Models

M. K. Chung†, S. Kim‡ and D. S. Bloswick§

†Department of Industrial Engineering, Pohang University of Science and Technology, Hyoja San 31, Pohang 790-784, Korea

‡Department of Industrial and Systems Engineering, Kumoh National University of Technology, Shinpyungdong 188, Kumi 730-701, Korea

§Department of Mechanical Engineering, University of Utah, 50 S. Central Campus Drive, Salt Lake City, UT 84112-9208, USA

1. INTRODUCTION

Computerized biomechanical models to predict the trunk muscle forces have been developed for many years because there is no other way directly to measure spinal loading *in vivo*. Most biomechanical models aimed at reducing the risk of musculoskeletal injuries have focused on the activities of the trunk musculature, which is important because trunk muscles counteract external moments imposed upon the trunk during manual materials handling (MMH) activities. The forces generated by these muscles can easily become excessive given that their moment arms between the spine are at a much shorter distance compared with the distance between the spine and the object being handled. Since these internal forces are often the primary loading components of the spine, it is extremely important to understand how the trunk muscles collectively load the spine under occupational conditions. Therefore, a desirable feature of a biomechanical model is that it should predict both the synergistic actions of the various internal forces acting upon the spine as well as the loading due to external forces.

Unfortunately the problem of muscle force prediction is not easy, partially because a large number of muscles cross each joint in the body. In a situation where more muscles cross each joint than mechanical equilibrium conditions, the mechanical representation of the joint is statically indeterminate. In such a situation, there are an infinite number of possible muscle force combinations that can satisfy the mechanical equilibrium conditions. Thus, a biomechanical model that incorporates many muscles must adopt a strategy to resolve the issue of static indeterminacy together with obtaining the simultaneous solution of moment constraints about the joint axes.

2. ANATOMY AND FUNCTIONS OF TRUNK MUSCLES

It is generally accepted that activities of the erector spinae (ES), rectus abdominus (RA), internal oblique (IO) and external oblique (EO) muscles are critical during flexion, extension, rotation and lateral bending of the torso. The general locations of the EO, IO, RA, and ES muscles are illustrated in Figure 1.

The effect of non-symmetric and symmetric hand forces on a three-dimensional torso model requires an initial analysis of four factors:

- Lines of action and level of activity of the IO and EO muscles as a function of body loading and position, and how these muscles relate to ES and RA activity.
- Effect of non-symmetric hand loading on ES and RA muscles.
- Significance of IO and EO muscle activity in relation to the ES and RA muscle groups presently included in the model.
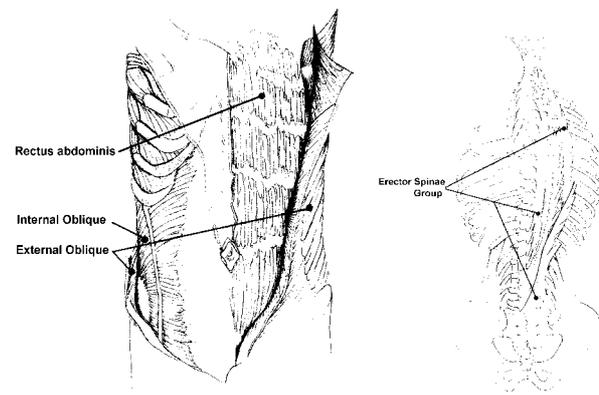
Figure 1. General location of external oblique, internal oblique, rectus abdominis and erector spinae.

- Confirmation of the assumption that antagonist muscle activity is zero. This assumption is frequently made to make the system of equilibrium equations determinant.

2.1. Lines of Action and Activity Level of Oblique Muscles

2.1.1. Flexion versus extension

It is clear that activity of the EO will cause trunk flexion since activity in any part of the muscle will tend to pull the trunk down. The activity of the IO will also cause flexion. The line of action of the posterior part of the muscles may oppose this however. Farfan (1975), in fact, contends that the posterior two-thirds of the IO causes an extensor moment about the L5 vertebrae. This is supported in a model developed by Takashima *et al.* (1979) in which the lines of action of the EO would cause flexion while the lines of action of the posterior part of the IO would cause trunk extension about the L5 vertebrae. Through the analysis of multiple cross sections through a frozen cadaver trunk, Rab *et al.* (1977) found that the lines of action of both the EO and IO would tend to cause trunk flexion. Yettram and Jackman (1980) discuss a model in which the IO muscles are trunk flexors and are only active in the upright position and not in full forward flexion where they are at the short end of their range of motion (ROM).

It was initially hypothesized that, if the activity of the IO and EO could be consistently correlated with either the ES or RA, then (assuming synergistic activity) the flexor or extensor role of the IO and EO could be established. Such consistency was not found however. Eckholm *et al.* (1982) notes that when IO and EO EMG activity exists in a lifting task it usually occurs in conjunction with RA EMG activity (flexion) rather than ES activity (extension). Kumar and Davis (1978), however, found that during a stoop lift the EO EMG activity was related to ES EMG activity at the moment of maximum stress during extension. Ortengren and Andersson (1977) note that RA, IO, and EO EMG activity increased when these muscles were stretched during extension and that during flexion they were active only during the first few degrees of rotation. Schultz *et al.* (1980) found that during physical activity there was significant EMG activity in the IO and EO that occurred with activity in the ES. Schultz also found that the IO and EO muscles contract with the RA to develop trunk flexion. It may be that the IO and EO are trunk flexors but exhibit antagonistic activity during extension.. It is also possible that the

IO and EO must be active to stabilize the trunk during manual material handling activities and therefore must be active whether the torso muscle force is needed to flex or extend the trunk.

While the EO muscles are accepted to be trunk flexors, there does not appear to be a consensus as to the effect of the IO on trunk flexion or extension during sagittal plane lifting activities. One possible reason for the inconsistency of the results is that sagittal analysis is inadequate. Since IO and EO muscles also act to rotate and bend the body laterally, their activity during lifting activities may be to assist in maintaining overall body stability and coordination of body movement in all three directions and not just assist in the creation of moment about the L5/S1 disc or pelvis. There was also no consensus as to the contribution made by contractions of the IO and EO on the increase of abdominal pressure during lifting activities. Kumar and Davis (1978) and Gracovetsky *et al.* (1981) contend that the oblique muscles are instrumental in the increase of abdominal pressure, but in a review of relevant literature Eckholm *et al.* (1982) suggest that the role of the oblique muscles in creating abdominal pressure may have been overemphasized.

2.1.2. Lateral flexion and rotation about the vertical-axis

In the case of the EO muscle, when one side acts alone it tends to bend the vertebral column laterally and bring the shoulder of the same side forward. In the case of the IO, when one side acts alone it tends to bend the vertebral column laterally and bring the shoulder of the opposite side forward. When both sides of either the IO or EO act together they tend to cause trunk flexion. In a calculation of muscle power based on the previously noted cadaver analysis, Rab *et al.* found that lateral bending is primarily caused by IO and EO muscles, and that rotation is caused by activity in the ipsilateral IO muscles and activity in the contralateral EO muscles.

Yettram and Jackman (1980) predicted, using an LP model which summarizes the total force in 171 muscle forces in the human spinal column in lateral flexion, that the activity level of the IO on the convex side was 141.4 N versus an activity level of 24.3 N on the concave side, and 80.2 N on both sides in the upright position. Schultz *et al.* (1980) developed a model using LP (minimize F comp) and assuming a maximum contractile force of 100 N/cm². He predicted that in resisting lateral bending, extension and twisting, the maximum contralateral oblique force would be 420 N as opposed to 120 N ipsilateral forces. Ortengren and Andersson (1977) note that back muscles and abdominal muscles show higher EMG activity on the contralateral side than on the ipsilateral side during trunk lateral flexion (bending). They also note that the back muscles are active on both sides of the spine in trunk rotation but found no consensus in the literature as to the relative activity level on the two sides. They note only slight abdominal muscle activity during trunk rotation. The consensus appears to be that IO, EO and ES muscles have higher activity on the contralateral side during bending and twisting, but that ipsilateral activity is not zero. McNeill *et al.* (1980) assumed that in lateral bending, the A-P forces of the IO and EO cancel one another leaving only vertical forces in the same frontal plane as the L5 vertebrae. Schultz and Andersson (1981) also assume that the centroid of a cross section of the IO and EO muscles acts in the same frontal plane as the L5/S1 disc. But they do not assume the A-P forces of the IO and EO are zero. This facilitates the assumption that the IO and EO cause no sagittal flexion or extension moment about the L5 vertebrae. This assumption does not appear to be entirely valid in light of the previous discussion of IO and EO activity during trunk flexion or extension.

2.2. Non-symmetric Hand Loading versus ES and RA Muscles

It is found that for a seated operator holding a load in a hand extended to the side, the ES EMG activity on the side opposite the load was, in general, greater than the ES EMG activity on the same side as the load. For the standing operator, when weight was held in one hand, the contralateral side EMG in the lumbar region was higher than that of the ipsilateral side. This effect was more pronounced when combined with 20° lateral flexion and reduced when combined with trunk rotation. Ortengren and Andersson (1977) note that during lateral flexion, the EMG activity of the back muscles was higher on the contralateral side of the lumbar region. They also note that in trunk rotation the back muscles are active on both sides. They cite three studies that note that the activity is similar on both sides and four studies that note that the activity is different.

2.3. Significance of IO and EO versus ES and RA Muscles

Rab *et al.* used cross-sectional measurements from human cadavers as the basis for their contention that the RA, EO, and IO each contribute ~33% to flexion moment. Takashima *et al.* (1979) found that the cross-sectional area of the RA, EO, IO and transverse obliques are all ~3.5 cm² (approximately equal) and that the total of the erectors is ~48 cm².

In the previously noted LP model developed by Schultz *et al.* (1980) the maximum forces in resistance to lateral bending, rotation and twisting for the ES, RA, contralateral oblique, and ipsilateral oblique muscles were 890 N (on each side), 130 N (on each side), 420 N and 120 N respectively. It can be seen that the forces generated by the oblique muscles during some activities are not insignificant, especially when compared with the rectus abdominus.

2.4. Investigation of the "Zero Antagonist" Assumption

Many models of the lumbar region include the assumption that the antagonist muscle activity is zero. This assumption is required to reduce the number of unknowns in the equations of equilibrium and arrive at a determinant state. A review of some relevant literature suggests that this assumption may not be entirely valid. The model developed by Yettram and Jackman (1980) predicted an IO activity level of 80.2 N on both sides in the upright position as compared with 24.3 N on the concave side and 141.4 N on the convex side in lateral bending. The predicted antagonist concave muscle activity is lower but not zero.

Significant antagonist ES EMG activity was observed for a seated operator when holding a load in the outstretched hand. In fact, in the no-load case the right side activity (same side as load) was higher than the left side when the hand was held laterally out from the body. Kumar and Davis (1978) note that

the EO muscles are trunk flexors but found that the 'massive activity' of the EO coincides with the ES at the moment of maximum stress in the stoop lift task. They suggest that this EO activity may assist the deeper muscles in raising the abdominal pressure, although as noted earlier, the role of the EO and IO muscles in raising the abdominal pressure has been questioned. Schultz *et al.* (1980) found that during various types of physical activity there was significant antagonistic EMG activity. For example, both the EO and IO were sometimes active and antagonistic to the ES. Of course, as noted earlier, it is not certain whether the posterior two-thirds of the EO would cause a flexor moment (antagonistic to the ES) or an extensor moment (synergistic to the ES). Ortengren and Andersson (1977) note that the abdominal muscles (RA, IO, and EO) show increasing activity while being stretched in trunk extension. However, this does not agree with the work by Lee (1982) indicating no RA activity during the pull task.

It therefore appears that the requirement for stabilization and control during dynamic activity results in situations where the assumptions that antagonist muscle activity does not exist must receive careful analysis to prevent the underprediction of total system muscle forces.

3. BIOMECHANICAL TRUNK MUSCLE FORCE MODELS

3.1. Static versus Dynamic Models

Computerized biomechanical models for trunk muscle force prediction models have progressed from static, two-dimensional analysis, to more recent attempts at understanding dynamic, three-dimensional stresses on the spine. These models attempt to accurately and realistically represent the mechanical loading and behavior of the lower back, while refraining from as much unnecessary complexity as possible.

3.1.1. Static models

Most knowledge concerning the reaction of the trunk internal spine loading structures has been based on static exertions of the trunk. Chaffin and Baker (1970) were the first to create a model of loading on the lumbar spine under occupational conditions. Their model, which was two-dimensional and static, consisted of several links representing the major articulations of the body. They computed the torque imposed on each joint and the subsequent compression on the spine. The effects of internal trunk loadings were not considered other than from trunk extensor muscles represented by a single equivalent trunk muscle. Schultz and Andersson (1981) later created a lumbar spine model that could be used for workplace assessment. Their model considered the effects of the external moments on the activity of ten trunk muscles and on intra-abdominal pressure. This model computed the compression and shear forces on the spine primarily under static conditions, but it was not designed to calculate additional spine loading attributable to co-activation of the trunk muscles.

The most obvious criticism of the static models is that most MMH activities are dynamic, therefore static models tend to underestimate the forces and moments because the internal loads imposed by dynamic actions are ignored. Many of the dynamic models have estimated that loading of the lumbar spine under dynamic conditions is 22.5–60% greater than loading under static conditions.

3.1.2. Dynamic models

Most of the dynamic models are based on kinetic or kinematic body motion information. The researchers computed the spinal loads attributable to internal forces generated by the body segments during motion as well as spinal loading from static gravitational forces. McGill and Norman (1986) developed a model of spinal loading during sagittally symmetric trunk motion using EMG data as input. This model incorporated previously published relationships of internal structure responses to motion conditions in order to assess spinal loading while tracking body segment motions. However, limited empirical evidence of the model's validity was provided. Marras and Sommerich (1991) developed and validated a three-dimensional dynamic model for predicting sagittally symmetric and asymmetric lifting moments and associated spinal loading during lifting.

3.2. Optimization-based versus EMG-assisted Models

Two basic approaches to partitioning the moment-generating duties among the trunk musculature have been reported in the literature: optimization-based and EMG-assisted models.

3.2.1. Optimization-based models

The optimization approach attempts to satisfy the reaction moment requirement by recruiting the various muscular contributions based on an optimization criterion such as minimization of joint compression and shear load, or first minimization of muscle contraction intensity and then spine compression load. One of the major problems with this approach is selection of appropriate objective functions as well as constraints to solve a redundant problem. It does not fall within the deterministic realm of mechanics and requires physiological data that are yet unavailable.

The application of a linear optimization technique for the prediction of trunk muscle force was suggested first by Schultz and Andersson (1981). In this study, the objective function was to minimize the compressive force on the lumbar vertebra. This objective function allocates the most forces needed to resist the external load to the muscles with the largest moment arms. To prevent this situation, Bean and Chaffin (1988) suggested an alternative procedure that involves formulating and solving two linear programming models sequentially using maximum muscle intensity and the sum of muscle forces respectively as an objective function. Han *et al.* (1991) proposed a nonlinear optimization force model using a nonlinear objective function for predicting forces in the muscles, and force contributions of the disc at the L3/L4 level of spine subject to a combined flexion and left lateral bending load. In these models, a cost function minimizes the sum of the square of the difference between the actual muscle stress and the maximum permissible muscle stress for various muscles. Chung *et al.* (1999) also suggested an optimization model that was modified to take into account the muscle co-activity. The model used three types of linear and nonlinear objective functions (minimize maximum muscle intensity, minimize sum of magnitudes of the muscle forces raised to power 3, and minimize sum of the muscle intensities raised to power 3) in various asymmetric lifting conditions. According to their results, these models reflect the twisting effect of muscle force vectors for eight primary trunk muscles when trunk rotation is

involved, and give good predictions of the forces of contralateral erector spinae, latissimus dorsi and external oblique muscles when compared with the EMG signals obtained from experiments. Among the three objective functions, minimization of the maximum muscle intensity showed the best prediction capability.

A common feature among the diversity of the optimization model formulation is the tendency for many muscles to have zero predicted muscle forces. Researchers have shown that the assumption of a lack of co-activity is not justified under dynamic conditions. Optimization models usually cannot account for co-activity because the quantity of the model solutions is limited by the number of non-boundary solutions, or six of the muscle forces, since the functional constraints are usually driven from the force and moment equations. Optimization models that have managed to include co-activity have forced it to occur by imposing artificial model boundary constraints. Consequently, these models often cannot predict the same level of co-activity that is observed experimentally.

3.2.2. EMG-assisted models

The second modeling approach to determine the load sharing duties among the trunk musculature is based on EMG and knowledge of joint kinematics. The advantage of these models are (1) they are not limited by the constraints of optimization objective functions; (2) they account for muscle co-activation forces via measurement; and (3) they typically use predicted, physiological coefficients for instantaneous validity checking.

McGill and Norman (1986) proposed a dynamic model of the lumbar spine that attempted to determine the significant forces in both active tissues using an EMG-assisted strategy, and in passive structures from estimated strain. McGill (1992) utilized the same methodology in calculation and assessment of the loads in both passive and active lumbar tissues during dynamic lateral bending.

Marras and Sommerich (1991) developed an EMG-assisted model in which relative muscle forces are determined from EMG activities normalized to a maximum, however, maximal and submaximal EMG change significantly with trunk angle, isokinetic velocity, and acceleration. Granata and Marras (1993) resolved this problem by normalizing the EMG input as a function of both trunk angle and asymmetry, and modifying for length and velocity artifact.

However, the EMG-assisted models are not general-purpose models. Rather, they are intended for use under laboratory conditions, and to interface with heavy laboratory instruments (EMG, dynamometers, etc.) that can assess the influence of motion-related biomechanical factors. It is often impractical to collect EMG activity in industrial locations due to the hostility (such as presence of magnetic field) of work environments. In this respect, Mirka *et al.* (1996) developed a simulation-based model that can generate EMG signals given a set of environmental conditions such as weight, moment, trunk posture and trunk dynamics. The shapes of best fit distributions were developed with multiple runs of the simulation and then the estimated EMG values were generated for bending and lifting activities by the multivariate Johnson distribution method.

Lee (1998) proposed another way for predicting the trunk muscle activities, named the fuzzy logic-based human expert EMG prediction model (FLHEPM), which utilizes two physical variables (trunk moment and trunk velocity) as inputs and ten trunk muscle activities as output. In this model, human expert knowledge obtained from task observations is utilized as model inputs in terms of linguistic values in order to predict the EMG signals.

The basic idea of the above two approaches is to develop an EMG prediction model for generating EMG data under given task conditions, which will be utilized as input to any EMG-assisted model, without direct measurement.

4. SUMMARY AND CONCLUSION

The assets and liabilities of state-of-the-art biomechanical models for predicting spinal loadings during MMH tasks have been discussed. Many of the biomechanical models that are currently available are still in their evolutionary stages and many researchers are trying to enhance the accuracy of models by removing several unrealistic conditions from them. In recent years, most research efforts have focused on investigating the dynamic conditions relative to static conditions, the effects of asymmetric conditions relative to symmetric conditions, and the interaction effect of dynamic and asymmetric conditions on the amount of stresses imposed on the spine. This raises a problem concerned with the trade-off between accuracy and applicability. Like any modeling approach, biomechanical models attempt to represent the workload accurately and realistically while refraining from as much unnecessary complexity as possible. The more input parameters are incorporated in the model, the more valid estimates for the workload might be obtained. As the models become more complex, however, their suitability for application in the occupational field declines considerably.

REFERENCES

BEAN, J.C. and CHAFFIN, D.B., 1988, Biomechanical model calculation of muscle contraction forces: a double linear programming method, *Journal of Biomechanics*, 21, 59–66.

CHAFFIN, D.B. and ANDERSSON, G.B.J., 1991, *Occupational Biomechanics* (2nd ed.) (Chichester: Wiley).

CHAFFIN, D.B. and BAKER, W.H., 1970, A biomechanical model for analysis of symmetric sagittal plane lifting, *AIIE Transactions*, 2, 16–27.

CHUNG, M.K., SONG, Y.W., HONG, Y. and CHOI, K.I., 1999, A novel optimization model for predicting trunk muscle forces during asymmetric lifting tasks, *International Journal of Industrial Ergonomics*, 23, 41–50.

ECKHOLM, J., ARBORELIUS, U.P. and NEMETH, G., 1982, The load on the lumbo-sacral joint and trunk muscle activity during lifting, *Ergonomics*, 25, 145–61.

FARFAN, H.F., 1975, Muscular of the lumbar spine and the position of power and efficiency, *Orthopedic Clinics of North America*, 6, 135–43.

GRACOVETSKY, S., FARFAN, H.F. and LAMY, C., 1981, The mechanism of the lumbar spine, *Spine*, 6, 249–62.

GRANATA, K.P. and MARRAS, W.S., 1993, An EMG-assisted model of loads on the lumbar spine during asymmetric trunk extensions, *Journal of Biomechanics*, 26, 1429–38.

HAN, J.S., GOEL, V.K. and KUMAR, S., 1991, A nonlinear optimization force model of the human lumbar spine, *International Journal of Industrial Ergonomics*, 8, 289–301.

KUMAR, S. and DAVIS, P.R., 1978, Interrelationship of physiological and biomechanical parameters during stoop lifting. In *Biomechanics of Sports and Kinanthropometry*, 181–91.

LEE, K.S., 1982, 'Biomechanical analysis of cart pushing and pulling', University of Michigan, Industrial and Operations Engineering, unpublished manuscript.

LEE, W.G., 1998, A neuro-fuzzy human expert model for predicting the EMG magnitude of selected trunk muscles based on lifting task variables, unpublished doctoral dissertation, University of Louisville.

MARRAS, W.S. and SOMMERICH, C.M., 1991, A three-dimensional motion model of loads on the lumbar spine: 1. Model structure, *Human Factors*, 33, 123–37.

McGILL, S.M., 1992, A myoelectrically based dynamic three dimensional model to predict loads on lumbar spine tissues during lateral bending, *Journal of Biomechanics*, 25, 395–414.

McGILL, S.M. and NORMAN, R.W., 1986, Partitioning of the L4–L5 dynamic moment into disc, ligamentous, and muscular components during lifting, *Spine*, 11, 666–78.

McNEIL, T., WARWICK, D., ANDERSSON, G.B.J. and SCHULTZ, A.B., 1980, Trunk strengths in attempted flexion, extension, and lateral bending in healthy subjects and patients with low-back disorders, *Spine*, 5, 529–38.

MIRKA, G.A., GLASSCOCK, N.F., STANFIELD, P.M., PSIHOGIOS, J.P. and DAVIS, J.R., 1996, The use of the multivariate Johnson distribution to model trunk muscle coactivation. In *Proceedings of HFES 40th Annual Meeting*, 584–88.

ORTENGREN, R. and ANDERSSON, G.B.J., 1977, Electromyographic studies of trunk muscles with special reference to the functional anatomy of the lumbar spine, *Spine*, 2, 44–52.

RAB, G.T., CHAO, E.Y.S. and STAUFFER, R.N., 1977, Muscle force analysis of the lumbar spine, *Orthopedic Clinics of North America*, 8, 193–9.

SCHULTZ, A.B. and ANDERSSON, G.B.J., 1981, Analysis of loads on the lumbar spine, *Spine*, 6, 76–82.

SCHULTZ, A.B., ANDERSSON, G.B.J., ORTENGREN, R., HADERSPECK, K. and NACHEMSON, A., 1980, Loads on the lumbar spine: validation of a biomechanical analysis by measurements of intradiscal pressures and myoelectric signals, University of Illinois at Chicago circle, unpublished manuscript.

TAKASHIMA, S.T., SINGH, S.P., HADERSPECK, K.A. and SCHULTZ, A.B., 1979, A model for semi-quantitative studies of muscle actions, *Journal of Biomechanics*, 12, 292–339.

YETTRAM, A.L. and JACKMAN, M.J., 1980, Equilibrium analysis for the forces in the human spinal column and its musculature, *Spine*, 5, 402–11.

Visual Perception, Age, and Driving

I. Noy

330 Sparks St, Tower 'C', Ottawa, Ontario, K1A 0N5, Canada

1. INTRODUCTION

It is widely held that vision provides 90% of driving-relevant information to the driver, though this has not been quantified. Over the past 25 years, there have been many studies which have investigated the relationship between specific visual attributes and driving performance in order to determine the visual requirements of the driving task or develop more effective driver vision tests. Many of these were correlational studies in which the visual attributes of drivers were measured and related to traffic convictions and accident experience (Burg 1968, 1971, 1975). Although these studies only produced weak correlations, researchers were able to speculate about the relative importance of various visual attributes to different driving behaviors (Henderson and Burg 1974). Shinar (1977) evaluated a battery of vision tests (including static visual acuity under normal illumination, dynamic visual acuity, detection-acquisition-interpretation, movement-in-depth in the central visual field, central and peripheral angular movement thresholds, field of view in the horizontal meridian and static acuity in presence of glare), and found that all of the attributes significantly correlated with accident rates in at least one of the subsets (based on age and time of accident), but none correlated in all the subsets. The results indicated that many of these tests had less than desired repeatability and none, including the SVA (static visual acuity), could reliably predict driving performance. For a review of the literature in this area, readers should consult Bailey and Sheedy (1988) and Shinar (1977).

Other researchers have found stronger correlations. For example, Hofstetter (1976) provided some evidence that poor visual acuity discriminated between drivers with multiple accidents and those with one or no accidents. Johnson and Keltner (1983) reported that drivers with binocular field loss had traffic violations and accident rates twice that of drivers with normal visual field. Interestingly, the acuity and visual field tests employed in these studies used conventional clinical tests rather than the more exotic tests favored by vision scientists.

In summary, however, correlational studies have generally failed to provide consistent evidence for a strong relationship between measures of visual capacity and driving. There are a multitude of good reasons for the weak statistical relationships found between visual measures and driving performance (such as citations and accidents). These include driver compensation for poor visibility, high individual differences in driving ability, diversity of driving situations and effects of other visibility factors such as windshield dirt, solar glare, fog and rain, obstructed signs, worn lanes, and driver compensation. Furthermore, most drivers with severe visual deficits have been denied licenses so that they are not represented in aggregate road safety statistics. Another important factor is that it is difficult to separate purely visual limitations from cognitive processes involved in areas such as attention, perception, recognition, and decision-making.

2. VISION SCREENING FOR ELDERLY DRIVERS

2.1. Current State of Driver Licensing

Driver visual screening is based primarily upon measures of acuity (e.g. using Snellen charts). A survey of Canadian provincial and US state licensing requirements reveals that the most common standard is 6/12 (20/40), with a range of 6/6 (20/20) to 6/21 (20/70). In addition, some jurisdictions have a standard for monocular and binocular horizontal visual field, color vision, and diploplia.

Briggs (1983) asserts that traditional acuity testing by itself or supplemented with simple color, peripheral vision, or other tests may be inadequate as a basis for visual standards because:

- the test is irrelevant to many real-world tasks such as driving
- there is little correlation between acuity and task performance
- acuity testing is often a clinical first step in discovering anomalies in visual function; this is irrelevant to performance screening
- there is no consistency in applying a standard for acuity
- criterion levels are often based upon the consensus of a panel
- experts from different jurisdictions do not agree upon any acuity standard or test procedure
- there is disagreement on the use of corrective lenses to meet visual acuity standards

2.2. Effects of Age on Visual Capacity

The visual characteristics of the elderly and the effects of aging on visual capacity is well documented. Nearly every aspects of visual function decreases with age, including:

- loss of accommodation
- decrease in color discrimination due to loss of transparency of lens
- decreased sensitivity to light
- decreased static and dynamic visual acuity (to the point where corrective lenses cannot completely overcome refraction losses)
- decreased visual field
- decrease contrast sensitivity
- increased susceptibility to disability glare due to increased intraocular light scatter
- decreased ability of pupil to dilate resulting in decreased scotopic vision
- retinal disorders include degeneration of macular retina, development of cataracts, retinal detachments, retinitis pigmentosa, degenerative myopia, diabetic retinopathy, hypertensive retinopathy, arteriosclerotic retinopathies and glaucoma.

In addition, higher-order functions have also been shown to decline with age. These include:

- visual spatial judgments (e.g. depth and speed) and organization skills
- psychomotor ability
- focused and divided attention
- short-term memory
- complex problem-solving abilities

2.3. Prospects for Improving Driver Vision Testing

There are seven useful criteria described for establishing visual standards: (1) visual attribute must be linked to safe driving; (2) pass/fail criterion should allow individual consideration; (3) vision

screening test should be reliable and valid; (4) the imposition of standard should aim to identify those whose vision can be significantly improved through proper refractive correction and eliminate those that cannot achieve minimum standards even with the best of correction; (5) the safety benefits must justify the costs of administrating the standard; (6) there should not be engineering or other interventions which are more cost-beneficial than the imposition of the standard; (7) the standard should permit imposition of restrictions, where appropriate (Bailey and Sheedy 1989).

Most, if not all, of suggested alternative standards fail to meet these criteria. A major difficulty has been the inability to identify the visual attributes most important for safe driving and determine criterion levels of performance. Moreover, most of the alternate tests of vision that have been suggested have serious shortcomings (e.g. low inherent repeatability of measures, high cost of equipment, high administrative burden associated with implementing new standards such as the need to provide space and trained personnel and the need to deal with a greater number of failed drivers).

Notwithstanding, the lack of success of past efforts in this area, researchers continue to seek improved techniques for screening drivers (young and old). Readers should refer to contrast sensitivity and effective visual field.

2.4. Accommodating age effect in the design of highways and vehicles

Since, for the foreseeable future, the prospects for improving driver vision testing are not high, it may be more appropriate to concentrate efforts on interventions which are likely to reduce the visual requirements of drivers in an attempt to achieve a better match between the demands of the driving task with the capabilities and limitations of elderly drivers. The most critical problems associated with advancing age are (1) declining visual acuity and contrast sensitivity, (2) glare sensitivity, and (3) decreasing ability to accommodate. Many of these problems can be alleviated by improved designs for highway signs, in-vehicle displays, and improved headlight systems. It is recognized that these interventions are not trivial or inexpensive, but they may offer more effective solutions for accommodating the growing population of elderly drivers.

3. USEFUL FIELD OF VIEW

Driving is a skill that requires processes beyond just visual input. Other central functions such as attention also play a key role in driving performance. Some of the most promising work relating vision and aging to driving safety has emerged from research on the Useful Field of View (UFOV). The UFOV is a measure of visual function, which takes into account the cognitive and the perceptual aspects of the driver's performance. The UFOV measure integrates both visual and attentional processes by investigating localization, detection, and identification performance of visual information obtained from the external environment.

Research on the UFOV has consistently shown that driving performance declines for the elderly when faced with additional task demands beyond driving (i.e. steering and button pressing versus just steering) (Schieber 1999). Older adults appear to experience difficulties in absorbing all the aspects of divided attention. In addition, the area of the visual field over which attention may operate decreases with age. It has been suggested that perhaps the elderly have a difficult time coordinating two or more different motor processes required to execute dual task responses into one effective motor scheme. For instance, it has been found that under divided attention conditions, steering accuracy improved significantly when a vocal response (as opposed to a manual response) was required on the secondary task. The size of the age-related differences due to dual task responses decreased when the several tasks performed required the use of different attentional resources. It appears that the functional field of view is narrowed in older adults due to the attentional demands imposed by driving (Schieber 1999).

REFERENCES

BAILEY, I.L. and SHEEDY, J.E., 1988, *Vision Screening for Driver Licensure. Transportation in an Aging Society: Improving Mobility and Safety for Older Persons.* Transportation Research Board Special Report 218.

BRIGGS, R.P., 1983, New Visual Validation Strategies: Some Reactions to the UC, *Berkeley Workshop. Commission on Peace Officer Standards and Training,* California.

BURG, A., 1968, *Vision Test Scores and Driving Record: Additional Findings.* Report 68-27, Department of Engineering, University of California.

BURG, A., 1971, Vision and Driving: A Report on Research. *Human Factors* 13(1) 79–89.

BURG, A., 1975, The Relationship Between Visual Ability and Road Accidents. *Proceedings First International Congress on Vision and Road Safety,* Paris.

HENDERSON, R.L. and BURG, A., 1974, *Vision and Audition in Driving. Final Report,* Systems Development Corporation.

HOFSTETTER, H.W., 1976, Visual Acuity and Highway Accidents. *Journal of American Optometric Association,* 47, 887–93.

JOHNSON, C.A. and KELTNER, J.L., 1983, Incidence of visual field loss in 20,000 eyes and its relationship to driving performance. *Archives of Ophthalmology,* 101, 371–75.

SCHIEBER, F., 1999, *Beyond TRB 218: A Select Summary of Development in the Field of Transportation and Aging Since 1988.* http://www.usd.edu/~schieber/pdf/Beyond218.doc.

SHINAR, D., 1977, Driver Visual Limitations, Diagnosis and Treatment. NHTSA.

Workload and Electro-encephalography Dynamics

G. Sammer

Center of Psychobiology and Behavioral Medicine, University of Giessen, D-35394 Giessen, Germany

1. WORKLOAD

Workload is a rather ambiguous concept. Subconcepts like cognitive load, informational load, attentional load and emotional load are not based on uniformly used theoretical considerations. The distinction between stress, strain and workload is difficult to make, especially in real work settings. Within these concepts psychophysiological measurements are influenced by a number of psychological and physiological factors due to intrasubject and task parameters in the interaction of subject with task. The importance of motivational processes, emotional processes and the total workload, resulting from both paid and unpaid work in its entirety, has been discussed within cognitive–affective–physiological frameworks. Moreover, there is evidence that time-related factors, individual differences, cognitive styles and coping strategies have effects on the psychophysiological measurements and cause changes in the dynamic properties of the physiological systems.

The term "workload" frequently describes concepts that focus on the costs necessary for a subject to perform a task. Within selective and divided attention theories the costs are seen as a problem of the limited capacity of cognitive resources. Therefore, load means costs as a function of available resources. The experimental manipulation of the resources during the performance of tasks by variations of modality, physical properties of the stimuli, the required reaction, variation of speed-features of a task, dual tasks and consequences from the quality of performance are standard factors in the investigation workload.

Mental workload describes the amount of cognitive resources involved in performing a task. The concept refers to a dynamic equilibrium between requirements of the task and available cognitive resources that are necessary for the task performance. Mental load is higher if the task requirements are larger than the available resources and vice versa. Mental load cannot be measured only by investigating one of both task requirements and the outcome of performance but also has to take into account the relationship between requirement and costs.

An alternative approach for the investigation of workload is to investigate single subprocesses of human information processing under workload conditions. Assuming independent resources, modality specific subsystems are postulated, which can be described by different cognitive operations. Such cognitive operations, which are included in practically all relevant theoretical frameworks, are stimulus encoding, stimulus analysis, semantic processing, recovering of information, decision-making and response selection. In a review on mental load Gopher (1994) suggested including automatic versus controlled performance,

implicit versus explicit processes, influence of capacity and interference processes, conscious attention control and others. In addition, Gopher also distinguishes between short-time load and vigilance, such as are of importance in different kinds of monitoring tasks in industrial production or traffic professions. It is of interest that these terms represent different kinds of cognitive processing which do not necessarily affect the outcome of a task.

2. PSYCHOPHYSIOLOGICAL APPROACH

As stated, workload research concentrates on the costs necessary for a subject to maintain its task performance. According to Gopher (1994) there are two classes of physiological cost functions: (1) general arousal and energy modulation measures and (2) correlates of specific brain activity involved mental load processes. The underlying assumption in point 1 is that task demands activate body systems to service the increased requirements for efficient task processing.

Although the psychophysiological approach seems to have satisfactory validity, serious theoretical problems still remain. Psychophysiological measurement requires the relationship between the underlying construct of workload and the physiological indicator-system to be defined. Usually this cannot be achieved by a single model or hypothesis that integrates all psychological and psychophysiological results. More often psychological models try to explain cognitive or behavioral relationships, and physiological and biophysical models try to explain the relationship between the activity of a physiological system and their measurements. Explicit psychophysiological hypotheses only exist for a few psychological constructs, such as stress. However, one difficulty with these psychophysiological hypotheses lies in the fact that different patterns of psychophysiological measurements for the same psychological construct are found between individuals.

Further, it is difficult to integrate concepts, which have their roots in different disciplines. For instance, from the neuropsychological point of view it is necessary to understand *where* and *how* information is processed in the brain. Cognitive modeling is based on empirical evidence without being mainly interested in the issue of localization; for neuropsychological modeling neurobiological evidence has to be added. It follows (1) that the concepts concerned are quite different and (2) that psychological constructs can not easily be mapped neither on neuronal structures nor on neurophysiological activity. Instead of mapping "workload" on cerebral structures or on, for example, electrophysiological patterns it appears to be useful to extract basic functions in order to gain concepts of lower complexity.

2.1. Electro-encephalography, Information Processing and Workload

Electro-encephalography (EEG) is a multivariate method (e.g. different recording leads, different frequency bands). Both tonic and phasic components can be recorded with high-time resolution. Recording is non-invasive and may be implemented in the laboratory as well as in field studies. The reactivity is considered to be very low.

EEG represents macropatterns of cortical cell-assembly activity. This activity is strongly influenced by subcortical structures. For example, a pacemaker-function of some specific

thalamic nuclei for the EEG a-activity is discussed frequently. However, empirical and theoretical evidence is not sufficient to explain the "secrets" of the brain rhythmic activity. Different from attempting to identify some cortical or subcortical generators it might be useful to relate the macroscopic EEG spatio-temporal dynamics to well defined brain states.

The toolbox for the analysis of EEG dynamics includes spectral analysis and auto-regressive methods, non-linear methods (dimensional complexity measures, entropy measures), coherence-analysis, the computation of the event-related de-synchronization and even more methods.

In many studies, mental workload assessment was investigated with event related potentials (ERP) of the EEG. From the literature, there is a good evidence that different components of the ERP are workload sensitive. In fewer workload studies the ongoing EEG was investigated. This fact might be due to two disadvantages of ongoing EEG analysis.(1) Spectral analysis requires the data to be stationary. This prerequisite condition does not meet psychological assumptions about human information processing. Brain states are discussed to be stationary in the range between 10 and a few hundreds of milliseconds. (2) The relatively short duration of stationary brain states reduces both the time precision and the correlation to any brain state. That makes it difficult to relate EEG measures to time-structured models of human information processing (as often done with ERP parameters).

3. WORKLOAD ASSESSMENT BY EEG

For EEG studies workload can be understand as changes in both quality and quantity of cognitive and cerebral processes caused by an increased task-demand under the premise of keeping task-performance constant. This definition makes possible (1) multivariate designs for the investigation of cognitive workload by variations of both quality and quantity features of task-demands, (2) relating psychological or psychophysiological indicators to concepts of human information processing and workload in neuropsychology, cognitive psychology and ergonomics, and (3) the inclusion of results from EEG studies on subsystems of human information processing (not on workload in general).

3.1. Spectral Analysis

Spectral analysis was the most frequently used method in studies of workload and ongoing EEG. The spectral power is computed for the usually inspected EEG frequency bands d (0.5–3.5 Hz), q (4–7 Hz), a (8–13 Hz) and b (14–30 Hz). Sometimes the a band is subdivided into a lower a_1-band (8–10.5 Hz) and an upper a_2-band (11–13 Hz). Similarly the b frequency range is split into several sub-bands. The meaning of the bands is well documented for arousal, relating slower waves, which usually have larger amplitudes, to low arousal states and sleep. However, activity in all of the frequency bands can be observed even during task-performance, where the relationship to functional states of the brain is manifold and presently is under intense discussion.

3.1.1. Alpha- and Theta-EEG

Studies gave evidence for a decrease of activity in the a-band during cognitive task performance. This was especially true for the a_2 frequency range in tasks such as concept learning, memory scanning, mental arithmetic, aviation (flight performance) or

playing computer games. Moreover, there is good agreement that an increase of q activity occurs during cognitive task performance. This applies especially for anterior electrode locations. Similar results on both frequency ranges were gained from computations of the event-related de-synchronization (ERD). However, there are as well reports about increased a-activity with workload. Increased a-power was shown to come up with mental arithmetic, imagery, mental rotation and others. There is some evidence that the faster a_2 frequency band is rather load-sensitive, while a_1-power is in correlation with changes of the attentional demand of both external and mental tasks. Owing to the highly complex structure of attention, the topographical distribution depends on the specific characteristics of the task. For instance, selective attention and controlled processing were shown to be in correlation with frontal activity, whereas the transition from controlled toward automatic processing is accompanied by an activity shift towards parieto-temporal regions. Another reason for conflicting results may be found in different implementations of the baseline recordings. It makes an important difference if eyes are open or closed, if subjects are instructed to just relax or if they are asked to perform a reference task. It is also strongly recommended that multiple baselines are implemented to monitor time-related changes. Various frequency band definitions contribute to a reduced comparability of results. If appropriate, this could be faced computing individual frequency peaks for each subject.

Topographical mapping of EEG parameters give information about any subsystem-configuration which is specific for a certain task-induced workload. Mostly, hemispherical lateralization, lobe specific distributions and coherence between various cortical areas are evaluated. Beside the rather coarse-grained topographical findings mentioned above, which were replicated many times, topographical distributions show a strong dependency from specific characteristics of a task. The investigation of hemispherical asymmetries did not reveal significance as an indicator for cognitive workload. Mostly, hemispherical asymmetries can be explained with well-known material-specific (verbal, figurative) effects.

Not always the appearance of slow EEG frequencies (e.g. q-EEG) is an indicator of increased workload. Long-distance train drivers also elicited increased power in slow EEG frequency bands during low-workload stages (Cabon *et al*. 1993). This effect was explained with low vigilance that is well known from monotone sustained attention tasks with a very low target probability. Similarly, if human limits are overtaxed, not only is a high cognitive workload stage is induced, but also there may emerge serious changes in emotion, motivation or arousal. Dependent on these factors it was shown that EEG parameters can shift toward the baseline values or return to the baseline. Therefore, it is important to collect behavioral data to discriminate high workload from overtaxing stages.

3.2. DC–EEG

The investigation of EEG dynamics is not limited to the conventional frequency range of the EEG. DC-EEG (often the term steady potential is used) is defined within a frequency range below d. For recording DC-EEG it is necessary to use adequate DC-amplifiers. Phasic changes of the DC-potential (DC-potential shifts) represent processes of cortical neuronal excitability and changes of arousal. DC-potential shifts appear spontaneously as

well as event-related. Negative DC-shifts were found in various workload conditions, e.g. traffic-noise exposure, tracking- and reaction-time tasks, mental arithmetic.

3.3. Non-linear Analysis

Recently, methods of the non-linear systems theory have been applied to the ongoing EEG. The same time series fed into the spectral analysis can be used for the computation of the so-called correlation dimension. This method is used to evaluate the geometrical representation of the reconstructed system dynamics. The correlation dimension is an estimation of the lower threshold of the number of degrees of freedom for the underlying system. Therefore, the correlation dimension is a measure for the overall complexity of the brain system dynamics estimated from the (EEG) time series. Following Pritchard and Duke (1995) the term dimensional complexity is now used for estimations of the correlation dimension.

There are few studies on workload and non-linear EEG dynamics. Sammer (1996) carried out studies on working memory-induced load and the dimensional complexity of the EEG. The main results showed an increased dimensional complexity with higher workload over frontal areas. Often, dimensional complexity was in negative correlation to the q-activity, which varies with task induced load. Results from a few studies on dimensional complexity and task-induced workload are very similar.

No differences were found between mental, physical or emotional workload indicating cognitive task involvement even under conditions of physical and emotional load. As for spectral parameters of the EEG, hemispherical asymmetries were dependent on characteristics of the stimuli.

4. SUMMARY

Both the theoretical framework and the practical measurement of mental workload are far from being standardized rules for the measurement or fundamental general theoretical modeling of mental workload. There is a remarkable overlap of concepts. Moreover, the theoretical gap between psychological concepts of mental workload and (psycho-)physiological measurement is far from being closed. Given the rich and elaborate psychological hypotheses that exist for some aspects of mental workload, priority should now be assigned to developing hypotheses to link these with psychophysiological phenomena to enhance the reliability and validity of psychophysiological measurement in the investigation of mental workload.

EEG dynamics were shown to vary with workload. Results from the application of different methods are in good agreement. Other than heart rate variability measures, which are load-sensitive but do not allow the discrimination of different amounts of workload (with the exception of memory load), there are several findings indicating that EEG dynamics as well allow a discrimination of hard and less hard workload stages. Moreover, topographical analysis in connection with a focus on cognitive subprocesses of workload may reveal both a better diagnosis of actual workload and may raise new questions of interest. From these results ongoing-EEG analysis can be especially recommended for investigations in which interest centers on different levels of task-induced cognitive workload, the discrimination of cognitive load related to different stimulus-types or differences between cognitive workload and attentional processes.

However, the relationship between EEG dynamics and workload is manifold and complex depending on specific characteristics of the workload setting. Attention must be paid to a number of factors that have an uncertain impact on the EEG dynamics. Such factors are for instance monotonous low-load work, overtaxing of individual abilities, interindividual differences (e.g. personality, individual physiological responses) as well as effects of cardiovascular activity on brain activity (e.g. baroceptor activity influences thalamocortical pathways, e.g. in connection with the transition of the body into orthostasis), where further research still is needed.

To conclude, there are many studies suggesting that EEG dynamic analysis may essentially contribute to powerful strategies of workload assessment. Nevertheless, a lot of both basic research and applied studies must be carried out to close the gap between theoretical frameworks and applied ergonomics.

REFERENCES

CABON, P., COBLENTZ, A., MOLLARD, R. and FOUILLOT, J. P., 1993, Human vigilance in railway and long-haul flight operation. *Ergonomics*, 36, 1019–33.

GOPHER, D., 1994, Analysis and Measurement of Mental Load. In D'Ydewalle, G., Eelen, P. and Bertelson, P. (eds) *International Perspectives on Psychological Science*, vol. 2: *The State of the Art* (Hillsdale: LEA), pp. 265–92.

PRITCHARD, W.S. and DUKE, D.W. 1995, Measuring "chaos" in the brain: a tutorial review of EEG dimension estimation. *Brain and Cognition*, 27, 353–97.

SAMMER, G. 1996, Working memory load and dimensional complexity of the EEG. *International Journal of Psychophysiology*, 24, 173–82.

Part 3

Performance Related Factors

Activity

J. Leplat

Ecole Pratique des Hautes Etudes, Laboratoire d'Ergonomie Physiologique et Cognitive, F-75005 Paris, France

1. INTRODUCTION

The notion of activity tends to be more and more central in ergonomics. One of its origins can be found in Russian psychological research (Bedny and Miller 1997). It became very important in European ergonomics, first in the French-speaking countries, then in the Scandinavian countries. American research, influenced by the trend called "situated action" accords it a growing interest. In France, a well-known ergonomist, Wisner (1995), wrote that "the central and original part of the ergonomic analysis of work is the analysis of activities." This ergonomics of activity is well illustrated into a French manual of ergonomics (Guérin *et al.* 1997). In a recent "Vocabulary of Ergonomics," de Montmollin (1997) distinguishes this ergonomics of activity from the ergonomics of human factors. Schematizing the difference a little, it can be said that the latter views ergonomics as the application of general knowledge to human functioning, which knowledge is, for the most part, acquired outside of the work situation. Emphasis is placed on the acquisition of this general knowledge: its application is not considered as a preferred object of research. Ergonomics of activity is centered on the activity of the person in his/her work situation, on what s/he really does. This activity is viewed as a complex process, which comprises essential dynamic and temporal aspects and which integrates the effect of multiple conditions. Activity must be distinguished from behavior that only constitutes the observable facet of activity: activity includes behavior and its regulating mechanisms.

Activity is often associated with the notion of task, which is defined as a goal to be achieved under determined conditions. Activity answers task goals and demands, but it answers them in a non-passive way. An operator receives the task and answers it according to his own purposes which can lead him to re-define the task, its goals and conditions. Operator, task and activity form a triad in which the terms are in a relationship of co-determination. Let one take the simple (?) example of using a video cassette recorder. The user's task is to obtain a correct recording, with the technical conditions defined by the properties of the instrument and various constraints. The user has a certain level of competence with regard to the instrument, as well as a more or less great desire to succeed alone, etc., which will influence his activity. The activity will express itself by observable trials and errors, by comments concerning the task, satisfaction (or annoyance!), etc. All of this cannot be straightforwardly and completely inferred from previous knowledge about the task and the user. Interaction between the three terms can be easily imagined. The situation can be made more complicated by introducing a collective component into the triad, with a person who comes to give advice.

To study activity is to study the organization of the terms of the triad. To do this, one can take the task or operator point of view, it being clear that the one involves the other as shown by the discussion below.

2. ACTIVITY FROM TASK POINT OF VIEW

In adopting this point of view, it should not be forgotten that the task is always for a given operator, in other words, the relationship between activity and task is conditioned by the operator who carries out the task.

2.1. Prescribed Task

The definition of "task" is often taken from Leontiev (1972) who considers a task as a goal to be achieved under determined conditions. "Conditions" are to be understood as everything that must be taken into account for the activity to be carried out. When these conditions are deemed to have negative effects, they are called "constraints." The task is called "prescribed" when the goal and conditions are fixed by hierarchical, thus external, authority (business, administration, etc.). They can be fixed with variable degrees of precision. In complex work, such as the nuclear field, goals and conditions are minutely defined by a procedure. At the other extreme, the task is only made more explicit by its goal or goals. In this case, one often speaks of a *mission* and the operator must define his own task himself, which task will guide his activity: this is the case, for example, with design activities.

A prescribed task is itself the result of an activity, designer's or organizer's activity. It depends on both what the designer or organizer wants to achieve and on the model he has build of the operator's activity. The designer or organizer more or less explain the task according to the competence the operator is deemed to have, more if the operator is judged less competent. The operator must be provided with the information he his lacking by means of external conditions for realization (instructions, diverse aids).

There are various nature of task conditions: physical (i.e. thermic, luminous ambiance, etc.), technical (i.e. type of tool, machine, etc. Rabardel, 1995), organizational (i.e. isolated work, type of management, span of discretion, etc.), social (belonging to an occupational or labor group, social protection, etc), etc.

2.2. From Prescribed Task to Activity

A certain number of steps can be distinguished in the activity which leads from the prescribed task to the final activity: prescribed task → task deemed prescribed by the operator → task re-defined by the operator → task truly realized (corresponding to the final activity). Final activity can be characterized by the analyst (realized task for the analyst) or by the operator himself (realized task for the operator).

Deviations between these tasks reveal activity characteristics and can be used as guidelines for analysis. For example, a deviation between the prescribed task and the task deemed prescribed can come from a poorly formulated task or from a lack of operator's competence.

2.3. Prescribed Task Analysis and Activity Analysis

Several authors talk about and use the essential distinction "task"/"activity," but employ terms other than those presented. For example, Hollnagel (1993) calls prescribed task analysis "task analysis" and activity analysis "task description or performance analysis."

Activity analysis is oriented by task analysis, but must be carefully distinguished from it. A pitfall to be avoided is that of

considering prescribed task as a reference by assimilating the deviations between the real and prescribed task to error. Deviations are very important for activity analysis, but they must be interpreted: the mechanisms which produce deviations as well as how they are involved in the activity must be discovered. Deviation raises a question, but does not constitute a response.

3. ACTIVITY FROM THE OPERATOR'S POINT OF VIEW

Activity is always the activity of a person often called "operator" in ergonomics. In addition, a good understanding of the activity depends on the operator's knowledge relative to the tasks to be realized. Two means by which the operator intervenes in the activity can be distinguished: as a task treatment system and as a generator of activity purposes.

3.1. The Operator as a Task Treatment System

To carry out his task and achieve his own objectives, the operator disposes of means, instruments and resources linked to his human quality and to acquisitions obtained during his life. Thus, he possesses a number of characteristics, in particular physiological and cognitive, which condition his activity and are present in the handbooks and manuals of ergonomics and human factors. For example, properties of perceptive, motor and cognitive systems, chronobiological and chronopsychological factors, modifications connected with age, postural factors, etc. will be looked at. Ergonomics studies (or should study!) these traits in relationship to work. For example, visual discrimination will be linked to instrument panel reading, nycthemeral variations with shiftwork, etc.

Cognitive ergonomics has lead to the giving of much importance to competence or the system of knowledge which enables the generation of activity in response to the demands of a class of tasks. Knowledge of this cognitive architecture underlying activity is essential for the fitting out and design of numerous types of work.

A growing interest is given to meta-knowledge, in other words, to knowledge the operator has of his own competence, knowledge which is essential for the management of his activity. It is actually necessary that he knows what he is able to do, what he can do under such and such conditions, and the limits beyond which it is no longer possible to respond to task demands correctly.

Among operator traits that condition the activity, personality must not be forgotten, even if ergonomics has only attached little importance to it up to now, perhaps because its study involves methods not familiar to ergonomics and its role is difficult to bring to the foreground and to take into account.

3.1.1. Activity cost and resources

Activity represents a cost (effort, constraint, load) to the organism and requires resources: here, the matter is the energetic or intensive component of the activity. The evolution of activity lead to an interest in mental load. Evaluation of this load and its nature has been the subject of much research. Regulation mechanisms of load are important in the work situation and were particularly brought to the foreground in the context of task allocation in collective work.

3.2. Operator as Generator of Activity Purposes

Activity has not as its only function to respond to task demands, and an operator has not as his only role to exploit his possibilities to respond to task demands. Through his activity, the operator aims at achieving his own purposes: to build his identity, to improve his self-image, to be recognized by his peers, to express certain values, etc. Therefore, activity responds to a double rationality, instrumental and subjective, and finds its meaning in relation to external demands (the task) and in relation to the author's own purposes. Work psychologists and sociologists emphasize the importance of the second type of rationality by noting, for example, "the commitment of the subject's values in his activity" (Clot 1995) and that "technical activity interferes with moral concerns" (Dodier 1995). Until recently, this facet of the activity takes little importance in ergonomics. The first traces of it can be found through the notion of motivation, understood as the value or interest the operator attaches to his activity. The agent's purposes were also approached from the point of view of physical and mental health protection. Thus, it could be shown that the risks incurred in activity arouse defense mechanisms which, paradoxically, can be expressed by an increase in risk (Dejours 1993).

4. ACTIVITY AS A PROCESS

Activity has a manifest, therefore observable, component and an internal, therefore hidden and unobservable component. The observable component corresponds to behavior while the second concerns the mechanisms that direct and regulate this behavior: cognitive, affective and cognitive mechanisms. This component, inferred from the first one, always maintains a hypothetical character. The difficulty lies in the fact that some mechanisms can only and very indirectly be expressed into observable activity. As a result, inference of activity mechanisms is a basic goal of the analysis of the same. Thus, the importance of the methods of activity analysis can be understood. These methods are, in general, derived from classical methods, but adapted to work situations. Let us take, as example, methods of aided observation (with event counters, videotapes), of aided consecutive verbalization (self-confrontation), of simulation, of errors, incidents, accidents analysis and of analysis of long term effects of activity (fatigue, psychological troubles), exploratory sequential data analysis, etc. Models also convey and favor certain types of methods.

The importance of *collective activity* is growing in ergonomics. This is the activity of a group of persons, often called team, who interact in order to realize the same task: what one does depends, in part, on what the other does. In relation to individual activity, this activity raises new problems: nature of the work group (composition, organization, competence, etc.), link between group structure and task structure, mode of task allocation within the group. It gives prominence to the main role of coordination and communication. It leads to interest in the common system of reference or shared mental model which is worked on to make communication easier (for example, language and gestural work systems).

Activities can also be characterized over diverse dimensions and the models proposed for them are themselves diverse and often non-exclusive. These models can be seen as revealing the

variety of activity properties: some types of these properties are mentioned below by emphasizing the activity characteristics they favor.

4.1. Models of Activity Structure

Several principles have been put forward to structure activity. First, one encounters the methods proposed for task analysis, such as the hierarchical analysis (Patrick 1992) and the goals-means task analysis (Hollnagel 1993). Russian psychology is the origin of an often exploited model which distinguishes three phases or functionings in activity: an *orientation phase*, preparatory to execution and aiming at establishing a model of the situation likely to serve as a base for the activity which follows; an *executory phase*, which allows the passage from the initial state to the final state to be obtained; and a *control phase*, which monitors the progress of the previous phases and evaluates the conformity of the result with the targeted goal.

Another framework of analysis, often cited, is from Rasmussen (*et al.* 1994) called the "*decision ladder.*" In the production of activity, a certain number of steps are distinguished: activation, observation, identification, interpretation, evaluation, task definition, procedure definition, execution. All these steps are not necessarily found in a particular activity, and typical short cuts enable activity levels to be characterized (see below).

4.2. Models of Organizational Level of Activity

An often-used model from Rasmussen (*et al.* 1994) distinguishes three modes of activity control: knowledge-, rule- and skill-based. The distinction "controlled activities/automated activities," which comes from experimental psychology, is also very popular. The operator often has the possibility to regulate his activity at one or the other of these different levels, which is what gives him a large capacity to adapt and, therefore, to be flexible in carrying out tasks.

4.3. Model of Regulation: Emphasis on Goal and Evaluation

This model, popularized by cybernetics, emphasizes the role of goal and of the interpretation of deviations from this goal, in other words the process that leads from deviations to the actions to be taken in order to reduce them. Activity can implement several regulatory loops, which can be fitted into each other or not. Some loops enable mental simulation of the controlled system and make anticipation of phases of functioning of this system easier, thus, the operator can prevent dysfunctions.

4.4. Systemic Models: Widening of the Field taken into Account by the Activity

Activity can be considered in relation to a more or less wide set of conditions and one can go more or less far back in the analysis of these conditions. Systemic perspective of activity analysis aims at articulating these different conditions. In particular, it has developed through the introduction of collective and organizational components: for example, let us look at the processes of intervention described and illustrated by Guérin *et al.* (1997) and Rasmussen's (*et al.* 1994) model with its "nested loops of social control." The systemic point of view leads to a modification of the methodologies which is less oriented toward an analysis into elements than to the overall functioning within various contexts.

5. CONCLUSION

To place activity in the center of ergonomics leads ergonomics to focus on its essential object: anthropocentered improvement of work conditions. Individual or collective activity expresses the quality of this improvement, and it is through a better knowledge of activity that technical and organizational improvements, in particular, can be made more adapted to the operators.

REFERENCES

BEDNY, G. and MEISTER, D., 1997, *The Russian Theory of Activity* (Mahwah: Lawrence Erlbaum).

CLOT, Y., 1995, *Le travail sans l'homme?* (Paris: la Découverte).

GUÉRIN, F., LAVILLE, A., DANIELLOU, F., DURAFFOURG, J. and KERGUELEN, A., 1997, *Comprendre le travail pour le transformer* (Paris: ANACT).

HOC, J.-M., 1988, *Cognitive Psychology of Planning* (London: Academic Press).

HOLLNAGEL, E., 1992, *Human Reliability Analysis: Context and Control* (London: Academic Press).

LEONTIEV, A., 1972, *Le développement du psychisme* (Paris: Sociales).

LEPLAT, J., 1997, *Regards sur l'activité en situation de travail* (Paris: PUF).

MONTMOLLIN, M. DE. (ed) 1997, *Vocabulaire de l'ergonomie* (Toulouse: Octares).

RABARDEL, P., 1995, *Les hommes et les technologies* (Paris: A. Colin).

RASMUSSEN, J., PERJTERSEN, A.M. and GOODSTEIN, L.P., 1994, *Cognitive Systems Engineering* (New York: John Wiley).

Activity Theory

G. Bedny

Essex County College, Newark, NJ07079, USA

1. INTRODUCTION

It becomes apparent that overcoming the traditional separation between studies of cognition, external behavior and motivation is essential for increasing the efficiency of ergonomics. A general approach that attempts to integrate cognitive, behavioral and motivational components of behavior into a holistic system is called the theory of activity. The conceptual apparatus in this approach differs significantly from those applied in behavioral or cognitive psychology. The term "activity" here derives from this conceptual apparatus and is exclusively restricted to human behavior and cognition. Psychologists who take part in the development of this approach emphasize that there is a big difference between human and non-human psychic processes and behavior. Humans, because of their ability to use language, can acquire knowledge of events that they have not actually experienced. Language allows an individual to consciously evaluate potential consequences of a specific behavior. The psychic processes of humans are developed not only according to the biological laws but also according to the laws of social–historical evolution (Vygotsky 1956). Through evolution, work activity and social interaction shaped human consciousness (Rubinshtein 1959, Leont'ev 1977). Tools made by people determine the specificity of their actions. One generation can transfer its experiences to another through actions and operations, which are to be carried out by means of specific tools. The theory of activity uses mental and motor actions performed by the person as the major units of analysis in the study of human behavior.

2. MOTIVES AND GOALS AS COMPONENTS OF ACTIVITY

An important aspect of the theory of activity is its analysis of the relationship between motive and goal as component of activity. Motives and motivation in general are considered a source of energy that drives the activity. Goal is the conscious and desired result of activity and it represents information that includes imaginative and logical components of the future results of a person's own actions. Not all mental reflections of future results are goals. Goals are future results that are connected to motives and can be attained as a result of person's activity. This brings us to a different understanding of goal from the one that exists in Western psychology (Bedny and Meister 1997). For example, according to Lee *et al*. (1989) goal has the attribute of intensity. The more intense the goal the more an individual will strive to reach it. Hence, goals "pull" activity. In the theory of activity the goal does not have such an attribute as intensity. A goal is only cognitive and informational in nature and can be precise, clear, and totally or partially conscious. The motive, on the other hand, is an energetic component, and may be more or less intense. The more intense the motivational process, the more desirable the goal, the more efforts a person will expend to achieve that goal. The goal is connected with the motives and creates the vector "Motive → Goal" that lends activity a goal-directed character. Psychological differences between the desired goal and current

situation are evaluated from the viewpoint of personal significance. The more significant the difference, the more the individual is motivated to act to attain the desired goal. The level of motivation depends not only upon the psychological distance between the goal and the current state of the situation, but also on the significance of the goal, consequences of failing to attain the desired goal, and obstacles between the current situation and the desired goal (Bedny and Meister 1999).

The analysis of the above material allows us to distinguish two forms of behavior. The more basic form has a reactive character. It occurs only in response to presented stimuli. Activity is a more complicated form of behavior that has goal-directed character. There are various approaches to understanding the concept of goal. For the behaviorist approach the goal is not a scientific notion. A person's behavior is described in terms of stimulus, reaction and reinforcement. In the field of cybernetics goal is a final situation, which can be achieved during the functioning of technical or biological systems. For psychologists who study purposeful behavior the goal is the end state toward which motivated behavior is directed. In the study of activity however, the goal is always a conscious mental representation of the future result of a person's own action and is connected with motive.

3. GENERAL CHARACTERISTICS OF SYSTEMIC–STRUCTURAL ANALYSIS OF ACTIVITY

One important aspect of the systemic approach is that the object of study is considered as a system that consists of elements and their connections. This allows us to describe the structure of the object under study. Representation of the object as a system depends not only on the specificity of the object, but also on what method of study is used to analyze this object. According to the systemic approach the same object can be described from different points of view, using different models that represent distinct aspects of the object or phenomena. Systemic analysis of a real object or phenomena with a description of its structure is called systemic–structural analysis.

Activity is not an aggregation of responses to diverse stimuli. Rather it is an organized system of distinct units with specific relationships among these units. This is why systemic–structural analysis constitutes a basic paradigm for approaching the study of activity. A system is goal directed if it continues to pursue the same goal through different strategies of behavior while environment conditions change. This system can also change its goal while functioning. Activity is a goal-directed system. Here we use the systemic–structural approach to study human behavior rather than the man–machine system.

We can outline the following methods in the study of activity: (1) cognitive analysis of performed activity, where human information processing is focused on the decomposition of activity into sensation, perception, memory, thinking and so forth; and (2) the activity approach, which is based on the systemic–structural analysis, and includes the parametric, functional and morphological methods of describing, and studying activity. The parametric method has received wide recognition in practice. It is the study of the aspects or elements of human behavior most important to the given research. These elements can include time of task performance, speed of separate reactions, duration of visual fixations, etc. This method of study is useful but very often not sufficient in the theory of activity. Other methods of study of

human behavior that allow scientists to describe activity as a system were developed. The concept of self-regulation of activity allows the integration of separate elements of activity into a system and makes possible a functional and morphological analysis of activity.

The basis of the functional description of activity is the analysis of the mechanisms of self-regulation. The concept of self-regulation is widely used in contemporary psychology. This term has been made synonymous with such notions as willpower, ego strength, and volition (Kuhl 1992). Kanfer (1996) considered self-regulation as a motivational mechanism that sustains attention and effort over time in the face of difficulty and failures. This understanding of self-regulation is not accurate. Self-regulation is a process related not only to living beings but to inanimate objects as well although they lack consciousness, will and motivation. We can give the following definition of self-regulation. Self-regulation is an influence on the system that derives from the system to correct its behavior or activity. In humans there are two types of self-regulation processes: physiological and psychological. The physiological self-regulation system is based on homeostasis. The psychological self-regulation system is directed to achieving an established goal. Self-regulation enables the person to change the strategies and goals of activity. For example, it was discovered that even in strictly predetermined stimulus–response situations, the strategies developed by the subjects varied with the conditions and instructions presented to subjects. This suggests that complicated self-regulative processes take place even in the simplest reaction tasks.

More complicated processes of self-regulation have been observed during the performance of more elaborate tasks. For example, we studied the interaction of complex choice reactions performed sequentially (Bedny 1987). The left-hand reaction was performed first in response to a sound stimulus. It followed immediately the right-hand reaction performed in response to visual stimulus. The auditory stimulus varied from one to four. The visual stimulus varied from one to eight. By changing the amount of stimuli possibilities for acoustical and visual tasks, performed sequentially, the researcher can determine how the subject allocates attention between these two tasks, depending on their complexity and on what strategies the subjects are using in general. The study showed that in spite of the fact that the two reactions were performed sequentially they could not be considered as independent of each other. The time necessary to perform the second reaction depends not only on its own complexity but also on the complexity of the previous reaction. The more complicated the first reactions the more time is necessary to perform the second reaction. This experiment demonstrates that subjects develop complicated strategies of activity that can be explained in terms of self-regulation mechanisms. Independent reactions when performed in sequence are integrated into a holistic system and influence each other. Hence, the study of activity as a system with a particular structure through systemic–structural methods is essential for ergonomics.

4. FUNCTIONAL DESCRIPTION OF ACTIVITY

Activity can be described as a functional system that consists of functional mechanisms. These mechanisms are integrated with the help of the feedforward and feedback interconnections to achieve the accepted goal. Graphical representations of such a
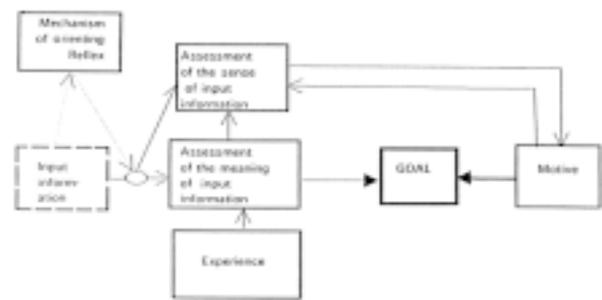
Figure 1. Functional model of goal formation process.

system are functional models of activity. Figure 1 presents an example of a model of the functional system that describes the goal-formation process according to Bedny and Meister (1997). This model is a dynamic system that consists of different function blocks, which are the major units of the activity analysis. The concept of a function block is very important in functional analysis and will be examined further in more detail. Here we just want to specify that the model of goal-formation processing consists of the following function blocks: mechanisms of the orienting reflex; assessment of the meaning of the input information; the assessment of the sense of input information; motive; goal and experience.

Let us briefly consider the above listed function blocks. The orientating reflex alters the sensitivity of different sense organs, changes blood pressure, the electrophysiological activity of the brain etc. The orienting reflex plays an important role in the functional mechanism of involuntary attention.

The function block "assessment of the meaning of the input of information" refers to a functional mechanism that represents reality in our consciousness. "Meaning" is the person's interpretation of the situation. It must be noted that meaning provides not only orientation in a situation, but also regulates the executive actions of the operator. The meaningful interpretation of information is possible if the operator has adequate past experience and the ability to use the required skills and knowledge in the present situation. Depending on the operator's past experience and abilities s/he can infer different meanings and understandings of the same situation. To account for this the function block "experience" is presented in Figure 1.

The function block "assessment of the sense of the input information" refers to objective meaning being translated into internal subjective sense. "Sense" is an evaluation of the significance of information for the operator, and represents the emotionally evaluative components of activity. The "meaning" defines the position and function of an object among other objects. The "sense" determines the relationship between external objects and the needs of the person.

The function block "motivation" is distinct from the function block "sense" because of its essential link with goal directness. The presence of the motivational function block allows emotionally evaluative components of activity to be transformed into inducing components. This gives the human self-regulative system its goal-directed quality, which can be represented as the vector "motive–goal" (designated by a bold arrow on Figure 1).

The model in Figure 1 explains how a person accepts or internally formulates a goal. The system described in this model is an entity, which consists of a combination of functional mechanisms directed to achieving the established goal. Each functional mechanism is designated as a function block, the contents and importance of which depends on the specific task.

The functional structure of activity can be defined as a system of interconnected and interrelated mechanisms or components. Functional mechanisms and function blocks are fundamental concepts in functional analysis. Studies in cognitive psychology as well as in theory of activity demonstrate that processing of information, which appears instantaneous to the individual, actually occurs as a series of subprocesses or stages of information processing that are short but measurable in time. In the theory of activity these stages are called functional mechanisms. These mechanisms are not something physical or directly observable, but rather constructs that are inferred as a result of certain chronometrical experiments and qualitative analysis. Examples of such mechanisms are sensory memory, iconic memory, scanning, recognition buffer, etc. (Sperling 1970, Zinchenko 1972). When functional mechanisms are presented as components of the functional model of activity with their feedforward and feedback interconnections, they are defined as function blocks. In functional models of activity such as the one presented in Figure 1, the function block has a much more complex architecture and takes much more time than in a microstructural analysis of the cognitive processes. In the functional model function blocks represent a coordinated system of subfunctions, which has a specific purpose in the structure of activity. For example, a function block can represent the creation of a goal, dynamic image of the situation, or program of execution, or complex functions of control and corrections. Any function block, which is introduced in a functional analysis, is the product of elaborate experimental and theoretical studies. Functional models composed of such blocks should be distinguished from multiple figures and schemes that also consist of boxes and are informally introduced by researchers to describe their results. To introduce a new block into a functional model one has to prove experimentally and theoretically that a certain functional mechanism does exist and then determine its interconnections with the other functional mechanisms of the developed model.

The functional analysis of activity describes the phenomena of cognition more comprehensively than existing theories. For example, the concept of situation awareness (SA) in ergonomics can be understood more precisely as a function block in the activity structure (Bedny and Meister 1999). This function block is responsible for developing a dynamic picture of the world. Functional models of activity allow a more detailed analysis of the strategies of and the prediction of an operator's behavior. The explanatory and predictive features of these models are facilitated by the approach that considers motive, goal, meaning, sense, etc. as parts of a system of interconnected mechanisms, that have certain functions in the structure of activity and influence each other. Further while many theorists assume the existence of a fixed goals, the functional model on the other hand emphasizes the process of goal acceptance or formation and its specific functions in the performance of a particular task. The functional analysis of activity allows comprehensive evaluation of the operator's performance. For instance it explains why an operator can neglect the safety requirements when he has high aspirations to reach the goal. Here is another example. The accuracy with which a pilot can read aviation instruments often depends more on the significance of the instrument than on its visual features. Another illustration would be that a subjective standard of success can influence the precision of performance.

5. MORPHOLOGICAL DESCRIPTION OF ACTIVITY

The morphological description of activity entails the description of the structure of activity in which the major units of analysis are actions and their components are called operations. We can give the following definition of the structure of activity in conducting a morphological analysis. The structure of activity is a logical and spatio-temporal organization of actions and operations performed to achieve a given goal. To describe the structure of activity it is necessary to subdivide it into tasks, which must then be individually described in terms of mental and motor actions. Each action has a separate, intermediate goal, which must be reached to attain the final and overarching goal of the entire task. An action is a relatively bounded element of activity that fulfills an intermediate conscious goal. In light of the fact that actions and operations are the fundamental units of the morphological description of activity, it is necessary to briefly describe them from the perspective of the theory of activity. Actions that can change the states of objects in the external world are called object actions, or motor actions. These actions include different motions, which are called motor operations. For example, action "move arm and grasp lever" is composed of two operations "move arm" and "grasp." Mental actions are classified according to mental processes dominating at the present time. Such actions can be complex and simple. If they are complex they are composed of several mental operations. A mental operation is a relatively homogeneous mental act that has no conscious goal. We can outline the following mental actions. Sensory actions allow us to obtain information about separate qualities of external objects such as color, temperature, or sound pitch. Perceptual actions allow us to perceive whole qualities of objects or events. Recognition of a picture is an example of perceptual action. Mnemonic (memory) actions are composed of memory processes such as memorization of units of information, recollection of names and events etc. We can also isolate imaginative actions. For example, one can mentally turn the visual image of an object from one position to the next according to the goal of one's action. Verbal actions are tied to verbal expression of different statements. Finally we can outline thinking actions and decision-making actions. A simple example is making a decision according to "if–then" rules. Other methods of classifying actions exist; an example of another classification is object–practical actions, which are performed with real objects, object–mental actions that are performed mentally with images of objects. Sign–practical actions are performed mentally with images of objects. Sign–practical actions are performed with real signs like receiving symbolic information from different devices and transforming this information. Sign–mental actions are performed mentally by manipulating symbols. Each action is completed when the person reaches the intermittent conscious goal. Internal actions are formed on the basis of external actions. There are two theoretical concepts of this process. One conception is based on the theory of internalization, and the other on the theory of self-regulation of activity (Bedny and Meister 1997).

The criteria by which actions are isolated can be technological or psychological. The classification of actions by technological criteria entails the subdivision of the task into specific elementary parts or subtasks that allow attaining intermediate technological result. For example, "move the lever," "read the meter" are actions isolated according to the technological principle. However, this description of action is often insufficient. As a result, in the second stage it is necessary to apply the psychological method of description. In this stage psychological criteria are used (Bedny 1987). For example, the action "move the lever" can be very different depending on the necessity to move the lever to an exact location at a distance of 10 inches with the control resistance of 7 lb; we can characterize this movement according to MTM-1 rules as M10C7. Any specialist familiar with MTM-1 knows what kind of movement the operator will perform. The similar principle is used in describing mental actions. For example, the action "read the meter" in technological terms in normal visual conditions according to MTM-1 is called "eye focus time" in psychological terms. In more complicated situations we can use verbal terminology applied in cognitive psychology. For example, "recognition of a familiar visual signal in normal conditions" can be considered as a standard perceptual action, if one knows the specificity of the signal, the viewing distance, and the illumination. It also has to be determined when the action begins and when it ends. Actions are usually first described in technological terms and only after that in psychological terms because often in the beginning there is no data necessary for the psychological analysis of actions.

6. FOUR STAGES OF SYSTEMIC–STRUCTURAL ANALYSIS OF ACTIVITY

Systemic–structural analysis of work activity entails four stages to the loop structure principle (Figure 2). The later stages sometimes require reconsidering the preliminary stages and at any stage activity can be described in different levels of detail. Each stage involves a cognitive, functional and morphological analysis.

The initial qualitative stage of description includes a general description of the task. In this stage the researcher characterizes the task in general terms, gives its verbal description, and attempts
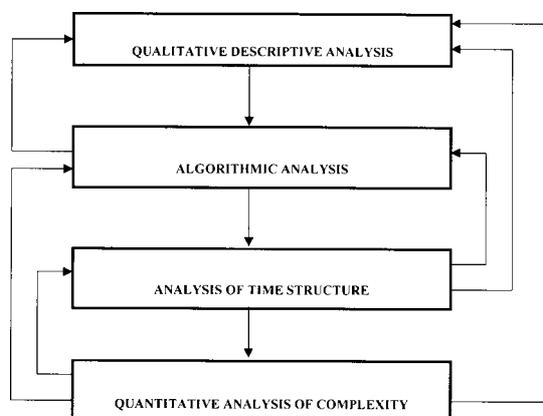
Figure 2. Four stages of systemic-structural analysis of work activity

to find out the most preferable performance strategies. The verbal description can be combined with the symbolic description of the task and can be done in different levels of detail. At the first stage of activity analysis the functional description plays the leading role. Here we study acceptance or formation of the goal, the significance of the task and its separate components and actions, subjective criteria of success etc. particular to the situation.

The second stage is the algorithmic description of activity. An algorithm of activity is a logically structured system of actions and operations, the order of which is determined by the set of rules defined by this algorithm. The human algorithms are significantly different from non-human algorithms. The algorithmic description of human activity derives from non-human algorithms. The algorithmic description of human activity divides the task into elementary actions that are called "operators" and "logical conditions." The latter determines the sequence of actions. At the same time several tightly interconnected actions can be described as one "operator." For example, the "operator" "move the organ of control to the specifically set position" is composed of two motor actions "move the hand and grasp the handle" and "move the lever to the specifically set position." The level of details of the algorithmic analysis often cannot be specified in advance and may have to be subsequently modified.

It is important to differentiate between prescribed and performed algorithms of activity. Prescribed algorithms are standardized while performed algorithms are those that are actually used by the operator. In the process of design the ergonomists develop prescribed algorithms. It is important that these prescribed algorithms of activity correlate as closely as possible probabilistic as well as deterministic (Bedny 1987). Deterministic algorithms have logical conditions with only two outputs, 0 and 1. Probabilistic algorithms can have more then two possible outputs each of which can vary in their probabilities.

The third stage of activity design is the description of the temporal structure of activity. It is possible to build a temporal structure of activity only after determining the duration of the actions involved in the task performance. The temporal structure of activity consists of the logical sequence of activity elements, their duration, and the possibility of them being performed simultaneously or sequentially (Bedny 1987, Bedny and Meister 1997). Before developing a temporal structure of activity it is necessary to describe the structure of activity algorithmically. Temporal structure of activity can not be developed until it is determined what elements of activity can be performed simultaneously, and what elements can be performed only sequentially.

The last stage of analysis is an evaluation of task complexity. The more complex the task itself the higher the probability that it will be difficult for the performer. The intervals of time devoted to standard elements of activity should be used as the units of measurement of complexity (Bedny 1987, Bedny and Meister 1997).

A quantitative evaluation of the complexity of task performance is possible only after developing a temporal structure of activity. A systemic–structural description of activity makes it possible to represent the structure of activity as a system of logically organized actions and operations. The goal of the design process is to create a system of interdependent models of the operator's activity. These models describe the structure of the operator's activity and can be used to design equipment and

develop efficient methods of performance. The main idea is that changes in equipment configuration probabilistically change the structure of activity. Suppose we wish to develop the time structure of activity during the performance of a particular subtask such as the following: An operator is simultaneously presented with a weak acoustic signal and a well-defined red light. The operator must detect the acoustical signal and perceive a red light at the same time. In response to this information he must grasp a lever on his right with his right hand, and move it straight forward to a specified position. The movement of the lever is complicated because he has to press a pedal with his right foot at the same time. On the other hand, if the acoustical signal is accompanied by a green light, the operator must move the same lever backwards, similarly accompanied by the same pedal movement with the right leg. We can determine the motor components of activity using system MTM-1. Based on an engineering psychology handbook or through chronometrical studies developed in cognitive psychology we may determine the duration of the cognitive components of activity. Based on this, the temporal structure of the activity may be developed (Figure 3). In our case we can extract the following actions in described subtask:

- Sensory action – detection of acoustic stimulus in threshold area;
- Perceptual action – recognition of visual stimulus in optimal visual conditions;
- Mental action – decision-making having two alternatives;
- Motor action – reach for lever and grasp it (method RA + G1A); and
- Motor action – move lever to exact location without releasing it (method MC);

The same rules that are used in MTM-1 representing particular motions apply. Symbol RA represents reaching the object in fixed location. G1A is to grasp the object when it can be easily done. MC means "move object to an exact position." FM means "press and release pedal with the foot." If we know the time structure of the activity we can calculate the complexity of task performance, including complexity of motor regulation of activity that has a cognitive nature (Bedny and Meister 1997). If one introduces some changes in equipment configuration, the time structure of the activity may change. Hence, one should evaluate the complexity of the task again. Comparing the complexity of task performance in both cases one can conclude which version is more difficult and demanding. Based on this, one can decide which physical configuration is superior.
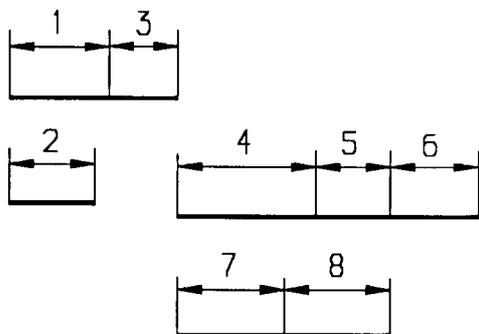
7. CONCLUSION

Activity places a great deal of emphasis on studying units of analysis of human behavior. It calls for the use of systemic–structural method of analysis of the operator's activity. As a result it has become possible to combine methods of study of cognitive psychology with methods developed in the theory of activity where the major concept is an action. The information processing stage of analysis of cognition should be integrated with activity analysis.

The description of the relationship between the actions of an operator and the physical configuration of equipment offers new opportunities in ergonomic design. Given the same equipment, this approach also allows the development of more efficient and safer methods of performance. It also allows one to study how the time-structure of activity changes during training, and based on it to develop more efficient skill acquisition. Owing to the orientation of the study of behavior as a system in activity theory, it enables the integration of cognitive, behavioral and motivational aspects of human activity, and as a result increases the efficiency of applied research.

REFERENCES

BEDNY, G., 1987, *The Psychological Foundations of Analyzing and Designing of Work Processes* (Kiev: Higher Education Publ.).

BEDNY, G. and MEISTER, D., 1997. In Koubek, K. and Salvendy, G. (eds) *The Russian Theory of Activity: Current Application to Design and Learning* (Mahwah: Lawrence Erlbaum Associates).

BEDNY, G. and MEISTER, D., 1999, Theory of activity and situation awareness. *International Journal of Cognitive Ergonomics*, 3, 63–72

KANFER, R., 1996, Self-regulation and other non-ability determinants of skill acquisition. In Golliwitzer and Bargh (eds) *The Psychology of Action* (New York: Guilford), pp. 404–23.KUHL, 1992.

LEONT'EV, A.N., 1977, *Activity, Consciousness and Personality* (Moscow: Political).

LEE, T.W., LOCKE, E.A. and LATHAM, G.P., 1989, Goal setting theory and job performance. In Pervin, (ed) *Goal Concepts in Personality and Social Psychology* (Hillsdale: Lawrence Erlbaum Associates), pp. 291–326.

RUBINSHTEIN, S.L., 1959, *Principles and Directions of Developing Psychology* (Moscow: Academic Science).

SPERLING, G., 1970, The information available in brief visual presentations. *Psychological Monographs*, 74.

ZINCHENKO V.P., 1972, *The Microstructural Method of Study of Cognitive Activity* (Moscow: VNITE), pp. 3–27.

Figure 3. Temporal structure of activity

Allocation of Functions: Past, Present and Future Perspectives

E.F. Fallon

National University of Ireland Galway, Ireland

1. INTRODUCTION

Allocation of functions is concerned with decisions regarding what agent, human or machine, or combination of agents will carry out individual functions in a system. It is regarded by ergonomics/human factors academics as a core activity in the ergonomics design process and university programs in human factors/ergonomics that do not encompass the topic would quite rightly be regarded as less than ideal. However, the importance attached to the topic in the classroom is not always reflected in the workplace. Reported experience tells us that that allocation of functions is not addressed as often as theory would suggest except in situations where regulatory requirements must be met and safety is paramount.

2. FROM PAST TO PRESENT

Historically, allocation of functions has primarily been explicitly considered in the context of large systems development. The first reference to the topic was by Fitts (1951) with respect to air traffic control systems. The approach to the problem proposed then was to assign functions to either men or machines based on their relative capabilities and limitations with respect to predefined performance criteria. The ensuing Men Are Better At/Machines Are Better At (MABA/MABA) tools were generally applied within the context of a rational, mechanistic, systems development model. Their application was extended to military and nuclear power systems, and also to decisions regarding allocation of functions between robots and people in manufacturing systems. The latter application resulted in the formulation of Men Are Better At/Robots Are Better At (MABA/RABA) lists. Typically

Table 1. Typical MABA/MABA and MABA/RABA list entries

Men (people) are Better At (MABA)	Machines Are Better At (MABA)
Detecting masked signals	Detecting signals from many sources
Pattern recognition in light and sound	Detecting signals at speed
Creative and inductive reasoning	Deductive reasoning
Responding to unexpected events	Processing large amounts of data quickly
Improvising using small – medium forces	Storing large amounts of data

Men (people) Are Better At (MABA)	Robots Are Better At (RABA)
Require less detail in job descriptions	Energy efficient
Flexible (reprogramming not required)	Operate continuously
Require less space	Can apply large forces
Require less capital investment	Have no social or personal needs
Are easily transported	Repeat operations accurately

MABA/MABA and MABA/RABA lists would consist of the following types of items:

In the intervening years there has been recognition that the comparative approach based on Fitts' Lists or derivatives of them are not particularly useful for a number of reasons:

i. Humans and machines are not alike and are therefore not comparable.
ii. A number of important criteria are not explicitly addressed as part of the allocation process, e.g. cost, reliability, job design, and organizational considerations.
iii. There is no consideration of changes in the relative capabilities of humans and machines, e.g. pattern recognition and advances in artificial intelligence advances in the case of machines.
iv. Allocations are static and do not take account of the changing nature of task scenarios or conditions, e.g. increases in human workload during disturbances.
v. The final allocation never seems to resemble that suggested by the MABA/MABA lists.
vi. Experience of addressing the issue in practice reveals that at the very least allocation of functions is an iterative process.
vii. The requirement for a formalized proscriptive approach to allocation of functions is not been justified.
viii. There is no explicitly agreed definition of the term function and how it differs from a task, i.e. is it functions or tasks that are being allocated?
ix. The emphasis in their application has tended to be on single operator machine systems or groups of operators and machines treated as several single operator–machine systems.
x. There is a growing realization that the nature of work to be carried out in human–machine systems can only be determined through doing that work and not by a priori specification.

Jordan (1963) viewed humans and machines as complementary to each other. In effect the decision to allocate to human or machine was not viewed as an either/or situation. While this concept has intuitive appeal, it has been difficult to operationalize in the context of system design and development.

There is little recorded evidence of the successful application of allocation of functions methods. Typically, the final allocation never looks like that envisaged by "Fitts' Lists" or indeed any other structured design method. It is generally acknowledged that in safety-related and large human–machine systems, some element of allocation of functions is carried out, however the process of allocation is regarded as much of an art as a science.

It has been argued that allocation of functions as a discrete design activity does not exist. At best allocations are outcomes of design and are regarded as implicit in the activity of designing. A concomitant view is that successful allocation of functions will evolve, partly by chance and partly to conform to our fancies of the moment.

More recently the requirement for iteration in the allocation process has been recognized. In practice this essentially means that following the activities of initial function allocation, task synthesis, task description and task evaluation — reallocation is carried out through resynthesis of tasks without necessarily referring back to the original functional specification. This has tended to draw attention to the lack of clarity in distinguishing between functions and tasks.

Those involved in the rapidly developing field of Computer

Supported Cooperative Work (CSCW) have begun to consider the allocation of functions concept in the context of group and teamwork and its social requirements. For this community, the key issue is whether allocation of functions as currently expounded is a useful construct as it is argued that many functions are, and can only be, allocated during the course of use of a system rather than during the design process. Consequently, the optimum or a satisfying human–machine mix cannot be achieved at the design stage. The argument is that the act of placing technology in a work environment changes work practices and that these changes only become apparent as problems or conflicts are revealed.

On the positive side there is a significant cohort of researchers and practitioners who hold the view that allocation of functions is a real design activity and that resources should be employed to try to achieve a fuller understanding of how it is carried out and to develop better methods for its implementation. Within this context, there is recognition of the need to understand the relationship between allocation of functions and the design process employed and to facilitate communication between other actors in the design process. There is also recognition of the need to widen the focus of allocation efforts from individual interactions with machines to team-based and organizational considerations such as job design, work organization, operator roles, and manpower planning.

3. THE CONTEXT OF ALLOCATION OF FUNCTIONS

3.1. Systems Development

The simplest and most primitive human–machine system can be analyzed in terms of its functions. However, in such cases, the allocation of these functions between human and machine is implicit in the design process, e.g. a spade or a hammer. Moreover, the choice of allocations is invariably restricted and often one which reduces human physical workload and extends or compensates for human capabilities or their deficiencies respectively. It is only when significant levels of automation are introduced that allocation of function issues have merited explicit consideration. Traditionally this was done within the framework of a systems development model at the earlier planning and conceptual design stages. The structure of such models was such that a single allocation was envisaged and the decision was normally an either/or decision — either to the human or to the machine. In theory, no iterations of the stages of the model were envisaged, though in practice the final allocation was rarely the initial one. The system development model initially employed by the US military industrial complex during the cold war period largely reinforced this rational, mechanistic approach to function allocation. In this model, detailed design of machine subsystem components was considered in parallel with detailed design of human subsystems. Such an approach requires a priori the allocation of functions to one or the other and implies subsequent independent consideration of each. In more recent years an understanding has developed that an iterative approach within the basic systems development framework is required.

3.2. Manufacturing

In the immediate post-World War II period, the approach adopted to allocation of functions in the context of discrete parts manufacturing systems was to automate those functions it was possible to automate and leave the remainder to the operator. Broadly speaking, the advent of Numerically Controlled (NC) machine tools resulted in the automation of machine tool control functions. The skill in determining cutter paths for tools was reallocated to programmers. The operator was left with responsibility for loading and unloading machines, turning them on and off, and intervening when required in order to avert or mitigate undesirable events. The initial development of Computer Numerically Controlled (CNC) machine tools, robotics, Flexible Manufacturing Systems (FMS) and Computer Integrated Manufacturing (CIM) saw little change in this approach. Ergonomists did make some attempts to consider function allocation in the context of work with robots, leading to Men Are Better At (MABA) Robots Are Better At (RABA) lists as outlined earlier. Despite this, there was little consideration of the broader manufacturing context in which manufacturing systems would operate.

The failure to achieve the anticipated benefits due to the application of technocentered CIM precipitated a rethink with regard to the role of humans in such systems. Efforts were focused on addressing the function allocation issue in the context of Human-Centered Computer Integrated Manufacturing (Human-Centered CIM). Human-Centered CIM advocates a work- or labor-oriented model of advanced manufacturing based on a strategy of group manufacturing in which operators are allocated the function of system control. The design approach employed involves the design of the technology, the social system, the interaction between the two, and the design process itself. System design is viewed as an evolutionary process involving user participation, prototyping, and continuous change. The approach includes as one of its main objectives a concern for the quality of working life. Consequently, in a human-centered system, human operators are regarded as an essential component of manufacturing systems and are given due consideration during the design process.

3.3. Supervisory Control

Supervisory control is a term used to describe the way in which operators and computers interact in controlling vehicles, continuous processes, robots, and also FMSs. Supervisory control activity is concentrated mainly at the human–machine interaction level and makes no pretences to explicit consideration of the quality of working life. The human operator normally sets the system goal by initiating set points, the computer effects some action to achieve the goal, measures its own performance relative to the goal, and subsequently adjusts its actions accordingly. The emphasis is on providing operators with the appropriate resources to carry out the following functions: planning, teaching, monitoring, intervening, and learning. In some respects, supervisory control is not an alternative to the earlier technocentric approach to human–machine systems. It recognizes that functions which can be automated will ultimately be allocated to machines, and that humans will carry out the remaining functions. However, within supervisory control, the human roles

required to carry out these remaining functions are identified and considered within a cognitive human factors framework.

3.4. Software Systems Design

Despite the fact that allocation of functions is viewed within the human factors/ergonomics community as a critical component in the design of human machine systems, it has not received similar consideration in the area of software systems design and HCI. This is surprising given that a significant number of Human Computer Interaction (HCI) textbooks contain sections on the topic. The predominant tendency in software systems design is to embody as many functions as possible in the software. However, unlike traditional systems development approaches, there is no explicit consideration of functions as mutually exclusive human or machine sets. This activity is implicit in the design process. There are a number of critical differences between the outputs and the context of each design process that may account for this. Some of the more important ones are illustrated below in table 2.

4. IMPLEMENTATION PHILOSOPHIES

In its initial inception allocation of functions was considered at the level of individual work activities at the immediate interface between human and machine. The main allocation criteria used were performance related, e.g. speed of operation, accuracy, and strength requirements. As outlined above, allocation decisions were made on the basis of comparative performance of human and machine with respect to these criteria. The process was viewed as a definite step within a sequential systems development framework. Given the rapid development in the capabilities and performance of machines, the logical consequence of this approach was that the majority of functions that machines could

carry out were allocated to machines and those functions that they were unable to carry out were allocated to humans.

Subsequent developments in the area were influenced by the realization that humans and machines were not comparable and should be viewed as complementary components within systems. They were also influenced by the recognition of the need to extend the allocation criteria to include, cost, reliability, maintainability, personnel requirements and safety.

Socio-technical theory provided the impetus for further developments. The locus of application of allocation of function decisions was extended from the immediate human machine interface to considerations of job design and organizational issues. The provision of meaningful work was endorsed as an acceptable allocation criterion.

Dynamic allocations in which functions are allocated to humans or a machine based on situation assessment has emerged as a possible approach to the issue. Successful application of this concept depends on the ability to identify situational factors that impact on allocation criteria and also to identify appropriate alternative human machine allocations. Typically, workload and predictions of human behavior are used to dynamically select the human–machine mix. The forms of dynamic allocation are varied and range from situations in which the machine is only used when the human requests help, to situations where the machine monitors the state of the human and intervenes when specified state variable thresholds are exceeded. As Older *et al.* (1997) point out, "Ironically it can be seen that dynamic allocation involves an allocation decision in itself, when deciding who/what should be responsible for initiating the change in allocation."

Some within the human factors community hold the view that allocation of functions is neither a useful nor pragmatic concept (Sheridan 1999; Fuld 1997). Sheridan argues that our

Table 2: Context and output differences between the human—machine design process and the software design process with respect to allocation of functions

Factors	Human machine system Design outputs and context	Software system Design outputs and context
Category	Bespoke and Specific	Generic
Operating Enviroment	Hazardous, can be safety-related or safety-critical	Less specific, rarely hazardous, hardly ever safety-related
Application areas	Aviation, Aerospace, Manufacturing (FMS), Medical systems, Process control — nuclear, power generation, petrochemical, pharmaceutical	Office enviroment. Home, Education establishments Manufacturing, e.g. MRP, Product database, Computer-Aided Design (CAD)
Configuration	Human as system components; Specialised software with relatively long time frame for replacement; Complex equipment operated and controlled through and by software	Human as user; Generic hardware platform; Software developed using generic programming languages or development enviroments; Replacement/re-release envisaged every 6–18 months
Quality standards	Comprehensive quality standards applied by outside bodies and client	Limited quality standards applied by outside bodies; Other standards selected by developer
Legislative contstraints	Possibly stringent legislative requirements imposed by local, national and international governments	Minimum health and safety standards imposed by national and international governments; Standards imposed by best practice and market demands
Consequences of error	Can be catastrophic; Significant loss of life; Significant loss of equipment; Reduced or no revenue.	Unlikely to be – catastrphic, loss of life. significant equipment losses. Short-term loss of revenue.
Design process	Tightly controlled, stage model or derivative; Possibly seperate contracts for individual stages; Multidisciplinary – hardware and software engineers; human factors engineers; Possibly significant human factors/ergonomics effort; Specified deliverables at decision points	Flexible stage model or derivative; Developer normally completes entire contract, largely software & hardware engineers; Minimum human factors/ergonomics effort; Specified deliverables at decision points; however the outcome is very flexible

ability to think through the human—machine mix in advance of trying it out is limited. McCarthy *et al.* (1999) paraphrase Sheridan's view as follows, "It is also unhelpful insofar as it attempts to limit the requisite variety which enables us to muddle through making sense of our experiences and learning from our mistakes, and in the longer term limits our understanding of both ourselves, the machines we make, and interactions between the two." Fuld, writing in the context of nuclear power, concludes that allocations are always implicit in design, but formal allocation models are not prescriptive models for design behavior. Moreover, he contends that function allocation provides a post hoc rationale for situated, pragmatic decision-making. For Fuld, human factors in design is concerned with building on acceptable solutions and weeding out the misfits between automation decisions and human capacities.

The CSCW perspective on allocation of functions is a much more radical departure from those previously described. It views formal system specification and design as no more than the commencement of a process of technology and practice in situ, which occurs mainly through developing and maintaining working divisions of labor. Essentially this acknowledges that work involves more than just transforming specified inputs to outputs and that operators are often engaged in workarounds and tailoring of technology to the working situation.

By definition working divisions of labor cannot be prespecified. They emerge through a process of work, negotiating, persuading, and coercing. The implications for design are that systems should have a high degree of tailorability and in addition to those functions assigned from without, should support mutual awareness and information sharing and self-organization. This view also has significant implications for the design process and procedures for quality and safety assurance.

5. METHODS AND CRITERIA

Though he has been predominantly associated with comparative performance of humans and machines, Fitts (1951) also gave other criteria due consideration in his paper, e.g. economic, manpower, and personnel issues. Despite this, they were largely ignored in subsequent proselytizing of his work. A short description of the main methods of allocation of functions that followed Fitts' pioneering work is included below: (Older *et al.* 1997; Fallon *et al.* 1997; Sharit 1997).

1. A quantitative evaluation of all reasonable function allocation arrangements with respect to selected criteria. It was envisaged that the criteria would be broader that those advocated in Fitts' lists and that relative weightings would be assigned to them for each arrangement. The weightings could then be summed or multiplied and the final decision on the human machine mix based on the outcome of the calculations.

2. An approach with four identifiable allocation strategies:
   - Mandatory allocations as dictated by legislative requirements or specific system requirements
   - A balance of value strategy in which allocations are made based on a relative goodness of fit of human or machine with respect to performance. The allocation assignment is not assumed to be an either or issue and the final outcome is dependent on the resources available to both human and machine:
   - Utilitarian and cost based approach — this is the opposite

of the technological imperative, however, initial selection of operators to carry out tasks must be subjected to relative cost considerations.
   - Allocations based on cognitive processing requirements and the quality of working life of the personnel involved.

3. A methodology with a sociotechnical perspective. This methodology encompasses many aspects of the strategies above, however it differs in that it includes organizational considerations and develops scenarios of prospective implementations. It also places less emphasis on performance criteria and comparability of humans and machines.

4. A computerized method that supports interdisciplinary design teams in developing complementary allocation of functions in Advanced Manufacturing Technology (AMT) systems. The emphasis is on ensuring that human flexibility is maintained in the final system through opportunities for the development and maintenance of both practical production skills and theoretical system knowledge. AMT system designers are provided with tools that enable them to analyze existing and envisaged tasks and provide support for human–automation interactions during the system design process. The method operates at three different levels: immediate human–machine interaction level, job level, and the work system level. The main criteria considered as part of the method include process transparency, dynamic coupling, decision authority and flexibility.

5. A team-based approach involving several stakeholders including designers, manpower-planning experts, operations staff, and human factors experts. The method involves capturing information requirements, allocation criteria, and operator and equipment requirements for each function in the system, and their subsequent consideration and allocation by the stakeholders. Significant state-of-the art knowledge of both hardware and software is required to successfully implement the approach.

6. An approach based on the configuration of knowledge required to carry out manufacturing functions. The basis of the approach is that organizations can be construed as a configuration of knowledge embodied in humans and machines that utilizes data to create information (e.g. the product data model) and its physical manifestations (products for sale). The problem is to optimize the configuration of knowledge and its allocation to humans and machines. The resulting output includes an allocation of functions, the definition of human roles and the distribution of management functions to these roles.

7. A method based on guidelines for allocating tasks within a manufacturing system. The method provides a series of detailed flowcharts for carrying out the allocations; however, it is recognized that much of this information is situation specific. There is also an acknowledgement that allocation of functions takes place within a multilevel environment and those initial allocations may be changed in the light of job design and work organization.

8. A language-based system for allocating cognitive and decision functions within a team supported by intelligent automation. It consists of three types of tools: a language for describing cognitive processes at the team level, a set of general principles for allocating cognitive functions to individual team members

and/or automation, and a set of measures relating the former to the later. The system develops a number of candidate solutions that are tested for their adherence to the general allocation principles. Iteration is implicit in this process.

9. A method that assesses the implications of human–machine allocation decisions on manning, recruitment and training policies. This method uses a socio-technical systems methodology to determine roles and responsibilities within a planned socio-technical system, and a scenario-based workshop approach to establish the implications of these.

10. A formal analysis method utilizing scenarios and a cost-benefit analysis of the impact of automation decisions on the performance of the primary role of operators and the relative demands placed on them within the generated scenario. The primary application area is aviation, however it could be extended to other areas.

11. A method that utilizes a joint cognitive system model of human and machine for specified target domains. Simple models of function allocation are rejected in favor of an emphasis on what is termed "function congruence" over a range of situations. Humans and machines are viewed as collaborators in achieving stated goals. The emphasis is on the shifting conditions of work, and the need to maintain the total equilibrium of the work situation. The *dynamic* aspects of allocation decisions are considered within a framework that includes human, technological, and (to some extent) organizational factors.

6. FUNCTIONS OR TASKS

One of the difficulties experienced in the application of allocation of functions is the question of what constitutes a function and how is it distinguished from a task? This question is similar to that often asked in formalized approaches to engineering component design, i.e. where does the artifact currently being considered fit in the hierarchy of assembly, sub-assembly, and component? Similar issues apply to decomposition of systems and sub-systems.

In the rational view of design, functions are viewed as activities that a system must carry out to achieve its goals. Such functions are considered to be independent of any implementation technology. In contrast, tasks are regarded as activities that people carry out in order to achieve a goal, subject to defined constraints and the availability of specified resources. This approach to functions depends for its validity on the assumption that systems can be formally and fully described in advance of building, testing or commissioning. It is largely supported and reinforced by stage models of the design process and sequential models of system development.

There are a number of difficulties with the approach. A number of researchers have used task descriptions as the unit of task allocation thus dispensing with the concept of a function as being independent of its implementation. More recent models of the design process such as the Spiral and Star models, and the user-centered design approach reject the sequential stage model view of design. While there is a recognition that the various stages exist in some form or other, there is emphasis on iteration, prototyping and user involvement throughout the design cycle. In other words, tasks are often considered before some functions are revealed through a process of testing prototype implementations in the workplace.

Different disciplines have different understandings of the term function. This lack of a common appreciation of the concept causes problems in communication between disciplines particularly in the context of the systems development process. For example, systems engineers tend to model functions using dataflow representations, while in contrast, human factors/ergonomics practitioners tend to model functions using task analysis techniques. It has been argued that the differences in representation account for the lack of consideration of inputs by human factors/ergonomics practitioners to the design process. A scheme to improve this situation has been proposed. In this scheme it is envisaged that logical functions (e.g. dataflow representations) would form inputs to the allocation of functions process and that tasks would be the output. Allocation of functions would be viewed as a linking activity between two competing requirements gathering methodologies in the design process, one used by engineers and the other by ergonomics/human factors practitioners.

The notion that design does not proceed in a rational formalized manner is a viable one. In particular, it can be argued that the concept of a function plays no worthwhile role in it. Design can be regarded as an evolutionary process in which failures are weeded out through a process of design and test. New designs are built on other previously successful designs. There can be little room for functional decomposition as part of this process, except as a component of a post hoc evaluation.

Researchers from the CSCW field question the notion that the essence of work can be described using functional decompositions. While recognizing that there are work allocations that can be made from without (during the design process), there is also an acceptance that a significant number of allocations can only be made from within (during system use). This latter concept has been referred to as "working divisions of labor" or the "work to make it work" and has been described earlier in section 4. At best, prespecified functions and allocations are regarded as one of many resources for situated decision-making.

7. STANDARDS AND REQUIREMENTS

Standards dealing with allocation of functions generally apply to large human–machine systems in the military and nuclear power domains. Typically, they provide a framework for carrying out allocation of functions, but do not necessarily require the use of specific allocation methods. Meister (1991) outlines a systematic and prescriptive method for carrying out allocation of functions in the context of military/naval systems design. The method concentrates on human–machine interactions and is summarized in section 5 above. The same method is recommended in UK Human Factors Defence Standards, though in more recently published defense guidelines greater emphasis is placed on human jobs and roles within the overall organization of the system. A similar emphasis is adopted within the Manpower and Personnel Integration (MANPRINT) military program. The MANPRINT framework recognizes that allocation of functions (tasks) takes place throughout the development cycle and is not necessarily a single identifiable event. In contrast to Fitts' Lists, MANPRINT considers the economics of tasks, the coherence of jobs and the flexibility in decision-making of human operators. It also evaluates operator workload and takes cognizance of the fact that military systems can be used in a wide range of scenarios and by a variety of personnel.

In contrast nuclear power standards have tended to be more prescriptive. Fuld (1997) provides a useful discussion of standards on allocation of functions to the design of new nuclear reactors. While lauding the descriptive evolutionary approach of the US Nuclear Regulatory Commission (USNRC), he criticizes the allocation process approach of the International Atomic Energy Authority (IAEA). He states, "This document still plainly suffers from FA's dichotomies; and though different sections of the document show conflicting philosophical perspectives, this at least adds some balance to the overall product. Its struggle for practicality, it is hoped, will go further in future revisions."

Allocation of functions is at the heart of recent draft European standards on work tasks with machines and human-centered design for interactive systems. In the former case it is part of a formal, systematic procedure for work task design. In the latter appropriate allocation of functions is described as one of four key principles of user-centered design. The standard emphasizes making sure human skill is used properly and that systems are effective, efficient and satisfying for their users.

8. SURVEYS OF USE

In the introduction to this paper, the bona fides of allocation of functions as part of human factors/ergonomics courses was alluded to; however, the extent to which it is practiced is not well documented. Older et al. (1997) reported on a survey of 32 practitioners in the systems development field. The methods reported as used for allocation of functions were not allocation of functions methods per se. They were components of larger systems development methodologies that did not cater explicitly for allocation of functions. They concluded that while the relevance of the allocation of functions process was obvious, it was rarely explicitly addressed in practice. Specific weaknesses in how to allocate tasks between humans, the consideration of social and organizational issues and the timing of allocation decisions were identified.

Bevis et al. (1996) reported on a NATO Workshop on "Improving Function Allocation for Integrated Systems Design" in a military context. They concluded that there was a real need for allocation of functions in systems development and that the most important outcome of doing so was the subsequent reduction in human cost and risk factors. The practical approaches reported at the workshop indicated that there was a departure in practice from the methods recommended in the humans factors literature. They also contradicted the popular view that organizational and social issues were neglected. The authors identified requirements for increased support of the creative design process through the provision of successful complete or partial solutions. They also identified the need for a comprehensive description of human factors criteria and their interrelationships including a trade-off matrix that helps with weighting of the relative contributions of the criteria.

Kearney and Fallon (1977) report on a survey of the role of the user in software systems design. Designers were asked what they understood by allocation of functions and at what stage in the design cycle it was carried out? They were also asked to identify the methods they use during this stage of design. The results revealed that the developers understanding of the allocation of function concept was limited. They were unable to identify any specific methods or tools they used for this activity

and where it fitted into the design cycle. The most favored approach was a "common sense" or "suck it and see" one.

Jenkinson (1997), in a paper on allocation of functions in industrial alarm systems, comments that, "in the majority of alarm systems studied by the author, alarm functions are not allocated according to any established ergonomics principles." The main reasons offered for this are that many of those who design industrial information systems are not aware of such principles or are insufficiently persuaded of their benefits. He also comments on the need for the availability of resources to implement available human factors/ergonomics techniques.

9. CONCLUSIONS

The study and understanding of allocation of functions has progressed from the early days of Fitts' Lists. There is widespread acceptance that a simple comparison between the respective capabilities and limitations of humans and machines at a functional level is an inappropriate basis for allocation decisions. The criteria for decision making have expanded to include economic, safety, job design and organizational criteria. The allocation of functions decision does not require assignment to either machine or human, rather it is a question of the selection of an appropriate human/machine mix. Allocation options are not regarded as exclusively static as technological developments have increased the possibility of practical realization of dynamic allocation. In theory the application domains of allocation of functions have broadened in scope due to developments in computing technology, nevertheless the practice does not reflect this. Allocation of functions can never be a single step in a sequential prescriptive design process. At worst it requires several iterations of allocate, test and reallocate. The nature of the human—machine system design process itself has changed and can involve extensive user participation and scenario modeling.

Despite the progress described above, the methods used by practitioners to implement allocation of functions have not advanced at the same rate or to the same degree as our understanding of the concept. It is clear from the literature that designers often completely ignore this phase of system development. Where it is explicitly addressed ad hoc methods or other vaguely related methods from extensive systems development methodologies are sometimes used. In other instances, designers rely on previous experience or what some have termed "common sense".

The relevance and usefulness of allocation of functions as proffered by the general ergonomics/human factors community has been questioned. One view disputes whether it exists as a separate stage in design. The main argument is that allocations that take place during design are implicit and that any attempt to impose a prescriptive design method on designers will impede them. The best one can do with allocation of functions is to use it as a means of identifying acceptable solutions and weeding out misfits between automation decisions and human capacities. A second view contends that it is futile to attempt to specify the human/machine mix in advance of trying it out. Allocation of functions limits the requisite variety needed to enable learning and to facilitate the emergence of satisfying solutions.

A less pessimistic but more radical perspective is the one that contends that complete allocation of functions cannot by definition take place prior to use of a system. Allocation of

functions during design is but the commencement of a process of technology and practice in situ which occurs mainly through developing and maintaining working divisions of labor. Allocations, it is argued, emerge through a process of work, negotiating, persuading, and coercing. Acceptance of this approach has implications for the nature of the design process and its outputs. For example, it would be reasonable to assume that once a system has been commissioned and is in use, the underlying user model utilized during the design process would no longer apply. Consequently, any a priori analysis relating to for example, human operator actions or safety analysis of human–machine interaction would no longer be necessarily valid. Ideally human–machine system designs should have a high degree of tailorability and in addition to those functions assigned from without, should support mutual awareness and information sharing and self-organization.

What is the outlook for the future of allocation of function? Older *et al.* (1997) have identified future directions for research and development in this area. They include:

i.   Development of methods for allocating tasks within teams.
ii. Integration of task allocation work from the ergonomics/ human factors and organizational psychology domains.
iii. Development of tools for generating viable alternative allocations.
iv. Integration of allocation of functions methods with existing system development methods.
v.   Evaluation of the effectiveness of explicit allocation of functions methods.
vi. Development of further decision criteria and identification of rationale for trade-offs between criteria.
vii. Involvement of users in all aspects of the development of methods and criteria.
viii. A contingency based method for dynamically allocating tasks.

If one were to accept that allocation of functions is not relevant to design, then it is recommended that no further development should be carried out. However, given its proposed usefulness for weeding out misfits, it would seem worthwhile expending some energy developing a design support tool containing details of design solutions containing acceptable allocations.

For those espousing technological determinism, it would be helpful if it could be established what role if any is envisaged for ergonomics/human factors in the machine–machine design process.

The CSCW perspective of the allocation of functions concept demands serious consideration. We need to explore the practical implications for design of adopting this view. In particular, issues pertaining to the scope and duration of design and the involvement of users should be addressed. It seems reasonable to suggest that any products of design should cater for and

facilitate as many permutations as possible for working divisions of labor. Alternatively, we should be able to bound the problem through the identification of viable candidate solutions.

REFERENCES

BEVIS, D., ESSENS, P., and SCHUFFEL, H., 1996, *State of the Art Report: Improving Function Allocation for Integrated Systems Design* (Wright-Patterson Air Force Base, Ohio: Crew Systems Ergonomics Information Analysis Center).

BOWERS, J., 1997, The allocation of function: a Computer Supported Cooperative Work (CSCW) perspective. Presented at ALLFN'97, Revisiting the Allocation of Functions Issue: New Perspectives, First International Conference on Allocation of Functions. 1—3 October, Galway, Ireland.

FALLON, E.F., BANNON, L., and McCARTHY, J., 1997, (eds), ALLFN'97, *Revisiting the Allocation of Functions Issue: New Perspectives. Proceedings of the First International Conference on Allocation of Functions Vol.1.* (Louisville: IEA Press).

FITTS, P.M., 1951, *Human Engineering for an Effective Air Navigation and Traffic Control System* (Washington DC: National research Council).

FULD, R.B., 1997, The fiction of function allocation, revisited. In Fallon, Bannon, McCarthy (eds), *ALLFN'97, Revisiting The Allocation Of Functions Issue: New Perspectives, Proceedings of the First International Conference on Allocation of Functions Vol.1.* (Louisville: IEA Press), pp. 17—30.

JENKINSON, J., 1997, Allocation of functions in industrial alarm systems. In Fallon, Bannon, McCarthy (eds), *ALLFN'97, Revisiting The Allocation Of Functions Issue: New Perspectives, Proceedings of the First International Conference on Allocation of Functions Vol.2.* (Louisville: IEA Press), pp. 27—38.

JORDAN, N., 1963, Allocation of functions between men and machines in automated systems, *Journal of Applied Psychology*, 47, 161—5.

KEARNEY, D. and FALLON, E.F., 1997, The role of the user in the Irish Software Industry. In Fallon, Bannon, McCarthy (eds*), ALLFN '97, Revisiting The Allocation Of Functions Issue: New Perspectives*, *Proceedings of the First International Conference on Allocation of Functions Vol.2.* (Louisville: IEA Press), pp. 167—76.

McCARTHY, J.C., FALLON, E.F., and BANNON, L., 1999, Dialogues on function allocation. *International Journal of Human Computer Systems*, In press.

MEISTER, D., 1991, *Psychology of System Design, Advances in Human Factors/Ergonomics, 17* (Amsterdam: Elsevier).

OLDER, M.T., WATERSON, P.E., and CLEGG, C.W., 1997, A critical assessment of task allocation methods and their applicability. *Ergonomics*, 40, 151—71

SHARIT, J., 1997, Allocation of functions. In Salvendy (ed.), *Handbook of Human Factors* (New York: Wiley), pp. 301—39

SHERIDAN, T.P., 1999, Function allocation: algorithm, alchemy or apostasy. *International Journal of Human Computer Systems*, In press.

Arousal States and Human Performance

G. Matthews

Department of Psychology, University of Cincinnati, OH 45221, USA

1. INTRODUCTION

Arousal may be defined as a continuum of states of mental activity, ranging from deep sleep to drowsiness, alert wakefulness and agitation. "Activation" is sometimes used as a synonym or related construct. The psychological theory of arousal originated in studies using the electroencephalograph (EEG) to assess central nervous system (CNS) function, pioneered by H. Berger in the 1920s. Researchers of the 1940s and 1950s, such as D. Lindsley, noticed a correspondence between the pattern of electrical activity recorded from scalp electrodes and the person's state of wakefulness (for a review of early studies, see Duffy 1962). Alert, mentally active subjects showed predominantly low amplitude waveforms, with a frequency of 13–30 Hz (b waves). As arousal decreased, amplitude tended to increase and frequency decreased, generating a waves (relaxed state 8–12 Hz) or q waves (drowsy state 4–7.5 Hz). At the same time, animal studies showed that stimulation of the reticular formation provoked both behavioral activation and concomitant changes in the EEG. Studies suggested that arousal may also be indexed by measures of autonomic nervous system (ANS) activity. Measures such as heart rate, electrical conductance of the skin and constriction of vascular muscles are linked to arousal associated with the "fight-or-flight" reaction, and the sympathetic branch of the ANS.

Hence, traditional arousal theory makes two key assumptions. First, physiological reactions can be described in terms of a single dimension of brain activity or excitement influencing both CNS and ANS measures. Second, this physiological dimension is psychologically meaningful: brain states correspond to mental states and to characteristic styles of behavior. It is often supposed that moderate levels of arousal promote effective action and optimal performance. There may be human factors problems deriving from the impaired performance of both tired, under-aroused operators, and excited, over-aroused operators.

2. MEASUREMENT OF AROUSAL

The physiological definition of arousal implies that a person's arousal level should be measurable from several alternative psychophysiological indices. A basic psychometric assumption is that alternative indices of the same underlying construct should be positively correlated. A person whose EEG shows b activity should also show an elevated heart rate, high skin conductance (due to perspiration), and other signs of CNS and ANS arousal. Unfortunately, alternative indices of arousal often fail to correlate with another, so that the construct resists reliable psychophysiological measurement. This finding is a serious problem that calls into question the validity of arousal theory. It has inspired three alternative research strategies: dealing with methodological difficulties in measurement, discriminating multiple dimensions of arousal and measuring arousal through self-report.

2.1. Psychophysiological Measurement

One possibility is that arousal is a valid construct, but it is difficult to measure. Psychophysiological indices may reflect factors other than brain state. For example, heart rate is influenced by muscle tension and activity, so an apparently aroused person may merely be fidgeting. Also, psychophysiological responses to increasing external stimulation may be non-linear. It is sometimes claimed that a process of "transmarginal inhibition" operates to decrease CNS arousal when the person is in danger of over-stimulation. Another problem is individual differences in the expression of arousal in psychophysiological measures. Some people might show arousal primarily as increased blood pressure, others as elevated skin conductance ("response stereotypy"). Improvements in psychophysiological methods may yet provide a reliable measure of overall brain activity. However, the complexity of psychophysiological responses demonstrated in existing research suggests this outcome is unlikely.

2.2. Multiple Dimensions of Arousal

One explanation for the observed dissociations between different psychophysiological indices is that they index different brain systems. The brain may have separate circuits controlling reactions to reward and punishment signals, for example, which may be independently aroused or de-aroused. There have been various attempts to develop multiple arousal theories, discriminating different systems, often on the basis of neurotransmitters. For example, it has been suggested that arousal associated with noradrenergic systems relates to selective attention, whereas dopaminergic arousal relates to motor activity. However, the impact of these theories on human performance studies has, so far, been limited.

2.3. Measurement of Subjective Arousal

Arousal may also be defined and measured psychologically rather than physiologically, through self-report measures of mental state. Thayer (1989) proposed that there are two dimensions of subjective arousal. A dimension of energetic arousal contrasts vigor with tiredness, whereas tense arousal contrasts nervousness with relaxation. Self-report arousal is reliably measured by questionnaire or adjective checklist. The disadvantage of such assessments is that respondents may deliberately or unconsciously distort their responses, but there is considerable evidence for the validity of self-report. Self-report scales are sensitive to external manipulations of stress and arousal, and predict performance in some paradigms. Self-report arousal correlates to a moderate degree with psychophysiological indices, although the neural basis for subjective arousal remains uncertain.

3. INFLUENCES ON AROUSAL

Arousal theory assumes that, at any given moment, the person's arousal level reflects both environmental factors such as sources of stress, and personality or dispositional factors. Indices of arousal are generally more sensitive to environmental stress factors than to personality. Environmental factors may be subdivided into those associated with external stimulation, and those related to the internal physiological environment. In contemporary research, environmental agents are conceptualized as "stressors" that influence a variety of aspects of state (Hockey 1984). Most researchers favor a multiple dimension conception of arousal,

within which different stressors may affect different physiological or subjective dimensions.

3.1. External Environmental Influences

Environmental stressors may influence arousal through both their physical and psychological properties. Loud noise, heat and cold all tend to increase arousal primarily because they are physically aversive. However, threat stimuli such as failure on a task or criticism from other people may raise arousal because of the way the stimulus is evaluated or appraised. The positive motivations and cognitions evoked by incentives for good performance may also be arousing. The distinction between physical and psychological stressors is not clear-cut, because response to physical stimuli is modified by the person's appraisals. Noise stimuli may be more "stressful" and arousing when they are appraised as uncontrollable and disruptive to ongoing activities.

3.2. Internal Influences

Various arousing or de-arousing agents influence the internal environment. The simplest example is psychoactive drugs. Stimulants such as amphetamine, caffeine and nicotine affect neurotransmitter action directly and tend to increase CNS and ANS arousal. Tranquilizers and sedatives decrease arousal. Another source of arousal variation is biological rhythms. There is pronounced variation in arousal with time of day, although the time at which arousal is highest depends on the arousal index used. Body temperature and indices of ANS arousal are maximal in the early evening, whereas subjective alertness and energy are highest at midday. Several distinct aspects of arousal may be controlled by internal "clocks" or oscillators driven in forced oscillation by external time-cues ("zeitgebers"). The phase of the menstrual cycle may also influence arousal, with women tending to experience higher tense arousal and lower energetic arousal premenstrually. A further type of "internal" influence is abnormal physiological states, such as disease. Sleep deprivation normally lowers physiological arousal, although arousal indices continue to show the normal circadian rhythm in the sleep-deprived person. Fatigue induced by prolonged or demanding task performance may also lead to loss of arousal. In these instances, low arousal is likely to be experienced as decreased energy; tension may even increase in some circumstances.

The mediating role of brain arousal systems in controlling the effects of stress factors is controversial. Some researchers believe that the person's cognitions and appraisals of physiological activity may be equally important. For example, placebos have been shown to influence arousal when the person is expecting a drug effect. Arousal is sensitive to suggestive techniques, such as the Velten technique which affect the person's mood and self-cognitions. It seems likely that arousal responses are determined by an interaction of physiological and cognitive factors but the nature of the interaction is not well understood. Some stress factors (e.g. circadian rhythms, sedative drugs) may be more strongly determined by physiology, whereas others (e.g. premenstrual tension, task-induced fatigue) may be more sensitive to cognitive factors.

3.3. Personality and Arousal

Arousal may be influenced by the interaction of environmental stressors with personality factors. Some individuals may be more prone to states of high or low arousal than others. The two personality factors most often implicated in arousal response are extraversion–introversion and anxiety. According to Eysenck and Eysenck (1985), introverts are more easily aroused than introverts, due to greater sensitivity of a reticulo-cortical circuit in the brain. In fact, the psychophysiological evidence for higher tonic cortical arousal in introverts is inconclusive. There is stronger evidence that extraversion–introversion influences transient, phasic arousal responses, but the behavioral implications of these effects remains unclear. The further personality dimension of trait anxiety is also said to relate to high arousal. In this case, the psychophysiological evidence fails to support the arousal hypothesis, and the behavioral consequences of trait anxiety are usually attributed to the worries, or interfering cognitions, that typically accompany anxiety. It remains likely that people differ in their arousability, but current personality theory has not been very successful in predicting which individuals will be most aroused in given circumstances.

4. AROUSAL AND PERFORMANCE

4.1. The Yerkes–Dodson Law

It is useful to distinguish traditional and contemporary perspectives on the relationship between arousal and performance. The traditional view is expressed by the "Yerkes–Dodson Law", which makes two statements. First, the relationship between cortical arousal and performance efficiency is described by an inverted-U function, so that moderate levels of arousal are optimal for performance, and states of low or high arousal are detrimental. Second, the precise optimal level of arousal is inversely related to task difficulty. Easier tasks are performed best when arousal is relatively high (alert, mentally active state), whereas a relatively low level of arousal is preferable for more difficult tasks (relaxed state). The law refers to a 1908 study conducted by R.M. Yerkes and J. D. Dodson that showed curvilinear relationships between strength of electric shock and learning of underwater mazes in the mouse, although the study was concerned with motivation rather than arousal. The relevance of this study to human performance has been questioned.

Testing of the Yerkes–Dodson Law has been impeded by the lack of a reliable psychophysiological index of arousal. Many researchers have followed a behavioral approach, advocated by D. E. Broadbent in the 1960s, that manipulates arousal through various stress factors such as noise, anxiety or time of day. Often, the aim is test to whether stress factors interact as predicted by arousal theory. For example, the combined effect of two arousing stressors, such as noise and anxiety, should be especially detrimental to performance. In general, such studies provide only patchy support for the Yerkes–Dodson Law (Matthews and Amelang 1993). Some pairs of stressors conform to the law, but others do not. The behavioral approach also has other weaknesses. A serious methodological difficulty for performance studies is that the interaction method does not test the Yerkes–Dodson Law stringently. With typical designs, it is possible to fit various patterns of data to the law *post hoc* (Hockey 1984). It is also difficult to distinguish multiple dimensions of arousal using this approach, although additive-factors methods have been used in attempts to do so. A further weakness of traditional studies is that they often fail to distinguish different components of performance. Cognitive psychological critiques of arousal studies

emphasize that arousing agents may have quite different effects on different processing functions.

4.2. Arousal and Cognitive Patterning

Contemporary studies typically acknowledge the multidimensionality of both arousal and performance. Hockey (1984) advocates a "cognitive patterning" approach in which the effects of stressors are evaluated across performance indices of various processing components, such as selective attention, speeded response and short-term memory. Different stressors tend to provoke different patterns of change in processing, which are only partially attributable to arousal change. Other mechanisms such as distraction, and motivational and cognitive change also contribute to stressor effects on performance. For example, fatigue states are de-arousing, but the detrimental effects of fatigue primarily reflect loss of task motivation and effort. Stressor effects may reflect both changes in basic information-processing functions, such as changes in attentional capacity, and changes in strategy, such as willingness to exert effort to maintain effective performance. The challenge for arousal theorists is to discriminate the subset of performance changes that directly reflect change in one or more dimensions of arousal, and to develop cognitive–psychological explanations. Hockey (1984) suggests that one of the principal effects of arousal is to increase attentional selectivity, so that primary tasks are prioritized at the expense of secondary tasks. Most arousing stressors provoke changes of this kind. Revelle (1993) proposes that arousal increases availability of resources for "sustained information throughput", required for attentionally demanding tasks, but decreases resources for short-term memory. The evidence that arousal improves attentional efficiency is stronger than evidence for detrimental effects of arousal on short-term memory. For example, many, though not all, arousing agents reduce vigilance decrement.

4.3. Individual Differences in Arousal and Performance

Contemporary studies also focus on individual differences in arousal, often using self-report assessment of arousal. Higher energetic arousal is associated with superior performance on a variety of demanding attentional tasks, including high event rate vigilance tasks and controlled search for various stimuli. High energy seems to become increasingly facilitative as task demands increase (Matthews and Davies 1998). These findings are consistent with Thayer's (1989) view that energy is a marker for a general psychological and physical mobilization for vigorous action. In contrast, the second major dimension of subjective arousal, tense arousal, does not relate consistently to performance. Personality effects on performance may sometimes be mediated by individual differences in arousal. Extraverts often perform better than introverts in stimulating or arousing conditions, which may reflect lower cortical arousability in extraverts. However, this explanation is challenged by the weakness of the association between extraversion and arousal, and by studies showing that extraversion effects vary with the information-processing demands of the task.

5. CONCLUSIONS

Arousal theory was at one time seen as the best explanation for stressor effects on human performance, and it still has value as a unifying principle. However, its application to performance studies has declined. The construct originated in psychophysiological studies, but it has proved to difficult to measure from CNS or ANS indices. It is also difficult to link psychophysiological indices to the multiple brain systems influencing performance. Furthermore, effects of arousing agents on performance vary with the precise information-processing demands of the task, indicating the need for a more differentiated, cognitive–psychological account of performance change, and a psychological rather than a physiological definition of arousal. Contemporary work generally focuses on specific stressors, and the multiple mechanisms that may contribute to the "cognitive patterning" of their effects on performance. Only some of these mechanisms reflect arousal processes. One such mechanism is the increased attentional selectivity often observed in states of high arousal. A second mechanism is the enhancement of attentional function associated with high subjective energy, and with agents which boost energy, such as stimulant drugs. Within this narrower view of arousal effects, research is exploring both the precise information-processing functions sensitive to arousal, and their neural underpinnings.

Traditional arousal theory is of very limited practical utility. It points to both under-and over-arousal as potential problems in operational settings, but it does not allow accurate prediction of performance failure. Contemporary research indicates the importance of fine-grained assessment of performance and state in operational environments: arousal change is embedded within often complex stress reactions. A detailed understanding of the operator's cognitions of the environment and its demands is required for designing systems for stress-tolerance, and for training operators to cope with stressors. In some settings, arousal may be much less significant than faulty cognitions, unwillingness to exert effort and distraction from worries and bodily sensations. Nevertheless, arousal effects such as neglect of secondary information sources (high arousal) and lack of attentional resources (low arousal) may sometimes be a focus for intervention. Good design of displays can mitigate against over-selectivity of attention, and attentional demands can be reduced to counter under-arousal. Future technology may provide 'neuroergonomic' interventions such as continuous psychophysiological monitoring for potentially maladaptive arousal changes.

REFERENCES

DUFFY, E., 1962, *Activation and Behavior* (New York: Wiley).

EYSENCK, H. J. and EYSENCK, S. B. G., 1985, *Personality and Individual Differences: a natural science approach* (New York: Plenum).

HOCKEY, G. R. J., 1984, Varieties of attentional state. In Parasuraman, R. and Davies, D. R. (Eds) *Varieties of Attention* (Orlando: Academic Press), pp. 449–83.

MATTHEWS, G. and AMELANG, M., 1993, Extraversion, arousal theory and performance: a study of individual differences in the EEG. *Personality and Individual Differences*, 14, 347–64.

MATTHEWS, G. and DAVIES, D. R., 1998, Arousal and vigilance: the role of task factors. In Hoffman, R. B., Sherrick, M. F. and Warm, J. S. (Eds) *Viewing Psychology as a Whole: the integrative science of William N. Dember* (Washington: American Psychological Association), pp. 113–44.

REVELLE, W., 1993, Individual differences in personality and motivation: 'non-cognitive' determinants of performance. In Baddeley, A. and

Weiskrantz, L. (Eds) *Attention: selection, awareness and control* (Oxford: Oxford University Press), pp. 346–73.

THAYER, R. E., 1989, *The Biopsychology of Mood and Arousal* (Oxford: Oxford University Press).

Attention and Human Performance: A Cognitive Neuroscience Approach

A. Berger and M. I. Posner

University of Oregon, Eugene Oregon and Sackler Institute for Human Brain Development, Cornell Medical College, New York, NY, USA

1. A HISTORICAL REVIEW: MAIN QUESTIONS AND PARADIGMS

While human capacity is impressive in many areas, there are basic limitations to human performance. Eyes cannot be moved to two different locations at the same time and one cannot pronounce two different phonemes simultaneously. Human limitations go far beyond these physical constraints. Even if two activities are not physically impossible, performance does deteriorate when two different operations must occur at the same time even when input and responses are quite separate such as in reading a book while listening to TV. It seems as if the mind is constantly selecting a subset of stimuli, and this selection affects mental processes and behavior. This basic feature of performance gives rise to one of the most widely used metaphors regarding attention, the metaphor of the light beam that selectively illuminates stimuli sources, spatial locations and mental structures.

The modern era of attention research began during World War II in England. Pioneers like Broadbent, Cherry and Welford began the information-processing approach. Human performance was analyzed according to different processing stages and the main question of attention was at which stage was selection done.

1.1. Selective Attention

At the perceptual level, a person can, for example, selectively attend to one specific voice, enhancing it from the noise background. In the 1950s the basic experimental paradigms concerning selective attention was dichotic listening and shadowing. The subject repeated back information from one ear, while additional information was presented to the other ear. The question was to what extent subjects process information on the *unattended* ear. The first modern theory of attention was Broadbent's Filter theory (1958) that arose from an analysis of selective listening. The main idea was that selection occurs early in the processing stages. All stimuli are analyzed at the level of physical attributes, then the attentional *filter* selects stimuli for further processing. Stimuli that are filtered out do not reach semantic identification. According to *early selection* theories the bottleneck is localized at an early stage of the perceptual process, whether it is in the auditory, visual or other modality.

Consistent with the notion of early selection was evidence that subjects are unable to report almost anything about properties and content of the ignored stimulus. Subjects could report some physical features of the unattended message. For example, whether it was a male or a female voice and if they noticed if the feature changed while they were shadowing the attended message. However, they did not notice if the message started, for example, in English and then switched to German.

Filter theory was an example of an extreme early selection theory. At the other extreme were late selection theories. According to this view the bottleneck is localized after the semantic identification stage. There are no capacity limitations on the recognition of objects, only on memory processes, awareness and responses. Classic examples that seemed to refute early selection theories were breaking through ignored and unattended stimuli, like noticing one's own name in an unattended channel (Moray 1959). Another example in favor of the late selection account was that while subjects attended and shadowed a message in one ear, non-attended words presented to the other ear elicited galvanic skin responses (GSR) if they were previously associated with electric shocks (Corteen and Wood 1972).

Forty years of intensive research has supported elements of both early *and* late selection. When the overall processing load is very light, late selection theories do well, but with high processing load little unattended information survives. Some of the theories that were clearly "early" or "late" have been updated to support elements of both. One example is Treisman's Feature Integration Theory (FIT) (Treisman and Gelade 1985, Treisman and Gormican 1988). This theory dealt with search for a target in a complex field. According to this theory only simple visual attributes are registered and stored in parallel. These features are combined and integrated to form an object representation. This integration requires attentional focus to the object's location. Thus, attention is the glue required for object perception. In this framework the selective focusing of attention is early and precedes object recognition since it is considered a prerequisite for the object's representation. However, more recent descriptions of FIT (Treisman 1993) include ideas of late selection based on object files and the top-down selection of objects to control responses.

In the context of neuroscience there is clear evidence that selective attention influences visual system processes perhaps as early as the primary visual cortex. In this sense there is clear evidence for early selection. However, behavioral studies show that information not selected by attention may reach high levels of analysis including semantic levels. Thus, in appropriate conditions late selection can also occur.

1.2 Divided Attention

As mentioned above, there are some tasks that people seem able to carry out at the same time, while for other tasks this seems very difficult to do.

In the 1960s dual-task studies took a somewhat different direction than the dichotic-listening studies. In dual-task studies of human performance subjects were asked to perform two independent tasks in close temporal proximity. A leading theory at that time was that the source of interference between tasks was access to a limited capacity central processor. This single processor must be engaged to complete certain mental operations as memory retrieval, mental transformation and response selection.

In response to the finding that interference between tasks depended on their capacity demands, *Resource Theories* were developed. The main idea of these theories was that resources can be flexibly allocated to different tasks. For each task, two different stages were defined: (1) as long as performance of the task improves as a function of resource investment, the task is called *resource limited.* (2) However, there is a point at which more resource investment will no longer improve performance. At this point the

task is said to be *data limited* (Norman and Borrow 1975). Under dual-tasks conditions one can plot a *Performance Operating Characteristic* (POC) that specifies quantitatively the possible tradeoff in allocating resources between the tasks.

An alternative view to the unified resource pool suggests that there are *Multiple Resources.* One such view suggested a three-dimensional matrix of separate resources for the different processing stages (encoding, processing, responding) and for the different modalities (visual versus auditory, verbal versus spatial) (Wickens 1980). The initial conceptualization of this theory suggested a set of non-overlapping pools, each defined by a combination of levels on the three orthogonal dimensions. Literally, it predicted that task demanding non-overlapping resources would time-share perfectly (e.g. walking while listening to music through Walkman earphones). Evidence was found that, tasks within the same cell in the matrix interfere more with each other than tasks in adjacent cells. However, the data did not support the idea that tasks in adjacent cells are perfectly time-shared. For example, monitoring simultaneously to visual and auditory channels for detecting a target is impaired in comparison with single-channel monitoring. This is found both when targets are defined semantically, and when targets are defined physically within each modality of input (Treisman and Davies 1973). Moreover, it was found that when subjects monitor many channels for a target, whenever a target is detected, the processing of targets of any other channel drops dramatically (Duncan 1980). This finding leads to the idea of hierarchical resources. This is a kind of compromise conceptualization, between resource and central capacity views. While different modalities may have basic separate resources, there is still a central resource that could be defined as executive attention. This central supervisory
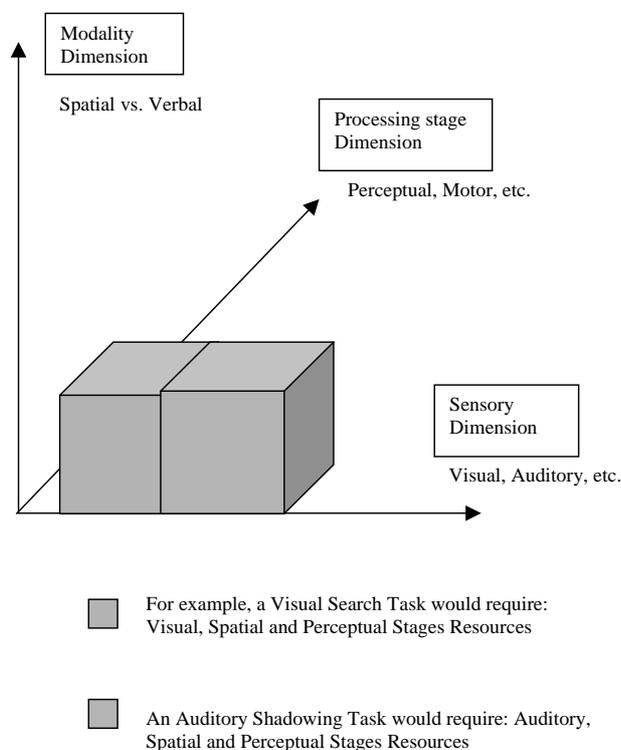
Figure 1. The idea of multiple resources, following Wickens (1980).

resource would be involved in the strategic allocation of other resources, monitoring, target detection, etc.

2. ATTENTION IN COGNITIVE NEUROSCIENCE

2.1. New Methodologies

The introduction of neuroscience methodologies to the study of attention have included the use of micro electrodes with alert primates, anatomical methods like computer tomography (CT) and magnetic resonance imaging (MRI), and functional methods like positron emission tomography (PET) and functional MRI (fMRI). These methods have permitted more meaningful investigation of the localization of cognitive functions in the brain. Recording from scalp electrodes can be used to provide clues regarding the temporal dynamics of anatomical areas involved in tasks. Although knowledge of the precise neural mechanisms responsible for attentional operations is still incomplete, many of the brain areas and networks involved have been identified.

2.2. Discovering the Neuro–anatomy of Attention

Based on behavioral data, neuropsychological data and the addition of the new imaging techniques data, Posner and Raichle (1996) summarized the three attentional networks and their localization in the brain. The anatomy of these three networks is presented in Figure 2.

2.2.1. Orienting network

The orienting network deals with shifting and focusing of visual attention. Orienting responses are usually made by movements of the eyes and head causing foveation of the stimulus. These are called *overt* shifts of attention. However, it is also possible to orient attention without eye movements. These are called *covert* shifts of attention. A second important discrimination; the one between *reflexive* orienting (meaning that a cue pulls attention to orient to the stimulus, e.g. a cursor blinking on PC screen) versus *voluntary* orienting of attention (meaning that the subject is given an instruction to attend to a location in the visual field, or that he has a internal motivation for doing so, e.g. a basketball player facing a player from the opposite team, would focus attention in order to pass him the ball accurately. This would probably be done covertly, otherwise his opponent would guess where the pass might go).

Neuropsychological evidence (including split-brain evidence) suggests that these attentional operations involve a brain network including the posterior parietal lobes, the pulvinar nucleus of the thalamus and the superior colliculus. PET studies have supported the importance of this functional network and have shown striking similarity between the anatomy of eye movements and of covert attention (Corbetta 1998).

2.2.2. Executive-control network

This network has been related to control of goal directed behavior, target detection, error detection, conflict resolution and inhibition of automatic responses. The executive control network seems to include the midline frontal areas including the anterior cingulate gyrus, supplementary motor area and portions of the basal ganglia. Neuro-imaging studies have shown activity in this area during tasks that require mental effort such as in dealing with conflict, handling novelty, developing anticipation and detecting errors.

Recently it has been shown that tasks involving both cognitive and emotional controls produce activation in the cingulate (Bush *et al.* 1998). There is some evidence that pathologies related to deficits in exercising executive control and monitor goal directed behavior, might be related to dysfunction and abnormalities of the executive network brain areas (Posner and DiGirolamo 1998, Berger and Posner 1999).

2.2.3. Alerting network

This network is involved in establishing a vigilant state and maintaining readiness to react to sensory events. Early attentional studies used long boring tasks to study the maintenance of the alert state. These tasks are important in tonic alertness (Parasuraman 1998). A second task used to study phasic alertness is readiness following a warning signal. Although these tasks are quite different they both seem to involve a strong right lateralized parietal and right frontal cortical network. The alerting network also involves the locus coruleus, which is the source of neurotransmitter norepinephrine in the brain. Norepinephrine is involved in alerting and may work by boosting the signal-to-noise ratio of the brain area where it is released.

It is important to mention that there are connections between these three networks and they interact with each other in their operations. These connections are both at the anatomical, and functional levels.

2.3. Discovering the Mechanisms of Attention

The progress in the field, and the introduction of new neuroscience techniques allow re-examination of some of the issues debated in the study of attention.

2.3.1. Early versus late selection – revised

The networks of attention that were briefly presented in the previous section, represent the *source* of the attentional selection. These are areas of the brain specifically activated during attentional processes. However, the *site of action* of the selection process, is the actual area of the brain where the computation is performed. For example, if the organism is monitoring a screen for detecting a specific target, one would expect that the site of the attentional effect would be an area related to visual processing.

How is this selection implemented? A basic finding from studies using microelectrode recording in alert animals is that attention modulation is observed at the level of individual neurons cells (Colby 1991). Motter (1998) presents an interesting example of attentional modulation of an individual cell's activity. A relevant and an irrelevant stimulus were both presented within the receptive field of a neuron recorded in a secondary visual cortex (V4). Subjects (alert monkeys) where attending the relevant stimulus in order to perform a "match to sample" task. This focal attention suppressed cells response to the irrelevant stimulus. Moreover, cells response to the relevant stimulus was unaffected by the presence or absence of the irrelevant stimulus appearing within the same receptive field. It is important to notice that when subjects were not focusing their attention on the relevant stimulus, the irrelevant stimulus did affect the cell response, since it was appearing within this cell receptive field.

The old *early versus late* question, is now rephrased to how early in the processing of stimuli, is it possible to observe attention modulation of cells activity? Continuing within the example of the visual modality, PET studies show that attention can influence the
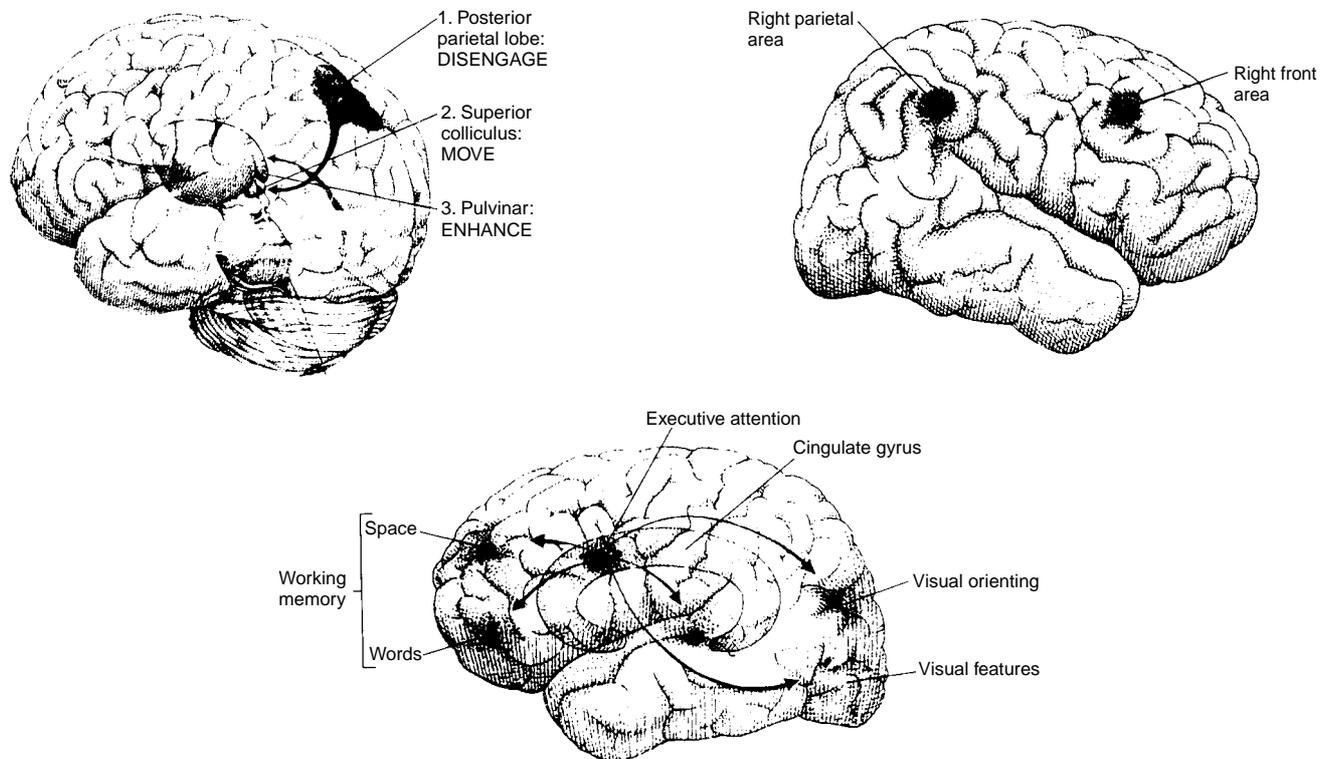
Figure 2. **Neuro-anatomy of the attentional networks. Start at 12 o'clock and moving clockwise: the orienting, executive and alerting networks. Adapted from Posner and Raichle (1996).**

activation of the secondary visual (extrastriate) cortex. Moreover, single cell recording suggests that under some conditions, the attentional modulation of cell activity can be observed as early as in the primary visual (striate) cortex Moran and Desimone (1985), which is the first cortical stage of stimuli analysis in the visual modality. These findings support the idea that attention *can* affect perceptual input at a very early stage of analysis.

2.3.2. Dual tasks – revised

As mentioned in the network description, there seem to be specific areas of the brain that show activity while dealing with monitoring of goal directed behavior. Interestingly, imaging studies show that these same areas are active during dual-task situations. This activation was observed in situations where the tasks did not share input nor response modalities, supporting the hypothesis that the interference between the tasks occurred at the level of a central resource (D'Esposito *et al*. 1995, Passingham 1996). These imaging findings seem to support the idea that during dual-tasks, the bottleneck occurs in the area of the brain related to the pragmatic allocation of attentional resources and the control of goal directed behavior. As mentioned in Section 1.2, subjects show impressive ability to allocate differentially their resources to the tasks and adjust their performance to the situation requirements (i.e. the experimenter's instructions).

According to the discrimination between *source* and *site of action* that was presented in Section 2.3.1, the cingulate gyrus and dorsolateral frontal areas would be the source of the attentional influence, while the site of this influence would be widely dispersed in the areas of the brain where the actual computations for the different tasks are made (i.e. working memory, memory retrieval, motor response planning, etc.) depending on the requirements of the involved tasks.

3. CONCLUSION

The study of attention is constantly making progress. Important advances include aspects of the neuroanatomy, neurochemistry, and temporal parametrics of attention. In this chapter we restricted ourselves to human ability to perform more that one task at the same time and to selectively attend to relevant information. We focused on these issues since they have had a major impact and implication within the ergonomics and human factors literature, especially in relation to questions about mental workload and humans capacity limitations. Although our understanding of the attention system is still far from being complete, it is important that human factor theories take into account the state-of-the-art, most updated knowledge in the cognitive neuroscience field, regarding this basic function of human cognition.

ACKNOWLEDGEMENTS

REFERENCES

BADDELEY, A. and WEISKRANTZ, E. (eds), 1993, *Attention: Selection, Awareness, and Control. A Tribute to Donald Broadbent* (Oxford: Clarendon).

BROADBENT, D.E., 1958, *Perception and Communication* (Oxford: Pergamon).

BUSH, G., WHALEN, P.J., ROSEN, B.R., JENIKE, M.A., MCINEREY, S.C. and RAUCH, S.L., 1998, The counting stroop: an interference task specialized for functional neuroimaging — validation study with functional MRI. *Human Brain Mapping*, 6, 270–82.

COLBY, C.L., 1991, The neuroanatomy and neurophysiology of attention. *Journal of Child Neurology*, 6, 90–118.

CORBETTA, M., 1998, Fronto-parietal cortical networks for directing attention and eye to visual locations: identical, independent or overlapping neural systems. *Proceedings of the National Academy of Sciences, USA*, 95, 831–8.

CORTEEN, R.S. and WOOD, B., 1972, Autonomic responses to shock-associated words in an unattended channel. *Journal of Experimental Psychology*, 94, 308–13.

D'ESPOSITO, M., DETRE, J.A., ALSOP, D.C., SHIN, R.K., ATLAS, S., and GROSSMAN, M., 1995, The neural basis of the central executive system of working memory. *Nature*, 378, 279–81.

DUNCAN, J., 1980, The locus of interference in the perception of simultaneous stimuli. *Psychological Review*, 87, 272–300.

MORAN, J. and DESIMONE, R., 1985, Selective attention gates visual processing in the extrastriate cortex. *Science*, 229, 782–4.

MORAY, N., 1959, Attention in dichotic listening: Affective cues and the influence of instructions. *Quarterly Journal of Experimental Psychology*, 11, 56–60.

MOTTER, B.C., 1998, Neurophysiology of visual attention. In R. Parasuraman (ed.)*The Attentive Brain* (Cambridge, MA: MIT Press), 51–69.

NORMAN, D.A. and BORROW, D., 1975, On data-limited and resource-limited processing. *Journal of Cognitive Psychology*, 7, 44–60.

PARASURAMAN, R. (ed.), 1984, *Varieties of Attention* (Orlando: Academic Press).

PARASURAMAN, R. (ed.), 1998, *The Attentive Brain* (Cambridge, MA: MIT Press).

PARASURAMAN, R., 1998, Brain systems of Vigilance. In R. Parasuraman (ed.) *The Attentive Brain* (Cambridge, MA: MIT Press), 221–56.

PASHLER, H.E., 1998, *The Psychology of Attention* (Cambridge, MA: MIT Press).

PASSINGHAM, R.E., 1996, Attention to action. *Proceedings of the Royal Society of London*, 351, 1473–9.

POSNER, M.I. and DIGIROLAMO, G.J., 1998, Executive attention: conflict, target detection, and cognitive control. In R. Parasuraman (ed.) *The Attentive Brain* (Cambridge, MA: MIT Press), 401–23.

POSNER, M.I. and RAICHLE, M.E., 1996, *Images of Mind*, revd (New York: Scientific American Library).

TREISMAN, A., 1993, The perception of features and objects. In A. Baddeley and L. Weiskrantz (eds) *Attention: Selection, Awareness, and Control. A Tribute to Donald Broadbent* (Oxford: Clarendon), 5–35.

TREISMAN, A. and DAVIES, A., 1973, Dividing attention to ear and eye. In S. Kornblum (ed.) *Attention and Performance IV* (New York: Academic Press), 101–17.

TREISMAN, A. and GELADE, G., 1980, A feature integration theory of attention. *Cognitive Psychology*, 12, 97–136.

TREISMAN, A. and GORMICAN, S., 1988, Feature analysis in early vision: evidence from search asymmetries. *Psychological Review*, 95, 15–48.

WICKENS, C.D., 1984, The structure of attentional resources. In R. Nickerson (ed.) *Attention and Performance VIII* (Hillsdale: Erlbaum).

Body Postures

K. H. E. Kroemer

Industrial And Systems Engineering, Virginia Tech, Blacksburg, VA, and Eri, Inc., A Consulting Company, PO Box 3019, Radford, VA 24143-3019, USA

To standardize anthropometric measurements (see the chapter on Anthropometry), the body is put into defined static postures:

- *Standing*: the instruction is "stand erect; heels together; rears of heels, buttock and shoulders touching a vertical wall; head at erect; look straight ahead; arms hang straight down (or upper arms hang, forearms are horizontal and extended forward); fingers extended."
- *Sitting*: on a plane, horizontal, hard surface adjusted so in height that the thighs are horizontal; "sit with lower legs vertical, feet flat on the floor; trunk and head erect; look straight ahead; arms hang straight down (or upper arms hang, forearms horizontal and extended forward); fingers extended."
- The *head* (including the neck) is held erect (or "upright")

when, *in the front view*, the pupils are aligned horizontally, and, *in the side view*; the ear–eye line is angled ~15° above the horizon (Figure 1).

People do not stand or sit naturally in these standard measurement postures. Thus, the dimensions taken on the body in the standardized positions must be converted to reflect real postures and motions at work, or leisure, which vary greatly. The designer has to estimate the corrections that reflect the anticipated movements.

For more information see Kroemer (1999) and Kroemer *et al.* (2000).

REFERENCES

KROEMER, K.H.E., 1999, Engineering anthropometry. In W. Karwowski and W.S. Marras (eds) *The Occupational Ergonomics Handbook* (Boca Raton: CRC Press), 139–65.

KROEMER, K.H.E., KROEMER, H.B. and KROEMER-ELBERT, K.E., 2000, *Ergonomics: How to Design for Ease and Efficiency* (2nd ed.) (Englewood Cliffs: Prentice-Hall).
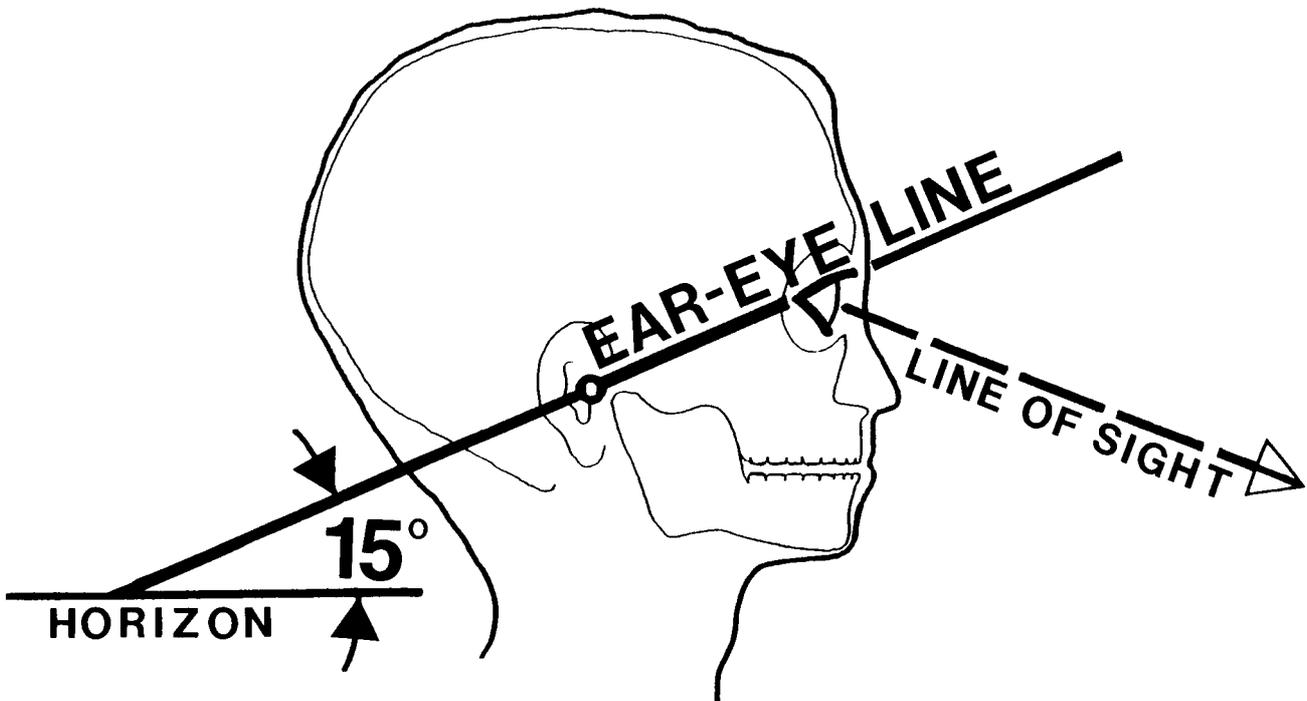
Figure 1. Head posture. The ear–eye line serves as a reference to describe head posture and the line-of-sight angle. The ear–eye line runs through the ear whole and the outside juncture of the eyelids.

Brain and Muscle Signal-based Control

G. R. McMillan

Air Force Research Laboratory, AFRL/HECP, 2255 H Street, Bd. 33, WPAFB, OH 45433, USA

1. INTRODUCTION

The activity of muscles and nerves can often be measured as electrical potentials at the surface of the skin. These potentials can, in turn, be used to control external devices. This chapter focuses on the use of two types of signals: the electromyographic (EMG) signals associated with the contraction of skeletal muscle and the electroencephalographic (EEG) signals associated with brain activity (Farry *et al.* 1996).

Control system designers in many fields have attempted to tap our natural physiological systems to achieve intuitive, non-fatiguing control of external devices. The notion of operating a device simply by thinking about it represents the ultimate in intuitive control. Although current technology limits our ability to achieve such natural control systems, several practical devices have been commercialized and other promising technologies are being evaluated in the research community. For example, prosthetic hands and wrists, controlled by electrical muscle signals, are of significant value to those with lower-arm amputations. Thousands of units have been fitted worldwide (Figure 1). This area represents the most significant current application of human physiological signals for control (McFarland *et al.* 1997).

2. CONTROL USING MUSCLE SIGNALS

2.1. Control Technology Description

The EMG signal measured at the skin surface resembles random noise that is modulated in amplitude by changes in muscle activity. It results from the asynchronous firing of hundreds of groups of muscle fibers. The number of groups and their firing frequency
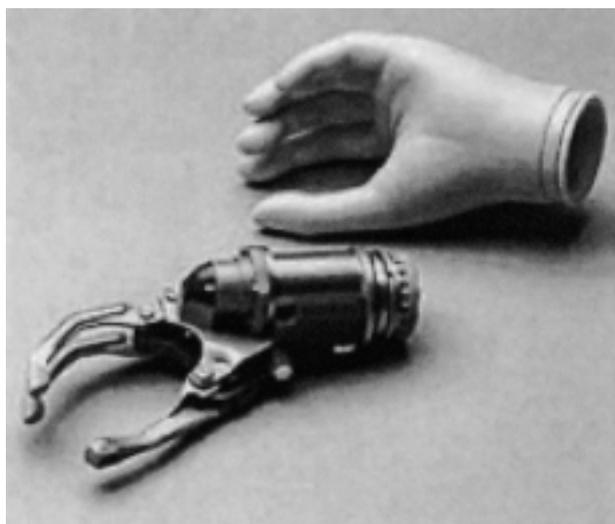
Figure 1. Prosthetic hand system. Photo: courtesy Otto Bock USA, Minneapolis.

controls the amplitude of this signal and the force produced by the muscle contraction. The EMG signals of interest range in amplitude from several hundred microvolts to a few millivolts. The EMG frequency spectrum has a peak at ~60 Hz and significant power from ~20 to 500 Hz. The small size of the EMG, and the fact that its frequency spectrum falls within the range of many common electrical devices, leads to signal acquisition challenges. However, these challenges have been successfully addressed in the laboratory and for numerous applications in the real world.

2.1.1. Signal acquisition

Prosthetic systems typically use stainless steel electrodes that are molded into the prosthesis and contact the skin surface. Because the EMG signal is most often measured at points on hairless skin, gel electrodes also represent a convenient option. Differential recording, with two active electrodes and one reference, is typically employed to minimize interference from external noise sources. Commercial biological signal amplifiers are well suited to the amplification and filtering of the EMG signal.

2.1.2. Signal processing

The raw EMG signal has an average value equal to zero; therefore this parameter cannot be used for control purposes. Mathematically, the mean absolute value of the signal is related to the strength of the muscle contraction, but the precise form of this relationship varies across muscles and individuals. In practice, rectification and low-pass filtering of the EMG signal can be used to estimate the strength of a muscle contraction. The most common prosthetic control algorithms employ on–off control based on the level or rate of change of this measure of EMG activity. If muscle activity at one electrode site exceeds some threshold, the prosthetic hand opens. Above-threshold activity at another site causes the hand to close. Hand movement stops when the EMG at both sites is below threshold. To permit control of functions such as grip force, some on–off algorithms employ "time proportional" techniques, i.e. grip force increases as long as the operator holds the closing signal above threshold. For multifunction prosthetic devices (e.g. simultaneous elbow, wrist and hand control), systems are being developed that use patterns of activity across multiple muscle groups or patterns extracted from the responses of individual muscles to control specific functions. In some cases these systems employ neural networks to recognize natural contraction patterns and produce a similar response in the prosthetic device.

2.1.3. Feedback requirements

EMG-based control provides muscle contraction feedback for operators with intact sensory systems. However, many other feedback channels are absent. Attempts to provide grip force feedback in prosthetic devices have most often employed vibratory or electrical cues. One study showed that pressure cues provided by an inflatable cuff permitted better grip force control than vibratory cues. However, visual cues alone appeared to be sufficient. To some extent this finding reflects limitations in the performance of current prosthetic devices, and it is generally believed that enhanced feedback will be required as the performance of prosthetic devices improves.

2.2. Other Applications and Evaluations

Numerous other applications of EMG signals have been

investigated, in addition to prosthetic devices. Early work by the US Air Force investigated the idea of assisting pilot arm movements during periods of high acceleration flight. An EMG-controlled, powered splint was developed and evaluated. Although the EMG signals provided an effective means of control, the system did not go into production.

Several studies by the National Aeronautics and Space Agency (NASA) have evaluated the feasibility of using EMG signals from an operator's arm to intuitively control a remote robot arm. If feasible, a low-cost EMG system might be used to replace or augment joysticks and exoskeletal instrumentation. Initial studies suggested that raw EMG time histories from normal hand and arm motion were not sufficient for controlling the complex movements of a robot arm. However, more recent work has found that a time–frequency analysis of the EMG patterns from forearm musculature can discriminate several different hand grasp types and thumb motions with a high degree of accuracy.

More recent work sponsored by the US Air Force has shown that subjects can use facial EMG signals, measured with electrodes on the forehead, to control a cursor and track computer-generated targets. For discrete responses, very rapid control inputs were achieved with little operator training. Early studies of an automotive application found that reaction time can be enhanced using an EMG trigger. A substantial reduction in operator reaction time was observed when an electronic braking system, triggered by EMG from the frontalis muscle, was used to augment a normal foot-activated automobile brake.

The EMG signal has additional potential to augment more traditional control methodologies, such as automatic speech recognition. For example, recent work has shown that certain facial EMG activity is highly correlated with speech. Follow-on studies at the University of New Brunswick in Canada demonstrated that a neural network could achieve up to 95% accuracy in recognizing the spoken words "zero" through "nine" based on EMG signals recorded from electrodes mounted in a flight oxygen mask.

3. CONTROL USING BRAIN SIGNALS

3.1. Control Technology Description

EEG recorded from the surface of the scalp represents a summation of the electrical activity of the brain. Although much of the EEG appears to be noise-like, it does contain specific rhythms and patterns that represent the synchronized activity of large groups of neurons. A large body of research has shown that these patterns are meaningful indicators of human sensory processing, cognitive activity and movement control. In addition, EEG patterns can be brought under conscious voluntary control with appropriate training and feedback. Although current EEG-based systems represent fledgling steps toward a "thought-based interface," significant long-term development is required to reach that goal. EEG-based control research is primarily confined to laboratory systems and is based on two general approaches:

- Learned control (self-regulation) of specific EEG characteristics. For example, voluntarily increasing the EEG activity in a specific frequency band might be used to turn a switch on.
- The use of involuntary EEG patterns, normally associated with sensory or motor activity, to produce an appropriate

response in an external device. For example, an operator might imagine moving their right hand to push a button. The computer system would recognize the EEG pattern associated with this movement preparation and operate the right-hand button without further operator action.

3.1.1. Signal acquisition

The most commonly used scalp electrodes are small (~1 cm diameter) gold or silver/silver chloride disks developed for EEG recording. Electrical contact is maintained with a conductive paste and the electrodes are affixed with adhesive rings, elastic caps and bands, or the conductive paste alone. Real-world applications will benefit from convenient dry electrode systems, but these are not yet commercially available. The EEG signals of interest are in the 1–40-Hz frequency range with amplitudes ranging from 1 to 50 mV. Although these signals are very small, appropriate biological signal amplifiers are commercially available. Despite a variety of signal acquisition challenges, EEG signals can be recorded from active operators, such as pilots.

3.1.2. Signal processing

After amplification and filtering to isolate the frequency range of interest, the first step in most systems is analog-to-digital signal conversion. The next step usually involves estimation of the EEG power in specific frequency bands using fast Fourier transforms, digital band-pass filters or some proprietary approach. Personal computer systems of the Pentium class are sufficient to support these computations, although digital signal processing boards are sometimes used. In several systems, the power in a specific frequency band is used as the control signal. For example, high power might move a computer cursor upward, while low power would move it downward. Some systems employ neural networks or other pattern discrimination techniques to determine the relationship between the processed EEG signals and specific body movements, joystick movements or simple utterances. Upon recognition of the required pattern the system performs the appropriate control action.

3.1.3. Feedback requirements

Biofeedback is one of the key technologies that enabled the development of systems based on learned control of EEG patterns. Operator feedback has been implemented in two ways: (1) as an inherent part of the task, e.g. movement of the computer cursor being controlled by EEG, or (2) as a separate display element when movement of the controlled element does not provide timely feedback. Biofeedback is not required in systems that use involuntary EEG responses. However, such feedback can be motivating to operators and informal observations suggest that it may useful in improving the quality and timeliness of the involuntary responses.

3.2. Applications and Evaluations

Several basic computer operations have been demonstrated with EEG-based control systems. Scientists at the Wadsworth Center, New York, have developed a system for computer cursor control that is based on self-regulation of mu (8–12 Hz) or beta (18–25 Hz) rhythms. In their two-axis control system, high mu or beta rhythm amplitude in both brain hemispheres produces upward

motion while low amplitude causes the cursor to move downward. Left and right motion is controlled by the difference in mu or beta rhythm amplitude between the two brain hemispheres. This represents a promising form of control for physically challenged individuals, despite the fact that speed and accuracy are limited. Scientists at Emory University, USA, are investigating cursor control using electrodes implanted in certain motor control areas of the brain. Simple word and message selection has been demonstrated in severely disabled patients.

Several researchers have also demonstrated alphanumeric input systems. Most approaches involve modulating (flickering) the individual elements of a simulated keyboard presented on a computer monitor. The resulting sensory-evoked responses are used to determine which character or word the operator wishes to select. Communication rates of up to 10–12 words per minute have been achieved. Scientists at the University of Tubingen, Germany, and the University of Alabama, USA, have employed the self-regulation of slow cortical potentials for alphanumeric input. By making these potentials negative, operators select one-half of the alphabet, while making them positive selects the other half. This halving process is repeated until the desired character is selected.

Scientists at Graz University, Austria, are developing the technology to "automate" simple control actions normally performed with the hands or feet. Neural networks are employed to recognize EEG patterns that precede specific body movements, such as finger, toe or tongue activity. No operator training is required, but the movements must be repeated 100–200 times for neural network training. After training, movement prediction is possible with only one second of EEG data. They have achieved > 90% accuracy in predicting button pushes with the left or right hand, for example. In addition, the neural networks can be trained with imagined rather than actual movements.

The US Air Force has investigated a variety of EEG-based control systems that involve the detection and use of evoked responses to flickering lights incorporated in a task display. One study addressed the control of a simple flight simulator. The simulator display included flickering lights at each side and presented a random series of commands requiring the operator to roll right or left to specific target angles. Operators controlled the simulator by raising their visual evoked response above a high threshold to roll right and suppressing the response below a low threshold to roll left. Typically, they were able to acquire 70–85% of the roll-angle targets after 5–6 h of training. A related study addressed the operation of a neuromuscular stimulator designed to exercise paralyzed limbs. Raising the evoked response above a high threshold turned the stimulator on and suppressing the response below a low threshold turned it off. This, in turn, caused the operator's knee to extend or flex in response to changes in stimulator current. The most recent Air Force work involves modulating selectable functions on a computer monitor. The

operator selects the desired function simply by looking at it. Over 90% accuracy with 2-s selection times was demonstrated for a two-choice task.

4. FUTURE DEVELOPMENT

With creative interface design, almost any discrete response task can be performed with EMG- or EEG-based systems. In certain cases, response time advantages have been demonstrated using EMG signals to replace the physical movement of a conventional control. The ability to perform other tasks is limited, and generally poorer than what can be achieved with conventional input devices. Simple EMG and EEG responses can be identified rapidly, but the recognition of complex patterns is more time consuming and constrains the speed of human–system interaction. Nevertheless, the pattern recognition approaches offer the greatest potential for discriminating signal from noise and for controlling multiple degree-of-freedom systems.

As discussed by McMillan *et al.* (1997) EMG and EEG signals may be most useful when applied in a manner that combines aspects of system control and operator state monitoring. Referred to as an intelligent controller paradigm, this approach employs an intelligent interpreter that monitors a range of human outputs (eye and head movements, gestures, behavior patterns, EMG and EEG signals). The interpreter combines this information to infer intent, and assists the operator, based upon this inference. While such a system has not yet been demonstrated, it may well represent the path to optimal utilization of human physiological signals for control (Sutter 1992).

REFERENCES

FARRY, K.A., WALKER, I.D. and BARANIUK, R.G., 1996, Myoelectric teleoperation of a complex robotic hand. *IEEE Transactions on Robotics and Automation*, 12, 775–8.

KALCHER, J., FLOTZINGER, D., NEUPER, C., GOLLY, S. and PFURTSCHELLER, G., 1996, Graz brain–computer interface II: towards communication between humans and computers based on online classification of three different EEG patterns. *Medical and Biological Engineering and Computing*, 34, 382–8.

MCFARLAND, D.J., LEFKOWICZ, A.T. and WOLPAW, J.R., 1997, Design and operation of an EEG-based brain–computer interface with digital signal processing technology. *Behavior Research Methods, Instruments and Computers*, 29, 337–45.

MCMILLAN, G.R., EGGLESTON, R.G. and ANDERSON, T.R., 1997, Nonconventional controls. In G. Salvendy (ed.) *Handbook of Human Factors and Ergonomics* (2nd ed) (New York: Wiley), 729–71.

SCOTT, R.N. and PARKER, P.A., 1988, Myoelectric prostheses: state of the art. *Journal of Medical Engineering and Technology*, 12, 143–51.

SUTTER, E.E., 1992, The brain–response interface: communication through visually induced electrical brain responses. *Journal of Microcomputer Applications*, 15, 31–45.

Burnout

W.B. Schaufeli

Department of Psychology, Utrecht University, The Netherlands

1. INTRODUCTION

Burnout is a metaphor that describes a state of mental exhaustion or occupational fatigue. Literally it means to fail, to wear out or to become exhausted by making excessive demands on energy, strength or resources. In the mid-1970s, the colloquial term "burnout" was introduced as a scholarly construct to denote a negative psychological condition mainly characterized by lack of energy, detachment, decreased motivation, distress and a sense of reduced professional efficacy. The New York psychiatrist Herbert Freudenberger and the Californian social psychological researcher Christina Maslach were the first to use "burnout" as a psychological notion. Burnout was particularly observed among those who worked with other people, i.e. human services professionals such as teachers, nurses and social workers, and correctional officers, poverty lawyers and pastors. Soon after its introduction, burnout became a very popular topic, not only in work organisations and in the mass media, but also in academic psychology. By the end of the twentieth century, over 5500 scholarly publications had appeared and the number increases by some 300 each year. Single occupational groups most often studied are teachers (22%), followed by nurses (17%) and social workers (7%). Despite the impressive quantity of the empirical publications, their methodological quality is often questionable: typically, studies on burnout are surveys that almost exclusively rely on self-reports, use non-random samples and one-shot research designs. However, particularly in recent years, the quality of burnout research increased.

2. SYMPTOMS

A wide range of *individual* burnout symptoms has been identified that can grouped into five main categories: (1) affective (e.g. depressed mood, emotional exhaustion, anxiety); (2) cognitive (e.g. sense of failure, forgetfulness, inability to concentrate); (3) physical (e.g. headaches, muscle pains, chronic fatigue); (4) behavioral (e.g. procrastination, hyperactivity, impulsiveness); and (5) motivational (e.g. loss of zeal, disillusionment, demoralization). In addition, *interpersonal* symptoms have been described: (1) affective (e.g. irritability, lessened emotional empathy with recipients, increased anger); (2) cognitive (e.g. cynical and dehumanizing perception of recipients, stereotyping, blaming the victim); (3) behavioral (e.g. social isolation and withdrawal, aggressiveness towards recipients, expressing sick humor); and (4) motivational (e.g. loss of interest in others, discouragement, indifference with respect to recipients). Finally, symptoms in relation to the *organisation* have been observed: (1) affective (e.g. job dissatisfaction); (2) cognitive (e.g. cynicism about work role, distrust in management, peers and supervisors); (3) behavioral (e.g. reduced effectiveness, poor work performance, tardiness); and (4) motivational (e.g. loss of work motivation, low morale, dampening of work initiative). Most symptoms stem from uncontrolled clinical observations or from interview studies rather then from rigorously designed and thoroughly conducted quantitative studies. Despite the over-inclusiveness of this list, it appears that burnout symptoms are typically at three levels (individual, interpersonal, organizational) and that they include motivational symptoms, in addition to affective, cognitive, physical and behavioral symptoms which are also found in job stress. Most important, however, is that virtually all authors agree that exhaustion, energy depletion or chronic work related fatigue is the core symptom of burnout.

3. DEFINITIONS

Traditionally, state and process definitions of burnout are distinguished. The former describes burnout as a negative psychological condition, whereas the latter emphasizes its development. The most popular state definition is by Maslach and Jackson (1986: 1): "Burnout is a syndrome of emotional exhaustion, depersonalization, and reduced personal accomplishment that can occur among individuals who do 'people work' of some kind." This definition lies at the core of the most widely used self-report burnout questionnaire; the Maslach Burnout Inventory (MBI). *Emotional exhaustion* refers to the depletion or draining of emotional resources: professionals feel that they are no longer able to give themselves at a psychological level. *Depersonalization* points to the development of negative, callous and cynical attitudes towards the recipients of one's services. Contrary to its use in psychiatry, depersonalization does not refer to the extreme alienation from the self, but to an impersonal or dehumanizing perception of others. *Lack of personal accomplishment* is the tendency to evaluate one's work with recipients negatively, which is accompanied by feelings of insufficiency and poor professional self-esteem. Initially, Maslach and Jackson (1986) claimed that burnout exclusively occurred among those who dealt with recipients, like students, clients and patients. However, recently, the burnout concept was expanded beyond the human services and was redefined as a crisis in one's relationship with work, not necessarily as a crisis in one's relationship with people at work. Accordingly, it includes three more general aspects: exhaustion, cynicism and professional efficacy.

Cherniss (1980: 5) proposed an influential process definition: 'Burnout refers to a process in which the professional's attitudes and behavior change in negative ways in response to job strain.' More specifically, he describes burnout as a three-stage process: (1) an imbalance between resources and demands (stress); (2) the development of emotional tension, fatigue, and exhaustion (strain); and (3) changes in attitudes and behaviors, such as a tendency to treat clients in a detached and mechanical fashion (defensive coping). Consequently, for Cherniss, burnout is a wrong way of coping with chronic job-related strain.

Despite obvious inconsistencies among burnout definitions Schaufeli and Enzmann (1998: 36) formulated a synthetic definition:

> Burnout is a persistent, negative, work-related state of mind in "normal" individuals that is primarily characterized by exhaustion, which is accompanied by distress, a sense of reduced effectiveness, decreased motivation, and the development of dysfunctional attitudes and behaviors at work. This psychological condition develops gradually but may remain unnoticed for a long time for the individual involved. It results from a misfit between intentions and reality at the job. Often burnout is self-perpetuating because

of inadequate coping strategies that are associated with the syndrome.

4. BURNOUT AND RELATED PSYCHOLOGICAL CONDITIONS

From the outset, the distinction between burnout and related psychological phenomena (e.g. job stress and depression) has been debated. Burnout is as a *particular kind* of prolonged job stress. An individual experiences job stress when job demands tax or exceed adaptive resources. That is, job stress refers to a temporary adaptation process accompanied by mental and physical symptoms, whereas burnout refers to a breakdown in adaptation that manifests itself in profound and chronic malfunctioning at work. Furthermore, the stress responses includes physical, psychological (affective, cognitive, motivational) and behavioral symptoms, whereas burnout is defined as a multidimensional syndrome that includes — in addition to such symptoms — the development of negative, dysfunctional attitudes and behaviors at work. Finally, everybody can experience stress, while burnout can only be experienced by those who entered their careers enthusiastically with high goals and expectations and who restlessly pursued success in their jobs. Empirically speaking, burnout — as measured with the MBI — is related with job stress, but can nevertheless be differentiated from it.

Despite clear similarities such as dysphoric mood, fatigue, and loss of motivation burnout and depressive illness (i.e. mood disorder) can be distinguished. First, the former is usually accompanied by guilt (leading to suicidal ideation), whereas burnout generally occurs in the context of anger and resentment. Moreover, burnout symptoms, at least initially, tend to be job-related and situation-specific rather than pervasive. In contrast, depressive illness is characterized by a generalization of the person's symptoms across all situations and spheres of life. Finally, typical symptoms of depression such as significant weight gain or weight loss and mood swings during the day are virtually absent in burnout. Not only the clinical picture of burnout and depressive illness differs, but also empirical research corroborates the discriminant validity of the two most used self-report indicators of both constructs (i.e. the MBI and Beck's Depression Inventory).

5. ASSESSMENT

Burnout is almost exclusively assessed by means of self-report questionnaires; no standardized observation or interview procedures exist. The MBI (Maslach *et al.* 1986) is almost universally used: > 90% of the empirical studies on burnout include this instrument. The second most widely used self-report scale is the Burnout Measure (BM), which exclusively assesses the employees' level of exhaustion. Three versions of the MBI exist: the Human Services Survey (HSS), the Educator's Survey (ES) and the General Survey (GS). All versions include three scales that correspond with the dimensions of the state definition of burnout. Instead of one total composite score, the MBI yields three scale scores. The MBI has been thoroughly investigated psychometrically and appears to be a reliable and valid indicator of burnout.

"Burnout" is not a formal, officially known diagnostic label. It comes most close to the diagnostic criteria of an Unspecified Adjustment Disorder (Diagnostic and Statistical Manual of Mental Disorders — DSM-IVr) or of Neurasthenia (International Classification of Mental and Behavioral Disorders — ICD10) that is work related. Principally, the MBI can be used to assess the individual's level of clinical burnout. However, the current classification of burnout levels is based on arbitrary statistical norms that have not been clinically validated in a sample that consists of burned out patients. Therefore, the MBI categorization of burnout into "high", "average" and "low" levels is not (yet) suited for individual assessment.

6. PREVALENCE

Since no clinically validated cut-off score is available to distinguish "burnout cases" from "non-cases", the absolute prevalence of burnout among the working population cannot be determined reliably. However, the relative prevalence in particular occupational groups in the USA shows the following pattern: levels of emotional exhaustion are clearly highest among teachers; intermediate levels are found in the social services and in medicine, whereas workers in mental health services experience the lowest levels. For depersonalization the picture is slightly different: social workers and teachers report the highest levels, whereas levels in the mental health services are lowest. Physicians and police officers exhibit particularly high levels of depersonalization, which might reflect their occupational socialization characterized by objectiveness and distance. Finally, reduced personal accomplishment is especially found in the social services and among nurses, police officers, and probation and correction officers. Not surprisingly, the most highly trained professionals (i.e. physicians and psychologists) experience the strongest sense of accomplishment in their jobs. There is some evidence that these specific occupational burnout profiles are consistent across countries.

7. CORRELATES OF BURNOUT

Tables 1 and 2 present an overview of the correlates of burnout. Since the vast majority of studies used a one-shot design, only very few cause–effect relationships could be established so far.

As can be seen in table 1, burnout is observed more often among younger employees than among those aged > 30–40 years. This is in line with the observation that burnout is negatively related to work experience. On balance, women tend to score slightly higher on emotional exhaustion, whereas males score significantly higher on depersonalization. Most likely this reflects differences in sex roles, i.e. women are more emotionally responsive, whereas men hold more instrumental attitudes. Singles seem to be more prone to burnout than those who are married (or divorced) persons. Persons with high burnout levels are characterized by: (1) low involvement in daily activities, poor control over events, and little openness to change ("hardiness"); (2) attributing events to powerful others or to change ("external locus of control"); (3) avoiding problems instead of tackling them ("avoiding coping style"); (4) low self-esteem; (5) emotional instability ("neuroticism"); and (6) low intensity of interpersonal interaction ("extroversion"). Furthermore, burnout is positively related to high or unrealistic expectations about one's job and to various job characteristics. Generally speaking relationships with job characteristics are stronger than with personality characteristics.

It appears from table 2 that burnout is particularly associated

Table 1: Correlates of burnout: possible causes*

biographic characteristics			
age	−	−	
gender			+
work experience		−	
marital status			−
personality characteristics			
hardiness	−	−	−
external control orientation		+	+
avoiding coping style		+	+
self-esteem		−	−
neuroticism	+	+	+
extroversion			−
job related attitudes			
high (unrealistic) expectations			+
job characteristics			
workload	+	+	+
time pressure	+	+	+
role conflict and ambiguity		+	+
direct client contact		+	+
social support from colleagues or superiors		−	−
lack of feedback		+	+
participation in decision making		−	−
autonomy			−

Notes: * Adapted from Schaufeli & Enzmann, 1998; p. 75; The number of minus or plus signs denotes the strength and the direction of the relationship based on: (1) the number of studies involved; (2) their methodological quality; (3) the consistency of results across studies.

Table 2: Correlates of burnout: possible consequences*

individual level			
depression	+	+	+
psychosomatic complaints	+	+	+
health problems			+
substance use			+
spillover to private life			+
work orientation and attitudes			
job satisfaction	–	–	–
organisational commitment		–	–
intention to quit		+	+
organisational level			
absenteeism and sick-leave			+
job turnover			+
performance and quality of services			–

Notes: * Adapted from Schaufeli & Enzmann, 1998; p. 86. The number of minus or plus signs denotes the strength and the direction of the relationship based on: (1) the number of studies involved; (2) their methodological quality; (3) the consistency of results across studies.

with depression, psychosomatic complaints and job dissatisfaction, whereas it is less strongly related with subjective health problems, substance use, spill-over, poor organizational commitment, intention to quit, absenteeism, job turnover and poor job performance.

As far as the three dimensions of the MBI are concerned, emotional exhaustion is strongest related to possible causes, and consequences of burnout. This applies especially to job characteristics, neuroticism, depression and psychosomatic symptoms. Generally, correlations with depersonalization are weaker. On balance, personal accomplishment is least strongly related to potential correlates of burnout with the exception of some personality characteristics.

8. THEORETICAL EXPLANATIONS

There is no general psychological theory to explain burnout. Rather, three different sets of approaches can be distinguished that are supplementary instead of mutually exclusive (Schaufeli *et al.* 1993, Schaufeli and Enzmann 1998). First, *individual approaches* emphasize the role of factors and processes within the person. For instance, burnout can be seen as the result of a pattern of wrong (i.e. high or unrealistic) expectations, or as a failure to retain one's idealized self-image, or as a narcissistic disorder, or as a failed quest for existential meaning, or as a disturbed action pattern. Second, *interpersonal approaches* focus on demanding relations with others at work (i.e. recipients and/ or colleagues). For instance, from this perspective burnout is considered to be the result of emotional overload due to the demanding interactions with difficult recipients, or of a lack of reciprocity between investments put into recipients and the outcomes received, or of the dissonance between the displayed and the experienced emotions, or of emotional contagion (i.e. being 'infected' by others with burnout). Third, *organizational approaches* stress the importance of the organizational context.

For instance, from this perspective burnout is seen as the result of a mismatch between person and job, or as 'reality shock' (the clash of personal ideals with harsh organizational reality). Generally speaking, most theoretical approaches are rather speculative and have no firm empirical bases, perhaps with the exception of some interpersonal explanations.

9. INTERVENTIONS

Basically, interventions to reduce burnout may be focus on: (1) the individual; (2) the individual/organisation interface (i.e. the interplay between individual and organisation); and (3) the organisation. Although numerous interventions have been proposed (Schaufeli and Enzmann 1998), few are specific for burnout. That is, most interventions are general approaches to reduce stress at the workplace. Examples of *individual level* interventions are: didactic stress management (i.e. providing information about burnout to increase awareness and improve self-care), promoting a healthy life style (i.e. physical exercise, weight control, smoking cessation), relaxation, and cognitive-behavioral techniques (e.g. redirection the individual's irrational thinking). Examples at the *individual/organisation interface level* include: time management, interpersonal skills training (e.g. assertiveness), balancing work and private life, coaching and consultation by peers and supervisors, career planning, and specialized counseling and psychotherapy. Finally, interventions at the *organizational level* include: improving the job content by job redesign, management development, introducing corporate fitness and wellness programs, organizational development, and the institutionalization of occupational health and safety services. Many interventions at this level are mere 'Band-Aids' primarily designed for increasing productivity, improving quality or reducing costs. Empirical studies have shown that particularly cognitive and behavioral approaches such as relaxation training, time-management, didactic stress management and cognitive restructuring are effective in reducing levels of exhaustion. Effects on depersonalization and reduced personal accomplishment are far less strong or even absent.

Some of these interventions — most notably at the individual or interface level — are combined in so-called *burnout workshops* in which small groups of workers participate. Typically, such preventive workshops rest on two pillars: increasing the participants' awareness of work-related problems and augmenting their coping resources by cognitive and behavioral skills training and by establishing support networks. More specifically, such workshops include self-assessment, didactic stress-management, relaxation, cognitive and behavioral techniques, time-management, peer support, and the promotion of a more realistic image of the job. Taking together the few studies on the effectiveness of such multi-faceted workshops, it seems that they are effective in reducing levels of emotional exhaustion, even across relatively long periods of time (up to a year). Other burnout dimensions are usually not affected, though. The fact that depersonalization and reduced personal accomplishment do not change is not very surprising because most techniques focus on reducing arousal and not on changing attitudes (depersonalization) or on enhancing specific professional skills or resources (personal accomplishment).

REFERENCES

Cherniss, C., 1980, *Professional Burnout in the Human Service Organizations* (New York: Praeger).

Maslach, C. and Jackson, S.E., 1986, *Maslach Burnout Inventory: Manual* (2nd ed.) (Palo Alto: Consulting Psychologists Press).

Maslach, C., Jackson, S.E. and Leiter, M., 1996, *Maslach Burnout Inventory: Manual* (3rd ed.) (Palo Alto: Consulting Psychologists Press).

Schaufeli, W.B., Maslach, C. and Marek, T. (eds), 1993, *Professional Burnout: Recent Developments in Theory and Research* (Washington, DC: Taylor & Francis).

Schaufeli, W.B. and Enzmann, D., 1998, *The Burnout Companion to Research and Practice: A Critical Analysis* (London: Taylor & Francis).

Cognitive Modeling in Human–Computer Interaction

W. D. Gray and E. M. Altmann

Human Factors and Applied Cognition, Department of Psychology, George Mason University, Fairfax, VA 22030, USA

"There is nothing so useful as a good theory" (Lewin 1951).
"Nothing drives basic science better than a good applied problem" (Newell and Card 1985).

1. INTRODUCTION

The quotations from Lewin and from Newell and Card capture what motivates those who apply cognitive modeling to human–computer interaction (HCI). Cognitive modeling springs from cognitive science. It is both a research tool for theory building and an engineering tool for applying theory. To the extent that the theories are sound and powerful, cognitive modeling can aid HCI in the design and evaluation of interface alternatives. To the extent that the problems posed by HCI are difficult to model or cannot be modeled, HCI has served to pinpoint gaps or inconsistencies in cognitive theory. In common with design, science is an iterative process. The symbiotic relationship between modeling and HCI furthers the scientific enterprise of cognitive science and the engineering enterprise of human factors.

Cognitive modeling is a form of task analysis and, as such, is congenial to many areas and aspects of human factors. However, the control provided by the computer environment, in which most dimensions of behavior can be easily and accurately measured, has made HCI the modeler's primary target. As modeling techniques become more powerful and as computers become more ubiquitous, cognitive modeling will spread into other areas of human factors.

We begin this article by discussing three cognitive models of HCI tasks, focusing on what the models tell us about the tasks rather than on the details of the models themselves. We next examine how these models, as well as cognitive models in general, integrate constraints from the cognitive system, from the artifact that the operator uses to do the task, and from the task itself. We then explore what sets cognitive modeling apart from other types of cognitive task analysis and examine dimensions on which cognitive models differ. We conclude with a brief summary.

2. THREE EXAMPLES OF COGNITIVE MODELING APPLIED TO HCI

The three examples span the gamut of how models are used in HCI. We discuss them in the order of most applied to most theoretical. However, it would be a mistake to think of these as application versus research as each has contributed strongly to theory and each has clear applications to HCI issues.

2.1 Project Ernestine: CPM–GOMS

In the world of the telephone company, time is literally money. In the late 1980s, NYNEX calculated that if the length of each operator-assisted call decreased by 1 sec. the company's operating costs would be reduced by $3 million per year. Potential savings
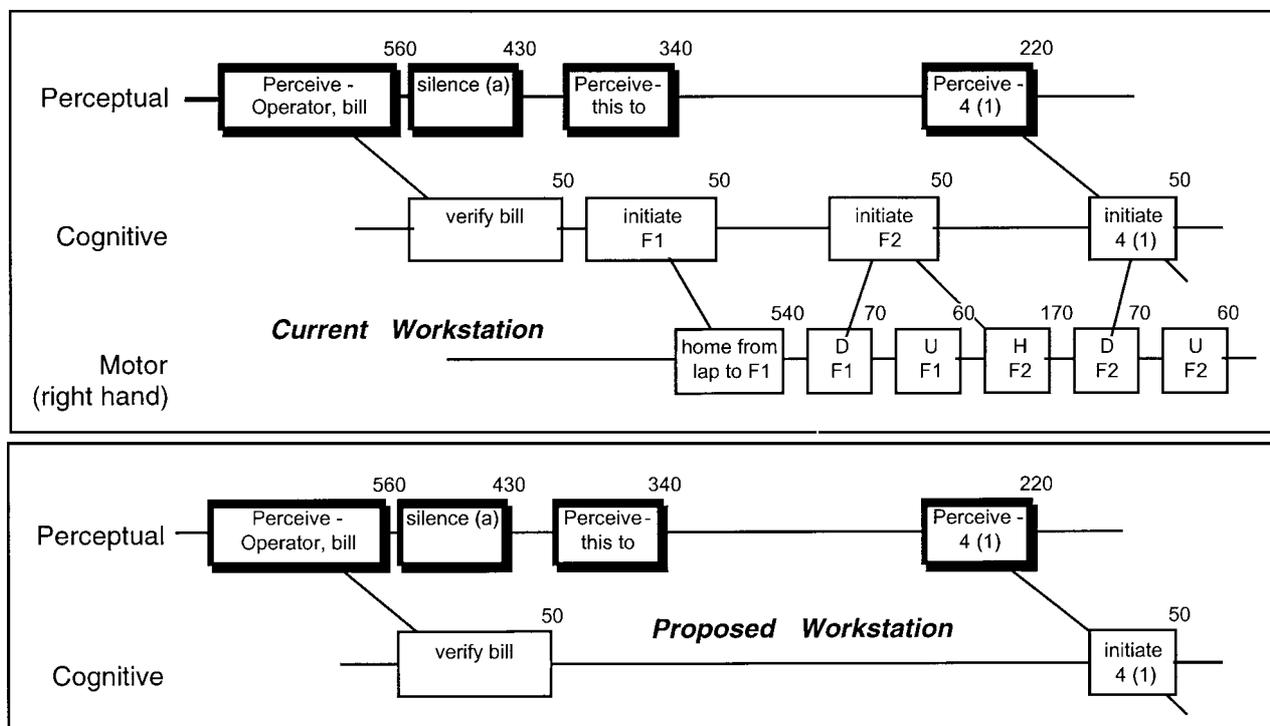
Figure 1. Section of CPM-GOMS analysis for an operator-assisted call. The proposed workstation (bottom) has two fewer keystrokes totaling 6 motor and 2 cognitive steps. However, deleting these steps did not alter the critical path (shown in bold).

on this scale provided an incentive to shave seconds from the time that toll and assistance operators (TAOs) spent on operator-assisted calls.

A major telecommunications equipment manufacturer promised to do just that. For an equipment investment of $60 to $80 million, the old TAO workstations could be replaced by new, ergonomically engineered workstations. The manufacturer's back-of-the-envelope style calculations predicted that the new workstations would shave about 4 sec. from the average call for an estimated savings of $12 million annually.

Project Ernestine involved a combination of cognitive modeling and field trial to compare the new workstations with the old (Gray, John, and Atwood 1993; Gray, John, Stuart, Lawrence, and Atwood 1995). The cognitive models created in Project Ernestine used the GOMS task analysis technique developed by Card, Moran, and Newell (1983). GOMS analyzes a task in terms of Goals, simple Operators used by the person performing the task, and sequences of operators that form Methods for accomplishing a goal. If alternative methods exist for accomplishing a goal, then a Selection rule is required to choose among them. GOMS is best suited to the analysis of routine, skilled performance, as opposed to problem-solving. The power of GOMS derives in part from the fine-grain level of detail at which it specifies the operators involved in such performance. (For a fuller exposition on GOMS, see John and Kieras, 1996a, 1996b).

Project Ernestine employed a GOMS variant, CPM-GOMS, to analyze the TAO's task. CPM-GOMS specifies the parallelism and timing of elementary cognitive, perceptual, and motor operators, using a schedule chart format that enables use of the critical path method to analyze dependencies between these operators.

Contrary to expectations, the cognitive models predicted that the new workstations would add about 1 sec. to the average call. Rather than reducing costs as predicted by the manufacturer, this increased time would result in $3 million in additional operating costs. This prediction was borne out empirically by a four-month field study using live telephone traffic. A sample of the CPM-GOMS model for the beginning part of one call type is shown in Figure 1.

Beyond its prediction, CPM-GOMS was able to provide explanation. The manufacturer had shown that the proposed workstation reduced the number of keystrokes required to process a typical call and from this inferred that the new workstation would be faster. However, their analysis ignored the context of the call, namely the interaction of customer, workstation, and TAO. CPM-GOMS captured this context in the form of a critical path of cognitive, perceptual, and motor actions required for a typical call. By filling in the missing context, CPM-GOMS showed that the proposed workstation *added* more steps to the critical path than it eliminated. This qualitative explanation made the model's prediction credible to telephone company executives.

2.2 Postcompletion Error

An adequate theory of error "is one that enables us to forecast both the conditions under which an error will occur, and the particular form that it will take" (Reason 1990, p. 4). Such a theory was developed by Byrne and Bovair (1997) for a phenomenon that they named *postcompletion error*.

The tasks that people want to accomplish are usually distinct from the devices used to accomplish them. For example, the task might be to withdraw cash from a bank account and the device might be an automated teller machine (ATM). From the perspective that task and device are distinct, any action that the device (the ATM) requires us to perform after we complete our task (withdrawing cash) is a postcompletion action. An omitted postcompletion action is thus a postcompletion error. Postcompletion errors include leaving the card in the ATM after taking the money; leaving the originals on the photocopier after taking the copies; and forgetting to set the video cassette recorder (VCR) to *record* after programming it to videotape a show.

The striking characteristic of postcompletion errors is that, although they occur, they do not occur often. Most people, most of the time, take both the money and the card from the ATM (else, we suspect many fewer of us would use ATMs). What, if anything, predicts the occurrence of a postcompletion error?

Byrne and Bovair's postcompletion error model is based on the notion of *activation* of memory elements. Activation is a hypothetical construct that quantifies the strength or salience of information stored in memory. The postcompletion error model was constructed using CAPS, a programmable model of the human cognitive architecture (Just, Carpenter, and Keller 1996). CAPS assumes that a memory element is accessible only if it has enough activation. It also assumes that total activation is limited. Activation flows from one memory element to another if the two are related and if one is the focus of attention. This spreading activation accounts for standard psychological effects like semantic priming, in which, for example, focusing on the notion of "doctor" might spread activation to related concepts like "nurse."

In Byrne and Bovair's error model, as long as the focus is on a task goal like getting money, related device actions like *take the card* continue to receive activation. However, when a task goal is accomplished, attention shifts away from it. When this shift occurs, the device actions associated with the task begin to lose activation. This is fine for actions like *take the money*, which are necessarily complete, but problematic for postcompletion actions like *take the card*. If these postcompletion actions lose enough activation, they will simply be forgotten.

Beyond its explanation, the postcompletion error model offered a prediction. Like most memory theories, CAPS assumes that unused memory elements decay and over time; that is, their activation decreases. Because activation in CAPS is a common resource, decay of one memory element makes more activation available for other elements. Commensurately, Byrne and Bovair found fewer postcompletion errors in a condition that included a prolonged tracking task. Apparently the tracking task, which involved no memory load itself, allowed completed actions of the main task to decay. Postcompletion actions continued to receive activation because the task goal was not yet accomplished. In addition, they received the activation lost by the actions that decayed. This additional activation reduced postcompletion error.

As an example of applied theory, the postcompletion error model is important for several reasons. First, its explanations and predictions flow from existing cognitive theory, not from ad hoc assumptions made by the analysts. The model functioned primarily as a means of instantiating a theory on a particular problem. Second, the prediction comes from the model not the modeler. Any analyst could run the postcompletion error model with the same outcome. The debate over this outcome is then

limited to and focused by the representational assumptions and parameter settings reified in a running computer program.

2.3 Information Access

Information in the world is useful only if we can find it when we need it. For example, an illustration in a book is helpful only if we know it exists, if we recall its existence when it is needed, and if we can find it. This view of information adds a cognitive dimension to research into information access (e.g, the HCI sub-areas of information retrieval and interface design). How do we recall the existence of the helpful illustration? What was stored in memory about the illustration, and exactly what is being recalled? What were the cues that prompted the recollection? From the cognitive perspective, the process of information access is complex. However, with a better understanding of the role of memory, we can engineer memory aids that support this process.

Altmann and John (in press) studied the behavior of a programmer making changes to code that had been written over a series of years by a team of which the programmer was a member. Verbal and action protocols (keypresses and scrolling) were recorded throughout an 80-min. session. During this session, the programmer would trace the program for several steps, stop it, interrogate the current value of relevant variables, and so on. Over the course of the session 2,482 lines of code were generated and displayed on the programmer's screen. On 26 occasions, she scrolled back to view information that had appeared earlier but had scrolled off the screen.

Of interest was the role of memory in these scrolling episodes. The volume of potential scrolling *targets* was huge and the programmer's need for any particular target was small. However, the protocol data revealed that scrolling was purposeful rather than random, implying a specific memory triggered by a specific cue. These constraints meant that the programmer's memory-encoding strategy must have been both sweeping in its coverage of potential targets and economical in terms of cognitive effort.

Altmann and John developed a computational cognitive model of *episodic indexing* that simulated the programmer's behavior. The model was developed using Soar, which, like CAPS, is a cognitive theory with a computational implementation (Newell 1990). Based on the *chunking* theory of learning, Soar encodes sweeping amounts of information economically in memory, but retrieval of this information depends on having the right cue. When the episodic-indexing model attends to a displayed item (e.g. a program variable), Soar creates an episodic chunk in memory that maps semantic information about the item to episodic information indicating that the item was attended. A second encounter with the item triggers recall of the episodic chunk, which in turn triggers an inference that the item exists in the environment. Based on this inference, the model decides whether or not to pursue the target by scrolling to it.

The episodic indexing model suggests that memory depends on attention, not intent. That is, episodic chunks are stored in memory as a by-product of attending to an object, with no need for any specific intent to revisit that object later. The implication is that people store vast amounts of information about their environment that they would recall given the right cues. This, in turn, suggests that activities like browsing are potentially much better investments than we might have thought. The key to unlocking this potential is to analyze the semantic structure of
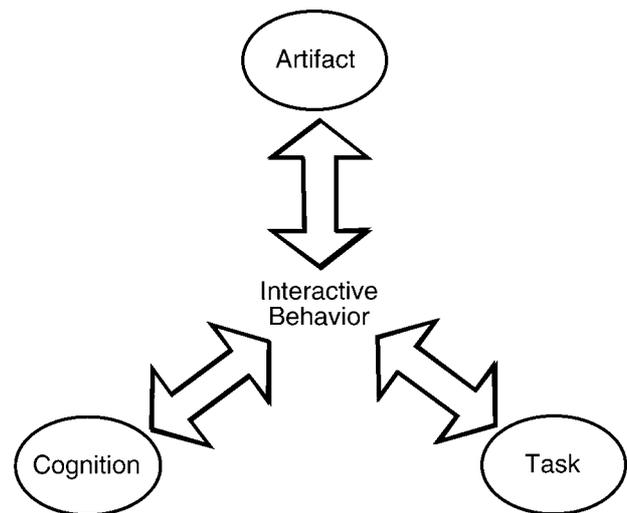
Figure 2. The cognition–artifact–task triad. The behavior of an operator using an artifact to perform a task arises from the limits, mutual constraints, and interactions between and among each member of the cognition–artifact–task triad.

the knowledge being browsed and to ask how artifacts might help produce good cues later when the browsed information would be relevant.

3. THE COGNITION–ARTIFACT–TASK TRIAD

Almost everything we do requires using some sort of artifact to accomplish some sort of task. As Figure 2 illustrates, the interactive behavior for any given artifact–task combination arises from the limits, mutual constraints, and interactions between and among each member of the cognition–artifact–task triad. Cognitive modeling requires that each of these three factors be incorporated into each model.

Traditional methodologies generally consider cognition, artifact, and task *pairwise* rather than altogether. For example, psychological research typically seeks experimental control by using simple tasks that require little external support, thereby focussing on cognition and task but minimizing the role of artifact. Industrial human-factors research often takes the artifact itself to be the task, largely ignoring the artifact's purpose. For example, the proposed TAO workstation had an ergonomically designed keyboard and display but ignored the TAO's task of interacting with the customer to complete a call. Finally, engineering and computer science focus on developing artifacts, often in response to tasks, but generally not in response to cognitive concerns. The price of ignoring any one of cognition, artifact, and task is that the resulting interactive behavior may be effortful, error-prone, or even impossible.

In contrast, cognitive modeling as a methodology is bound to consider cognition, artifact, and task as inter-related components. The primary measure of cognition is behavior, so analysis of cognition always occurs in the context of a task. Moreover, analyzing knowledge in enough detail to represent it in a model requires attention to where this knowledge resides — in the head or in artifacts in the world — and how its transmission between head and world is constrained by human perceptual/

motor capabilities. Indeed, computational theories of cognition are now committed to realistic interaction with realistic artifacts (see, for example, Anderson, Matessa, and Lebiére, 1997; Howes and Young, 1997; Kieras and Meyer, 1997). Thus, given that human factors must consider cognition, artifact, and task together, cognitive modeling is an appropriate methodology.

4. COGNITIVE MODELING VERSUS COGNITIVE TASK ANALYSIS

Cognitive task analysis, broadly defined, specifies the cognitive steps (at some grain size) required to perform a task using an artifact. Cognitive modeling goes beyond cognitive task analysis per se in that each step is grounded in cognitive theory. In terms of the triad of Figure 2, this theory fills in the details of the cognition component at a level appropriate to the task and the artifact.

In Project Ernestine, for example, the manufacturer's predictions about the proposed workstation came from a cognitive task analysis. However, this analysis specified the cognitive steps involved in using the workstation as the manufacturer saw them. The CPM-GOMS models, in contrast, took into account theoretical constraints on cognitive parallelism and made predictions that were dramatically more accurate. In the other two models the influence of theory is strong as well. Most any cognitive analysis would identify memory failure as the cause of postcompletion error. However, the model based on CAPS went further, linking decay of completed goals to improved memory for pending goals. Similarly, memory is clearly a factor in information access, but the model based on Soar detailed the underlying memory processes to highlight the potential of browsing and the importance of effective cues.

Our discussions of cognitive theory have focused on GOMS (John and Kieras 1996a, 1996b), ACT-R (Anderson and Lebiére 1998), CAPS (Just *et al.* 1996), EPIC (Kieras and Meyer 1997), and Soar (Newell 1990). These are broad and integrated theories that deal with cognitive control (GOMS, ACT-R, and Soar), learning and memory (ACT-R, CAPS, and Soar), and perception and action (ACT-R and EPIC). There is also the class of connectionist or neural network models that offers learning and memory functions that have been highly successful in accounting for lower-level cognitive phenomena like visual attention (e.g. Mozer and Sitton 1998). In sum, a broad range of theory is now available to elaborate the steps of a cognitive task analysis and thus produce cognitively plausible models of interactive behavior.

5. DIMENSIONS OF COGNITIVE MODELS

The models we have described are points in a much larger space. In general, a model simply represents or stands for a natural system that for some reason we are unable to study directly. Many psychological models, for example, are mathematical functions, like memory-retention functions or regression equations. These make accurate, quantitative predictions but are opaque qualitatively in that they provide no analysis of what lies behind the behavior they describe.

We have focused on models that characterize the cognitive processes involved in interactive behavior. Process models can make quantitative predictions, like those of the TAO model, but go beyond such predictions to specify with considerable precision the cognitive steps involved in the behavior being analyzed. To give a sense of the space of possibilities, we compare and contrast process models on two dimensions: generative vs. descriptive, and generality vs. realism.

5.1 Generative versus Descriptive

Two of our sample models are generative and one is descriptive. The postcompletion error and episodic indexing models actually generate behavior simulating that of human subjects. Generative models are implemented as executable computer programs (hence are often referred to as computational cognitive models) that take the same inputs and generate the same outputs that people do. The TAO model, in contrast, simply describes sequences of actions rather than actually carrying them out.

Generative models have several advantages. These include, first, proof of sufficiency. Running the model proves that the mechanisms and knowledge it represents are sufficient to generate the target behavior. Given sufficiency, evaluation can shift, for example, to whether the model's knowledge and mechanisms are cognitively plausible. A second benefit is the ability to inspect intermediate states. To the extent that a model is cognitively plausible, its internal states represent snapshots of what a human operator may be thinking. A third benefit is reduced opportunity for human (analyst) error. Generative models run on a computer, whereas descriptive models must be hand-simulated, increasing the chance of error.

5.2 Generality versus Realism

Models vary in their concern with *generality* versus *realism*. Generality is the extent to which a model offers theoretical implications that extend beyond the model's domain. Realism, in contrast, is the extent to which the modeled behavior corresponds to the actual interactive behavior of a particular operator performing a given task.

Project Ernestine showed high realism in that each model accounted for the behavior for an entire unit task; that is, one phone call of a particular call category for a particular workstation. These models were not general in that it would be difficult to apply them to any task other than the one modeled. For example, they could not be applied to model ATM performance or VCR programming. Indeed, the existing models apply only to a particular set of call categories. If another call category were to be modeled, another model would have to be built.

In contrast, the models of postcompletion error and episodic indexing lack realism in that their accounts of behavior are incomplete. Byrne and Bovair's model cannot perform the entire task, and Altmann and John's model cannot debug the code. However, the implications of these models extend far beyond the tasks in which they are based. Situations involving postcompletion actions are susceptible to postcompletion error. If postcompletion actions cannot be designed out of an interface then special safeguards against postcompletion error must be designed in. Likewise, episodic indexing suggests that human cognition reliably encodes a little information about whatever it attends to. With the right cue, this information can be retrieved. These hypotheses bear on any artifact–task combination in which memory is an issue.

6. SUMMARY

The space of cognitive process models, even within the space of

models in general, is quite large (see Gray, Young, and Kirschenbaum 1997 for an alternative cross-section). It used to be that developing process models required access to specialized hardware and software that was available only at certain locations. Fortunately, the technology of programmable cognitive theories has improved to the point where computational models can be run and inspected over the Web (e.g. most of the models discussed in Anderson and Lebiére 1998 are available on the web). Access to such models enables the analyst to study working copies of validated models and potentially to build on, rather than duplicate, the work of others.

Cognitive modeling is the application of cognitive theory to applied problems. Those problems serve to drive the development of cognitive theory. Some applications of cognitive modeling are relatively pure application with little return to theory. Of the three models we considered, the model of the TAO in Project Ernestine (Gray *et al.* 1993) best fits this characterization.

In contrast, the episodic indexing model (Altmann and John, in press) was driven by an applied question — how a programmer works on her system — but produced no new tool or concrete evaluation. Instead, it proposed a theory of how people maintain effective access to large amounts of information. This theory suggests a class of design proposals in which the artifact plays the role of memory aid.

In the middle, the model of postcompletion error (Byrne and Bovair 1997) used existing theory to predict when an applied problem (error) was most likely to occur. On this middle ground, where theory meets problem, is where cognitive modeling will have its greatest effect — first on HCI, then on human factors.

ACKNOWLEDGMENTS

REFERENCES

ALTMANN, E.M. and JOHN, B.E., in press, Modeling episodic indexing of external information. *Cognitive Science*.

ANDERSON, J.R. and LEBIÉRE, C. (eds), 1998. *Atomic components of thought* (Hillsdale, NJ: Erlbaum).

ANDERSON, J.R., MATESSA, M., and LEBIÉRE, C.,1997, ACT-R: A theory of higher-level cognition and its relation to visual attention. *Human-Computer Interaction* 12(4), 439–62.

BYRNE, M. D., and BOVAIR, S., 1997, A working memory model of a common procedural error. *Cognitive Science* 21(1), 31–61.

CARD, S. K., MORAN, T. P., and NEWELL, A., 1983, *The psychology of human-computer interaction*. Hillsdale, (NJ: Lawrence Erlbaum).

GRAY, W.D., JOHN, B.E., and ATWOOD, M.E., 1993, Project Ernestine: Validating a GOMS analysis for predicting and explaining real-world performance. *Human-Computer Interaction*, 8(3), 237–309

GRAY, W.D., JOHN, B.E., STUART, R., LAWRENCE, D., and ATWOOD, M.E., 1995, GOMS meets the phone company: Analytic modeling applied to real-world problems. In R. M. Baecker, J. Grudin, W.A.S. Buxton, and S. Greenberg (eds), *Readings in human-computer interaction: Toward the year 2000* (2nd edn,). (San Francisco: Morgan Kaufmann), pp. 634–9.

GRAY, W.D., YOUNG, R.M., and KIRSCHENBAUM, S.S. (1997). Introduction to this Special Issue on Cognitive Architectures and Human-Computer Interaction. *Human-Computer Interaction*, 12(4), 301–9.

HOWES, A. and YOUNG, R.M., 1997, The role of cognitive architecture in modelling the user: Soar's learning mechanism. *Human-Computer Interaction* 12(4), 311—43.

JOHN, B.E., and KIERAS, D.E., 1996a, The GOMS family of user interface analysis techniques: Comparison and contrast. *ACM Transactions on Computer-Human Interaction*, 3(4), 320—51.

JOHN, B.E., and KIERAS, D.E., 1996b, Using GOMS for user interface design and evaluation: Which technique? *ACM Transactions on Computer-Human Interaction* 3(4), 287–319.

JUST, M.A., CARPENTER, P.A., and KELLER, T., 1996, The capacity theory of comprehension: New frontiers of evidence and arguments. *Psychological Review*, 103(4), 773–80.

KIERAS, D.E., and MEYER, D.E., 1997, An overview of the EPIC architecture for cognition and performance with application to human-computer interaction. *Human-Computer Interaction*, 12(4), 391–438.

LEWIN, K., 1951, *Field theory in social science* (New York: Harper Row).

MOZER, M.C. and SITTON, M., 1998, Computational modeling of spatial attention. In H. Pashler (ed.), *Attention* (East Sussex, UK: Psychology Press), pp. 341–88.

NEWELL, A., 1990, *Unified theories of cognition* (Cambridge, MA: Harvard University Press).

NEWELL, A. and CARD, S.K., 1985, The prospects for psychological science in human-computer interaction. *Human-Computer Interaction* 1(3), 209–42.

REASON, J., 1990, *Human Error* (New York, NY: Cambridge University Press).

Cognitive Psychophysiology in Ergonomics and Human Factors

A. F. Kramer

Department of Psychology and Beckman Institute, University of Illinois at Urbana-Champaign, IL 16801, USA

1. INTRODUCTION

This paper will focus on recent applications of techniques and methods from Cognitive Psychophysiology to Human Factors and Ergonomics. It will begin by providing a definition of Cognitive Psychophysiology followed by a brief historical discussion of the use of psychophysiological techniques in Human Factors. It will then discuss the types of Human Factors questions and issues that might be addressed with Cognitive Psychophysiology and will provide illustrative examples of applications of psychophysiological measures to Human Factors issues. Finally, it will discuss potential future applications of Cognitive Psychophysiology to Human Factors.

Cognitive Psychophysiology entails the use of physiological measures that can be noninvasively recorded from humans in the study of mental processes including perception, cognition, memory and action. Over the past several decades a number of physiological measures have subserved this function. These measures can be subdivided into those that primarily index activity in the central nervous system (CNS) and those that are primarily sensitive to peripheral nervous system (PNS) activity. It is important to point out, however, that the CNS and PNS distinction is only a shorthand for the organization of the nervous system, since these two systems interact in the control of many physiological functions. The CNS contains all cells within the bony structures of the skull and the spinal column, including the brain, brain stem and spinal cord. The CNS measures that have been used in Cognitive Psychophysiology research and applications include the recording of electroencephalographic (EEG) activity, event-related brain potentials (ERP), magnetic activity of the brain (MEG), neuro-imaging measures including positron emission tomography (PET) and functional magnetic resonance imaging (fMRI), and optical and electro-oculographic (EOG) measurements of eye movements. Measures of PNS activity have, for the most part, concentrated on activity of a particular branch of the PNS system, the autonomic nervous system (ANS). The ANS controls the internal organs of the body by inneverating involuntary (smooth) musculature and serves multiple functions, such as the mobilization of the body to meet emergencies and the conservation and maintenance of bodily resources, through its subdivisions the sympathetic (SNS) and parasympathetic (PNS) nervous systems. The PNS, largely ANS, measures which have been used in Cognitive Psychophysiological research include cardiovascular (e.g. heart rate and heart rate variability measures), respiratory, pupil diameter and electrodermal measures. The detailed description and discussion of the individual measures is beyond the scope here but can be found in Cacioppo *et al.* (in press).

2. COGNITIVE PSYCHOPHYSIOLOGY AND HUMAN FACTORS: A BRIEF HISTORY

Psychophysiological measures, principally measures of gaze direction, played an important role in the examination of human performance in complex systems during the early years of Human Factors. In the 1950s Fitts *et al.* (1950) at Wright-Patterson Air Force Base used measures of gaze direction, gaze duration and the sequence of eye movements to examine the information extraction strategies employed by novice and experienced aircraft pilots during instrument flight. The information acquired from these studies was used to reconfigure instrument panels to optimize the speed and accuracy with which pilots could locate and extract flight relevant information.

The use of other psychophysiological measures to examine issues of interest to the Human Factors community soon followed the pioneering research of Fitts *et al.* For example, EEG activity was recorded from pilots as they flew missions of varying difficulty in simulated and actual flight in an effort to examine the utility of this psychophysiological measure for the assessment of the deleterious effects of high-G environments and mental and emotional workload. Measures of heart rate and heart rate variability have long been used to provide a continuous record of the cardiorespiratory function and mental workload of operators in complex simulated and real-world systems. Heart rate measures have been recorded as aircraft pilots execute a number of maneuvers in simulated and actual aircraft such as landing, refueling during long haul flights, performing steep descents and flying combat missions. Such measures continue to be used today, often in the context of other psychophysiological measures and along with a more sophisticated appreciation for the underlying physiology and the relevant cognitive psychology.

In summary, although the application of psychophysiological techniques to Human Factors issues has a relatively recent history these measures have provided and continue to provide useful insights into human performance and cognition in laboratory and extra-laboratory (i.e. simulator and operational contexts) environments. One can now turn to a discussion of recent applications of psychophysiological measures to topics of interest to the Human factors community.

3. APPLICATIONS OF COGNITIVE PSYCHOPHYSIOLOGY TO HUMAN FACTORS ISSUES AND PROBLEMS

Given that there are a multitude of techniques available to address Human Factors problems and issues one must ask what role psychophysiology might play in Human Factors research and application. Certainly if the information that is gained through psychophysiological measurement is redundant with that obtained from other measures employed by the Human Factors community it is unlikely that psychophysiological measures will gain wide acceptance. This is likely to be the case for the foreseeable future given the relatively high cost and substantial amount of expertise required for the collection, analysis and interpretation of psychophysiological measures relative to the more traditional subjective rating and performance-based measurement

techniques employed by Human Factors practitioners and researchers.

Thus, for psychophysiological measurement techniques to gain acceptance in the Human Factors community these measures must (1) prove to be more valid or reliable indices of relevant psychological (or in some cases physiological) constructs than traditional behavioral and subjective measures, or (2) enable the measurement of constructs that are difficult or impossible to measure with traditional performance-based or subjective measures, or (3) enable the measurement of relevant constructs in situations in which other types of measures are unavailable. Indeed, there is evidence that each of these three criteria has been or can be met within a subset of Human Factors contexts and with subsets of psychophysiological measures.

One additional issue that merits some discussion is the applicability of psychophysiological measurement techniques to high fidelity simulation and operational environments. Certainly, the great majority of psychophysiological research has focused on explicating the functional significance of different measures and components with relatively simple tasks in well controlled laboratory environments. Even in such well controlled environments, a great deal of effort has been expended on the elimination of potential artifacts (e.g. from ambient electrical fields, contamination from other physiological signals that may mask the signal of interest, individual differences in baselines or a measure's morphology or topography, etc.). Given the diversity and magnitude of the artifacts that are encountered in such well controlled settings, is it a reasonable expectation to collect valid and reliable psychophysiological data in less well controlled environments such as high fidelity simulators or operational environments? Although the collection of psychophysiological data in such environments clearly still provides a considerable technical challenge, there have been a number of promising developments in the design of miniaturized recording equipment that can withstand the rigors of operational environments. There have also been developments of pattern recognition and signal analysis techniques that enhance the detection of some physiological signals in noise, and the development of automated artifact rejection procedures. Barring technological roadblocks, such developments should continue to increase the potential for recording psychophysiological signals in a number of complex environments.

One major area in which Cognitive Psychophysiology has made a substantial contribution to Human Factors is in the understanding and measurement of mental workload. Mental workload has been conceptualized as the interaction between the structure of systems and tasks on one hand, and the capabilities, motivation and state of the human operator on the other (Kramer 1991). More specifically, mental workload has been defined as the "costs" a human incurs as tasks are performed. Early views of the human side mental workload suggested that the "costs" of performing tasks could be conceptualized in terms of an undifferentiated capacity or resource. The performance of a difficult task (e.g. driving an automobile in heavy traffic) would leave fewer resources for performing additional tasks (e.g. talking to a passenger) than the performance of an easy task (e.g. driving an automobile on a deserted highway). More current views have argued that multiple resources and mechanisms are required to perform most complex tasks and the extent to which performance

suffers depends on the overlap between resources required for concurrently performed tasks.

Psychophysiological measures have been instrumental in defining the nature of the resources and mechanisms which underlie complex task performance while also enabling the measurement of dynamic changes in mental workload. For example, a number of components of the ERP have been useful in both characterizing resource tradeoffs between concurrently performed tasks and also distinguishing between different types of resources and processes required for complex task performance. The ERP is a transient series of voltage oscillations in the brain which can be recorded from the in response to discrete stimuli and responses. Specific ERP components, usually defined in terms of polarity and minimum latency with respect to a discrete stimulus or response, have been found to reflect a number of distinct perceptual, cognitive and motor processes thereby proving useful in decomposing the processing requirements of complex tasks.

For example, the P300 component of the ERP, a positive going deflection in the ERP which occurs anywhere from 300 to 1000 ms following the occurrence of a discrete stimulus (e.g. the appearance of an aircraft in a pilots field of view), has been found to provide a reliable index of the perceptual/cognitive resources required to perform a set of tasks. Sirevaag *et al.* (1987) found that P300s elicited by changes in the position of a target in a tracking task increased in amplitude with increases in the difficulty of the task. Correspondingly, the amplitude of P300's elicited by events in a task of lesser importance, an auditory discrimination task, decreased with increases in the difficulty of the tracking task. Thus, it would appear that changes in P300 amplitude in dual-task studies mimic the resource reciprocity effects predicted by resource models of mental workload, that is, when more resources are required to perform one task fewer resources are available for the performance of other tasks. The P300 reciprocity effects are important for several reasons. First, they provide converging support for the notion of resource trade-offs among tasks, support that is independent of the performance effects that have traditionally been used to define resource allocation policies. Second, the P300 data can be obtained in the absence of overt responses, thus enabling the assessment of resource allocation policies in situations where behavior occurs infrequently (e.g. a quality control inspector monitoring a visual display, a pilot monitoring the status of the aircraft while flying on autopilot, etc.). Third, the P300 appears to index a specific variety of resources, that is, resources required for perceptual and cognitive processes but not response selection and execution processes.

Other ERP components as well as aspects of heart rate variability have also shown to be sensitive to specific aspects of mental workload. For example, the P100 and N100 components of the ERP appear to reflect early selective attention processes and more specifically the distribution of attentional resources in visual space. The 10 Hz component of heart rate variability appears to be sensitive to working memory demands but insensitive to response or motor demands.

Interestingly, while some psychophysiological measures appear to be diagnostic of particular varieties of processing resources, other measures are less diagnostic and instead appear to reflect general or undifferentiated processing demands imposed upon the human. Psychophysiological measures that fall into this latter category include respiration, heart rate, eye blinks,

electrodermal activity and some components of EEG activity. The sensitivity of these measures to general or undifferentiated processing demands can have both advantages and disadvantages. On the positive side such measures can potentially be utilized in a wide variety of settings and across a number of different systems to provide a general indication of the mental workload experienced by the human operator. In many cases, such information can be extremely valuable to system designers who are interested in the overall magnitude of processing demands imposed upon the human operator. However, if more specific information concerning the type of processing demands is needed, for example to discern whether the response demands of a new control system or the perceptual demands of a new display configuration, are responsible for increased mental workload, then more diagnostic measures are necessary.

Recently there have been attempts to examine the extent to which psychophysiological measures can be used to track dynamic changes in mental load and operator alterness in an effort to enhance overall system effectiveness. For example, Makeig and Jung (1996) investigated the efficacy of EEG-based alertness detection systems for the prediction of missed responses during simulated sonar tasks using US Navy Sonar operators. In their studies, sonar operators attempted to detect 300 ms noise burst targets embedded in a white noise background. In some of the studies, particularly those in which ERP were of interest, brief task-irrelevant tones were also occasionally presented. Operator specific EEG algorithms were derived by employing several different frequencies in the d, q and a bands to predict vigilance decrements. These multiple regression-based algorithms were quite successful. The algorithms developed on data from one experimental session were capable of accounting for between 75 and 85% of the variance in error rates (i.e. response omissions or lapses) obtained in a second experimental session. Other results suggested that on-line algorithms, which detected changes in q and g (> 35 Hz) activity, were capable of predicting missed target detection responses up to 10 s in advance of the occurrence of a target. Finally, ERP components and, more specifically, N200 amplitude and latency and P100/N100 amplitude difference, elicited by task-irrelevant auditory probes did a reasonable job of predicting changes in error rates within a 32-s moving window.

The results from the Makeig and Jung and other recent studies suggest that it is now possible, in simulated real-world tasks, to detect and predict changes in alertness or task engagement that have important implications for system performance. One wonders, however, whether physiologically based alertness detection and prediction systems could be further improved by incorporating a number of psychophysiological measures. Indeed, a number of recent studies suggest that this may be the case. For example, studies have reported that changes in the a and q bands of the EEG along with slow eye movements detected in the EOG can reliably predict that a target would be missed a minute in the future. Thus, it would appear that a promising area of future research is the incorporation of a number of different physiological measures into alertness detection and mental workload assessment systems. Clearly, another important direction is the transition of psychophysiologicaly based alertness detection systems out of the laboratory and into operational environments. This is a particularly important step since many of the vigilance decrements that have been observed in laboratory research have not been reported in the field, which is likely due to the differential incentives to maintain adequate performance in these two settings.

REFERENCES

CACIOPPO, J., TASSINARY, L. and BERTSON, G., (eds) in press. *Handbook of Psychophysiology* (New York: Cambridge University Press).

FITTS, P., JONES, R. and MILTON, J., 1950. Eye fixations of aircraft pilots III: Frequency, duration and sequence of fixations while flying Air Force Ground Controlled Approach System (GCA). Air Material Command Technical Report. USAF Technical Report No. 5967 (Ohio: Wright Patterson AFB).

JONES, R., MILTON, J. and FITTS, P., 1950. Eye Fixations of Aircraft Pilots IV: Frequency, Duration and Sequence of Fixations During Routine Instrument Flight. USAF Technical Report No. 5795 (Ohio: Wright Patterson AFB).

KRAMER, A.F, 1991. Physiological metrics of mental workload: a review of recent progress. In D. Damos (ed.) *Multiple Task Performance* (London: Taylor & Francis), pp. 279-328.

KRAMER, A. F. and WEBER, T., in press. Application of psychophysiology to human factors. In J. Cacioppo, L. Tassinary and G. Bertson (eds), *Handbook of Psychophysiology* (New York: Cambridge University Press).

MAKEIG, S. and JUNG, T.P., 1996. Tonic, phasic and transient EEG correlates of auditory awareness in drowsiness. *Cognitive Brain Research*, 4, 15-25.

SIREVAAG, E., KRAMER, A., COLES, M. and DONCHIN, E. 1987. Resource reciprocity: An event-related brain potentials analysis. *Acta Psychologica*, 70, 77-90.

Combination Manual Handling Tasks

L. M. Straker

Curtin University of Technology, Shenton Park, WA 6008, Australia

1. INTRODUCTION

Manual handling is thought by many to be an important contributor to musculoskeletal disorders such as work-related back disorder. The ergonomics profession has expended considerable resources, both in research and practice, attempting to both: understand the role of manual handling in the etiology of work-related musculoskeletal disorders and determine what ergonomic interventions can lessen the negative consequences of manual handling.

Combination manual handling tasks (CMHT) probably form the majority of manual handling tasks. Therefore, to be successful in preventing musculoskeletal disorders, effective assessment and reduction of the risk associated with CMHT must be achieved.

Combination manual handling tasks are composed of some sequence of single manual handling tasks such as lifting, lowering, pushing, pulling, holding or carrying. For example, a worker in a paint warehouse may be required to move 20-liter containers of paint from a shoulder-height shelf on one side of an aisle to a shoulder-height shelf on the other side of the aisle during stock taking. To achieve this the worker would commonly pull the container to the edge of the shelf, lower it to waist height, carry it to the new location, lift it to shoulder height and finally push the container into place on the shelf.

This chapter reviews the current state of knowledge on risk assessment for CMHT, presents recommendations for the assessment of risk in CMHT and suggests how to prevent musculoskeletal disorders associated with CMHT.

2. RISK ASSESSMENT FOR COMBINATION MANUAL HANDLING TASKS

The approaches available for the assessment of risk from CMHT are the same as those for single manual handling tasks: psychophysical, physiological, biomechanical, psychological, performance and combined approaches (Straker 1995). Within most of these approaches there are commonly four methods used: direct measurement, direct estimation from single-task values, estimation using models based on single-task values, and estimation using models based on worker and work factors. Although there has been considerable investigation and use of these approaches for single manual handling tasks, there is a scarcity of information concerning their use with CMHT. The current state of knowledge is summarized under the approach headings below.

2.1. Psychophysical Approach

Gallagher (1991), Rodrigues *et al.* (1989) and Straker *et al.* (1996) measured maximum acceptable weights (MAW) from people performing CMHT thus demonstrating that direct measurement using psychophysical measures can be used to assess the risk in CMHT.

Jiang and Smith (1985) stated that the only way to analyze CMHT was to analyze each component of the combination and use the most critical (strenuous) component task as the limiting factor. Thus, for a CMHT involving a lift, carry and lower, single-task MAW tables for lift, carry and lower could be viewed and the lowest single-task component MAW, say lift, could be used as an estimate, or substitute, of the CMHT MAW. A number of researchers have compared single and combination-task MAW to evaluate use of estimates of CMHT risk from component single-task risks. Taboun and Dutta (1984), Jiang and Smith (1985), Ciriello *et al.* (1993) and Straker *et al.* (1996) reported results that provide evidence that substituting single-task MAW for CMHT MAW may result in significant errors in risk estimation.

In response to this problem, Taboun and Dutta (1984) recommended that data bases be generated on CMHT capacities and presented one such table in their 1986 paper (Taboun and Dutta 1986). Jiang and Smith (1985) also present a table of MAW for various CMHT that could form part of a database. A number of problems could be anticipated with tables generated specifically for CMHT, the most obvious one being that there are so many potential combinations of tasks, when task type, height, frequency etc. are considered, that it may be impractical to develop and use tables for specific CMHT.

Besides the use of tables, modeling CMHT MAW has also been investigated. Jiang *et al.* (1986) concluded that if the limiting (most strenuous) single-task MAW is known, the CMHT MAW can be accurately predicted. However, modeling CMHT MAW from single-task MAW assumes a stable relationship between the single tasks and CMHT. The veracity of this assumption is questionable given the variety of CMHT and the evidence from the Straker *et al.* (1996) study.

Jiang *et al.* (1986), Taboun and Dutta (1986), Fredericks *et al.* (1992) and Rodrigues *et al.* (1992) have reported attempts at modeling CMHT MAW from work, worker and worker–work interaction characteristics. Although the models had reasonable coefficients of determination, they were unique to each task and again assumed a consistent relationship between CMHT MAW and work, worker and worker–work interaction characteristics across combinations of tasks. Without evidence of a stable relationship the generalizability of these models is questionable.

2.2. Physiological Approach

In the studies by Gallagher (1991) and Straker *et al.* (1997) direct measurement of physiological costs involved in performing CMHT was shown to be possible using measures like heart rate, oxygen consumption and ventilation.

Studies by Jiang and Smith (1985), Ciriello *et al.* (1990, 1993), Taboun and Dutta (1984) and Straker *et al.* (1997) have fairly consistently reported increases in physiological risk measures for CMHT compared with component single-tasks in the order of 15%.

Garg *et al.* (1978) developed models of energy expenditure for manual handling based on the assumption that a job can be divided into simple tasks. Taboun and Dutta (1989) tested the additivity assumption by calculating the extra cost of performing a CMHT over the sum of its components. The extra oxygen cost varied from –40 to +60%. Therefore, the additivity assumption appears not to be a valid assumption for every type of CMHT.

Taboun and Dutta (1984, 1989) report the development and

testing of models for estimating physiological cost during CMHT based on worker and work characteristics. The models had reasonable coefficients of determination (~0.8); however, each model was for a specific CMHT, limiting their utility. Also, the reports of modeling of MAW from worker and work characteristics showed that this method was often not as accurate as modeling from single-task MAW.

2.3. Biomechanical Approach

Despite the importance of the biomechanical approach, and its common usage for assessing manual handling tasks risk, little information on assessing CMHT risk using this approach could be found.

Of the commonly used biomechanical methods only the Force Limits method (Davis 1980) mentions CMHT, suggesting they can be divided and analyzed as separate single tasks. However, the relationship between IAP (intra-abdominal pressure) and lumbar stress is probably even more complex for CMHT than for single tasks. Therefore, this method is probably of limited utility given the validity concerns.

The two widely used biomechanical methods based on kinematic and kinetic data, SSPP (Center for Ergonomics 1990) and WATBAK (Norman and McGill 1989) models, do not mention assessing CMHT. Presumably a CMHT could be segmented into a number of single tasks and a static analysis performed on each component. However, Straker et al. (1997) compared estimates of peak lumbar stresses (compression force, shear force and torque) for a range of single tasks (lift, lower, push, pull, carry) and combinations of these single tasks and showed mean absolute differences between single and CMHT of ~22%.

2.4. Psychological Approach

2.4.1. Discomfort

Studies by Rodrigues (1990) and Straker et al. (1997) both illustrate the potential for direct measurement of discomfort to provide useful information for risk assessment. Straker (1995) suggests that although the use of tables and modeling is possible, given the difficulties in using potentially less labile dependent variables like heart rate and energy cost, it seems unlikely that these would be useful options.

2.5. Ratings of Perceived Exertion

Exertion ratings have been used in studies on CMHT risks by Taboun and Dutta (1984) and Straker et al. (1997), showing that the direct measurement option is possible. The study by Straker also explored the option of substituting single-task measures for CMHT measures. The results paralleled those for discomfort, with mean absolute errors for exertion rating of 19%. No reports of modeling in CMHT could be found.

2.6. Work Satisfaction

No reports could be found of the use of ratings of satisfaction for CMHT risk assessment. However, ratings of satisfaction with work tasks have the potential to provide useful information for risk assessment.

2.7. Performance Approach

Similarly, no reports could be found of the use of performance measures to assess the change in risk of CMHT following intervention to control risk.

2.8. Combined Approaches

Many combined approaches (National Institute for Occupational Safety and Health equation 1983, the tables produced by Asfour et al. 1985, and the modeling of Kim and Ayoub 1991) are only suitable for lifting. The tables prepared by Mital et al. (1993) do cover a range of single tasks, but not CMHT. Mital et al.'s tables could therefore be used to substitute single-task values for CMHT values. Checklists such as that of Worksafe Australia (1990) could be used for combination though are probably not sensitive enough.

3. RECOMMENDATIONS FOR ASSESSMENT OF COMBINATION MANUAL HANDLING TASKS

Direct measurement, using the psychophysical approach, is expensive, provides results which are probably not generalizable and is vulnerable to extraneous influences such as industrial relations issues. However, direct measurement of MAW for CMHT does not require the assumptions that the other methods do. It may therefore be useful for specialized tasks where unique characteristics make the extrapolation of other methods precarious.

The estimation of CMHT MAW from single-task tables or modeled on single-task values, assumes a consistent relationship between single-task MAW and CMHT MAW. This has been shown to be not the case (Jiang and Smith 1985, Taboun and Dutta 1984, Straker et al. 1997) with a mean error of estimation around ±20%. Modeling using linear models appears to be not sufficiently more accurate than substitution to warrant the extra complication. Thus, the best option for the assessment of risk in CMHT, when resources are not adequate to do direct measurement, may be to use substitution of MAW for critical component tasks (commonly lift or carry). However, this should only be used with the knowledge of an error margin of ±20%.

Physiological direct measurement appears a viable risk assessment option for CMHT. Its main disadvantage is the resource cost if used to evaluate every CMHT. As for MAW, this method is likely to be the most accurate. Estimation from single-task tables appears to be hazardous as there is consistent evidence for CMHT having different physiological costs to component tasks. Estimation from modeling from worker and work characteristics is likely to be less successful than modeling from single-task values. However, these conclusions are drawn from a small number of studies with limitations. Until more definitive research is available, the best option for assessment of CMHT using the physiological approach, when direct measurement is not practicable, may be to use substitution of single-task measures. However, considerable error margins need to be acknowledged if this method is used: ~60% when absolute loads are compared and ~20% when MAW loads are compared.

Biomechanical modeling based on kinematic and kinetic data appears the only viable biomechanical method available for assessing the risk in a CMHT. The literature suggests the most accurate method is by dynamic modeling from CMHT data. However, where this is not practicable, static analysis of critical parts of the CMHT may be of some use, although as for other approaches the errors are likely to be significant.

The use of discomfort, exertion, satisfaction or performance measures as risk indicators would probably be most accurate if done using the direct measurement method. Further developmental work is required to produce a reliable measures. However, when used in conjunction with other measures, it would provide a more holistic appraisal of risk.

No methods that combine approaches appear to be available for assessing the risk of CMHT. It is therefore probably important that measures from a number of approaches are gained as no one approach seems to adequately describe the risk.

Current methods for the assessment of risk in CMHT tend to suffer from limited validity or utility. Effective risk assessment is important to the ergonomics profession, both to evaluate the success of interventions and to defend the decisions of ergonomist in litigation. Practitioners and researchers therefore need to improve the current methods.

4. PREVENTION OF MUSCULOSKELETAL DISORDERS FROM COMBINATION MANUAL HANDLING TASKS

In view of the limited knowledge on risk assessment of CMHT, ergonomists should be tentative in their recommendations for preventing musculoskeletal disorders associated with CMHT. However, the following principles are likely to assist prevention:

- No component of a CMHT should exceed single-task recommendations.
- Loads for CMHT should err on the conservative side to allow for cumulative effects.

Principles of risk reduction for single tasks should be applied to CMHT (for example eliminate manual handling were possible, eliminate handling below knee and above shoulder level were possible, facilitate minimal load–worker distance, facilitate secure grips, facilitate secure footings, encourage smooth steady movements).

REFERENCES

ASFOUR, S.S., GENAIDY, A.M., KHALIL, T.M. and GRECO, E.C., 1985, A combined approach for determination of lifting capacity. In Eberts, R.E. and Eberts, C.G. (eds) *Trends in Ergonomics/Human Factors II* (Amsterdam: North-Holland), pp. 617–23.

CENTER FOR ERGONOMICS, 1990, *User's Manual for the Two Dimensional Static Strength Prediction Program* (Version 4.2e ed.) (Ann Arbour: University of Michigan).

CIRIELLO, V.M., SNOOK, S.H., BLICK, A.C. and WILKINSON, P.L., 1990, The effects of task duration on psychophysically determined maximum acceptable weights and forces. *Ergonomics*, 33, 187–200.

CIRIELLO, V.M., SNOOK, S.H. and HUGHES, G.H., 1993, Further studies of psychophysically determined maximum acceptable weights and forces. *Human Factors*, 35, 175–86.

DAVIS, P.R. and THE MATERIALS HANDLING RESEARCH UNIT UNIVERSITY OF SURREY, 1980, *Force Limits in Manual Work* (Guildford: IPC Science and Technology Press).

EUROPEAN COMMUNITIES COMMISSION, 1990, *Manual Handling Directive 90/269/EEC* (Brussels: European Communities Commission).

FREDERICKS, T.K., FERNANDEZ, J.E. and RODRIGUES, C.C., 1992, Predicting the maximum acceptable weight of lift for an asymmetrical combination task. In *Proceedings of the Human Factors Society 36th Annual Meeting* (Atlanta: Human Factors Society), pp. 707–11.

GALLAGHER, S., 1991, Acceptable weights and physiological costs of performing combined manual handling tasks in restricted postures. *Ergonomics*, 34, 939–52.

GARG, A., CHAFFIN, D.B. and HERRIN, G.D., 1978, Prediction of metabolic rates for manual materials handling jobs. *American Industrial Hygiene Association Journal*, 39, 661–74.

JIANG, B.C. and SMITH, J.L., 1985, The comparison of stressfulness of manual materials handling activities. In Eberts, R.E. and Eberts, C.G. (eds) *Trends in Ergonomics/Human Factors II* (Amsterdam: North-Holland), pp. 577–89.

JIANG, B.C., SMITH, J.L. and AYOUB, M.M., 1986, Psychophysical modeling of manual materials handling capacities using isoinertial strength variables. *Human Factors*, 28, 691–702.

KIM, H.K. and AYOUB, M.M., 1991, A model for determining the permissible workload for lifting. In Queinnec, Y. and Daniellou, F. (eds) *Designing for Everyone: Proceedings of the 11th Congress of the International Ergonomics Association* (London: Taylor & Francis), 1, pp. 90–2.

MITAL, A., NICHOLSON, A.S. and AYOUB, M.M., 1993, *A Guide to Manual Materials Handling* (London: Taylor & Francis).

NATIONAL INSTITUTE FOR OCCUPATIONAL SAFETY AND HEALTH, 1983, *Work Practices Guide for Manual Lifting* (Akron: American Industrial Hygiene Association).

NORMAN, R.W. and McGILL, S.M., 1989, *WATBAK: User's Manual for a Computer Software Package for the Assessment of Low Back Injury Risk During Manual Handling Tasks* (Version 3.0 ed.) (Waterloo: University of Waterloo).

RODRIGUES, C.C., CONGLETON, J.J., KOPPA, R.J. and HUCHINGSON, R.D., 1989, Maximum acceptable weight of lift for an asymmetrical combination manual handling task. *International Journal of Industrial Ergonomics*, 4, 245–53.

RODRIGUES, C.C., 1990, Analysis of subjective responses and heart rates for a combination manual handling task. In Das, B. (ed) *Advances in Industrial Ergonomics and Safety II* (London: Taylor & Francis), pp. 619–26.

RODRIGUES, C.C., FREDERICKS, T.K. and FERNANDEZ, J.E., 1992, Prediction models for asymmetrical lifting combination tasks. In Kumar, S. (ed) *Advances in Industrial Ergonomics and Safety IV* (London: Taylor & Francis), pp. 875–82.

SNOOK, S.H. and CIRIELLO, V.M., 1991, The design of manual handling tasks: revised tables of maximum acceptable weights and forces. *Ergonomics*, 34, 1197–213.

STRAKER, L.M., 1995, A model of manual handling risk assessment. *Ergonomics Australia*, 9, 16–30.

STRAKER, L.M., STEVENSON, M.G. and TWOMEY, L. T., 1996, A comparison of single and combination manual handling tasks risk assessment: 1. Maximum Acceptable Weight measures. *Ergonomics*, 39, 127–40.

STRAKER, L. M., STEVENSON, M. G. and TWOMEY, L. T., 1997, A comparison of single and combination manual handling tasks risk assessment: 2. Discomfort, rating of perceived exertion and heart rate measures. *Ergonomics,* 40, 656–69.

STRAKER, L. M., STEVENSON, M. G., TWOMEY, L. T. and SMITH, L. M., 1997, A comparison of single and combination manual handling tasks risk assessment: 3. Biomechanical measures. *Ergonomics,* 40, 708–28.

TABOUN, S. and DUTTA, S.P., 1984, Prediction models for combined tasks in manual materials handling (CTMMH). In Attwood, D.A. and McCann, C. (eds) *Proceedings of the 1984 International Conference on Occupational Ergonomics* (Toronto: Human Factors Association of Canada), 1 pp. 551–55.

TABOUN, S.M. and DUTTA, S.P., 1986, Modeling psychophysical

capacities for combined manual materials handling activities. In Karwowski, W. (ed) *Trends in Ergonomics/Human Factors III, Proceedings of the Annual International Industrial Ergonomics and Safety Conference* (Amsterdam: North-Holland), pp. 785–91.

TABOUN, S.M. and DUTTA, S.P., 1989, Energy cost models for combined lifting and carrying tasks. *International Journal of Industrial Ergonomics*, 4, 1–17.

WATERS, T.R., PUTZ-ANDERSON, V. and GARG, A., 1993, *Applications Manual for the Revised NIOSH Lifting Equation* (Cincinnati, OH: US Department of Health and Human Services, National Institute for Occupational Safety and Health).

WORKSAFE AUSTRALIA, 1990, *National Standard for Manual Handling and National Code of Practice for Manual Handling* (Canberra: Australian Government Publishing Service).

Comfort–Discomfort Phase Change

R. S. Goonetilleke

Department of Industrial Engineering and Engineering Management, Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong

1. INTRODUCTION

Pressure at the human interface is generally considered to be an important parameter in comfort evaluation. However, the ideal pressure distribution between the human body and any surface of a given application has yet to be defined. A hard-core materials perspective to loading would be to distribute the force over an area as large as possible to minimize stress concentration. However, products such as a bed of nails or a bed of springs, "health sandals," steering wheel covers with semi-spherical protrusions are very popular in some cultures. So, why is it that people tend to like products with localized force? Researchers have pointed out that poor biomechanics may turn comfort into discomfort. Why does such a transition take place over time? The objective of this article is to explain why some design strategies work for human–product interfaces and why some others "slip" from their desirable state or change phase.

2. INTERFACE PRESSURE

Pressure at the human–product interface is unavoidable, and it has received considerable attention over the years because it can cause injury, pain, and discomfort. Pressure is defined as follows:

Pressure (Pascal) = force/area (Newton/m^2).(1)

Alternatively, stress has a similar formulation and is generally in the form of compression, tension or shear:

Stress (Pascal) = force/area (Newton/m^2).   (2)

Interface pressure can play an important role in the development of products and devices such as seat cushions (Figure 1), shoe insoles, handles and beds. An understanding of the pressure patterns that are appropriate for the human body (that is, the patterns that reduce discomfort or improve what is called comfort) can make product design and usage very satisfying and fulfilling. One primary problem related to this aspect stems from the fact that the word *comfort* has many definitions. For example, comfort has been defined as the lack of discomfort. More recently, it has been associated with feelings of *relaxation* and well-being (Zhang *et al.* 1996). A good example is the use of a beanbag. Generally, a sense of comfort is associated with a beanbag comprising macrospheres of rigifoam or the like. Are beanbags really comfortable because of their pressure distribution or are they comfortable because we use them in environments where we are already in a relaxed state?

As shown in Figure 1, the pressure pattern between the two supporting surfaces is somewhat different. Researchers have shown that this pressure pattern is related to the level of comfort and discomfort (Lueder and Noro 1992). Generally, discomfort may be quantified using a combination of the following:
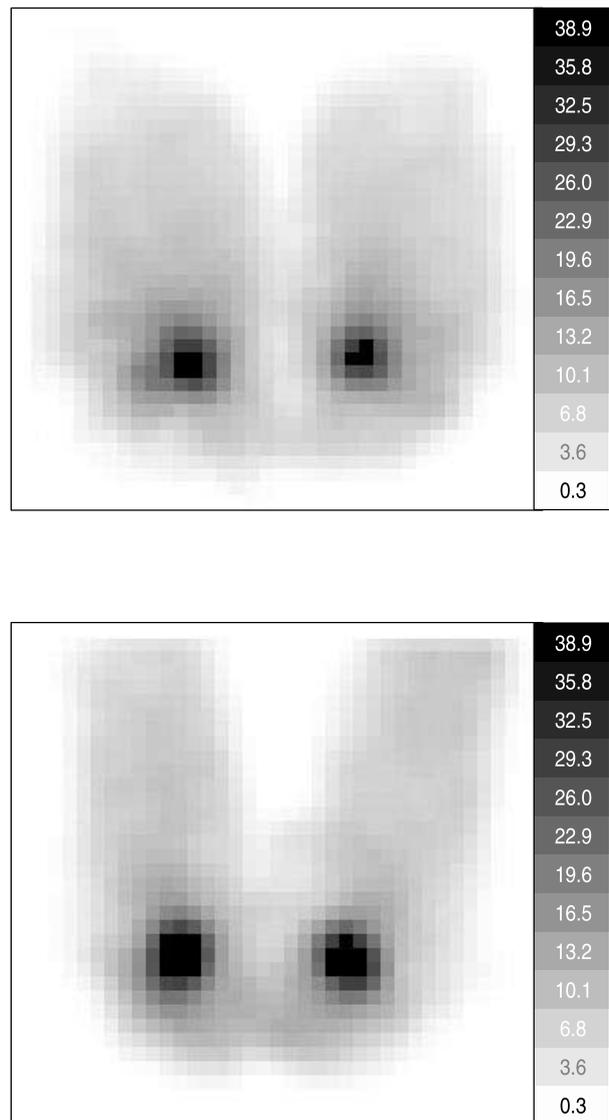
Figure 1. Pressure on buttock during sitting. (a) Pressure pattern on a laboratory chair; (b) pressure pattern on a hard surface (in kPa).

- Peak pressure.
- Pressure gradients.
- Contact area.

It is quite unfortunate that the most appropriate pressure pattern for each individual or individual–product interface is still unknown. If it were known, product design for the human body would be so much easier. Part of the problem stems from the fact that contact area has been neglected in the past.

In general, forces can be supported in two ways:

- distribute "Uniformly," or
- concentrate (i.e. load the "stronger" parts of the structure) to reduce "breakage."

Which one of the two strategies is best to maximize comfort or even minimize discomfort on the human body and why? Could it be that both strategies are applicable depending on the loading scenario? The most common and sometimes naive approach is to distribute force, as much as possible, to achieve the uniform condition. However, some consumer products and existing

research suggest a concentrated strategy. For example, Krouskop *et al.* (1985) has shown that mattresses with a uniform pressure distribution make people restless thereby opposing the distributed theory of force. In addition, products such as a bed of nails or a bed of springs, "health sandals," shoe insoles or steering wheel covers with semi-spherical protrusions, "massage" mats made of wooden slats and cane chairs are very popular in some cultures. All such devices induce localized force, rather than "distributed" force, supposedly creating desired sensations. Active cushions, which operate on the basis of periodic pressure relief in the form of a pressure wave, are also popular among paraplegic individuals. In some sense, these active cushions may be viewed as units that impose "concentrated" loading of a certain level with reduced duration of exposure. In other words, they create a concentrated strategy of short duration. Hence, it appears that concentrated force also has certain advantages and possibly a sensation of comfort or relaxation. If uniform pressure is the ideal pressure distribution for optimal comfort, interface design should be relatively easy, especially since pneumatic or hydrostatic balloons or bladders can be used to give this "constant" or uniform pressure at the interface. What could be affecting the pressure–comfort or pressure–discomfort relationship?

3. PHYSICAL SENSATION

Discomfort or pain originates when special nerve endings, called nociceptors, detect an unpleasant stimulus. There are millions of nociceptors in the skin, bones, joints, muscles, and internal organs. These nociceptors use nerve impulses to relay pain messages to networks of nearby nerve cells (peripheral nervous system). Each cell-to-cell relay is almost instantaneous and is facilitated by neurotransmitters. Messages then travel along nerve pathways to the spinal cord and brain (central nervous system). Scientists believe that pain signals must reach a *threshold* before they are relayed. *Gate control theory* explains how specialized nerve cells in the spinal cord act as gates that open to allow messages to pass, depending on the intensity and nature of the pain signal.

4. SPATIAL SUMMATION THEORY

The *spatial summation theory* (SST) states that simultaneous stimulation of many sensory receptors is required to arouse stimulation (Hardy and Oppel 1937). In simple terms, it means that the larger the area stimulated, the greater the sensory response experienced. For example, the sensation induced by a hand would generally be greater than that induced by a finger alone. The spatial summation theory has important implications for force distribution. Consider the case of a pleasant sensation gradually moving towards discomfort when the applied pressure is increased. At the limit, when sensations tend toward the so-called discomfort experience, a force distributed over a large area may induce greater discomfort than the same force over a small area. Thus, the extension of the spatial summation theory may be used to explain the discomfort experience as well. In this article, the validity of the SST extension is shown using maximum pressure tolerance (MPT). In addition, I would use an example of footwear to show why "comfort" changes phase to discomfort in the long-term.

Researchers have shown that the skin blood flow changes are influenced by three factors: the ratios of bone depth, the ratios of indentor diameter to bone diameter, and percentage compression of the tissue overlying the bone. "Indentor" or loading area is a factor neglected by many and its effect on discomfort can explain perceived sensations of interface designs having concentrated loading.

5. MAXIMUM PRESSURE TOLERANCE

Goonetilleke and Eng (1994) showed that the maximum pressure tolerance (MPT = applied force/probe area) is strongly related to the probe or indentor size or the contact area of the stimulus. The mean MPT with a probe of 5 mm diameter (831 kPa) was 3.3 times that with a probe of 13 mm diameter (249 kPa). The results for two locations on the dorsum (top) side of the foot are shown in Figure 2.

The measurement procedure is described in Goonetilleke and Eng (1994). There were no statistically significant ($p < 0.05$) differences between locations or between genders. In general, it was found that the values of pressure tolerance with the two probes were related by:

$$MPT_{5\text{ mm diameter}} = 3.3 * MPT_{13\text{ mm diameter}}. \qquad (3)$$

It is logical to look at the maximum *force* (or maximum force tolerance, MFT = MPT* area) that can be supported rather than the pressure. The force relation between the two indentors would be:

$$MFT_{5\text{ mm diameter}} = 3.3 * (5/13)^2 * MFT_{13\text{ mm diameter}} \qquad (4)$$

or

$$MFT_{5\text{ mm diameter}} = 0.5 * MFT_{13\text{ mm diameter}}. \qquad (5)$$

In other words, at the maximum tolerable value, the force that a probe with a 5 mm diameter can exert is half of what a probe with a 13 mm diameter can exert. This suggests that even though the MPT is lower with a 13 mm diameter probe, the force that the 13 mm probe can support at the tolerable threshold is twice the force that can be supported by a 5 mm probe. This is a surprising result when the MPT is compared to a "dead" material. For any material other than the human tissue, the equivalent "MPT" at breaking point (or breaking strength) should be independent of area since it is *constant* for a given ("dead") material.

If the maximum force is such that even though only half of
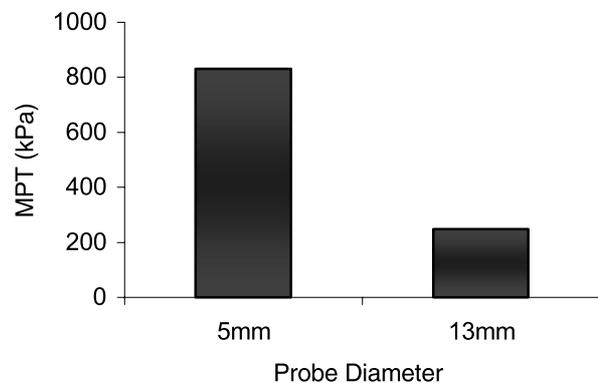
Figure 2. Effect of probe size on the mean value of maximum pressure tolerance.

the force can be supported with a 5 mm probe, the probe area results in 3.3 times the tolerance of the 13 mm probe area, then it implies a counter-intuitive suggestion for loading on the human body. Consider the force supported at the tolerance level (249 kPa) of the 13 mm probe. The maximum force corresponding to the MPT with the 13 mm probe (249 kPa) is 33 Newtons and the corresponding area of support is 132.7 mm². The area of the 5 mm probe is 19.6 mm² (that is 6.8 times smaller). If six (for convenience, rather than 6.8) load bearing areas of 5 mm diameter are chosen to carry the load corresponding to the MPT of the 13 mm probe, the load on each 5 mm probe will be 5.5 Newtons or a pressure of 281 kPa (will be 249 kPa if 6.8 was used instead of 6), which is far below the MPT with one 5 mm probe. In other words, the load, which may have caused someone to indicate that it is the maximum tolerable force over an area of 132.7 mm², could now be shared among a number of smaller areas (localized) without experiencing any such maximum tolerable value. The advantage of smaller areas to support loads is clear when the loads are high.

6. THE PHASE CHANGE

Since the maximum pressure tolerance is dependent on the contact area, it may be concluded that, at high forces, a larger area may cause a higher level of discomfort than a smaller area when stimulated with the same magnitude of pressure. The aforementioned results suggest that localized pressure regions may in fact prove to be less discomforting when compared to "distributed-moderate" pressures. However, we do not know whether distributing force over a larger area increases comfort at low force values even though spatial summation theory indicates that the sensation will be higher. We may thus conclude that perceived sensation and contact area appear to have a relationship similar to that shown in Figure 3. The term sensation is used since it is unclear whether comfort can be equated to positive sensations. However, a negative sensation may be viewed as discomfort. The traditional thinking of distributing forces may be successful only in the upper half of Figure 3, when forces are very low or below a critical value, F_{crit}. The bottom half of Figure 3 will not be seen at very low forces or forces below F_{crit}. Hence, the decision to distribute or concentrate forces really depends on the magnitude of the pressure that exceeds a critical or threshold pressure (P_{crit}) for a given surface area.

7. EXAMPLES

The extension of the spatial summation theory, the *discomfort hypothesis*, will now be illustrated using two examples. Some "high-end" commercial footwear have adopted "dynamic fit sleeves" made of an elastic material, which may convince consumers that such footwear conform to the foot resulting in a better "fit" and a uniform pressure, and thereby high sensation. In the short-term or at the point of purchase, these shoes are very comfortable. However, with prolonged use, due to foot swelling and foot deformations, shoes using a dynamic fit sleeve can be extremely uncomfortable. So how does the *discomfort hypothesis* explain such a change in sensation? At the point of purchase, the foot covering conforms to the foot giving a stronger sensation as predicted by the spatial summation theory since there are more sensory receptors being stimulated at one time (upper half of Figure 3). However, with activity, the foot swells and deforms, and then the pressure induced is greater at each of the
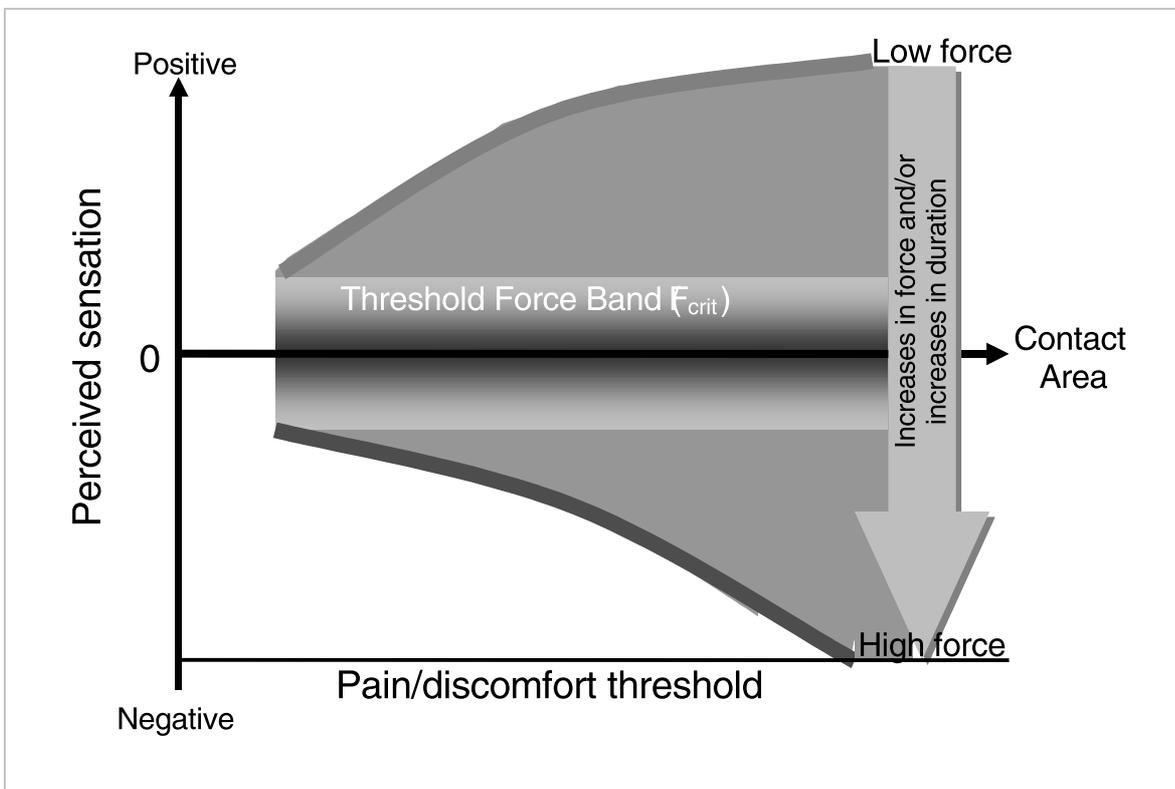
Figure 3. Hypothetical relationship between perceived sensation and contact area.

receptors. At this point, the wearer is at the other end of the sensory experience. More receptors are stimulated over a greater area, giving the wearer greater discomfort than if he/she wears shoes that do not have dynamic fit sleeves (bottom half of Figure 3). In other words, over time, the perceived sensation changes phase (or "slips") from the top half to the bottom half, transforming the "high" intensity pleasant sensations (at $t = 0$), to a "high" intensity unpleasant (or discomfort) experience.

A different type of example is a pilot headset. A conventional headset has cushioned material against the ears. It is generally comfortable in the short-term. As the flying time increases, however, such a headset causes a high level of discomfort as a result of a phase change from positive sensations to negative sensations. The discomfort on the ears can be relieved even in the long-term with the use of slow recovery (or open cell) foam such as Confor™ foam. With this type of foam, the force acting on the ears is significantly reduced so that even with increases in time, no phase change occurs.

8. RECOMMENDATIONS

Based on the above discussion, the following suggestions are made to designers of human–product interfaces:

- Identify the threshold force or pressure (F_{crit} or P_{crit}) to delineate between the experience of a positive sensation and discomfort.
- If the pressure is below P_{crit}, then it would be best to distribute the forces.
- When the pressures are 'high' and close to the MPT, the designer needs to consider a more localized or concentrated force (preferably of short duration) to relieve discomfort caused by simultaneous neuron firings over larger areas.

No doubt, it is sometimes impossible to concentrate pressures over small regions if there is a high risk of damage in terms of pressure sores or ulcerations (Webster 1991), especially for paraplegics who have no sensation or reduced sensation. The ideal pressure profile generally comes from a combination of distribution and concentration, which gives a showroom or point-of-purchase feel with a distributed force but a less discomforting support with concentrated force. Successful examples include mattress overlays and shoe insoles.

REFERENCES

GOONETILLEKE, R.S. and ENG, T.J., 1994, Contact area effects on discomfort. In *Proceedings of the Human Factors and Ergonomics Society 38th Annual Meeting*, 688–90.

HARDY, J.D. and OPPEL, T.W., 1937, Studies in temperature sensation. The sensitivity of the body to heat and the spatial summation of the end organ responses, *Journal of Clinical Investigation*, 16, 533–40.

KROUSKOP, T., KREBS, M., HERSKOWICZ, I. and GRABER, S., 1985, Effectiveness of mattress overlays in reducing interface pressure during recumbency, *Journal of Rehabilitation Research and Development*, 22, 7–10.

LUEDER, R. and NORO, K., 1992, *Hard facts About Soft Machines: The Ergonomics of Seating* (London: Taylor & Francis).

WEBSTER, J.G., 1991, *Prevention of Pressure Sores* (Philadelphia: Adam Hilger).

ZHANG, L., HELANDER, M.G. and DRURY, C.G., 1996, Identifying factors of comfort and discomfort in sitting, *Human Factors*, 38, 377–89.

Constraints in Design Decision-making

D.L. Day

Computer and Information Sciences Department, Towson University, Towson, MD 21252, USA

1. INTRODUCTION

Since the day in the 1960s when Stanley Kubrick's film *2001* was first screened in the United States, the role and impact of tools has been a central feature of popular American technology. Of course, tools had played a major role in the development of industrial society since long before the Industrial Revolution in England. As demonstrated by the popular BBC television series "Connections" and its sequel "The Day the Universe Changed", there has been a long-standing fascination with tools as an extension of man's inherent capabilities.

Beyond this fascination, however, lies a very real and very key fact for civilization, whether it be European, South American, or Asian: tools improve man's understanding of the natural world around him, and leverage his capabilities. They exaggerate man's command of his environment, taking advantage of his unique capability of storing, processing and applying information. Man's ability to shape useful artifacts in his environment to his purposes has been a central facet of life since Kubrick's proto-humans metaphorically grasped a jaw bone, thrust it into the air, and created a space station.

Tools are extensions of mankind. However, tools are a double-edged sword. The key benefit to tools is their capacity to be customized to a task at hand. The more appropriate a tool is for its intended task, the better it is deemed to be. However, the "appropriateness" of a tool always involves an estimate of its utility, its cost, and its alternatives within the natural environment. More, assessment of whether to use a tool inevitably rests on whether the tool benefits or harms the purpose at hand. If the learning necessary to use tools is taken into account, often the disadvantages outweigh promised benefits.

In the current century, the ultimate tool of mankind may be the computer workstation. It has become for many citizens the ultimate tool: a device that can be acquired cheaply, operated with minimum inconvenience and cost, and used to magnify individual skills to an extent that few could anticipate when Kubrick sent his modern apes to Jupiter.

2. CONSTRAINT OBJECTIVES AND IMPACTS

The benefit and bane of computerized systems is that, based upon the assessment of "professionals" in a particular domain (to which the systems are to be applied), they restrict the user to particular behaviors. These behaviors include only those actions that professionals in the field believe are rational within the context of the application. Irrational behaviors are held to be counter-productive, and are therefore discouraged (if not prohibited outright). The imposition of controls to prevent irrational behaviors within a specific application context is known as "constraint".

Constraint is a tricky concept. It has been defined variously as a principle that should be followed to control the selection and ordering of actions during design, and as a rule that defines the range of options available in performance of a task. In any case, constraints guide tool users in creation of quality outcomes by delimiting the problem space and recommending (with various degrees of force) preferred procedures.

In the Western cultures, concerned as they are with individual freedoms, the term "constraint" has a negative connotation. It implies an inappropriate control of the individual to further the undefined purposes of someone or something else. However, in other cultures, the idea of constraint is positive. It is the mechanism by which informed society communicates appropriate and necessary guidance to its members.

Regardless of context, however, the preferred behaviors encouraged or dictated by computerized systems are an important part of modern life. The question that remains, however, is how well the constraints imposed by computerized systems match the needs of people to use such systems as effective tools in their everyday lives. Unlike use of the jaw bone in Kubrick's film, the application of computers to domain tasks is very much dictated by the many highly technical individuals who created the tool — not by the user, who merely tries to operate an appropriate device to accomplish the tasks at hand.

Constraints are intended to make users more productive. Constraints limit wasted effort: they define the problem solution space within which the individual should work, based on the perceptions of people who developed their computerized tools. The problem, of course, is that many users do not agree with the perspectives of tool developers.

This chapter addresses the potential mismatch between computerized tools and their users. It examines the lack of understanding between highly skilled artisans of sophisticated systems and the perhaps less sophisticated but numerous users who must apply computerized systems for economically productive work. The key feature of this mismatch is the insensitive implementation of constraints for whatever it is that the user does with the system.

3. CONSTRAINTS IN DESIGN ENGINEERING

The examination of constraints as part of tool-assisted problem solving is relatively new. Relevant theory has been developed in several areas, including decision support, cognitive psychology, software engineering, and human–computer interaction.

Although the intent, implementation, and effects of constraints in computerized tools can be of interest in any activity, the context of interest here is design. Design engineering is fertile ground for the examination of constraints because design inherently involves individual creativity, typically affords and even requires independent decision-making, and has strong secondary effects. (The constraints applied to design activities directly impact work product quality and utility, as well as costs of product generation.)

Computerized design tools form a common medium through which the design preferences of tool developers are communicated to tool users. Communication takes place via constraints that developers feel are important in the application domain. Constraints impact design not only because of the types of controls developers chose to include in tools, and the strength with which such controls are enforced, but also because of the styles of implementation chosen. In practice, tools that apply constraints

to design tasks act as decision support systems. The decisions being supported are those intrinsic to the task at hand, to the domain environment, and to design engineering.

3.1. Modeling Constraints

One view holds that constraints naturally express the many dependencies, restrictions, and high-order characteristics that are intrinsic to design. The basic interdependence of decisions made during design can be represented as a semantic network. In this model, each node represents a significant decision, linked to others by arcs representing the potential sequence of decisions. Nodes in the network are viewed as frames, with slots composed of various attributes. Some of these are related to general decision-making style, while others reflect concerns in the specific task domain. Constraints and their characteristics form an important slot within the node frames. As decisions are taken, the links between nodes are fired, causing related further decisions to be considered in the context of decisions already made and of slot contents at the new node. A completed design is the sum of all decisions, including the extent to which standards and procedural best practice were followed (i.e. constraints adhered to) in creating the work product.

3.2. Constraint Characteristics

In order to be of practical use, concepts and approaches in any field must be operationalized. Only then can key elements and their impacts be assessed consistently and reliably. Mere definition, no matter how precise, is insufficient. In one respect, operationalization clarifies definition by providing examples.

If the domain of interest were computer-aided systems engineering, the following examples might be used to clarify the concept of constraint:

- Every process in a data flow model must have at least one input data flow and at least one output data flow.
- A parent process must be specified before its child processes.

Another way operationalization can clarify definition is by identifying descriptors that characterize instances of a concept in the given domain. The set of descriptors so identified are arrayed in a taxonomy, allowing the categorization (and subsequent measurement) of the phenomena of interest. In the case of constraints, such a taxonomy has been developed. It includes the following descriptors.

1. *Source of authority*. The justification for a constraint, usually stemming from the role, position, or recognized expertise of the authority that decided a constraint should exist.
2. *Weight*. The importance of a constraint, judged either as the extent to which appropriate process is considered vital or as the anticipated effects of constraint violation upon work product quality.
3. *Balance of control*. The degree to which a tool rules the design process, allows user intervention, and provides notice to users of decisions taken by the tool. This affects the kind of options given to users who encounter constraints, and the tool's response when constraints are violated.
4. *Constraint type and target*. The aspect of design activity (process or product) affected by a constraint, and the types of objects impacted. A constraint may affect the structure of the design artifact (product) or the way in which the artifact is produced (process).

5. *Means of implementation*. The style chosen by a tool builder to implement a constraint. Style is important because the same constraints implemented with different styles may result in different user responses, and therefore have different impacts on product and/or process.

Each of these descriptors is an architectural feature of decision node constraint slots, using the semantic network model of design decision making. A descriptor serves as a classification that may be used to group or separate constraints, based on whether they share significant common values. A profile of each constraint may be created, by assessing the values for all descriptors.

3.3. Constraint Negotiation Behaviors

Examination of the role that constraints play in design engineering stems from concerns about how constraints impact the behavior of tool users. That behavior impacts valued outcomes, such as work product quality and user productivity.

The need for a taxonomy of constraints was explained in the preceding section. A taxonomy of common user behaviors is also needed. Constraints are a stimulus to users as they operate their tools; user behavior when users encounter constraints is a response to that stimulus. In order to understand the impact of constraints, it is necessary to describe and measure behavioral responses.

A taxonomy of user behaviors in response to constraints has been developed. It includes the following descriptors:

- *Avoidance*. The user modifies his task approach pre-emptively because he knows that doing otherwise would trigger a constraint.
- *Compliance*. The user modifies his task approach to suit the limitations imposed by a constraint.
- *Subversion*. The user modifies his task approach to take advantage of known weaknesses in the tool, overriding the spirit but not the mechanism by which the constraint is implemented. This strategy also is known as 'workaround'.
- *Path-seeking*. The user modifies his task approach to find the least constraining path.
- *Deferral*. The user declines to modify his task approach, with the knowledge that subsequent system or human review may reverse, modify, or accept his decision to override the constraint.
- *Negation*. The user declines to modify his task approach, unconditionally overriding the constraint.

In categorizing any given behavior using this taxonomy, it is key whether the constraint is accepted, whether the attempted task approach is modified, and whether the user's decision is subject to later review (and possible reversal).

4. MEASURING THE IMPACT OF CONSTRAINTS

Taxonomies of constraint and user behavior can be applied empirically to assess the impact of constraints upon design process. The method has been standardized over many years of human factors research, aimed at improving the usability of computerized systems.

The method requires use of a human factors laboratory, in order to control for the potentially confounding variables in a typical workplace. The use of a lab reduces the ability to generalize findings to the workplace, but the precision of procedure and subsequent clarity of analysis more than compensate. Subsequent

in situ measurement can be useful, however, in fine tuning results to take into account contextual, socio-technical factors.

First, users are surveyed to assess individual differences relevant to computer use, such as experience in the task domain and experience with computerized design tools. Surveys also establish users' attitudes about and perceptions of constraint as it is encountered in the creation of work products.

Next, users are observed in the lab as they perform carefully specified scenarios with a given tool. These scenarios involve the design of artifacts typical of those created in the workplace. The observation includes automated monitoring of all keystrokes, mouse-clicks and system states, plus the recording of monitor displays. It also involves audio- and videotaping of user utterances and behavior.

Finally, users are debriefed to help determine reasons for specific behaviors noted during the observation.

During data analysis, transcripts from audio recordings are subjected to content analysis. Time on task and error rate data are used to identify portions of video recordings of interest, which are examined to categorize behaviors. Survey data and audio recordings are used to construct user mental models. Finally, metrics specific to the task domain are used to measure work product quality and user productivity. User decisions are mapped into a semantic network, to better understand design process and the impact of specific constraints.

5. OUTCOMES

Findings about the impact of constraints upon user behavior can be used to calibrate the means that computerized tools employ to guide the design process. An improved understanding of design decision-making can increase the ability of such tools to improve user productivity while protecting work product quality. Behavioral data in the context of constraint may even be useful in engineering adaptive tools that accommodate users' decision-making styles, experience levels, and expectations. The improvement of user satisfaction with their tools also is a secondary, but not insignificant, potential outcome.

REFERENCES

DARSES, F., 1991, The constraint satisfaction approach to design: a psychological investigation. *Acta Psychologica*, 78, 307–25.

DAVIS, F., 1993, User acceptance of information technology: system characteristics, user perceptions and behavioral impacts. *International Journal of Man-Machine Studies*, 38, 475–87.

JANKOWSKI, D., 1997, Computeer-aided systems engineering methodology support and its effects on the output of structured analysis. *Empirical Software Engineering*, 2, 11–38.

SCOTT, L., HORVATH, L., and DAY, D., 1999, Characterising CASE constraints. Forthcoming in *Communications of the ACM*.

SHERIDAN, T., 1980, Computer control and human alienation. *Technology Review*, 83(1), 65–7, 71–3.

SILVER, M., 1991, *Systems that Support Decision Makers* (Chichester, UK: Wiley).

WAERN, K.-G., 1988, Cognitive aspects of computer-aided design. In Helander (ed.) *Handbook of Human-Computer Interaction* (Amsterdam: North-Holland), 701–8.

Cross-cultural Comparison of Computer Anxiety: Concept, Measures and Related Variables

B. Han

Institute of Psychology, Chinese Academy of Sciences, Beijing 100101, China

1. INTRODUCTION: CA AND RELATED CONCEPTS

Computer anxiety (CA) was studied > 20 years ago. Whether it is called "CA," "technostress," "technophobia" or "computerphobia," it all refers to an aversion to computers and computer technology. Estimates indicate that as many as one of three adults suffers from the condition, and that this figure will not vanish with an increasing familiarity with computers (Lalomia and Sidowski 1993). Bradley and Russell (1997) identified three distinctive types of CA: damage anxiety, task anxiety and social anxiety. Extreme manifestations of CA accompanied by physiological reactions can be classified as computerphobia, which is overtly or covertly manifested in behaviors of rejection and avoidance, ranging from a simple unwillingness to work with computers to the act of vandalism (Dorinina 1995).

CA substantially reduced the effectiveness of mastering and using computers. Failure to get computer literacy (CL) may become a "critical filter" for those who will, therefore, not have access to many careers. As the most important predictor of computer performance (CP; Rosen and Weil 1995), CA reminds educators of its presence and impact. Studies on CA can also provide instruction to the marketing of technology production from the consumers' perspective, because "ownership of a technological device does not necessarily guarantee its use." Grouping of the consumers can be very helpful to technology producer.

Studies on CA in the 1980s were well reviewed by Rosen and Maguire (1990), Lalomia and Sidowski (1993) and Maurer (1994) from different perspectives, but there is no review on the research conducted in the 1990s, especially cross-cultural ones. This chapter will emphasize the measures developed and validated in different cultural settings; other related studies will also be reviewed briefly.

2. MEASURES: FACTOR STRUCTURE, VALIDITY AND VERIFICATION

Dozens of constructed measures have been developed, but only some have been validated with different subject groups, and the factor structure varied between measures (see Lalomia and Sidowski 1993 for a review of 12 scales developed in 1980s). Rosen and Maguire (1990) included 79 empirical studies (with 66 different measures) in their meta-analysis. Only five measures have been used by more than one researcher. Of these, only three have reported data concerning their factor structure. Only two studies have compared the factor structure of CA between groups. Cross-cultural validation of developed measure is still very rare to date. Given that the concept of computer technology is changing, some new items (e.g. Internet, information highway or national information infrastructure) should be included in a scale for assessing subjects' living in today's environment. Therefore, factors extracted from measures developed 10 years ago may not be valid at the present time. In addition, the changing trend may be different between countries.

Intercorrelations of scales purporting to measure CA have shown that CA is a very robust concept, and its various operational definitions (measures) exhibit a high degree of convergent validity (Gardner et al. 1993) and concurrent validity (Harrison and Rainer 1992). However, some scales measure many dimensions; others assess one more than the other. Studies on the predictive validity of CA on computer usage (CU) or CP are inconsistent (Szajna 1994), despite the demonstrated convergent validity of the CA construct. One possible reason of the conflict is the different factor structure between measures; another is the difference between subjects groups (e.g. culture, education and CL). The effect of CA on academic achievement in a computer course can also probably be overcome by other factors (e.g. motivation of students, good instruction, etc.; Maurer 1994).

3. RELATED VARIABLES

It is well documented that CA correlated negatively with CU and CP, but the effects of related variables were mixed. Although females have less positive attitudes, prior computer experience (CE) and access to computer than males (Maurer 1994), gender in general was not a predictor of CA (Anderson 1996). Older adults have more positive attitudes and less CA than young adults, and the factor structure was similar for younger and older adults (Dyck et al. 1998).

Education major students had significantly higher CA than computer and business major students (Maurer 1994). Rosen and Weil (1996) found that the CA in clinical psychologists, elementary school and secondary school teachers is higher than it in university students. Cognitive style is an important determinant of CA, with intuitive and thinking individuals exhibiting lower anxiety then their sensing and feeling counterparts (Shermis and Lombard 1998).

CE is the most prominent predictor of technophobia among many others (Anderson 1996). Qualitative (Todman and Monaghan 1994) and quantitative (Charness et al. 1992) differences between CE will lead to different levels of CA. The character of the initial stage of the person's interaction with the computer is a key factor in the etiology of CA (Dorinina 1995). Quality of prior CE is beyond simple exposure to computers in the forming of CA (see Chorpita and Barlow 1998 for a review). Hemby (1998) found that gender, keyboarding skill, age, socio-economic status and self-directedness are adequate predictors of CA in business communication students.

Tseng et al. (1998) found that self-rating of mood measured by the three methods (pen-based personal digital assistants, conventional computer and paper assessment) co-varied divergently with measures of CA and private self-consciousness. The attitude toward computers (ATC) is related but different with CA. ATC measures were developed based on a three-component (behavioral, affective, cognitive) model of attitudes (Charlton and Birkett 1995), or were concerned solely with the affective domain (Evans et al. 1995). There is only a small overlap between computerphobia and other anxiety measures (e.g. mathematical,

state, trait, test), but they are a separated construct (Shermis and Lombard 1998), and these relationships rarely account for > 10% of the variance in predicting computerphobia (Rosen and Maguire 1990).

4. CROSS-CULTURAL DIFFERENCE

Most of the studies on CA are based on subjects from Western countries; only several studies have assessed CA or ATC across countries (see Rosen and Weil 1995, Weil and Rosen 1995 for a short review), and four studies compared Chinese subjects with their Western counterparts directly. Do the conclusion and treatment in these studies stay valid in other cultural settings? Is there any difference for the occurrence and effect of CA on performance? Given that there exist differences in educational systems and technological policies between countries, one can predict some cross-cultural difference. Despite the popularity of CA and technology improvement everywhere, cross-cultural study on CA was rare and the findings were mixed.

Allwood and Wang (1990) showed some differences between college students majoring in Psychology and Computer Science in their conceptions of computers, with the Chinese students showing a more optimistic view of the future impact of technology than Swedish students. Collis and Williams (1987) found that the Chinese students (8th and 12th grade) were more positive in their ATC than their Canadian counterparts.

Marcoulides (1989, 1991) and Marcoulides and Wang (1990) proved the factor invariability of their Computer Attitude Scale (CAS). They found that the Chinese and American college students have similar degree of CA and structural component, named "general CA" and "computer equipment anxiety."

Rosen, Sears and Weil (1992) developed measures for diagnosing the technophobic, Computer Anxiety Rating Scale (CARS), Computer Thought Survey (CTS) and General Attitudes toward Computers Scale (GATCS). Each subscale has 20 items. The three factors — interactive computer learning anxiety, consumer technology anxiety, observational computer learning anxiety — for their CARS seemed not valid for subjects from other cultural settings.

Han *et al.* (1999) administered Rosen, Sears and Weil's (1992) CARS to 126 Chinese college students majored in English (grade 2) and found five factors interpretable as interactive computer learning anxiety, consumer technology anxiety, observational computer learning anxiety, general CA and computer equipment anxiety. The mean score of Chinese students is 54.27 (SD 14.41). It is higher than the American students' score of 41.46 (14.25). The difference of factor structure and mean score between Chinese and American college students showed the cultural effect on CA.

Comparison between 10 countries (Australia, Czechoslovakia, Germany, Hungary, Israel, Italy, Japan, Spain, Yugoslavia) showed that each possessed an unique culture-dependent model of CA (Rosen and Weil 1995). People in those countries (e.g. Singapore and Israel) where governments value technology very high, and where the computer was introduced by confident, well-trained teachers, were very comfortable with computers and technology. Of Japanese university students, 60% was tested as technophobic, no matter the popularity of advanced technical products in their environment. The lack of an early infusion of technology could be the possible reason since the Japanese government decided not to integrate computers at the primary level.

In a study involving 23 countries (USA, Yugoslavia, Thailand, Spain, Singapore, Saudi Arabia, Poland, North Ireland, Mexico, Kenya, Japan, Italy, Israel, Indonesia, India, Hungary, Greece, Germany, Egypt, Czechoslovakia, Belgium, Australia, Argentina), Weil and Rosen (1995) found that many countries showed a majority of technophobic students while others showed very few technophobes. CE was negatively related to technophobia in the majority of country samples. A Discriminant Function Analysis measure and a composite technophobia score were sufficient to provide maximal discrimination between the 23 country samples.

Dyck *et al.* (1998) administered the CAS (Marcoulides 1989) to 311 younger and 324 older subjects, and found that the construct of computer anxiety may have changed, since they found two factors (direct and indirect involvement with computer). These were different from Marcoulides' (1989) two factors (general CA and equipment anxiety), but similar to the first two factors of Rosen and Weil (1995) (interactive computer learning anxiety and observational computer learning anxiety), while items consisting of "consumer technology anxiety" factor were not included in their study.

Nine thousand people from eight countries (France, Germany, England, Italy, Norway, Spain, USA, Japan) were surveyed for their opinions concerning computers, with particular emphasis on the impact of technology in the workplace (Vine 1985). "There are very great differences in country attitudes toward information technology. These differences reflect a whole range of cultural attitudes and the data must be interpreted with that in mind" (cited by Rosen and Weil 1995: 46). Omar (1992) indicated that the US college students' ATC were more positive than those of Kuwaiti students.

Combined with other mixed features concerning CA and ATC from other studies, further study needed. There were several other studies conducted in countries other than the USA (see Weil and Rosen 1995 for a short review), but the results of these investigations do not permit direct cross-cultural comparisons due to the use of different measurement instruments and different research designs.

5. THEORY AND INTERVENTION

In overcoming the negative effect of CA and, even more importantly, preventing it from developing, specialized user training is essential. Several models were proposed to explain the phenomenon and to develop the intervention program (see Rosen, Sears and Weil 1992 for a short review, Maurer 1994, Todman and Monaghan 1994). Treatment should be in accordance with the different type and degree of CA in the population or culture. Individual computer training has proved quite effective in the prevention and cure of computer discomfort. Schoolteachers may be the ones with the power to change that trend, because the attitude of the technology "introducer" was important in predicting later psychological reactions to computers. Individuals' computerphobia problems can be resolved within time (from 5 h to 2 years) and users can approach technology in a confident manner (Russell and Bradley 1996). Programming resulted in the lowest degree of CA and increased problem-solving (Liu 1997) compared with general CL instruction.

6. DISCUSSION: CROSS-CULTURAL STUDY ON CA

Most of studies focused on the diagnostic issues of CA (measurement development and effect of related variables investigation); others focused on the intervention in practice, while very few have proposed theoretical models. Unfortunately, "much of the related studies is significantly flawed and making it difficult to support any particular claim" (Maurer 1994: 369).

Fear of technology has been a human concern for hundreds of years, and discourse on this has been firmly embedded in a culture that simultaneously embraces and rejects machines. Technophobia is not only a consequence of the competition of machines, but also is a cultural construct. Such a kind of fear was apparently improved given the much higher capacity and intelligence of the modern computer. A fear of computers has complex cultural and historical antecedents. When students enter a course of teacher training they bring with them a set of values and attitudes towards technology that will partly determine their approach to computers in education. Only through an examination of cultural and historical antecedents concerning fears of technology can student teachers' anxiety about computers be properly understood (Russell and Bradley 1996). Teaching the myths can teach CA to people, although the validity of common myths may varied in different culture (Rosen and Maguire 1990, Yeamen 1993).

Wording changes can make significant differences in the factor structure of scales and item validities. This should be investigated under more subject groups for checking the stability of earlier inference and conclusion, and for providing reliable measures for the future study. Strict procedure (double blind and back translation) should be applied when translating measures to another language for cross-cultural comparison.

CA is definitely implicated in performance (Anderson 1996). The extent to which CE and CA influence one's willingness to use a computer may depend upon the task to be accomplished with the computer. For the sake of comparability between studies and to stabilize the conclusion, on one hand a more clarified definition between the characteristics of task and required skills should be given, and on the other hand the possible effective but uncontrolled variables should be given more attention.

Nearly every segment of society is at risk for computerphobia. This were proved by the fact that technophobic students were less interested than non-phobic students in any form of computer interaction, both in their personal lives and in their academic careers, despite their age (old/young), ethnic (Asian/black/Hispanic/white), major (social science/humanities/business/science) and experience (have/have not). Most of the CA studies were conducted in Western countries, especially in the USA. Many studies used a certain and often small number of subjects, typically those enrolled for a psychology course. This induced less control to the variables like gender, age, education, academic major, profession or culture. It is obvious that the generalization of cross-cultural comparison can only based on a certain number of subjects that can represent the population in a given culture with respect to both demographic and cultural variables.

More carefully defined research questions and treatment should be used to answer larger questions (Maurer 1994). The differences between results imply the need to explore all these variables within a broader model encompassing a range of computer variables, rather than by studying each variable in isolation (Levine and Donitsa-Schmidt 1998).

This chapter is one part of an extensive review on studies of CA in the 1990s, which is beyond the limited space here. Further study needs to be made on the theoretical background with respect to the related variables and mechanisms, at both personality and cross-cultural level. Cultural specific intervention or coping strategy to CA should be developed. The fast population of NII and related computer technology may have an effect on CA by affecting experience, education, cognition and CU, etc.

7. CONCLUSION

As a phenomenon with as high as 33% morbidity rate in the population, CA attracted attention of researchers. Two decades of studies have accumulated findings that have led to a clearer understanding about CA and its effect on CU than ever before. The quality of CE and introducers' ATC are important to current CA. A treatment program has been developed and practiced successfully for > 10 years, but only in Western cultural settings. Given the more and more popular of CU at home, class, office (either intuitively or passively), CL is becoming a basic requirement for today's job seeking. The CA (or technophobia more generally) will be the key issues among educators, employers and business people in all cultural settings. However, systematic developing and validating of measures through carefully manipulating demographic and cultural variables are rare.

ACKNOWLEDGEMENT

REFERENCES

ALLWOOD, C.M. and WANG, Z., 1990, Conceptions of computes among students in China and Sweden. *Computers in Human Behavior*, 6, 185–99.

ANDERSON, A.A., 1996, Predictors of computer anxiety and performance in information systems. *Computers in Human Behavior*, 12, 61–77.

BRADLEY, G. and RUSSELL, G., 1997, Computer experience, school support and computer anxieties. *Educational Psychology*, 17, 267–84.

CHARLTON, J.P. and BIRKETT, P.E., 1995, The development and validation of the Computer Apathy and Anxiety Scale. *Journal of Educational Computing Research*, 13, 41–59.

CHARNESS, N., SCHUMANN, C.E. and BORITZ, G.M., 1992, Training older adults in word processing: Effects of age, training technique, and computer anxiety. *International Journal of Technology and Aging*, 5, 79–106.

CHORPITA, B.F. and BARLOW, D.H., 1998, The development of anxiety: the role of control in the early environment. *Psychological Bulletin*, 124, 3–21.

COLLIS, B.A. and WILLIAMS, R.L., 1987, Cross-cultural comparison of gender differences in adolescents' attitudes toward computers and selected school subjects. *Journal of Educational Research*, 81, 17–27.

DORININA, O.V., 1995, Fear of computers. *Russian Social Science Review*, 36, 79–96.

DYCK, J.L., GEE, N.R. and SMITHER, J.A.A., 1998, The changing construct of computer anxiety for younger and older adults. *Computers in Human Behavior*, 14, 61–77.

EVANS, T.E., KATZ, Y.J. and FRANCIS, L.J., 1995, Psychometric properties of the Hebrew version of the Francis Attitude Toward Computers Scale. *Psychological Reports*, 77, 1003–10.

GARDNER, D.G., DISCENZA, R. and DUKES, R.L., 1993, The measurement of computer attitudes: An empirical comparison of available scales. *Journal of Educational Computing Research*, 9, 487–507.

HAN, B.X., YING, X.P. and LUO, J., 1999, Technophobia in Chinese college students. Unpublished research manual.

HARRISON, A.W. and RAINER, R.K., 1992, An examination of the factor structures and concurrent validities for the Computer Attitude Scale, the Computer Anxiety Rating Scale, and the Computer-Self-Efficacy Scale. *Educational and Psychological Measurement*, 52, 735–45.

HEMBY, K.V., 1998, Predicting computer anxiety in the business communication classroom *Journal of Business and Technical Communication*, 12, 89–109.

LALOMIA, M.J. and SIDOWSKI, J.B., 1993, Measurements of computer anxiety: a review. *International Journal of Human Computer Interaction*, 5, 239–66.

LEVINE, T. and DONITSA-SCHMIDT, 1998, Computer use, confidence, attitudes, and knowledge: a causal analysis. *Computers in Human Behavior*, 14, 125–46.

LIU, M., 1997, The effects of HyperCard programming on teachers education. *Journal of Research on Computing in Education*, 29, 248–63.

MARCOULIDES, G.A., 1989, Measuring computer anxiety: the Computer Anxiety Scale. *Educational and Psychological Measurement*, 49, 733–9.

MARCOULIDES, G.A., 1991, An examination of cross-cultural differences toward computers. *Computers in Human Behavior*, 7, 281–9.

MARCOULIDES, G.A. and WANG, X., 1990, A cross-cultural comparison of computer anxiety in college students. *Journal of Educational Computing Research*, 6, 251–63.

MAURER, M.M., 1994, Computer anxiety correlates and what they tell us: a literature review. *Computers in Human Behavior*, 10, 369–76.

OMAR, M.H., 1992, Attitudes of college students towards computers: a comparative study in the United States and the Middle East. *Computers in Human Behavior*, 8, 249–57.

ROSEN, L.D. and MAGUIRE, P.D., 1990, Myths and realities of computerphobia: a meta-analysis. *Anxiety Research*, 3, 175–91.

ROSEN, L.D., SEARS, D.C. and WEIL, M.M., 1992, *Computerphobia Measurement: A Manual for the Administration and Scoring of Three Instruments: Computer Anxiety Rating Scale (Form C), Attitudes Toward Computers Scale (Form C), and Computer Thoughts Survey (Form C)* (Carson: California State University).

ROSEN, L.D. and WEIL, M.M., 1995, Computer anxiety: a cross-cultural comparison of university students in ten countries. *Computer in Human Behavior*, 11, 45–64.

ROSEN, L.D. and WEIL, M.M., 1996, Easing the transition from paper to computer-based systems. In T. Trabin (ed.) *The Computerization of Behavioral Health* (San Francisco: Jossey Bass), 87–107.

RUSSELL, G. and BRADLEY, G., 1996, Computer anxiety and student teachers. *Asia-Pacific Journal of Teacher Education*, 24, 245–58.

SHERMIS, M.D. and LOMBARD, D., 1998, Effects of computer-based test administrations on test anxiety and performance. *Computers in Human Behavior*, 14, 111–23.

SZAJNA, B., 1994, An investigation of the predictive validity of computer anxiety and computer aptitude. *Educational and Psychological Measurement*, 54, 926–34.

TODMAN, J. and MONAGHAN, E., 1994, Qualitative differences in computer experience, computer anxiety, and students' use of computers: a path model. *Computers in Human Behavior*, 10, 529–39.

TSENG, H.M., TIPLADY, B., MACLEOD, H.A. and WRIGHT, P., 1998, Computer anxiety: a comparison of pen-based personal digital assistants, conventional computer and paper assessment of mood and performance. *British Journal of Psychology*, 89, 599–611.

WEIL, M.M. and ROSEN, L.D., 1995, The psychological impact of technology from a global perspective: A study of technological sophistication and technophobia in university students from twenty-three countries. *Computers in Human Behavior*, 11, 95–133.

YEAMAN, A.R.J., 1993, Whose technology is it, anyway? [cover story]. *Education Digest*, 58, 19–24.

Cybersickness in Virtual Reality

K. M. Stanney† and R. S. Kennedy‡

†Department of Industrial Engineering & Management Systems, University of Central Florida, Orlando, FL 32816, USA

‡RSK Assessments Inc. 1040 Woodcock Road, #227, Orlando, FL 32803, USA

1. INTRODUCTION

Motion sickness is familiar to most individuals. A common experience while sitting in a car or plane is suddenly to feel like you are moving as the vehicle beside your stationary plane or car starts backing out. This can lead to a somersault sensation in the stomach or to a sense of postural instability. Another familiar problem is for a person to experience, at times being severe and long-lasting, sickness when attempting to read in a car. While the consequences of these two experiences are oftentimes similar sensations of motion sickness, this common outcome is rendered by quite different sensory conflicts. In the first instance the vestibular system senses that the person is physically stationary but the visual system senses motion as the adjacent vehicle moves. In the second instance the vestibular system senses that the person is physically moving in the vehicle while the visual system indicates non-motion as the stationary reading matter is viewed. The motion sickness experienced in virtual environments (VE) can be of either type but is more commonly of the former, where the visual system senses motion without corresponding stimulation of the vestibular system. More specifically, when one dons a head-mounted device, one is often inundated with a plethora of visual stimuli that indicate motion throughout a virtual world. Yet, there is oftentimes a lack of concordant physical motion to stimulate the vestibular system. Thus, a sensory conflict occurs between what one sees as motion and what one feels as non-motion. Such sensory conflicts are thought to lead to motion sickness. This sickness can be of many forms, including cerebral (e.g. headache), gastrointestinal (e.g. nausea, burping, emesis), psychological (e.g. anxiety, depression, apathy) and other less characteristic indicators of motion sickness such as fullness of the head (Kennedy *et al.* 1993).

2. TYPES OF MOTION SICKNESS

Sensory conflicts have been used formally to classify motion sickness into three main types (Reason 1978). The first presents conflicting signals A and B together, such as if a VE system were to be used on a ship for training or entertainment purposes. In this case the true bodily motions as influenced by the ship would not correspond to the virtual motion of the individual throughout the virtual world. The second type of sickness occurs when signal A occurs in the absence of signal B and was exemplified above by the airplane scenario. The third type involves signal B occurring in the absence of signal A, of which the car reading example was representative. This type of mismatch could occur in a VE that had a motion-base where a user was looking at instrument panels (e.g. "flying" on instruments only) and had no visual motion

stimulation to correspond to the vestibular stimulation provided by the motion-base.

Why do these sensory conflicts lead to motion sickness? Reason (1978) suggests that the conflicts that arise between the sensory signals of the body, eye and image as compared with those conditioned by past experiences are an inherent characteristic of all forms of motion sickness. The degree of sickness experienced is generally proportional to the severity of the sensory discordance. Through continued exposure to, and particularly interaction with, a displaced environment the plasticity of the human sensory system allows humans to adapt to the new stimuli. This adaptation will then allow the individual to perform effectively in the displaced environment. This can, in turn, lead to after effects when an individual returns to normal conditions, such as on terminating exposure to a virtual environment. In this case the individual, whose functioning was modified to accommodate the virtual world, may be compromised for a period of time when returning to the "real" world. These after effects, which can take the form of balance and visual disturbances or displaced hand–eye coordination, can present safety concerns.

3. INDIVIDUAL AND TECHNOLOGICAL FACTORS INFLUENCING MOTION SICKNESS

There is concern that continued development of VE technology may be compromised by the presence of these maladies and safety concerns, often referred to as "cybersickness". As much as 95% of those exposed to VE systems experience some level of discomfort and ~15% of those exposed must curtail their exposure, some to as little as 10 min. While there is no definitive determinant of who will become ill, age, gender, motion sickness histories, prior experience and individual factors have been shown to be useful in identifying persons who are susceptible to provocative motion environments (Stanney *et al.* 1998). With the high incidence rate and diversity of individual factors contributing to cybersickness, the practicality of screening for susceptibility is compromised. Thus, developers often look to technological solutions. Means of assessing the strength of a VE stimulus (i.e. how likely it is to cause cybersickness and after effects) are currently not available but there are several contributing factors that have been identified including update rate, lag, system consistency, mismatched interpupillary distance, field-of-view and spatial frequency content. While the contribution of many of these variables to cybersickness has been evaluated, to date there are no definitive technological solutions.

4. MEASURING CYBERSICKNESS

Systematic means of measuring cybersickness are needed to facilitate the development of technological or adaptive countermeasures (Stanney *et al.* 1998). It is important for developers of VE systems to determine if there will be after effects from the use of their systems and what percentage of the user population will be affected. The most common form of measurement is subjective self-report of symptomatology after exposure. Another form is objective measurement of physiological adaptation of the vestibular, visual and proprioceptive systems.

4.1. Subjective Measures

The Simulator Sickness Questionnaire (SSQ) (Kennedy *et al.* 1993) is the most commonly used tool to assess subjective symp-

tomatology from VE exposure. The SSQ consists of a checklist of 26 symptoms, each of which is designated in terms of degree of severity (none, slight, moderate, severe), with the highest possible total score being 300. A weighted scoring procedure is used to obtain a global index, known as the Total Severity (TS) score, which reflects the overall total discomfort level, as well as scores along three subscales (i.e. nausea, oculomotor disturbances, disorientation). By means of these subscales cybersickness can be more fully characterized. The average SSQ score for HMD-based VE systems is ~29 (with a range over several different VE systems of 19–55) on the TS scale (Stanney *et al.* 1998). In general, VE users report nearly 2.5 times more sickness after exposure than do users of military flight simulators and nearly twice more than astronauts during space travel. The first impulse might be to attribute these severity differences solely to susceptibility differences between the select military and space flight individuals as compared with the general population. An examination of the profile of SSQ symptoms engendered by these systems, however, indicate some interesting differences in symptomatology (Stanney *et al.* 1998). VE systems using HMD generally tend to produce more disorientation (D) than neurovegative (N) symptoms, and least of oculomotor-related (O) disturbances. This D > N > O profile does not match the profiles of either simulators (i.e. O > N > D) or space flight (i.e. N > D > O). Thus, VE systems tend to engender both a greater severity, as well as a different profile of symptoms as compared with other provocative environments. Means of managing these ill-effects are thus essential if VE technology is to fulfill its potential.

4.2. Objective Measures

Objective measures of after effects from VE exposure include measuring postural equilibrium (i.e. sway magnitude), adaptation of the vestibulo-ocular reflex (i.e. the ability to stabilize an image on the retina), changes in the kinesthetic position sense (i.e. hand–eye coordination), visual functioning (i.e. visual acuity, heterophoria, near point of convergence, amplitude of accommodation), as well as changes in EEG, salivary cortisol composition, and heart rate levels and variability (Stanney *et al.* 1998). These measurement approaches could be used both to determine if a user has adapted physiologically during VE exposure and to determine at what point the effects have dissipated after exposure. It may also be possible to take several of these measures during VE exposure, and to determine the relative sequence of physical versus subjective effects. This might permit development of on-line measures that could be used to predict and minimize sickness during interaction with a virtual environment.

While sickness as measured by the SSQ and objective measures of after effects are usually statistically increased after exposure, correlations between these measures are generally insignificant (Stanney *et al.* 1998). This implies that to characterize comprehensively cybersickness it is necessary to obtain both types of measures. Once cybersickness can be accurately measured, efforts to manage this malady can be implemented.

5. MANAGING CYBERSICKNESS

Currently, probably the simplest way to control for the problems associated with cybersickness is to institute usage protocols. Both exposure duration and number of repeat exposures have been shown to affect the level of motion sickness experienced by users,

with sickness increasing with duration and decreasing across repeated exposures (Stanney *et al.* 1998). Based on studies in our laboratory we have found that if exposure is limited to < 10 min ~95% of the population should endure exposure. Further, if short duration repeated exposures are provided, with a 2–7-day intersession interval, tolerance should be extended by the fourth to sixth exposure. Unfortunately, repetition may not be an efficient or practical means of adaptation because the time required for re-adaptation to one's normal environment is proportional to the time spent in adapting to the virtual world (Baltzley *et al.* 1989). Thus, if an individual must undergo numerous exposures to a VE to achieve adaptation a considerable amount of time may be necessary for re-calibrating to normal functioning once the exposure ceases. In many operational settings this may not be viable. Also, when used for entertainment purposes users may not have the opportunity for repeated exposure to the same virtual environment. The goal, therefore, should be to achieve adaptation in as little time as possible, while attempting to lessen the severity of the initial sickness experienced by users. This may be achieved through easing user interaction. Means of easing VE interaction include streamlining the degrees of freedom through which VE users can move about the virtual world, blanking the visual stimulus and manipulating image abstractness (Stanney *et al.* 1998). To ease interaction, tasks requiring high rates of linear or rotational acceleration and extraordinary maneuvers (e.g. flying backward) should be avoided during early exposure. Blanking involves graying-out a head-mounted display above a threshold level of head velocity, which selectively eliminates the effect of visual update delays during head movements. The degree of abstractness of virtual images can be manipulated through the use of polygonal versus texture gradients. In general, the greater the abstractness the more severe the sickness experienced. It is important to note that reducing the information content of a VE, either through blanking, manipulating abstractness, or other means, may compromise task performance such as the ability to detect direction of heading, steer effectively or make precise judgments of impeding arrival or collision in VE settings. Any changes in the information content targeted at reducing cybersickness must thus be balanced against removing the information that users would ordinarily rely upon for accurate judgments.

As tolerance is gained through repeated exposures or interaction techniques and exposure duration is extended, after effects may become more prevalent. Thus, before the identification of technological solutions that minimize after effects a means of managing after effects must be identified. The best way to eradicate VE after effects is to require users who have just exited the VE to engage in the same activities performed within the VE, but without the sensory conflicts imposed by the VE (Stanney *et al.* 1998). This should provide users with the opportunity to readapt their functioning to their natural environment by "unlearning" the adaptation they acquired in the VE (see the sensory rearrangement research of Welch 1978).

6. CONCLUSIONS

The cybersickness associated with VE technology is characterized by sickness and discomfort during or after VE exposure, postural, visual and coordination disturbances upon post-exposure, and interference with performance within the VE or on other tasks afterwards. While it is undeniable that some individuals

experience some after effects within some VE systems, it is not necessarily the case that VE technology is implicitly harmful. Careful consideration of individual and technological factors, usage protocols and design of VE interaction techniques may minimize any potential for sickness or harm. The contributions from ergonomics and human factors practitioners will be of particular relevance in addressing the multiplicity of human issues central or related to cybersickness in virtual environments.

REFERENCES

BALTZLEY, D.R., KENNEDY, R.S., BERBAUM, K.S., LILIENTHAL, M.G. and GOWER, D.W., 1989, The time course of postflight simulator sickness symptoms. *Aviation, Space, and Environmental Medicine*, 60, 1043–8.

KENNEDY, R.S., LANE, N.E., BERBAUM, K.S. and LILIENTHAL, M.G., 1993, Simulator sickness questionnaire: An enhanced method for quantifying simulator sickness. *International Journal of Aviation Psychology*, 3, 203–20.

REASON, J.T., 1978, Motion sickness adaptation: a neural mismatch model. *Journal of the Royal Society of Medicine*, 71, 819–29.

STANNEY, K.M., SALVENDY, G., DEISIGNER, J. *et al*., 1998, Aftereffects and sense of presence in virtual environments: Formulation of a research and development agenda. Report sponsored by the Life Sciences Division at NASA Headquarters. *International Journal of Human–Computer Interaction*, 10, 135–87.

WELCH, R.B., 1978, *Perceptual Modification: Adapting to Altered Sensory Environments* (New York: Academic Press).

Databases for Psychophysically Acceptable Maximum Weights and Forces in Manual Handling Tasks Developed by Liberty Mutual

W. Karwowski and R. Jang

Center for Industrial Ergonomics, University of Louisville, Louisville, KY 40292, USA

1. BACKGROUND

Over the past 30 years Dr S. Snook and colleagues have performed numerous studies of human capabilities in the performance of manual handling tasks using the psychophysical approach (Snook and Irvine 1967, Snook *et al.* 1970, Snook 1971, 1978, 1987, Snook and Ciriello 1974, 1991, Ciriello and Snook 1978, 1983). This research resulted in development of the comprehensive databases for evaluation and design of manual handling tasks in terms of the maximum acceptable weights (MAW) and forces for lifting, lowering, pushing, pulling, and carrying tasks.

The databases presented here are based on integration of the results of several experiments that utilized the psychophysical methodology, along with measurements of oxygen consumption, heart rate, and consideration of the human anthropometric characteristics (Snook and Ciriello 1991). All of the experiments were conducted in an environmental chamber, with the dry bulb temperature of 21°C and relative humidity of 45%. The studies employed industrial workers who were instructed to work on an incentive basis, working as hard as they could without straining themselves, or without becoming unusually tired, weakened, overheated or out of breath.

2. SCIENTIFIC BASIS FOR THE DATABASES

The presented databases are based on a sample size of 119 participants (68 males, 51 females) who participated in the experiments for a minimum of 10 weeks. The original tables (Ciriello and Snook 1983) were based on a sample of 57 particpiants (42 males, 15 females).

The presented databases are based on the series of studies addressing the following issues:
- Effect of task frequency on the maximum acceptable weights and forces (Ciriello and Snook 1983).
- Female responses to the object size (width), vertical distance of lift, and height of push/pull (Ciriello and Snook 1983).
- Effects of task duration on the maximum acceptable weights and forces and a combination task, consisting of a lift, carry, and lower, performed separately and in combination (Ciriello *et al.* 1990).

- Effects of using box handles, and lifting with the extended horizontal reach, on the MAW values at different frequency and height of lift (Ciriello *et al.* 1991).
- Effect of different box sizes for carrying tasks, and a longer distance (15.2 m), for the pulling task.

To complete the developed databases several assumptions were made to in order to approximate the values of response variables for the task conditions that have not been studied experimentally (Snook and Ciriello 1991). The databases include six distances for the pulling tasks, and the pushing and pulling frequencies that are the same for both males and females. It was noted that some of the weights and forces in tables exceed recommended physiological criteria when performed continuously for 8 h or more.

According to Snook and Ciriello (1991), the revised databases represent the best estimates of maximum acceptable weights and forces for industrial workers, and are intended to help industrial practitioners in the evaluation and design of manual handling tasks. As such, these databases should contribute to the reduction of disability from low back pain, and help in the evaluation of disabled workers in rehabilitation programs (Snook 1987).

REFERENCES

CIRIELLO, V.M. and SNOOK, S.H., 1978, The effects of size, distance, height, and frequency on manual handling performance. In *Proceedings of the Human Factors Society's 22nd Annual Meeting*, Detroit, MI, 318–22.

CIRIELLO, V.M. and SNOOK, S.H., 1983, A study of size, distance, height, and frequency effects on manual handling tasks, *Human Factors*, **25**, 473–83.

CIRIELLO, V.M., SNOOK, S.H., BLICK, A.C. and WILKINSON, P., 1990, The effects of task duration on psychophysically determined maximum acceptable weights and forces, *Ergonomics*, **33**, 187–200.

CIRIELLO, V.M., SNOOK, S.H. and HUGHES, G.J., 1991, Further studies of psychophysically determined maximum acceptable weights and forces. Unpublished manuscript.

SNOOK, S.H., 1971, The effects of age and physique on continuous work capacity, *Human Factors*, **13**, 467–79.

SNOOK, S.H., 1978, The design of manual handling tasks, *Ergonomics*, **21**, 963–85.

SNOOK, S.H., 1987, Approaches to the control of back pain in industry: job design, job placement, and education/training, *Spine: State of the Art Reviews*, **2**, 45–59.

SNOOK, S.H. and IRVINE, C.H., 1967, Maximum acceptable weight of lift, *American Industrial Hygiene Association Journal*, **28**, 322–9.

SNOOK, S.H. and CIRIELLO, V.M., 1974, Maximum weights and workloads acceptable to female workers, *Journal of Occupational Medicine*, **16**, 527–34.

SNOOK, S.H. and CIRIELLO, V.M., 1991, The design of manual handling tasks: revised tables of maximum acceptable weights and forces, *Ergonomics*, **34**, 1197–213.

SNOOK, S.H., IRVINE, C.H. and BASS, S.F., 1970, Maximum weights and work loads acceptable to male industrial workers, *American Industrial Hygiene Association Journal*, **31**, 579–86.

Table 1. Maximum acceptable weight of lift for males (kg); from Snook and Ciriello (1991).

			Floor level to knuckle height — One lift every								Knuckle height to shoulder height — One lift every								Shoulder height to arm reach — One lift every							
W	D	P	5 s	9 s	14	1	2 min	5	30	8 h	5 s	9 s	14	1	2 min	5	30	8 h	5 s	9 s	14	1	2 min	5	30	8 h
75	76	90	6	7	9	11	13	14	14	17	8	10	12	13	14	14	16	17	6	8	9	10	10	11	12	13
		75	9	11	13	16	19	20	21	24	10	14	16	18	18	19	21	23	8	10	12	14	14	14	16	17
		50	12	15	17	22	25	27	28	32	13	17	20	22	23	24	26	29	10	13	15	17	17	18	20	22
		25	15	18	21	28	31	34	35	41	16	21	24	27	27	28	32	35	11	16	18	21	21	22	24	27
		10	18	22	25	33	37	40	41	48	19	24	28	31	32	33	37	40	14	18	21	24	24	25	28	31
	51	90	6	8	9	12	13	15	15	17	8	11	13	15	15	16	18	19	6	8	9	10	10	12	13	15
		75	9	11	13	17	19	21	22	25	11	15	17	20	20	21	23	25	8	11	12	14	15	16	17	20
		50	13	15	18	23	26	28	29	34	14	19	21	25	25	26	29	32	10	14	16	18	18	20	21	25
		25	16	19	22	29	33	35	36	42	17	23	26	30	31	32	36	39	13	17	19	23	23	24	27	30
		10	19	22	26	34	38	42	43	50	20	26	30	35	36	37	41	45	15	19	22	26	28	29	32	35
	25	90	6	8	9	11	13	15	16	17	10	10	13	15	18	19	21	23	7	10	11	14	14	14	14	16
		75	11	13	15	19	22	24	24	28	13	17	20	23	24	25	27	30	10	13	15	18	19	19	19	21
		50	15	18	21	26	29	32	33	38	17	22	25	30	30	31	35	38	12	16	19	23	24	24	24	27
		25	18	22	26	33	37	40	41	48	20	27	30	36	36	38	42	46	15	20	22	28	28	29	29	33
		10	22	26	31	38	44	47	49	57	23	31	35	42	42	44	49	53	17	23	26	32	33	34	33	36
49	76	90	7	8	10	13	15	16	17	20	8	10	12	13	14	14	16	17	7	9	11	14	14	13	16	16
		75	10	12	14	18	22	24	24	28	10	14	16	18	18	19	21	23	9	11	14	18	18	17	21	21
		50	14	16	19	26	29	32	33	38	14	19	20	22	23	24	26	29	11	15	18	23	23	22	26	26
		25	17	20	24	33	37	40	41	48	17	24	24	27	27	28	32	35	14	18	21	28	28	26	32	31
		10	20	24	28	38	43	47	48	57	20	28	28	31	32	33	37	40	16	21	25	32	33	30	37	36
	51	90	7	9	10	14	16	17	18	20	8	11	13	15	15	16	18	19	7	9	11	14	14	13	16	18
		75	10	13	14	20	23	25	25	30	11	15	17	20	20	21	23	25	9	12	14	18	19	17	21	23
		50	14	17	19	27	30	33	34	41	14	19	21	25	25	26	29	32	11	15	18	23	23	21	26	29
		25	18	21	24	34	38	42	43	50	17	23	26	30	31	32	36	39	14	19	21	28	28	26	32	35
		10	21	25	28	40	45	49	50	59	20	26	30	35	36	37	41	45	16	22	25	33	33	30	37	41
	25	90	8	10	12	16	18	19	20	23	10	13	15	18	18	19	21	23	9	11	12	16	16	17	19	21
		75	12	15	16	23	26	28	29	33	13	17	20	23	24	25	27	30	11	14	16	21	21	22	25	27
		50	16	20	21	30	34	37	38	45	17	22	25	30	30	31	35	38	14	18	21	27	27	28	32	35
		25	21	25	26	38	43	47	48	56	20	27	30	36	36	38	42	46	16	22	25	33	33	34	38	42
		10	24	29	30	45	51	56	57	67	23	31	35	42	42	44	49	53	19	25	29	38	38	40	44	48
34	76	90	8	10	11	15	17	19	19	23	8	11	13	15	15	16	18	19	8	10	12	14	14	15	16	16
		75	12	14	17	22	25	28	28	33	11	15	17	20	20	21	23	25	10	14	16	18	19	19	22	24
		50	16	19	22	30	34	37	39	46	14	19	21	25	25	26	29	32	13	17	20	24	24	25	27	30
		25	20	24	28	37	42	47	47	57	17	23	26	30	31	32	36	39	16	21	24	29	29	30	33	36
		10	24	28	33	44	50	56	58	68	20	26	30	35	36	37	41	45	18	24	28	33	33	34	38	42
	51	90	9	10	12	16	18	20	20	24	9	12	14	17	17	18	20	22	8	11	13	16	16	17	18	20
		75	12	15	18	23	26	29	29	34	12	16	18	22	23	23	26	29	11	14	17	21	21	22	24	26
		50	17	20	24	31	35	39	39	46	15	20	23	28	29	30	33	36	14	18	21	26	26	28	31	34
		25	21	24	30	39	44	49	49	57	18	24	27	34	35	36	40	44	17	22	25	32	31	33	37	41
		10	25	30	35	46	52	58	58	68	21	28	32	40	40	42	46	51	19	26	29	37	37	39	43	47
	25	90	10	12	14	18	20	22	23	27	11	14	16	20	20	21	23	26	10	13	15	19	19	19	22	24
		75	15	18	21	26	30	32	33	38	14	18	21	26	27	28	31	34	13	17	20	24	25	26	29	31
		50	20	24	28	35	40	43	44	52	18	23	27	33	34	35	39	43	16	22	25	31	31	33	36	40
		25	26	30	35	44	50	54	55	65	21	28	32	40	41	42	47	52	20	26	30	37	37	39	44	46
		10	29	35	41	52	59	64	66	76	25	33	37	47	47	49	55	60	23	30	35	43	44	45	51	55

W: Box width (the dimension away from the body) (cm).
D: Vertical distance of lift (cm).
P: Percentage of industrial population.
Italicized values exceed 8 h physiological critiera.

Table 2. Maximum acceptable weight of lift for females (kg); from Snook and Ciriello (1991).

W	D	P	Floor level to knuckle height — One lift every								Knuckle height to shoulder height — One lift every								Shoulder height to arm reach — One lift every							
			5	9 s	14	1	2 min	5	30	8 h	5	9 s	14	1	2 min	5	30	8 h	5	9 s	14	1	2 min	5	30	8 h
75	76	90	5	6	7	7	8	8	9	12	5	6	7	9	9	9	10	12	4	5	5	6	7	7	7	8
		75	7	8	9	9	10	10	11	14	6	7	8	10	11	11	12	14	5	6	6	7	8	8	8	10
		50	8	10	10	11	12	12	13	17	7	8	9	11	12	12	13	16	6	7	7	8	9	9	10	11
		25	9	11	12	13	14	14	15	21	8	9	10	13	14	14	15	18	7	7	8	9	10	10	11	13
		10	11	13	14	14	15	16	17	23	9	10	11	14	15	15	17	20	7	8	9	10	11	11	12	14
	51	90	6	7	8	8	9	9	10	14	6	7	8	9	10	10	11	13	5	6	7	7	7	7	8	9
		75	7	9	9	10	11	11	13	17	7	8	9	11	12	12	13	15	5	6	8	8	9	9	9	11
		50	9	10	11	12	13	14	15	21	9	9	11	13	14	14	15	17	7	8	9	9	10	10	11	13
		25	10	12	13	15	16	16	18	24	10	11	12	14	16	16	17	20	8	9	10	10	11	11	12	14
		10	11	14	15	17	18	18	20	27	11	12	14	16	17	17	19	22	9	10	11	12	13	13	14	16
	25	90	6	8	8	9	9	9	11	14	6	7	8	10	11	11	12	14	5	6	7	8	8	8	9	10
		75	8	10	11	11	12	12	13	18	7	8	9	12	13	13	14	17	6	7	8	9	9	9	10	12
		50	10	12	13	13	14	14	16	21	9	10	11	14	15	15	16	19	7	8	9	10	11	11	12	14
		25	11	14	15	15	16	17	19	25	10	11	12	16	17	17	19	22	8	9	10	12	12	12	14	16
		10	13	16	17	17	19	19	21	29	11	12	14	18	19	19	21	24	9	10	11	13	14	14	15	17
49	76	90	5	6	7	8	8	8	9	13	5	6	7	9	9	9	10	12	4	5	5	7	7	7	8	9
		75	7	8	9	10	10	10	12	16	6	7	8	11	11	11	12	14	5	6	6	8	8	8	9	11
		50	8	10	10	12	12	13	14	19	7	8	9	13	14	14	14	16	6	7	7	9	10	10	11	12
		25	9	11	12	14	15	15	17	22	8	9	11	14	15	15	16	18	7	7	8	10	11	11	12	14
		10	11	13	14	15	17	17	19	25	9	10	12	16	17	17	18	20	7	8	9	11	12	12	13	15
	51	90	6	7	8	9	10	10	11	15	6	7	8	11	10	10	11	13	5	6	7	8	8	8	9	10
		75	7	9	9	11	12	12	14	18	7	8	9	13	12	12	13	15	6	7	8	9	9	9	10	12
		50	9	10	11	13	15	15	16	22	9	9	11	14	14	14	15	17	7	8	9	10	11	11	12	14
		25	10	12	13	16	17	17	19	26	10	11	12	16	16	16	17	20	8	9	10	11	12	12	13	15
		10	11	14	15	18	19	20	22	30	11	12	14	18	17	17	19	22	9	10	11	13	14	14	15	17
	25	90	6	8	8	9	10	10	11	15	6	7	8	10	11	11	12	14	5	6	7	8	9	9	10	11
		75	8	10	10	12	12	13	14	19	7	8	9	12	13	13	14	17	6	7	8	9	10	10	11	13
		50	10	12	12	14	15	15	17	23	9	10	11	14	15	15	16	19	7	8	9	11	12	12	13	15
		25	11	14	15	16	18	18	20	27	10	11	12	16	17	17	19	22	8	9	10	12	13	13	15	17
		10	13	16	16	19	20	21	23	31	11	12	14	19	19	19	21	24	9	10	11	14	15	15	16	19
34	76	90	7	8	9	9	10	10	11	18	6	7	8	9	10	10	11	13	5	6	7	8	8	8	9	11
		75	9	10	11	12	12	13	14	22	7	8	9	11	12	12	13	15	6	7	8	9	10	10	11	13
		50	11	12	13	14	15	15	16	27	9	9	11	13	14	14	15	17	7	8	9	11	12	12	13	15
		25	13	14	14	17	18	18	20	32	10	11	12	14	16	16	17	20	8	9	10	12	13	13	15	17
		10	14	16	17	19	20	21	23	36	11	12	14	16	18	18	19	22	9	10	11	14	15	15	16	19
	51	90	7	9	9	11	12	12	13	18	8	8	9	10	11	11	12	14	7	7	8	9	10	10	11	12
		75	9	11	12	14	15	15	16	22	9	10	11	12	13	13	14	17	8	8	9	11	11	11	12	14
		50	11	13	14	16	18	18	20	27	10	11	13	15	15	15	17	19	9	10	10	12	13	13	14	16
		25	13	15	17	19	21	21	24	32	12	13	14	17	17	17	19	22	10	11	12	14	15	15	16	19
		10	14	18	19	22	24	25	27	36	13	14	16	19	19	19	21	24	11	12	14	15	17	17	18	21
	25	90	8	10	11	11	12	12	14	19	8	8	9	12	12	12	14	16	7	7	8	10	11	11	12	14
		75	10	12	13	14	15	15	17	23	9	10	11	13	14	14	16	18	8	8	9	12	12	12	14	16
		50	12	15	16	17	18	19	21	28	10	11	13	16	17	17	18	21	9	10	11	13	14	14	16	18
		25	14	17	19	20	22	22	24	33	12	13	14	18	19	19	21	24	10	11	12	15	16	16	18	21
		10	16	20	21	23	25	25	28	38	13	14	16	19	21	21	23	27	11	12	14	17	18	18	20	23

W: Box width (the dimension away from the body) (cm).
D: Vertical distance of lift (cm).
P: Percentage of industrial population.
Italicized values exceed 8 h physiological critiera.

Table 3. Maximum acceptable weight of lower for females (kg); from Snook and Ciriello (1991).

W	D	P	Floor level to knuckle height One lift every								Knuckle height to shoulder height One lift every								Shoulder height to arm reach One lift every							
			5	9 s	14	1	2 min	5	30	8 h	5	9 s	14	1	2 min	5	30	8 h	5	9 s	14	1	2 min	5	30	8 h
75	76	90	7	9	10	12	14	15	16	20	10	11	14	14	15	15	16	19	6	7	9	9	10	10	11	13
		75	10	13	14	18	20	22	22	29	13	16	18	18	21	21	21	26	9	10	12	12	14	14	14	18
		50	14	17	19	23	27	29	30	38	18	20	24	24	27	27	28	34	11	13	15	16	18	18	19	23
		25	17	21	24	29	33	36	37	47	21	25	29	29	34	34	34	42	14	16	19	20	23	23	23	28
		10	20	25	28	34	39	42	44	56	25	29	34	34	39	39	39	49	16	19	22	23	26	26	27	33
	51	90	8	10	11	13	15	16	17	21	11	12	15	15	17	17	18	22	7	8	9	10	12	12	12	15
		75	11	14	15	18	21	23	23	30	14	17	21	21	24	24	24	30	9	11	13	14	16	16	16	20
		50	14	18	20	24	28	30	31	40	19	21	25	27	31	31	31	38	12	14	16	18	21	21	21	26
		25	18	22	25	30	34	37	39	49	23	26	31	33	38	38	38	47	15	17	20	22	25	25	26	32
		10	21	26	29	36	41	44	46	58	27	31	36	38	44	44	44	55	17	20	24	26	30	30	30	37
	25	90	9	11	12	15	17	18	19	24	12	14	17	18	21	21	21	26	8	9	11	12	14	14	14	17
		75	13	16	17	21	24	25	26	34	17	20	23	24	28	28	28	35	11	13	15	16	19	19	19	24
		50	17	21	23	27	31	34	35	45	22	25	30	32	36	36	37	45	14	16	19	21	24	24	25	31
		25	21	26	29	34	39	42	44	56	27	31	37	39	44	44	45	56	17	20	24	26	30	30	30	38
		10	24	31	34	40	46	49	51	66	31	36	43	45	52	52	52	65	20	23	28	30	35	35	35	44
49	76	90	8	10	11	15	17	18	19	24	10	11	14	14	15	15	16	19	7	8	10	11	12	12	12	15
		75	12	15	16	21	24	26	28	4	13	16	18	18	21	21	21	26	10	11	14	15	17	17	17	21
		50	15	19	21	27	31	35	35	45	18	20	24	24	27	27	28	34	13	15	17	19	22	22	22	27
		25	19	24	26	34	39	42	44	56	21	25	29	29	34	34	34	42	16	18	21	23	27	27	27	33
		10	25	28	31	40	46	49	51	65	25	29	34	34	39	39	39	49	18	21	25	27	31	31	31	39
	51	90	9	11	12	15	17	19	19	25	11	12	15	15	17	17	18	22	8	9	10	12	14	14	14	17
		75	12	15	17	22	25	26	28	35	14	17	21	21	24	24	24	30	10	12	14	16	19	19	19	24
		50	16	20	22	29	33	35	37	47	19	21	25	27	31	31	31	38	14	16	18	21	24	24	25	31
		25	20	25	27	36	41	44	46	58	23	26	31	33	38	38	38	47	17	19	23	26	30	30	30	37
		10	23	29	32	42	48	51	54	68	27	31	36	38	44	44	44	55	19	22	26	30	35	35	35	44
	25	90	10	13	14	17	20	21	22	28	12	14	17	18	21	21	21	26	9	10	12	14	16	16	16	20
		75	14	18	19	24	28	30	31	40	17	20	23	24	28	28	28	35	12	14	17	19	22	22	22	28
		50	19	24	26	32	37	40	41	54	22	25	30	32	36	36	37	45	16	18	22	25	29	29	29	36
		25	23	29	32	40	46	49	51	65	27	31	37	39	44	44	45	56	20	23	27	31	35	35	36	44
		10	27	34	38	47	54	58	60	77	31	36	43	45	52	52	52	65	23	27	31	36	41	41	42	52
34	76	90	10	12	13	17	19	21	21	27	11	13	15	15	17	17	18	22	9	10	12	12	14	14	14	18
		75	14	17	19	24	27	29	30	39	15	18	21	21	23	24	24	30	12	13	16	17	19	19	19	24
		50	18	23	25	32	36	39	40	51	20	23	27	27	30	31	31	38	15	17	21	22	25	25	25	31
		25	23	29	31	39	44	48	50	64	24	28	33	33	37	38	38	47	19	21	25	27	31	31	31	38
		10	27	34	37	46	53	57	59	75	28	33	39	38	43	44	44	55	22	25	30	31	36	36	36	45
	51	90	10	13	14	17	20	21	22	29	12	13	15	17	20	20	20	24	9	10	12	12	14	14	14	20
		75	14	18	20	25	28	30	32	40	17	18	21	23	27	27	27	33	12	13	16	17	19	19	19	27
		50	19	23	26	33	37	40	42	53	22	23	27	30	35	35	35	43	15	17	21	22	25	25	25	35
		25	24	29	33	41	47	50	52	67	23	27	32	37	42	42	43	51	19	21	25	27	31	31	31	43
		10	28	34	38	48	55	59	62	78	28	33	39	43	49	49	50	62	23	27	31	35	40	40	40	50
	25	90	12	15	16	20	23	24	25	32	13	15	18	20	23	23	23	29	11	12	15	16	19	19	19	23
		75	17	21	23	28	32	34	36	46	18	21	25	27	31	31	32	39	15	17	20	22	26	26	26	32
		50	23	28	31	37	42	46	47	60	23	27	32	35	41	41	41	51	19	22	26	29	33	33	33	41
		25	28	35	38	46	53	57	59	75	29	33	39	43	50	50	50	63	23	27	32	35	41	41	41	51
		10	33	41	45	54	62	67	70	89	33	39	46	51	58	58	59	73	27	31	37	41	47	47	48	59

W: Box width (the dimension away from the body) (cm).
D: Vertical distance of lift (cm).
P: Percentage of industrial population.
Italicized values exceed 8 h physiological critera.

Table 4. Maximum acceptable weight of lift for females (kg); from Snook and Ciriello (1991).

W	D	P	Floor level to knuckle height — One lift every								Knuckle height to shoulder height — One lift every								Shoulder height to arm reach — One lift every							
			5 s	9 s	14 s	1 min	2 min	5	30	8 h	5 s	9 s	14 s	1 min	2 min	5	30	8 h	5 s	9 s	14 s	1 min	2 min	5	30	8 h
75	76	90	5	6	7	8	8	8	9	12	6	6	7	8	9	10	10	13	5	5	5	6	7	7	7	9
		75	6	8	9	10	10	10	11	14	7	8	8	10	11	12	12	15	5	6	6	7	8	9	9	11
		50	7	9	11	11	12	12	13	17	8	9	10	12	13	14	14	15	7	7	8	8	9	10	10	13
		25	9	11	12	14	14	15	15	20	10	11	13	13	15	16	17	21	8	8	9	10	11	12	12	15
		10	10	13	13	15	15	16	17	23	11	12	15	15	17	19	19	24	9	10	10	11	12	14	14	17
	51	90	6	7	7	8	9	10	10	14	7	8	8	9	10	11	11	14	5	5	6	6	7	8	8	10
		75	7	8	9	10	11	12	13	17	8	9	10	11	12	13	13	17	7	7	8	8	9	10	10	12
		50	8	10	11	12	14	14	15	20	10	11	13	13	15	16	16	20	8	9	9	11	12	12	13	15
		25	10	12	13	14	16	17	18	24	12	13	15	15	17	19	19	23	9	10	11	12	13	13	15	17
		10	11	13	14	16	18	19	20	27	13	15	17	17	19	21	21	26	10	12	12	13	15	15	17	19
	25	90	6	8	8	9	10	10	11	14	7	8	10	10	11	12	12	15	5	6	6	7	8	9	9	11
		75	8	10	10	11	12	12	13	17	8	9	11	12	13	15	15	19	7	7	8	9	10	11	11	13
		50	9	11	11	13	14	15	16	21	10	11	13	14	16	18	18	22	8	9	9	11	12	13	13	16
		25	11	13	14	15	17	17	19	25	11	13	15	16	18	20	20	26	9	10	11	12	13	15	15	19
		10	12	15	16	17	19	20	21	28	13	15	17	19	21	23	23	29	10	12	12	13	15	17	17	21
49	76	90	5	6	7	8	8	9	10	13	6	6	8	8	9	10	10	13	5	5	5	6	7	8	8	10
		75	6	8	8	10	10	11	12	16	7	8	9	10	11	12	12	15	5	6	6	7	9	9	9	12
		50	8	9	10	11	13	13	14	19	8	9	11	13	13	14	14	18	7	8	8	9	10	11	11	14
		25	9	11	12	13	15	16	17	22	9	11	13	15	15	17	17	21	8	9	9	11	12	13	13	16
		10	10	13	13	15	17	18	19	25	11	12	15	17	17	19	19	24	9	10	10	12	13	15	15	19
	51	90	6	7	7	9	10	10	11	15	7	8	9	10	11	13	13	14	5	6	6	7	8	9	9	11
		75	7	8	9	11	12	13	14	18	8	9	11	12	13	15	15	17	7	8	8	9	10	11	11	13
		50	8	10	11	13	15	16	16	22	10	11	13	15	15	16	16	20	8	9	9	11	12	13	13	16
		25	10	12	13	15	17	18	19	26	11	13	15	17	17	19	19	23	9	11	11	13	14	15	15	18
		10	11	15	14	17	19	20	22	29	13	15	17	19	19	21	21	26	11	12	12	14	16	18	18	21
	25	90	6	8	8	9	10	11	12	15	7	8	10	10	11	12	12	15	6	6	6	8	9	9	9	12
		75	8	9	10	11	13	13	14	19	8	9	12	12	13	15	15	19	7	8	8	9	10	12	12	14
		50	10	11	13	14	15	16	17	23	10	11	13	14	16	18	18	22	8	9	9	11	13	14	14	17
		25	11	14	14	16	18	19	20	27	11	13	15	17	19	20	20	26	9	11	11	13	15	16	16	20
		10	13	16	16	18	20	21	23	30	12	14	17	19	21	23	23	29	11	12	12	14	16	18	18	23
34	76	90	7	9	9	11	12	13	14	18	9	9	11	12	14	15	15	17	6	7	7	8	9	11	11	13
		75	9	11	11	13	15	16	17	22	11	11	13	14	16	18	18	21	8	8	8	10	12	13	13	16
		50	10	13	14	15	18	19	20	27	13	13	16	17	19	21	21	25	10	10	11	12	14	15	15	19
		25	12	15	16	18	21	22	24	31	15	15	18	18	21	23	23	29	11	12	12	14	16	18	18	22
		10	14	17	18	20	24	25	27	35	16	17	21	21	23	26	26	32	13	14	15	16	18	20	20	25
	51	90	8	10	9	11	13	13	14	19	9	9	10	11	12	13	13	17	7	8	8	9	11	12	12	13
		75	10	12	11	13	15	16	17	23	11	11	12	13	15	16	16	21	8	9	10	11	13	14	14	16
		50	12	14	14	16	18	20	20	28	13	13	14	16	18	20	20	25	10	11	11	13	15	17	17	19
		25	14	16	17	19	21	23	24	33	15	15	17	18	21	23	23	29	11	13	13	15	18	19	19	22
		10	15	18	18	21	24	26	28	37	16	17	19	21	23	26	26	32	13	15	15	18	20	22	22	25
	25	90	8	10	10	11	12	13	14	19	9	9	11	11	12	13	13	15	7	8	8	9	11	12	12	15
		75	10	12	13	14	15	16	17	23	11	11	13	13	15	16	16	18	8	9	10	11	13	14	14	18
		50	12	14	15	17	19	20	21	28	13	13	16	16	18	20	20	21	10	11	11	13	15	17	17	21
		25	14	17	18	20	22	23	24	33	15	15	18	18	21	23	23	25	11	13	13	15	18	19	19	24
		10	15	19	19	22	25	26	28	37	16	17	20	21	23	26	26	29	13	15	15	18	20	22	22	28

W: Box width (the dimension away from the body) (cm).
D: Vertical distance of lift (cm).
P: Percentage of industrial population.
Italicized values exceed 8 h physiological critiera.

Table 5. Maximum acceptable forces of push for males (kg); from Snook and Ciriello (1991).

Initial forces*

H	P	2.1m push One push every							7.6m push One push every							15.2m push One push every							30.5 push One push every					45.7 push One push every					61.0m push One push every			
		6 s	12 s	1 min	2 min	5 min	30 min	8 h	15 s	22 s	1 min	2 min	5 min	30 min	8 h	25 s	35 s	1 min	2 min	5 min	30 min	8 h	1 min	2 min	5 min	30 min	8 h	1 min	2 min	5 min	30 min	8 h	2 min	5 min	30 min	8 h
144	90	20	22	25	25	26	26	31	14	16	21	21	22	22	26	18	19	19	19	20	21	25	15	16	19	19	24	13	14	16	16	20	12	14	14	18
	75	26	29	32	32	34	34	41	18	20	27	27	28	28	34	23	25	25	25	26	27	32	19	21	25	25	31	16	18	21	21	26	16	18	18	23
	50	32	36	40	40	42	42	51	23	25	33	33	35	35	42	29	31	31	31	33	33	40	24	27	31	31	38	20	23	26	26	33	20	22	22	28
	25	38	43	47	47	50	51	61	27	31	40	40	42	42	51	35	37	37	37	40	40	48	28	32	37	37	46	24	27	32	32	39	23	27	27	34
	10	44	49	55	55	58	58	70	31	35	46	46	48	49	58	40	43	43	43	45	46	55	32	37	42	42	53	28	31	36	36	45	27	31	31	39
95	90	21	24	26	26	28	28	34	16	18	23	23	25	25	30	21	22	22	22	23	24	28	17	19	22	22	27	14	16	19	19	23	14	16	16	20
	75	28	31	34	36	36	36	44	21	23	30	30	32	32	39	27	28	28	28	30	30	36	21	21	28	28	35	18	21	24	24	30	18	21	20	26
	50	34	38	43	43	45	45	54	26	29	38	38	40	40	48	33	35	35	35	37	38	45	27	30	35	35	44	23	26	30	30	37	22	26	26	32
	25	41	46	51	51	54	55	65	31	35	45	45	48	48	58	40	42	42	42	45	45	54	32	36	42	42	52	27	31	36	36	45	27	31	31	38
	10	47	53	59	59	62	63	75	35	40	52	52	55	56	66	46	49	49	49	52	52	62	37	41	48	48	60	32	36	41	41	52	31	35	35	44
64	90	19	22	23	23	25	26	31	13	14	20	20	21	21	26	17	19	19	19	20	20	24	14	16	19	19	23	12	14	16	16	20	12	14	14	17
	75	25	28	31	31	33	33	40	16	19	26	26	27	28	33	21	24	24	24	26	26	31	18	21	24	24	30	16	18	21	21	26	15	18	18	22
	50	31	35	39	39	41	41	50	20	23	32	32	34	35	41	27	30	30	30	32	33	39	23	26	30	30	37	20	22	26	26	32	19	22	22	28
	25	38	42	46	46	49	50	59	25	28	39	39	41	41	50	32	36	36	36	39	39	47	28	31	36	36	45	24	27	31	31	39	23	26	26	33
	10	43	48	53	53	57	57	68	28	32	45	45	47	48	57	37	42	42	42	44	45	54	32	36	41	41	52	27	31	36	36	44	26	30	30	38

Sustained forces**

H	P	2.1m push One push every							7.6m push One push every							15.2m push One push every							30.5 push One push every					45.7 push One push every					61.0m push One push every			
		6 s	12 s	1 min	2 min	5 min	30 min	8 h	15 s	22 s	1 min	2 min	5 min	30 min	8 h	25 s	35 s	1 min	2 min	5 min	30 min	8 h	1 min	2 min	5 min	30 min	8 h	1 min	2 min	5 min	30 min	8 h	2 min	5 min	30 min	8 h
144	90	10	13	15	16	18	18	22	8	9	13	13	15	16	18	8	9	11	12	13	14	16	8	10	12	13	16	7	8	10	11	13	7	8	9	11
	75	13	17	21	22	24	25	30	10	13	17	18	20	21	25	11	13	15	16	18	18	22	11	13	16	18	22	10	11	13	15	18	9	11	13	15
	50	17	22	27	28	31	32	38	13	16	22	23	26	27	32	14	17	20	20	23	24	28	15	17	20	23	28	12	14	17	19	23	12	14	16	19
	25	21	27	33	34	38	40	47	16	20	28	29	32	33	39	17	20	24	25	28	29	34	18	21	25	28	34	15	18	21	24	28	15	17	20	24
	10	25	31	38	40	45	46	54	19	23	32	33	38	39	46	20	24	28	29	33	34	40	21	25	29	33	39	18	21	24	28	33	17	20	23	28
95	90	10	13	16	17	19	19	23	8	10	13	13	15	15	18	8	10	11	12	13	13	16	8	10	12	13	16	7	8	9	11	13	7	8	9	11
	75	14	18	22	22	25	26	31	11	13	17	18	20	21	25	11	13	15	16	18	18	21	11	13	16	18	21	9	11	13	15	18	9	11	12	15
	50	18	23	28	28	33	34	40	14	17	22	23	26	27	32	14	17	19	20	23	23	28	15	17	20	23	28	12	14	17	19	23	12	14	16	19
	25	22	28	34	35	40	41	49	17	21	27	29	32	33	39	18	21	24	25	28	29	34	18	21	25	28	33	15	18	21	24	28	15	17	20	23
	10	26	33	40	41	46	48	57	20	24	32	33	37	38	45	20	25	28	29	33	33	40	21	25	29	33	39	17	20	24	27	32	17	20	23	27
64	90	10	13	16	16	18	19	23	8	10	12	13	14	15	18	8	10	11	11	12	13	15	8	9	11	13	15	7	8	9	11	13	7	8	9	10
	75	14	18	21	22	25	26	31	11	13	17	17	19	20	24	11	13	14	15	17	17	21	11	13	15	17	20	9	11	12	14	17	9	10	12	14
	50	18	23	28	29	32	33	39	14	17	21	22	25	26	31	14	17	19	19	22	22	27	14	16	19	22	26	12	14	16	18	22	12	14	15	18
	25	22	28	34	35	39	41	48	17	21	26	27	31	32	37	18	21	23	24	27	28	33	17	20	24	27	32	14	17	20	23	28	14	17	19	22
	10	26	32	39	41	46	48	56	20	25	30	32	36	37	44	21	25	27	28	31	32	38	20	24	28	32	38	17	20	23	26	31	16	19	22	26

H: vertical distance from floor to hands (cm).
P: Percentage of industrial population.
*The force required to get an object in motion.
**The force required to keep an object in motion.
Italicized values exceed 8 H physiological criteria.

Table 6. Maximum acceptable forces of push for females (kg); from Snook and Ciriello (1991).

Initial forces*

H	P	2.1m push — One push every							7.6m push — One push every							15.2 push — One push every							30.5 push — One push every					45.7 push — One push every					61.0m push — One push every			
		6s	12s	1	2	5	30	8h	15s	22s	1	2	5	30	8h	25s	35s	1	2	5	30	8h	1	2	5	30	8h	1	2	5	30	8h	2	5	30	8h
135	90	14	15	17	18	20	21	22	15	16	16	16	17	18	20	12	14	14	14	15	16	17	12	13	14	15	17	12	13	14	15	17	12	13	14	15
	75	17	18	21	22	24	25	27	18	19	19	20	22	23	24	15	17	17	17	19	20	21	15	16	17	19	21	15	16	17	19	21	14	15	17	19
	50	20	22	25	26	29	30	32	21	23	23	24	26	27	29	18	20	20	20	22	23	25	18	19	21	22	25	18	19	21	22	25	17	18	20	22
	25	24	25	29	33	33	35	37	25	26	27	28	31	32	34	20	23	23	24	26	27	29	20	22	24	26	29	20	22	24	26	29	20	21	23	26
	10	26	28	33	34	38	39	41	28	30	30	31	34	36	38	23	26	26	26	29	31	32	23	25	27	29	33	23	25	27	29	33	22	24	26	29
89	90	14	15	17	18	20	21	22	14	15	16	17	19	19	21	11	13	14	14	16	16	17	12	14	15	16	18	12	14	315	16	18	12	13	15	16
	75	17	18	21	22	24	25	27	17	18	20	20	22	23	25	14	16	17	17	19	20	21	15	16	18	19	21	15	16	18	19	21	15	16	18	19
	50	20	22	25	26	29	30	32	20	21	23	24	27	28	30	16	19	20	21	23	23	25	18	20	21	23	26	18	20	21	23	26	18	19	20	23
	25	24	25	29	30	33	35	37	23	25	27	28	31	33	34	19	22	23	24	27	28	29	21	23	24	26	30	21	23	24	26	30	20	22	24	27
	10	26	28	34	34	38	39	41	26	28	31	32	35	37	39	22	24	26	27	30	31	33	24	26	28	30	33	24	26	28	30	33	23	24	26	30
57	90	11	12	14	14	16	17	18	11	12	14	14	16	16	17	9	11	12	12	13	14	15	11	12	13	13	15	11	12	12	13	15	10	11	12	13
	75	14	15	17	17	19	20	21	14	15	17	17	19	20	21	11	13	14	15	16	17	18	13	14	15	16	18	13	14	15	16	18	12	13	14	16
	50	16	17	20	21	23	24	25	16	18	20	21	23	24	25	14	15	17	18	19	20	21	15	17	18	19	22	15	17	18	19	22	15	16	17	19
	25	19	20	23	24	27	28	30	19	21	23	24	27	28	29	16	18	20	20	23	24	25	18	19	21	22	25	18	19	21	22	25	17	19	20	23
	10	21	23	26	27	30	31	33	22	23	26	27	30	31	33	18	20	22	23	25	26	28	20	22	23	25	28	20	22	23	25	28	19	21	23	25

Sustained forces* *(italicized values)*

H	P	2.1m push — One push every							7.6m push — One push every							15.2 push — One push every							30.5 push — One push every					45.7 push — One push every					61.0m push — One push every			
		6s	12s	1	2	5	30	8h	15s	22s	1	2	5	30	8h	25s	35s	1	2	5	30	8h	1	2	5	30	8h	1	2	5	30	8h	2	5	30	8h
135	90	6	8	9	10	11	12	14	6	7	7	7	8	9	11	5	6	6	6	7	7	8	5	6	6	6	8	5	5	5	6	8	4	4	4	6
	75	9	12	14	14	16	17	21	9	10	11	11	12	13	16	7	8	9	9	10	11	12	7	8	9	9	12	7	7	8	8	11	6	6	6	9
	50	12	16	19	20	21	23	28	12	14	14	15	16	17	21	10	11	12	12	14	14	16	10	11	12	12	16	8	10	11	11	15	8	8	9	12
	25	16	20	24	25	27	29	36	15	17	18	18	20	22	27	12	14	15	16	17	18	21	13	14	15	15	21	11	13	13	14	19	10	10	11	15
	10	18	23	28	29	32	34	42	18	20	21	22	24	26	32	14	17	18	18	20	22	25	15	17	17	18	25	14	15	16	17	22	12	12	13	17
89	90	6	7	9	9	10	11	13	6	7	8	8	9	9	11	5	6	6	7	7	8	10	5	6	6	7	9	5	6	6	6	8	4	4	5	6
	75	8	11	13	13	15	16	19	9	10	11	11	13	13	17	7	8	9	10	11	11	13	8	9	9	10	13	7	8	8	9	12	6	6	7	9
	50	11	15	18	18	20	21	26	12	13	15	15	17	18	22	9	11	13	13	14	15	17	10	12	12	13	17	10	11	11	12	16	8	9	9	12
	25	14	18	22	23	25	27	33	15	17	19	19	21	23	28	12	14	16	16	18	19	22	13	15	15	16	22	12	14	14	15	20	11	11	12	15
	10	17	22	26	27	30	32	39	17	20	22	23	25	27	33	14	17	19	19	21	23	26	16	18	18	19	26	14	16	17	18	24	13	13	14	18
57	90	5	6	8	8	9	9	12	6	7	7	7	8	9	11	5	6	6	6	7	7	8	5	6	6	6	8	5	5	5	6	7	4	4	4	6
	75	7	9	11	11	13	12	17	8	10	10	11	12	12	15	7	8	9	9	10	10	12	7	8	8	9	12	7	7	8	8	11	6	6	6	8
	50	10	13	15	16	17	18	23	11	13	14	14	16	17	21	9	11	12	12	14	14	16	10	11	11	12	16	9	10	10	11	15	8	8	8	11
	25	12	16	19	20	22	23	29	14	17	18	18	20	21	26	12	14	15	15	17	18	21	12	14	14	15	20	11	13	13	14	19	10	10	11	14
	10	15	19	22	23	26	28	34	17	20	21	21	23	25	31	14	16	17	18	20	21	26	15	16	17	18	24	13	15	16	16	22	12	12	13	17

H: vertical distance from floor to hands (cm).
P: Percentage of industrial population.
*The force required to get an object in motion.
**The force required to keep an object in motion.
Italicized values exceed 8 H physiological criteria.

Table 7. Maximum acceptable forces of pull for males (kg); from Snook and Ciriello (1991).

Initial forces*

H	P	2.1m push — One push every							7.6m push — One push every							15.2 push — One push every							30.5 push — One push every					45.7 push — One push every					61.0m push — One push every			
		6 s	12 s	1 min	2 min	5 min	30 min	8 h	15 s	22 s	1 min	2 min	5 min	30 min	8 h	25 s	35 s	1 min	2 min	5 min	30 min	8 h	1 min	2 min	5 min	30 min	8 h	1 min	2 min	5 min	30 min	8 h	2 min	5 min	30 min	8 h
144	90	14	16	18	18	19	19	23	11	13	16	16	17	18	21	13	15	15	15	16	17	20	12	13	15	15	19	10	11	13	13	16	10	11	11	14
	75	17	19	22	22	23	24	28	14	15	20	20	21	21	26	16	18	19	19	20	20	24	14	16	19	19	23	12	14	16	16	20	12	14	14	17
	50	20	23	26	26	28	28	33	16	18	24	24	25	26	31	19	21	22	22	24	24	29	17	19	22	22	27	15	16	19	19	24	14	16	16	20
	25	24	27	31	31	32	33	39	19	21	28	28	29	30	36	22	25	26	26	28	28	33	20	22	26	26	32	17	19	22	22	28	16	19	19	24
	10	26	30	34	34	36	37	44	21	24	31	31	33	33	40	24	28	29	29	31	31	38	22	25	29	29	37	20	22	25	25	31	18	21	21	27
95	90	19	22	25	25	27	27	32	15	18	23	23	24	24	29	18	20	21	21	23	23	28	16	18	21	21	26	14	16	18	18	23	13	16	16	19
	75	23	27	31	31	32	33	39	19	21	28	28	29	30	36	22	25	26	26	28	28	33	20	22	26	26	32	17	19	22	22	28	16	19	19	24
	50	28	32	36	36	39	39	47	23	26	33	33	35	35	42	26	29	31	31	33	33	40	24	27	31	31	38	20	23	27	27	33	20	23	23	28
	25	33	37	42	42	45	45	54	26	30	39	39	41	41	49	30	34	36	36	38	39	46	27	31	36	36	45	24	27	31	31	38	23	26	26	33
	10	37	42	48	48	51	51	61	30	33	43	43	46	47	56	33	38	41	41	43	44	52	31	35	40	40	50	27	30	35	35	43	26	30	30	37
64	90	22	25	28	28	30	30	36	18	20	26	26	27	28	33	20	23	24	24	26	26	31	18	21	24	24	30	16	18	21	21	26	15	18	18	22
	75	27	30	34	34	37	37	44	21	24	31	31	33	34	40	24	28	29	29	31	32	38	22	25	29	29	36	19	22	25	25	31	19	21	21	27
	50	32	36	41	41	44	44	53	25	29	37	37	40	40	48	29	33	35	35	37	38	45	27	30	35	35	40	23	26	30	30	37	22	26	26	32
	25	37	42	48	48	51	51	61	30	34	44	44	46	47	56	34	39	41	41	43	44	52	31	35	41	41	50	27	30	35	35	43	26	30	30	37
	10	42	48	54	54	57	58	69	33	38	49	49	52	53	63	38	43	46	46	49	49	59	35	39	46	46	57	30	34	39	39	49	30	34	34	42

Sustained forces**

H	P	2.1m push — One push every							7.6m push — One push every							15.2 push — One push every							30.5 push — One push every					45.7 push — One push every					61.0m push — One push every			
		6 s	12 s	1 min	2 min	5 min	30 min	8 h	15 s	22 s	1 min	2 min	5 min	30 min	8 h	25 s	35 s	1 min	2 min	5 min	30 min	8 h	1 min	2 min	5 min	30 min	8 h	1 min	2 min	5 min	30 min	8 h	2 min	5 min	30 min	8 h
144	90	8	10	12	13	15	15	18	6	8	10	11	12	12	15	7	8	9	9	10	11	13	7	8	9	11	13	6	7	8	9	10	6	6	7	9
	75	10	13	16	17	19	20	23	8	10	13	14	16	16	19	9	10	12	12	14	14	17	9	10	12	14	16	7	9	10	11	14	7	8	10	11
	50	13	16	20	21	23	24	28	10	13	16	17	19	20	23	11	13	14	15	17	17	20	11	13	15	17	20	9	11	12	14	17	9	10	12	14
	25	15	20	24	25	28	29	34	12	15	20	20	23	24	28	13	15	17	18	20	21	24	13	15	18	20	24	11	13	15	17	20	11	12	14	17
	10	17	22	27	28	32	33	39	14	17	22	23	26	27	32	14	17	19	20	23	24	28	15	17	20	23	27	12	14	17	19	23	12	14	16	19
95	90	10	13	16	17	19	20	24	8	10	13	14	16	16	19	9	10	12	12	14	14	17	9	10	12	14	17	7	9	10	12	14	7	9	9	12
	75	13	17	21	22	25	26	30	11	13	17	18	20	21	25	11	14	15	15	18	18	22	12	13	16	18	21	10	11	13	16	18	9	11	13	15
	50	16	21	26	27	31	32	37	13	17	21	22	25	26	31	14	17	19	19	22	23	27	14	18	19	22	27	12	14	16	19	22	12	14	16	18
	25	19	26	31	33	37	38	45	16	20	26	27	30	31	37	17	20	22	23	26	27	32	17	20	23	27	32	14	17	19	22	26	14	16	19	22
	10	22	29	36	37	42	43	51	18	23	29	31	34	36	42	19	23	26	27	30	31	37	19	23	27	31	36	16	19	22	25	30	16	19	21	25
64	90	11	14	17	19	20	21	25	9	11	14	15	17	17	20	9	11	12	13	15	15	18	9	11	13	15	18	8	10	11	12	15	8	9	10	12
	75	14	19	23	23	26	27	32	11	14	19	19	22	22	26	12	14	16	17	19	19	23	12	14	17	19	23	10	12	14	16	19	10	12	13	16
	50	17	23	28	29	32	34	40	14	18	23	24	27	28	33	15	18	20	21	23	24	28	15	18	21	24	27	13	15	17	20	23	12	14	16	20
	25	20	27	33	35	39	40	48	17	21	27	27	32	33	39	18	21	24	25	28	29	34	18	21	25	28	33	15	18	21	24	28	15	17	20	23
	10	23	31	38	40	45	46	54	19	24	31	32	37	38	45	20	24	27	28	32	33	39	21	24	28	32	38	17	20	24	27	32	17	20	23	27

H: vertical distance from floor to hands (cm).
P: Percentage of industrial population.
*The force required to get an object in motion.
**The force required to keep an object in motion.
Italicized values exceed 8 H physiological criteria.

Table 8. Maximum acceptable forces of pull for females (kg); from Snook and Ciriello (1991).

H	P	2.1m push — One push every							7.6m push — One push every							15.2 push — One push every							30.5 push — One push every					45.7 push — One push every					61.0m push — One push every			
		6 s	12 s	1 min	2 min	5 min	30 min	8 h	15 s	22 s	1 min	2 min	5 min	30 min	8 h	25 s	35 s	1 min	2 min	5 min	30 min	8 h	1 min	2 min	5 min	30 min	8 h	1 min	2 min	5 min	30 min	8 h	2 min	5 min	30 min	8 h
																			*Initial forces**																	
135	90	13	16	17	18	20	21	22	13	14	16	16	18	19	20	12	12	13	14	15	16	17	12	13	14	15	17	12	13	14	15	17	12	13	14	15
	75	16	19	20	21	24	25	26	16	17	19	19	21	22	24	14	14	16	16	18	19	20	14	16	17	18	20	14	16	17	18	20	14	15	16	18
	50	19	22	24	25	28	29	31	19	20	22	23	25	26	28	16	16	19	19	21	22	24	17	18	20	21	24	17	18	20	21	24	16	18	19	21
	25	21	25	28	29	32	33	35	21	23	25	26	29	30	32	18	19	21	22	25	26	27	19	21	23	24	27	19	21	23	24	27	19	20	22	25
	10	24	28	31	32	36	37	39	24	26	28	29	32	34	36	21	21	24	25	27	29	30	22	24	25	27	31	22	24	25	27	31	21	23	24	27
89	90	14	16	18	19	21	22	23	14	15	16	17	19	20	21	12	12	14	14	16	17	18	13	14	15	16	18	13	14	15	16	18	12	13	14	16
	75	16	19	21	22	25	26	27	16	18	19	20	22	23	25	15	15	17	17	19	20	21	15	17	18	19	21	15	16	18	19	21	15	15	17	19
	50	19	23	25	26	29	30	32	19	21	23	24	26	27	29	17	17	19	20	22	23	25	18	19	21	22	25	18	19	21	22	25	17	18	20	22
	25	22	26	29	30	33	35	37	22	24	26	27	30	31	33	20	20	22	23	26	27	28	20	22	24	25	29	20	22	24	25	29	20	21	23	25
	10	25	29	32	33	37	39	41	25	27	29	30	33	35	37	22	23	25	26	29	30	32	23	25	26	28	32	23	25	26	28	32	22	24	25	29
57	90	15	17	19	20	22	23	24	15	16	17	18	20	21	22	13	13	15	15	17	18	19	13	14	15	17	19	13	14	15	17	19	13	14	15	17
	75	17	20	22	23	26	27	28	17	19	20	21	23	24	26	15	15	17	18	20	21	22	16	17	18	20	22	16	17	18	20	22	15	16	18	20
	50	20	24	26	27	30	32	33	20	22	24	25	28	29	30	18	18	20	21	23	24	26	18	20	22	23	26	18	20	22	23	26	18	19	21	23
	25	23	27	30	31	35	36	38	23	25	27	29	32	33	35	21	21	23	24	27	28	30	21	23	25	27	30	21	23	25	27	30	21	22	24	27
	10	26	31	34	35	39	40	43	26	28	31	32	35	37	39	23	23	26	27	30	31	33	24	26	28	30	34	24	26	28	30	34	23	25	27	30
																		*Sustained forces***																		
135	90	*6*	*9*	*10*	*10*	*11*	*12*	13	*7*	*8*	*9*	*9*	*10*	11	13	*6*	*7*	*8*	*8*	*8*	9	11	*6*	*7*	*7*	8	10	*6*	*6*	*7*	7	9	*5*	*5*	*5*	7
	75	*8*	*12*	*13*	*14*	*15*	*16*	18	*9*	*11*	*12*	*12*	*13*	14	18	*7*	*9*	*10*	*10*	*11*	12	15	*8*	*9*	*10*	10	14	*8*	*9*	*11*	11	12	*7*	*7*	*7*	10
	50	*10*	*16*	*16*	*17*	*19*	*19*	22	*12*	*13*	*15*	*16*	*17*	18	22	*9*	*11*	*13*	*13*	*14*	15	19	*11*	*12*	*12*	13	18	*10*	*11*	*12*	12	16	*8*	*9*	*9*	12
	25	*13*	*19*	*21*	*21*	*23*	*23*	27	*14*	*16*	*18*	*18*	*21*	22	27	*11*	*14*	*15*	*15*	*17*	19	23	*13*	*15*	*15*	16	21	*12*	*13*	*14*	14	15	*10*	*11*	*11*	15
	10	*15*	*22*	*24*	*25*	*27*	*29*	31	*16*	*19*	*21*	*21*	*24*	26	32	*13*	*16*	*17*	*17*	*20*	22	27	*15*	*17*	*17*	18	25	*14*	*15*	*16*	17	17	*12*	*12*	*13*	17
89	90	*6*	*9*	*10*	*10*	*11*	*12*	13	*7*	*8*	*9*	*9*	*10*	10	13	*5*	*6*	*7*	*7*	*7*	9	10	*6*	*7*	*7*	7	10	*5*	*6*	*6*	7	9	*5*	*5*	*5*	7
	75	*8*	*12*	*13*	*13*	*15*	*16*	17	*8*	*10*	*11*	*11*	*13*	14	17	*7*	*8*	*9*	*9*	*10*	12	13	*8*	*9*	*9*	10	13	*7*	*8*	*9*	9	12	*6*	*7*	*7*	9
	50	*10*	*15*	*16*	*17*	*19*	*18*	22	*11*	*13*	*15*	*15*	*16*	18	22	*8*	*11*	*12*	*12*	*13*	15	17	*10*	*12*	*12*	13	17	*8*	*10*	*10*	11	15	*8*	*8*	*9*	12
	25	*12*	*18*	*20*	*20*	*23*	*21*	27	*14*	*16*	*18*	*18*	*21*	20	27	*11*	*13*	*15*	*15*	*17*	18	21	*12*	*14*	*15*	15	21	*11*	*13*	*13*	14	19	*10*	*10*	*11*	15
	10	*14*	*21*	*23*	*24*	*26*	*25*	31	*16*	*18*	*21*	*21*	*23*	23	31	*13*	*15*	*17*	*17*	*20*	21	24	*15*	*16*	*18*	18	24	*13*	*15*	*16*	16	22	*12*	*12*	*13*	17
57	90	*5*	*8*	*9*	*9*	*10*	*11*	13	*6*	*7*	*8*	*8*	*9*	10	12	*5*	*6*	*7*	*7*	*7*	8	9	*6*	*6*	*6*	7	9	*6*	*6*	*6*	8	8	*4*	*5*	*5*	6
	75	*7*	*11*	*12*	*12*	*13*	*14*	18	*8*	*9*	*11*	*11*	*12*	13	16	*7*	*8*	*9*	*9*	*10*	11	12	*7*	*8*	*8*	9	11	*7*	*8*	*8*	9	11	*6*	*6*	*6*	9
	50	*9*	*14*	*15*	*16*	*17*	*18*	23	*10*	*12*	*13*	*14*	*15*	16	20	*8*	*10*	*11*	*11*	*12*	13	16	*9*	*11*	*11*	12	17	*9*	*10*	*10*	11	14	*8*	*8*	*8*	11
	25	*11*	*17*	*18*	*19*	*21*	*22*	27	*12*	*15*	*16*	*16*	*19*	20	24	*10*	*12*	*14*	*14*	*16*	17	21	*11*	*13*	*13*	14	19	*11*	*12*	*12*	13	17	*9*	*10*	*10*	13
	10	*13*	*20*	*21*	*22*	*24*	*26*	32	*15*	*17*	*18*	*19*	*22*	23	28	*12*	*14*	*16*	*16*	*18*	19	24	*13*	*15*	*15*	16	22	*12*	*14*	*14*	15	20	*10*	*11*	*12*	16

H: vertical distance from floor to hands (cm).
P: Percentage of industrial population.
*The force required to get an object in motion.
**The force required to keep an object in motion.
Italicized values exceed 8 H physiological criteria.

Table 9. Maximum acceptable weight of carry (kg); from Snook and Ciriello (1991).

H	P	2.1m carry One carry every 6s	12s	1 min	2 min	5 min	30 min	8h	4.3m carry One carry every 10s	16s	1 min	2 min	5 min	30 min	8h	8.5m carry One carry every 18s	24s	1 min	2 min	5 min	30 min	8h
														Males								
111	90	10	14	17	17	19	21	25	9	11	15	15	17	19	22	10	11	13	13	15	17	20
	75	14	19	23	23	26	29	34	13	16	21	21	23	26	30	13	15	18	18	20	23	27
	50	*19*	25	30	30	33	38	44	17	20	27	27	30	34	39	17	19	23	24	26	29	35
	25	*23*	30	37	37	41	46	54	*20*	25	33	33	37	41	48	*21*	*24*	29	29	32	36	43
	10	*27*	*35*	43	43	48	54	63	*24*	29	38	39	43	48	57	*24*	*28*	34	34	38	42	50
79	90	13	17	21	21	23	26	31	11	14	18	19	21	23	27	13	15	17	18	20	22	26
	75	18	23	28	29	32	36	42	16	19	25	25	28	32	37	17	20	24	24	27	30	35
	50	*23*	30	37	37	41	46	54	*20*	25	32	33	36	41	48	*22*	*26*	31	31	35	39	46
	25	*28*	*37*	45	46	51	57	67	*25*	30	40	40	45	50	59	*27*	*32*	38	38	42	48	56
	10	*33*	*43*	53	53	59	66	78	*29*	*35*	47	47	52	59	69	*32*	*38*	44	45	50	56	65
														Females								
105	90	11	12	13	13	13	13	18	9	10	13	13	13	13	18	10	11	12	12	12	12	16
	75	*13*	14	15	15	16	16	21	*11*	12	15	15	16	16	21	*12*	*13*	14	14	14	14	19
	50	*15*	16	18	18	18	18	25	*12*	13	18	18	18	18	24	*14*	*15*	16	16	16	16	22
	25	*17*	*18*	20	20	21	21	28	*14*	15	20	20	21	21	28	*15*	*17*	18	18	19	19	25
	10	*19*	*20*	22	22	23	23	31	*16*	17	22	22	23	23	31	*17*	*19*	20	20	21	21	28
72	90	*13*	14	16	16	16	16	22	10	11	14	14	14	14	20	*12*	12	14	14	14	14	19
	75	*15*	*17*	18	18	19	19	25	11	13	16	16	17	17	23	*14*	*15*	16	16	17	17	23
	50	*17*	*19*	21	21	22	22	29	*13*	15	19	19	20	20	26	*16*	*17*	19	19	20	20	26
	25	*20*	*22*	24	24	25	25	33	*15*	17	22	22	22	22	30	*18*	*19*	21	22	22	22	30
	10	*22*	*44*	27	27	28	28	37	*17*	*19*	24	24	25	25	33	*20*	*21*	24	24	25	25	33

H: vertical distance from floor to hands (cm).
P: Percentage of industrial population.
Italicized values exceed 8 H physiological criteria.

Databases for Psychophysically Acceptable Maximum Wrist Torques in Manual Tasks for Females Developed by Liberty Mutual

W. Karwowski and R. Jang

Center for Industrial Ergonomics, University of Louisville, Louisville, KY 40292, USA

1. BACKGROUND

Snook *et al.* (1995) used the psychophysical approach to determine maximum acceptable forces for various types and frequencies for repetitive wrist motion, grips, and repetition rates that would not result in significant changes in wrist strength, tactile sensitivity, or number of symptoms reported by the female subjects.

A total of 29 paid subjects participated in the study (the number of female subjects for experiments 1 and 2 was 15 and 14 respectively). All participants were instructed to work as if they were on an incentive basis, getting paid for the amount of work they performed. They were asked to work as hard as possible without developing unusual discomfort in the hands, wrists or forearms.

The reported soreness, stiffness and numbness symptoms were recorded every hour by indicating the intensity of each symptom on a scale of 0–3 at a given body location where each symptom occurred.

2. DATABASES

Three levels of wrist motion were used: (1) flexion motion with a power grip, (2) flexion motion with a pinch grip, and (3) extension motion with a power grip. The dependent variables were the maximum acceptable wrist torque, maximum isometric wrist strength, tactile sensitivity and symptoms. Experiment 1 utilized exposure of 2 days per week, while experiment 2 utilized 5 days per week.

The maximum acceptable wrist torque (MAWT) was defined as the number of Newton-meters of resistance set in the brake by the participants (averaged and recorded every minute). The maximum isometric wrist strength (MAWS), a measure of maximum voluntary contraction (MVC), was recorded in Newton-meters of torque with the handle fixed in the horizontal position (maximum exertion over 5 s).

3. DATABASES
3.1. Two Days per Week Exposure

Experiment 1 utilized exposure of 2 days per week. The maximum acceptable wrist torque for each repetition, rate and type of motion, as well as the combined means and standard deviations for each type of motion (regardless of the repetition rate), and for each repetition rate (regardless of the type of motion) are

Table 1. Maximum acceptable torque (Nm) (two days per week exposure); from Snook *et al.* (1995)

		Repetition rate					
		2/min	5/min	10/min	15/min	20/min	combined*
Flexion-	Mean	4.06	4.10	3.68	3.31	2.81	3.59 (a)
power grip	SD	2.00	2.04	1.75	1.39	1.43	1.80
Flexion-	Mean	3.82	3.54	3.08	3.08	2.49	3.20 (a)
pinch grip	SD	2.00	1.93	1.59	1.60	1.06	1.72
Extension-	Mean	2.39	2.39	2.13	1.91	1.51	2.07
power grip	SD	0.98	0.97	1.09	0.81	0.72	0.98
Combined	Mean	3.42(b)	3.42(b)	2.96(c)	2.77(c)	2.27	2.95
	SD	1.87	1.85	1.63	1.44	1.23	1.68

* Combined means with the same letter are not significantly different (p>0.05).

Table 2. Maximum isometric strength (Nm) (two days per week exposure); from Snook *et al.* (1995)

		Repetition rate					
		2/min	5/min	10/min	15/min	20/min	combined*
Flexion-	Mean	7.35	7.40	7.24	7.39	6.84	7.24
power grip	SD	3.28	3.13	2.67	3.16	2.76	2.99
Flexion-	Mean	6.32	6.07	6.21	6.13	6.03	6.15
pinch grip	SD	2.60	2.37	2.85	1.87	2.28	2.40
Extension-	Mean	5.80	4.97	4.94	4.81	4.91	5.08
power grip	SD	2.33	1.40	1.52	1.35	1.70	1.72
Combined	Mean	6.49(b)	6.15(ab)	6.13(ab)	6.(b)	5.93(b)	6.16
	SD	2.82	2.59	2.58	2.48	2.41	2.58

* Combined means with the same letter are not significantly different (p>0.05).

Table 3. Tactile sensitivity (vibration units) (two days per week exposure).*; from Snook *et al.* (1995)

		Repetition rate					
		2/min	5/min	10/min	15/min	20/min	combined*
Flexion-	Mean	0.79	0.74	0.78	0.75	0.71	0.75
power grip	SD	0.36	0.30	0.35	0.30	0.31	0.32
Flexion-	Mean	0.80	0.74	0.82	0.72	0.72	0.77
pinch grip	SD	0.29	0.32	0.36	0.29	0.28	0.35
Extension-	Mean	0.81	0.74	0.80	0.77	0.74	0.76
power grip	SD	4.00	0.37	0.30	0.34	0.34	0.31
Combined	Mean	0.80(a)	0.74(b)	0.80(a)	0.75(ab)	0.73(b)	0.76
	SD	0.35	0.33	0.34	0.31	0.31	0.33

* Combined means with the same letter are not significantly different (p>0.05).

presented in Table 1. The maximum isometric wrist strength and tactile sensitivity for each type of motion and repetition rate are shown in Tables 2 and 3. The symptom data are presented in Tables 4–6. The results for each hour during the day are presented in Table 4 for all dependent variables (regardless of the repetition rate and type of motion).

3.2. Five Days per Week Exposure

Experiment 2 utilized exposure of 5 days per week. Here, in addition to the four dependent variables measured in the 2 days per week exposure, the performance errors and duration of force were also measured. The results for each of the six dependent variables are shown in Table 5. The results for each hour during the day (across all days of exposure) as well as the significance

Table 4. Hourly results (two days per week exposure); from Snook et al. (1995).

	Hour							
	0	1	2	3	4	5	6	7
Maximum acceptable torque (Nm)*		3.20 (a)	3.14 (ab)	3.03 (bc)	2.94 (cd)	2.86 (de)	2.75 (e)	2.75 (e)
Maximum isometric wrist strength (Nm)*	6.21 (f)		6.12 (f)		6.15 (f)			
Tactile sensitivity (vibration units)*	0.82		0.74 (g)		0.73 (g)			
Number of symptoms		120	188	241	254	249	276	285

* Values with the same letter are not significantly different (p>0.05)

Table 5. Daily results for wrist flexion at 15 motions per minute (five days per week exposure); from Snook et al. (1995)

Day	MAT Mean*	SD	MIWS Mean*	SD	TS Mean*	SD	DOF Mean*	SD	Number of symptoms	Error rate (errors/100 motions)
5	2.03	0.82	6.67	3.11	1.27	0.43	0.85	0.19	152	0.61
6	2.10	0.85	6.36	2.69	1.29	0.35	0.83	0.20	178	0.45
7	2.15	0.92	6.61	2.92	1.35	0.44	0.77	0.17	133	0.49
8	2.05	0.86	7.02	2.85	1.39	0.55	0.75	0.17	154	0.48
9	2.14	0.97	6.80	3.41	1.40	0.54	0.78	0.19	131	0.39
10	2.14	0.97	7.42	2.84	1.35	0.46	0.76	0.19	120	0.41
11	2.05	0.83	6.71	3.26	1.32	0.43	0.79	0.24	104	0.41
12	2.11	0.91	6.84	2.70	1.34	0.40	0.82	0.26	104	0.37
13	2.04	0.91	6.14	2.66	1.35	0.46	0.80	0.27	83	0.46
14	2.05	0.90	6.69	2.86	1.34	0.44	0.82	0.26	101	0.51
15	2.03	0.92	6.67	2.76	1.43	0.56	0.82	0.27	74	0.46
16	2.05	0.81	6.75	2.95	1.36	0.40	0.80	0.25	105	0.40
17	2.11	0.82	7.05	3.22	1.30	0.35	0.83	0.25	88	0.44
18	2.07	0.80	7.12	2.86	1.33	0.48	0.80	0.25	90	0.48
19	2.15	0.84	7.26	3.26	1.39	0.58	0.80	0.24	96	0.48
20	2.16	0.74	7.00	2.93	1.39	0.52	0.83	0.28	83	0.52
21	2.20	0.94	6.49	3.02	1.34	0.51	0.82	0.28	88	0.51
22	2.29	1.02	7.14	3.07	1.33	0.62	0.83	0.24	86	0.48
23	2.25	0.97	6.89	3.11	1.39	0.57	0.81	0.23	81	0.47
Grand means and SDs										
	2.11	0.89	6.82	2.89	1.35	0.48	0.81	0.24	108	0.46

* No significant difference from day to day.
MAT: Maximum acceptable torque (Nm).
MIWS: Maximum isometric wrist strength (Nm)
TS: Tactile sensitivity (vibration units)
DOF: Duration force(s).

Table 6. Number of symptoms by type and location (two days per week exposure); from Snook et al. (1995)

	Type of symptom									
	Soreness			Stiffness			Numbness			
Location	Flex	Pinch	Exten	Flex	Pinch	Exten	Flex	Pinch	Exten	Total
Palmar side										
Fingers/thumb	38	98	75	25	71	27	0	4	29	367
Hand/wrist	79	50	92	13	12	7	0	0	1	254
Forearm	24	28	66	1	1	2	0	1	0	123
Total	141	176	233	39	84	36	0	5	30	744 (46.1%)
Dorsal side										
Fingers/thumb	34	52	52	59	111	71	0	7	26	412
Hand/wrist	64	39	94	11	15	56	0	2	0	281
Forearm	39	27	86	9	0	13	0	2	0	176
Total	137	118	232	79	126	140	0	11	26	869 (53.9%)
Grand total	278	294	465	118	210	176	0	16	56	1613 (100%)

Flex: Flexion with power grip
Pinch: Flexion with pinch grip
Exten: Extension with power grip

testing for maximum acceptable torque, tactile sensitivity and duration of force are shown in Table 6.

4. DATABASE EXTENSIONS

The data for maximum acceptable wrist torques for the 2 days per week exposure was used to estimate the maximum acceptable torques were estimated for different repetitions of wrist flexion (power grip) and different percentages of the population. This was done by using the adjusted means and coefficients of variation from the 2 days per week exposure. The original torque values were converted into forces by dividing each torque by the average

Table 7. Maximum acceptable forces for female wrist flexion (power grip) (N); from Snook et al. (1995)

Percent of Population	Repetition rate				
	2/min	5/min	10/min	15/min	20/min
90	14.9	14.9	13.5	12.0	10.2
75	23.2	23.2	20.9	18.6	15.8
50	32.3	32.3	29.0	26.0	22.1
25	41.5	41.5	37.2	33.5	28.4
10	49.8	49.8	44.6	40.1	34.0

length of the handle lever (0.081 m). The estimated values for the maximum acceptable forces for female wrist flexion (power grip) are shown in Table 7.

Similarly, the estimated maximum acceptable forces were developed for the wrist flexion (pinch grip; Table 8), and wrist

Table 8. Maximum acceptable forces for female wrist flexion (pinch grip) (N); from Snook et al. (1995)

Percent of Population	Repetition rate				
	2/min	5/min	10/min	15/min	20/min
90	9.2	8.5	7.4	7.4	6.0
75	14.2	13.2	11.5	11.5	9.3
50	19.8	18.4	16.0	16.0	12.9
25	25.4	23.6	20.6	20.6	16.6
10	30.5	28.2	24.6	24.6	19.8

Table 9. Maximum acceptable forces for female wrist flexion (power grip) (N); from Snook et al. (1995)

Percent of Population	Repetition rate				
	2/min	5/min	10/min	15/min	20/min
90	8.8	8.8	7.8	6.9	5.4
75	13.6	13.6	12.1	10.9	8.5
50	18.9	18.9	16.8	15.1	11.9
25	24.2	24.2	21.5	19.3	15.2
10	29.0	29.0	25.8	23.2	18.3

extension (power grip; Table 9). The torques were converted into forces by dividing by 0.081 m for the power grip, and 0.123m for the pinch grip.

Snook et al. (1995) note that the estimated values of the maximum acceptable wrist torque presented in Tables 7–9 do not apply to any other tasks and wrist positions except those that were used in their study.

REFERENCE

SNOOK, S.H., VAILLANCOURT, D.R., CIRIELLO, V.M and WEBSTER, B.S. 1995, Psychophysical studies of repetitive wrist flexion and extension. *Ergonomics*, 38, 1488–507.

Design Cognition

M. G. Helander

Nanyang Technological University, Singapore 639798

1. INTRODUCTION

Ever since Herbert Simon's *The Science of the Artificial* (1969), there has been an increasing interest in design research. His expression "The proper study of mankind is the science of design" remains a challenge to ergonomics since ergonomics has not produced much research. Cognitive scientists have performed most of the research on design cognition. This chapter reviews design cognition. It draws from studies of designers, as they were involved in the design of software, architecture or mechanical artifacts.

Much of the research has addressed either *descriptive models* or *prescriptive models* of design. Descriptive models analyze design behavior and cognition. Contrary to what one may believe, design is not a very systematic activity. There are few procedures and no established logic — not even in engineering design. Design work is an activity that is driven by the knowledge and experiences of the designer, and to the outside observer it is a fairly chaotic process. Design is basically an opportunistic activity.

Several prescriptive models have been proposed in engineering research. The purpose is to give structure to design; what should a designer do first and what activities should follow? A good prescriptive model could have many advantageous side effects, such as improved creativity in design solutions and streamlining the communication processes in concurrent design. The aim is to describe prescriptive models and descriptive models in design. Following this description, I will argue that prescriptive models are useful, since they structure not only design activity, but also communication among collaborators in design. Finally there is an overview of the most common research methods in design cognition.

2. PRESCRIPTIVE MODELS OF DESIGN ACTIVITY

Prescriptive models for design is a very large area of research, particularly in mechanical engineering. Since the primary concern of design is to meet the functional requirements of the intended use(r), most design methodologies emphasize functional requirements or user requirements. One of the earliest and most recognized is the Structured Analysis and Design Technique, which was developed by Ross and Schoman (1977) as an aid in software design. SADT was a precursor to object-oriented design. It uses graphical techniques to provide a basic set of primitive constructs from which analysts and designers can compose orderly structures for the system or task to be designed. The system is decomposed top-down in several abstraction hierarchies — from the highest level of System Purpose to the lowest level Components.

Mainly based on the graphic representation of SADT, the US Air Force developed a system, which has become the industry standard: IDEF$_0$ Functional Modeling Methodology. As in SADT, each function in IDEF$_0$ contains inputs (material or information to be processed), outputs (outcome of the processing function), mechanism (resource) and constraints (conditions the function is subject to). At any level of the hierarchy, a function consists of three to six subfunctions, each having the same structure as its parent function. In this way, the graphic model of the hierarchical structure of a system can be decomposed continuously until sufficient details are revealed so that the decomposed tasks become manageable.

Another set of notable development efforts in design theory is the German School of Mechanical Engineering Design. Total Design was presented by Hubka and Eder (1988) and the Systematic Approach by Pahl and Beitz (1988). The methods are similar in that they both partition the design life cycle into four stages: (1) design problem definition; (2) conceptual design; (3) design embodiment; and (4) detail design. Specifying the relationships between functions to be designed and function carriers supports the selection at the conceptual stage.

Suh (1990) introduced Axiomatic Design (AD). This was an attempt to develop a design methodology, which specifies a complete set of functional requirements at the early stages of design. With its clear and strong emphasis on Customer Needs, AD offers a systematic approach, which is particularly well suited to design problems in ergonomics.

2.1. Use of Abstraction Hierarchies, Functional Requirements and Design Parameters

According to Jens Rasmussen any system can be described using at the most five levels of abstraction: purpose, abstract function, generalized function, physical function and physical form. Figure 1 illustrates the design of an ergonomic workstation. Two levels of abstraction are illustrated. The top level is the "Abstract Function" and the lower level is the "Generalized Function." For each abstraction level functional requirements (FR) and corresponding design parameters (DP) are formulated. The FR for the two levels is (a) FR1 and (b) FR21, 22 and 23. Corresponding DP are (a) DP1 and (b) DP21, 22 and 23. Note that for each FR there is only one corresponding DP; DP1 corresponds to FR1, DP21 to FR21, etc. To arrive from FR1 to DP1 one asks the question WHAT. In other words, what type of design parameter DP1 would fulfill FR1.
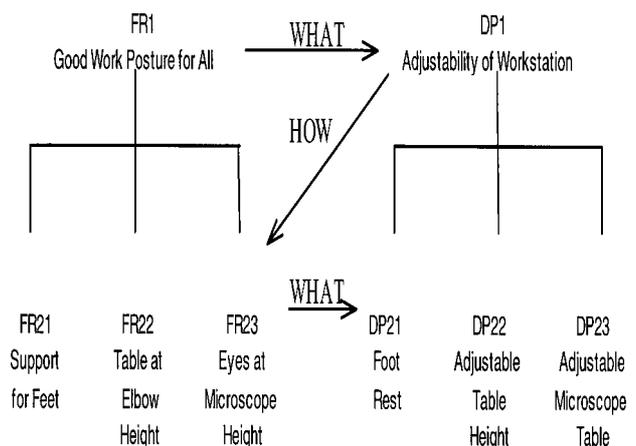
Figure 1. Use of functional requirements (FR) and design parameters (DP) in ergonomics design. There are two levels of abstraction. DP is derived from FR and FR at the lower level from a combination of DP and superordinated FR.

DP1 is then expanded at the lower level into three FR: FR21, 22 and 23. Considering the combination of the requirements stated in FR1 and DP1 derives this expansion. DP21, DP22 and DP23 are then derived from respectively FR21, FR22 and FR23 considering the constraining effect of DP1.

The structured, top-down approach does simplify design work, since the design parameters at the lower levels are easier to identify. Yet, the search for solutions at higher levels of abstraction will still require as much ingenuity as before; design synthesis is not simplified by design procedures.

FR1 could have been satisfied using other design parameters than DP1 — "Adjustability." For example one could have proposed DP1 — "Use standing postures" or DP1 — "Use reclining furniture." In either case there would be totally different design solutions. The choice of DP1 therefore has a constraining effect — once DP1 has been specified many potential design solutions can be ignored (for the better or worse).

Although design synthesis is not simplified by AD, there are other advantages. The functional requirements are emphasized throughout the design work, and the resulting design will be more appropriate and more creative (see below).

Incorporating specifications at the levels FR3 and DP3 can extend the abstraction hierarchy in Figure 1. DP3 would entail a specification of design elements. Corresponding to FR331, 332 and 333, a microscope table has a surface (DP331), legs (DP332) and an adjustability mechanism (DP333). DP4 would specify the materials used in the design. These specifications at the lower level are of interest for manufacturing and determine the price of the finished product. Where such details are considered the ergonomist may have to negotiate design solutions in a concurrent engineering team. All members of the design team must then consider benefits and costs of adjustability, and expensive propositions are traded off against less costly design solutions.

To summarize:

- Design problems are partitioned in an abstraction hierarchy, with discernable input and output.
- Most design methodologies analyze user requirements or customer needs in order to generate a design that satisfies the functions of the intended use.
- To realize both the functionality and the ergonomic aspects of a design, a consistent mapping from user requirements to final design is indispensable.

3. DESCRIPTIVE MODELS OF DESIGN

Designers rarely use systematic top-down design analysis and synthesis, as described in the previous section. Below I will explain why this is the case. I will then propose that it would be a good idea deliberately to impose a top-down design procedure — at least for some types of design.

Several researchers have investigated the problem-solving behavior of designers (Guindon, Helander, Rasmussen). The common conclusion is that designers jump back and forth between abstraction levels driven by their associations and opportunities for problem-solving. Typically a designer may be working with a specific subproblem. She may then have an association related to another subproblem. As a result she will abandon the original subproblem and devote herself to the new subproblem. The alternative would be to take a note about the association and return at a later time. According to Guindon

(1990) this would imply too much trouble. We are reminded of Simon's parable: The complex path of an ant traversing a beach does not reflect complex behavior, rather the complex environment — and the short legs (author's addition). Likewise with design. The complex behavior is not because of complex goals and complex design procedures, but because of the interactions with colleagues and the limitations of the short-term memory. This will lead to situated cognition, with frequent reformulation of goals and constraints. Design solutions may thus be driven by social and mental constraints of the designer, rather than by design logic.

Associations drive the designer and Raymonde Guindon characterized design behavior as opportunistic — rather than systematic. Opportunistic behavior may, however, have detrimental effects on the quality of design. As mentioned, designers may be better off following a top-down design procedure.

Goel and Pirolli (1992) noted several strategies that designers adopt in problem formulation and generation of solutions.

- *The problem is better understood as the solution emerges*
  Designers often explore the problem-and-solution together. Chuck Eastman at Carnegie Mellon University noted that architects instead of generating functional abstractions first generated a design element through sketching and drawing and then determined the implications. The nature of the problem emerges as proposed solutions are examined.
- *The reluctance to discard a design proposition and start anew*
  One problem in engineering design is that premature solutions are adopted. Engineers must be encouraged to abandon the first design and produce alternatives. This helps in staking out the design space and fully understand different trade-offs in design solutions.

As Heckel (1984) noted: "If Ernest Hemingway, James Michener, Neil Simon, Frank Lloyd Wright, and Pablo Picasso could not get it right the first time, why should a designer?" Frank Lloyd Wright noted that the two most important tools an architect has are the eraser in the drawing room and the sledgehammer at the construction site.

However, even when designers encounter severe problems there is a considerable effort to make the initial idea work, rather than to stand back and adopt a fresh point of departure.

- *To arrive at design solutions designers change goals and impose additional constraints*

A design problem is often ill defined or only partially defined (Goel and Pirolli 1992). Many of the design requirements emerge during the design process — new goals are added and new constraints emerge.

Goals and constraints are often not sacrosanct. Many goals are negotiable, such as performance goals and safety goals. Designers understand that additional information must be added in order to generate and justify design solutions. Particularly at the low levels of the abstraction hierarchy, there are often missing constraints, such as geometrical form or color.

Waldron and Waldron (1988) noted in engineering design that "the premises that were used in initial concept generation often proved, on subsequent investigation, to be wholly or partially fallacious. Nevertheless, they provided a necessary starting point. The process can be viewed as entirely self-

correcting, since later work tends to clarify and correct earlier work."

Some researchers claim that design is merely a process to identify constraints. Given all the possible design solutions in a design space, constraints have the valuable property of reducing the design space. Constraints occur at different levels of the design hierarchy: The truck must be able to transport 50 tons (high level) The truck must be yellow (low level). The implicit knowledge of "what not to do" simplifies the design task, but it does not solve the problem in proposing good design solutions.

4. DESIGN CREATIVITY

Creative design is of great economic interest to industry, and has inspired many publications particularly in management and economics. This literature is largely phenomenological. Many procedures are proposed with the intention that they will enhance creativity in groups as well as individually. There are good reasons to be critical to many of these propositions (Finke *et al.* 1992). Designers with extensive experience and knowledge (a rich associational network) create good design and by individuals who possess certain cognitive abilities, which makes it easier to design. Individuals with long experience and knowledge have a greater base for associations, which is necessary for creative design.

According to Donald Norman successful design is driven by a rich set of associations which is coupled to a rich set of emotions. Creativity in design is not a logical undertaking — it is driven more by emotional processes. Philip Johnsson-Laird (1993), a noted researcher in artificial intelligence, pointed out that the emotional and associational nature of design is why artificial intelligence may not succeed. It would be impossible to program the millions of associations of a human designer, and such information would be necessary for creative design. Artificial intelligence will have to resort to common or simple solutions.

In addition to knowledge and experience Fricke (1996) suggested that some cognitive abilities are particularly important; spatial imagination and heuristic competence. The latter refers to the ability of a designer to plan her activities for new types of problems, to recognize subproblems and place them in a correct order of importance. Other factors of importance include the ability to produce sketches and drawings in problem-solving and the level of motivation for design problem-solving.

4.1. Importance of Top-down Procedures for Creative Design

Several researchers have pointed to the importance of adopting top-down procedures in design. Finke *et al.* at Texas A&M University conducted a series of experiments of factors that enhance creativity. Their recommendation is summarized as follows: in designing a new product the designer should dwell at the higher levels of abstraction as long as possible. The designer should formulate the purpose and the top functional requirements and design parameters. Then reformulate them a few times. Any temptation to be sidetracked by associations at the lower levels of the abstraction hierarchy should be avoided. Out of this focus on higher level goals will follow creative design.

It is detrimental to good design if designers in a concurrent team are drawn into discussions of "How it was done last time," such as materials that were chosen, machines that were used in manufacturing, esthetic form and the types of fasteners that were used. These aspects are irrelevant at the early stages. Besides they will eventually follow, once the design formulations at the higher level have been settled. I once observed a design team getting sidetracked by a designer who, without much reflection, offered the statement: "Last time we used spot welding." The ensuing discussion took several hours until one of the team members observed that the team was off track.

Fricke agrees with these recommendations. An experimental study in our own laboratory imposed different procedures for design of displays in an automobile (Caldenfors 1998). Test persons who followed a top-down design procedure and spent much time formulating the top FR and DP produced more creative design as compared with those who did not.

5. TOP–DOWN HIERARCHICAL DESIGN IN CONCURRENT ENGINEERING

One may then conclude that top-down hierarchical design is a process that promotes good design solutions and enhances creative solutions. It seems that these procedures would also be useful in engineering teams. In concurrent engineering several experts such as manufacturing engineers, marketing, electrical engineers, software engineers, human factors engineers, management, industrial design, etc. sit around a table to identify *satisfying* design solutions (Simon 1981). Among the team members there is a give-and-take to identify solutions that can satisfy several different criteria, such as customer needs, low costs, ease of manufacturing, usability, maintainability, etc. Although no research has yet compared different procedures in the concurrent engineering, it would seem that top-down design procedures would be even more important in this case. To be effective, a design team must establish a set of common rules or procedures for their design deliberations, and they may want to keep track of the "level of abstraction" as the discussion proceeds. Not only would this enhance creativity but it would also minimize mis-communication between team members.

5.1. Web-based Design

Web-based design will be common in the future. Here team members may be selected from various parts of the world based on their expertise and expected to interact on the web. If concurrent engineering is difficult around a common table, web-based design will be even more difficult.

There is a great need to establish procedures for web-based design. This would include using a moderator in the design team, who can take notes and monitor the level of abstraction and inform team members when they get off track. There is also a need to communicate sketches and drawings, which enhance communication between team members.

6. NEED FOR SKETCHES AND DRAWINGS

It is common practice to differentiate between "language-like" and "picture-like" representations and to assign them to different cognitive functions. During design work the designer will encounter both. The design brief is usually given as verbal document, but the output of the process is a set of documents including both words and figures.

Several researchers have investigated the needs for sketches and drawings. The conclusion is that designers often need to

make sketches by hand, which seems to clarify thought processes. Sketches develop slowly and support the step-wise progress in thinking about a design. Goel and Pirolli (1992) concluded that present CAD systems cannot be used for sketches. Their systematic symbolic notations to not promote free thinking and exploration of ideas. Sketches must be drawn by hand!

Sketches and drawings are very important for the concurrent engineering team to promote a common language and communication. The problem is that different professionals refer to design elements using different professional lingo. Visualization helps since team members can simply point at design features — there is no need to name them.

7. RESEARCH TECHNIQUES IN DESIGN COGNITION

Design activity is usually unstructured, and designers improvise as design progresses. Momentary intentions, new associations, modifications of constraints and goals will appear as the work progresses. Given the purpose of the research several different methods can be used to study capture and design activity: experiments, verbal protocols, ethnographic methods, observational studies, and the use of interviews and questionnaires. These methods have different advantages and disadvantages. In many cases they may complement each other.

7.1. Verbal Protocols

To formulate cognitive models Verbal Protocols may be the most appropriate technique. The research could involve the investigation of a specific cognitive skill — to be modeled, then a carefully selected design task, which will allow the skill to be observed and a method for recording performance. Verbal protocols provide richness in data, which makes it possible to study complex interactions, which are common in design. As an example Beth Adelson's (1989) protocol studies showed that designers undertake mental simulations to investigate the appropriateness of design features.

Design is inherently complex, and to use cognitive models it is necessary to select a simple design task. The task must fit the purpose of the investigation.

Yet there are difficulties with verbal protocols; they are typically not complete. Subjects must be taught to verbalize their thoughts. This is not easy; some thought processes are visual in nature and difficult to verbalize. Other thought processes may have become automated and are not available for inspection and verbalization. Finally there are problems in deciding what should be evaluated in the verbal stream of information. Verbal protocols are not easy to use.

7.2. Ethnographic Studies

Anthropologists and sociologists employ ethnographic methods to study people "in the wild," as they go about their everyday activities in offices, homes, schools, or wherever they live (Nardi 1997). The main methods are interviews, observation and participant-observation. The methodology developed in the beginning of this century for the study of Indian Native Americans. The researches will live together with the group and participate in activities in daily life.

Ethnographic studies are time consuming — in the case of Indians it took a year of daily interaction. Ethnographic studies have become increasingly used for design of office tasks and human–computer interaction. For a design task a trained ethnographer could spend about six weeks in the design office. One could, as Bucciarelli (1988), study how engineers communicate in design and how they make decisions. Bucciarelli found, among other things, that > 50% of the working time is used for verbal communication between designers.

Ethnographic studies are qualitative; the results are used to obtain information about structures of knowledge, attitudes and behavior. Once this is understood it may be possible to formulate theories of behavior and cognition. These can then be used for controlled experiments to collect quantitative data and perform statistical testing.

Many researchers are attracted to ethnographic studies because of their "political correctness." Nardi pointed out that ethnography is not an excuse to avoid education in statistical methods. Frequent mistakes are done by untrained "ethnographers," the most common of which is to generalize beyond the study's sample population.

7.3. Experimentation in Design Research

Experimentation has one major advantage over other research methods: it is possible to attribute cause and effect. Ideally a theory should be the basis of experimentation and should dictate the design of the experimental task as well as the selection of test persons. Typically one or several factors (independent variables) are investigated as a function of some dependent variable, and analysis of variance is used to evaluate results. The studies by Finke *et al.* (1992) are prominent examples of this approach. They performed a series of experiments to investigate how creativity (dependent variable) can be enhanced in design. Their conclusions were that designers should focus on the functional requirements at the top levels of the abstraction hierarchy (independent variable).

Three types of dependent variables are used in human factors experimentation:
- Human performance data — time to perform a task and Errors in performing a task.
- Physiological data (e.g. heart rate as indicators of workload and/or stress).
- Subjective data from interviews and questionnaires.

7.4. Observational Studies

Observational studies are often performed in ergonomics. Sometimes they are conducted as experiments; sometimes they are merely observations of a "problem" group. In the first case there could be several experimental groups; for example one group in which test persons are exposed to the experimental variable, and one control group.

Often, however, ergonomists perform simplified studies, and many of them are from a scientific perspective substandard (Helander *et al.* 1984). Frequently data is collected from a problem group, without the use of a comparison group. The results from such so called one-shot case studies can not be interpreted.

8. CONCLUSION

Design cognition may be characterized as a situated cognition. The designer relies on knowledge and experience to propose

appropriate and creative design solutions. The process of design, as observed, is seemingly chaotic.

Top-down procedural design offers several advantages, since design intentions are easier to communicate to other designers. In addition, a careful elaboration of functional requirements at the top level promotes creative design.

REFERENCES

ADELSON, B., 1989, Cognitive research: uncovering how designers design; cognitive modeling: explaining and predicting how designers design. *Research in Engineering Design*, 1, 135–42.

BUCCIARELLI, L.L., 1988, An ethnographic perspective in engineering design. *Design Studies*, 9, 159–68.

CALDENFORS, D., 1998, Top-down reasoning on design synthesis and evaluation. Licentiat thesis, Graduate School of Human–Machine Interaction, Linköping University.

FINKE, R.A., WARD, T.B. and SMITH, S.M., 1992, *Creative Cognition* (Cambridge, MA: MIT Press).

FRICKE, G., 1996, Successful individual approaches in engineering design. *Research in Engineering Design*, 8, 151–65.

GOEL, V. and PIROLLI, P., 1992, The structure of design problem spaces. *Cognitive Science*, 16, 395–429.

GUINDON, R., 1990, Designing the design process: exploiting opportunistic thoughts. *Human–Computer Interaction*, 5, 305–44.

HECKEL, P., 1984, *The Elements of Friendly Software Design* (New York: Warner).

HELANDER, M.G., BILLINGSLEY, P.A. and SCHURICK, J.M., 1984, An evaluation of human factors research on visual display terminals in the workplace. *Human Factors Review*, 1, 55–129.

HUBKA, V. and EDER, W.E., 1988, *Theory of Technical Systems: A Total Concept Theory for Engineering Design* (New York: Springer).

JOHNSSON-LAIRD, P.N., 1993, *Human and Machine Thinking* (Hillsdale: Lawrence Erlbaum).

NARDI, B.A., 1997, The use of ethnographic methods in design and evaluation. In M. Helander, Landauer and Prabhu (eds), *Handbook of Human–Computer Interaction* (Amsterdam: North-Holland).

PAHL, G. and BEITZ, W., 1988, *Engineering Design: A Systematic Approach* (New York: Springer).

ROSS, D.T. and SCHOMAN, K.E. Jr, 1977, Structured analysis for requirements definition. *IEEE Transaction on Software Engineering*, SE-3, 6–15.

SIMON, H.A., 1981, *The Science of the Artificial* (2nd ed.) (Cambridge, MA: MIT Press).

SUH, N.P., 1990, *The Principles of Design* (New York: Oxford University Press).

WALDRON. M.B. and WALDRON, K.J., 1991, A time sequence study of a complex mechanical system. *Design Studies*, 9.

Designing for Body Strength

K. H. E. Kroemer

Professor Emeritus of Industrial and Systems Engineering, Virginia Tech, Blacksburg, VA, USA

1. INTRODUCTION

The engineer or designer wanting to consider operator strength has to make a series of decisions. These include:

Is the use mostly static or dynamic?

If static, information about isometric capabilities, listed below, can be used. If dynamic, other considerations apply in addition, concerning for example physical endurance (circulatory, respiratory, metabolic) capabilities of the operator, or prevailing environmental conditions. Physiologic and ergonomic texts provide such information.

Is the exertion by hand, by foot, or with other body segments?

For each, specific design information is available. If a choice is possible, it must be based on physiologic and ergonomic considerations to achieve the safest, least strenuous and most efficient performance. In comparison to hand movements over the same distance, foot motions consume more energy, are less accurate and slower, but they are stronger.

Is a maximal or a minimal strength exertion the critical design factor?

"Maximal" user strength usually relates to the structural strength of the object, so that a handle or a pedal may not be broken by the strongest operator. The design value is set, with a safety margin, above the highest perceivable strength application.

"Minimal" user strength is that expected from the weakest operator which still yields the desired result, so that a door handle or brake pedal can be successfully operated or a heavy object be moved.

A "range" of expected strength exertions is, obviously, that between the considered minimum and maximum. "Average user" strength is usually of no design value.

Most body segment strength data are available for static (isometric) exertions. They provide reasonable guidance also for slow motions, although they are probably too high for concentric motions and a bit too low for eccentric motions. Of the little information available for dynamic strength exertions, much is limited to isokinematic (constant velocity) cases. As a general rule, strength exerted in motion is less than measured in static positions located on the path of motion (see this volume, Static and Dynamic Strength).

Measured strength data are often treated, statistically, as if they were normally distributed and reported in terms of averages (means) and standard deviations. This allows the use of common statistical techniques to determine percentiles — data points of special interest to the designer (see below). However, in reality, body segment strength data are often in a skewed rather than in a bell-shaped distribution. This is not of great concern, however, because usually the data points of special interest are the extremes. The maximal forces or torques that the equipment must be able to bear without breaking are those above or near the strongest measured data points. The minimal exertions, which even "weak" persons are able to generate, can be identified as given percentile values at the low end of the distribution: often the fifth percentile is selected.

Determination of percentiles can be done either by estimation or by calculation.

Estimation is used when the data set is not normally distributed or too small. In this case, the data point is estimated by counting, weighing, or sample measurement according to the best possible judgment.

Calculation is used based on statistical considerations.

A normally distributed set of n data is described by two simple statistics:

The 50th percentile is by definition the same as the mean m (also commonly called average)

$$m = Rx/n$$

where Rx is the sum of the individual measurements.

The Standard Deviation SD describes the distribution of the data

$$SD = [R(x - m)^2/n-1]^{1/2}$$

It is often useful to describe the variability of a sample by dividing the standard deviation SD by the mean m. The resulting Coefficient of Variation CV (in percent) is:

$$CV = 100 \, SD / m$$

To calculate a percentile value p of a normal distribution you simply multiply the standard deviation SD by a factor k, selected from the table below. Then subtract the product from the mean m if p is below the mean:

$$p = m - k \, SD$$

If p is above the average add the product to the mean:

$$m = m + k \, SD$$

Examples:
To determine 95th percentile, use k = 1.65.
To determine 20th percentile, use k = 0.84.

2. DESIGNING FOR HAND STRENGTH

The human hand is able to perform a large variety of activities, ranging from those that require fine control to others that demand

Table 1. Factor for computing percentiles from mean x and standard deviation s.

Percentile p associated with X		
K	X = X – KS	X = X + KS
2.576	0.50	99.5
2.326	1	99
2.06	2	98
1.96	2.5	97.5
1.88	3	97
1.65	5	95
1.28	10	90
1.04	15	85
1.00	16.5	83.5
0.84	20	80
0.67	25	75
0	0.50	0.50

large forces. (But the feet and legs are capable of more forceful exertions than the hand, see below.)

One may divide hand tasks in this manner:

- Fine manipulation of objects, with little displacement and force. Examples are writing by hand, assembly of small parts, adjustment of controls.
- Fast movements to an object, requiring moderate accuracy to reach the target but fairly small force exertion there. An example is the movement to a switch and its operation.
- Frequent movements between targets, usually with some accuracy but little force; such as in a assembly task, where parts must be taken from bins and assembled.
- Forceful activities with little or moderate displacement (such as with many assembly or repair activities, for example when turning a hand tool against resistance).
- Forceful activities with large displacements (e.g. when hammering).

Of the digits of the hand, the thumb is the strongest and the little finger the weakest. Gripping and grasping strengths of the whole hand are larger, but depend on the coupling between the hand and the handle. The forearm can develop fairly large twisting torques. Large force and torque vectors are exertable with the elbow at about right angle, but the strongest pulling/pushing forces toward/away from the shoulder can be exerted with the extended arm, provided that the trunk can be braced against a solid structure. Torque about the elbow depends on the elbow angle.

The literature (see below) contains many sources for data on body strength that operators can apply. While these data indicate "orders of magnitude" of forces and torques, the exact numbers should be viewed with great caution because they were measured on various subject groups of rather small numbers under widely varying circumstances. It is advisable to take body strength measurements on a sample of the intended user population to verify that a new design is operable.

Note that thumb and finger forces, for example, depend decidedly on "skill and training" of the digits as well as the posture of the hand and wrist. Hand forces (and torques) also depend on wrist positions, and on arm and shoulder posture. Exertions with arm, leg, and "body" (shoulder, backside) depend much on the posture of the body and on the support provided to the body (i.e. on the "reaction force" in the sense of Newton's Third Law) in terms of friction or bracing against solid structures.

It is obvious that the amount of strength available for exertion to an object outside the body depends on the weakest part in the chain of strength-transmitting body parts. Hand pull force, for example, may be limited by finger strength, or shoulder strength, or low back strength; or it may be limited by the reaction force available to the body, as per Newton's Third Law. Often, the lumbar back area is the "weak link" as evidenced by the large number of low-back pain cases reported in the literature.

3. DESIGNING FOR FOOT STRENGTH

If a person stands at work, fairly little force and only infrequent operations of foot controls should be required because, during these exertions, the operator has to stand on the other leg. For a seated operator, however, operation of foot controls is much easier because the body is largely supported by the seat. Thus, the feet can move more freely and, given suitable conditions, can exert large forces and energies.

A typical example for such an exertion is pedaling a bicycle: all energy is transmitted from the leg muscles through the feet to the pedals. For normal use, these should be located directly underneath the body, so that the body weight above them provides the reactive force to the force transmitted to the pedal. Placing the pedals forward makes body weight less effective for generation of reaction force to the pedal effort, hence a suitable backrest should be provided against which the buttocks and low back press while the feet push forward on the pedal.

Small forces, such as for the operation of switches, can be generated in nearly all directions with the feet, with the downward or down-and-fore directions preferred. The largest forces can be generated with extended or nearly extended legs; in the downward direction limited by body inertia, in the more forward direction both by inertia and the provision of buttock and back support surfaces. These principles are typically applied in automobiles. For example, operation of a clutch or brake pedal can normally be performed easily with an about right angle at the knee. But if the power-assist system fails, very large forces must be exerted with the feet: in this case, thrusting one's back against a strong backrest and extending the legs are necessary to generate the needed pedal force. The largest forward thrust force can be exerted with the nearly extended legs which leaves very little room to move the foot control further away from the hip.

Information on body strength has been compiled, for example, in NASA and US Military Standards; by Karwowski and Marras (1999); Kroemer, Kroemer, and Kroemer-Elbert (1997, 2000); Weimer (1993); and Woodson, Tillman, and Tillman (1991). However, caution is necessary when applying these data because they were measured on different populations under varying conditions.

REFERENCES

KARWOWSKI, W. and MARRAS, W.S. (eds), 1999, *The Occupational Ergonomics Handbook* (Boca Raton, FL: CRC Press) (pp. 205—27).

KROEMER, K.H.E., KROEMER, H.B., and KROEMER-ELBERT, K.E., 1997, *Engineering Physiology* (3rd ed.) (New York, NY: Van Nostrand Reinhold/Wiley).

KROEMER, K.H.E., KROEMER, H.B., and KROEMER-ELBERT, K.E., 2000, *Ergonomics: How to Design for Ease and Efficiency* (2nd ed.) (Englewood Cliffs, NJ: Prentice-Hall).

WEIMER, J., 1993, *Handbook of Ergonomic and Human Factors Tables* (Englewood Cliffs, NJ: Prentice Hall).

WOODSON, W.E., TILLMAN, B., and TILLMAN, P., 1991, *Human Factors Design Handbook* (2nd ed.) (New York, NY: McGraw-Hill).

Driver Perception and Response

P. L. Olson

3545 Sulgrave Pl., Ann Arbor, MI 48105, USA

1. INTRODUCTION

With some regularity drivers encounter situations that impose demands over and above those represented by routine operations. In each of these the driver must perceive a potential hazard, meaning they must become consciously aware of it, and then diagnose it. With that done they must then respond, the first phase of which is to arrive at a decision as to what, if any, response is called for. If a response is indicated, the driver must then carry out the required action. Owing to the mental processing involved, each of these encounters takes time, even in those cases where no discernible response is made. In most driving situations events are spaced sufficiently and there is adequate time available. In others they come thick and fast, placing the driver under some stress and increasing the likelihood of error.

The time that it requires for drivers to perceive and respond to unexpected hazards is of considerable interest to civil and traffic engineers, who must design roadways and traffic control systems that take into account human performance capability. Organizations such as The American Association of State Highway and Transportation Officials (AASHTO) have formulated guidelines for driver perception and response for various conditions. One example is stopping sight distance (SSD), which establishes such things as minimum radii for curves and hill crests to ensure that hazards will not be hidden from the driver until he/she is too close to stop. The equation for SSD includes an estimate of driver perception and response time, currently set at 2.5 s.

Not only does a driver face numerous events requiring his/her attention, but also these will differ in the ease with which they can be detected, identified and dealt with. That being the case, it should be clear that the time required for the driver to resolve each will vary significantly. The research on driver perception and response has focused on relatively straightforward situations, telling us something about the minimum times required to respond to unexpected events. Left largely unresolved is the question of times associated with those characterized by greater complexity.

This chapter is intended to provide an overview of what is known about driver perception and response. It will begin with a definition, then progress to a description of basic and applied research on the subject. There will also be a discussion of situations that do not fit the basic model of driver perception and response.

2. BACKGROUND

2.1. Definition

Perception–response time is an interval that starts when some object or condition enters the driver's visual field and ends when the driver has made a discernable response (e.g. foot on the brake pedal and/or the hands start to turn the steering wheel). The time to complete the elected maneuver is normally not included.

Some situations of interest are not characterized by a clearly defined entry of the hazard into the driver's field of view. These are discussed later. It is also recognized that the start point for a perception–response interval may not be a visual stimulus. For example, the driver may hear a sound or sense that his/her vehicle has struck or run over something. Sometimes a driver will perceive an immediate threat and initiate a prompt response to this type of stimulus. In the absence of such a perceived threat the usual response is a visual search. With one exception, all of the research on driver perception and response has used visual stimuli. In the one exception drivers anticipated that a sound stimulus would be used.

2.2. Stages of Perception and Response

The time between the appearance of some hazard and a response is filled with certain activities. It is helpful to divide the perception–response interval into stages as a means of understanding both the nature of these activities and how altered circumstances can change the time required to respond. Not all authors agree as to the number of stages and the labeling. What follows is commonly used.

1. Detection. The detection interval starts when a hazard enters the driver's field of view and ends when the driver has become consciously aware that something is present. It is possible for the hazard to be in the field of view for some time before the driver becomes aware of it. This is more likely to happen when the driver is very much burdened, when the stimulus is small, has poor contrast, or first appears well off to the side. In such cases the detection interval may be lengthened.

2. Identification. Having become aware of something, the driver must next acquire sufficient information to be able to decide what action is appropriate. Identification in detail is not necessarily required. The driver may simply decide it is something he/she doesn't want to hit. If the object is moving or capable of movement the driver must also make an estimate of its speed and probable trajectory as part of the identification process. The identification stage is a frequent source of difficulty, with consequent lengthening of response time, particularly under adverse visibility conditions such as nighttime, fog, blowing dust, snow and rain.

3. Decision. With identification complete the driver must decide what action, if any, is appropriate. In most of the perception–response studies there was a choice between steering and braking. Sometimes there is no real choice available, so the decision interval may be shorter. On the other hand, the choice might be complex. Where the driver shows evidence of considering alternatives the decision interval may be lengthened appreciably. If, for example the driver says that he/she wanted to steer but checked and found the adjacent lane occupied, a minimum of 1 s must be added for each mirror checked.

4. Response. In this stage instructions are issued by the brain to the necessary muscle groups to carry out the intended action. The perception–response interval ends when the foot contacts the brake pedal and/or the hands begin to turn the steering wheel. Because the hands are already on the steering wheel, steering response time is slightly shorter than in the case of braking, where the foot must be moved to the brake pedal from the accelerator. Research has shown that foot-

movement time under emergency conditions is typically in a range from 0.15 to 0.25 s.

The first two of these stages are commonly referred to as perception, the last two as response. Time estimates are often given for each separately. There is no justification for doing so, since everything is internalized between the appearance of a stimulus and the driver's response.

3. DRIVER PERCEPTION AND RESPONSE

3.1. Background

Investigations of human perception–response time have been carried out for nearly 200 years. Much of this work has been directed toward a variety of basic issues and has no direct bearing on motor vehicle operation. In the simplest of such studies the subject may rest a finger on a telegrapher's key with instructions to release it as soon as a stimulus (light or sound) is presented. Clearly, detection, identification and decision are reduced to minimum values by such a procedure. An approach such as described will typically yield response times ranging from 0.16 to 0.25 s.

Among the conditions examined have been so-called "choice" studies. For example, the subject may be given a number of possible stimuli and required to respond to only one of them, which requires identification. Or the subject may be required to respond differently to different stimuli, which requires decision. Expanding the identification and/or decision requirements lengthens response time, as would be expected. It has been found that one choice yields an average perception–response time of 0.2 s. Faced with two choices response time increased to an average of 0.35 s. Additional choices added ~0.05 s each, up to a maximum of nine, at an average of 0.65 s.

3.2. Driver Perception and Response

Early investigations of driver perception and response were often carried out in simple simulators or a vehicle operating in light traffic. A signal (light or sound) would be presented and the driver was instructed to respond by pressing on the brake pedal. It has typically been found that 85 percent of subjects in such investigations respond within 0.75 s. While it is true that this technique complicates detection by requiring the subject to cope with a driving task, the nature and location of the stimulus are known and the response has been specified by the experimenter. Because detection, identification and decision are greatly simplified in such a test, the response times measured should average shorter than under real-life conditions.

Most of the better investigations of driver perception and response have been carried out using one of two approaches:

- Take experimental subjects who have been misled concerning the purpose of the study and arrange for them to have a surprise encounter with some apparent hazard.
- Create some situation on a normal highway and measure the response of passing drivers.

A number of investigations have been reported using these approaches. The results of one of these are shown in Figure 1.

Three situations are depicted in Figure 1. The one labeled "Surprise" shows the range of response times from a group of subjects who had an unexpected encounter with an apparent hazard in the road over a hillcrest. The shortest time recorded was ~0.8 s in this test, and 95% of the subjects had responded
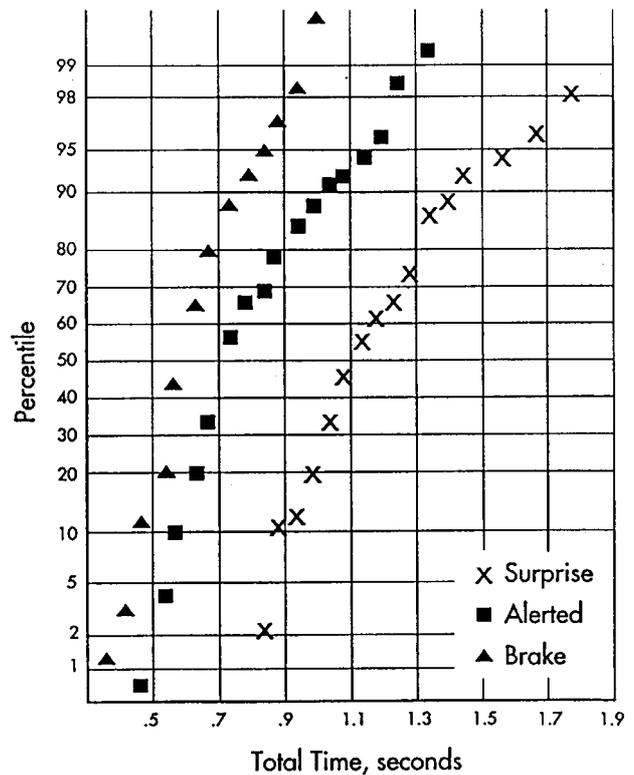
Figure 1. Distribution of response times from 49 young subjects fewer than three conditions. (Reproduced with permission from *Human Factors* 1986, 28. Copyright: Human Factors and Ergonomics Society. All rights reserved.)

by 1.6 s. The subjects were next told to retrace their path, and tap the brake pedal when they saw the hazard (alerted condition). There was some uncertainty in detection, because the hazard was moved each trial, but identification and decision were greatly simplified. Of the responses, the shortest was ~0.5 s and 95% of the responses were 1.15 s or less. Finally, in the so-called "brake" trials, a light was attached to the hood of the car and the subjects were instructed to tap the brake pedal each time it came on. Detection, identification and decision were reduced to minimums in this approach. The shortest response time recorded under this condition was 0.35 s and 95% of the responses were ~0.8 s or less.

The results of this investigation were typical of a number of others that have been reported. In general, they suggest that for a straightforward situation (i.e. a reasonably clear stimulus, and no problems in identification or decision) approximately 90% of drivers will respond within a range of 0.75–1.50 s. Obviously, not all situations are straightforward. In the discussion of the stages of perception and response a number of conditions that can affect the total time required to respond were briefly covered. The following discussion is intended to describe several driver-related factors that can affect performance.

4. DRIVER-RELATED FACTORS

4.1. Expectancy

Expectancy refers to a predisposition on the part of people to believe that something will happen or be configured in a certain way. All drivers have expectancies about roads, signing and the

behavior of other roadway users. So long as expectancies are met drivers will respond promptly and appropriately, resulting in greater throughput and fewer conflicts. Violations of expectancy adversely affect the detection and identification stages. As a result people take longer to respond to a given situation, thus increasing the likelihood of error.

4.2. Chemicals, Driver Fatigue

There are a great number of drugs, most of which are legal, that have an effect on the central nervous system and, potentially at least, on perception and response. They can affect detection, identification and decision. There is little research on most of these as concerns their effect on driving-related skills. An exception is alcohol, which can have a marked effect on information processing, and hence on perception and response. Alcohol has long been known to be a major factor in traffic collisions.

Fatigue is a common problem in motor vehicle operation. It happens that sometimes driving is a fairly boring activity that can continue for many hours at a stretch. A number of investigations have shown that response time does increase with driving exposure. There is also evidence that fatigue exacerbates the effects of alcohol.

4.3. Age and Sex

There is much information to suggest that perception–response time increases with age. Data from one very large study, using a simple simulator, found that average perception–response time ranged from 0.44 s in a 20-year-old group to 0.52 s in a 70-year-old group.

On average, women respond slower than men in standard perception–response tests, although the differences are small. For example, one large study found the average difference between males and females in a choice response time test was ~0.08 s.

5. WHEN THE CONDITIONS OF THE PERCEPTION–RESPONSE MODEL ARE NOT MET

Perception–response time was earlier defined as an interval that starts when some object or condition enters the driver's field of view and ends with a discernable response. Some situations do not fit this model. Probably the most common are cases where there is no clearly defined entry of the object or condition into the driver's field of view. This would most often occur under conditions that restrict visibility, such as nighttime. Typically under these conditions the hazard gradually becomes better defined as the distance closes. From the point where it just becomes possible to discern something to where all reasonably alert drivers will have detected it can cover hundreds of feet and a number of seconds. Picking a start point for the detection interval in such cases is an exercise in futility. In many such cases the simplest solution is to start the perception and response time after detection has been accomplished, i.e. start with the

identification stage. The question now becomes how much time should be allowed for perception–response absent a detection interval? Since there are no data on the length of the detection interval itself, it is necessary to estimate it. It is probable that under the circumstances of a straightforward situation the detection interval will be fairly short. Thus, a reasonable estimate of perception–response time absent the detection interval is 0.6–1.25 s.

The second case comes about generally from instances where the driver initially makes an erroneous identification. As an example imagine a vehicle proceeding along a through street, approaching an intersection. The driver sees a vehicle approaching on the cross street and notes that it slows to a stop, as appropriate. The next thing the driver realizes is that the vehicle is moving forward again and is on a collision course.

There are actually two perception–response intervals here. In the first the driver on the through street detected the car on the cross street, identified its actions as stopping, decided that it would properly yield right of way and no action was called for. The second interval starts with the realization that the car has not yielded right of way and a collision is imminent. The second interval actually started with an identification, i.e. that the car is coming and is on a collision course. All that remains for the driver is to decide what to do and do it.

In such cases, without detection or identification stages, the response time should be fairly short. Once again, absent data, one can only approximate what it would be. For straightforward situations a reasonable estimate would probably be in the range 0.40–0.75 s.

6. CONCLUSIONS

The time for a driver to perceive and respond to a hazard is a matter of some concern to persons involved in traffic engineering as well as those who are in the business of accident investigation. The matter is complex, due to the great variety of situations that can occur while operating a motor vehicle. It should be very clear that perception–response time is not a one-size-fits-all number.

Research on the time to respond to unexpected hazards has tended to concentrate on straightforward scenarios. The result is that there is information about the response of drivers to relatively simple situations. These data indicate that most drivers will respond to such situations within ~1.5 s. Many situations are not straightforward and, other than the fact that responses will take longer, little is known about how much longer. Unfortunately, given the complexity of carrying out realistic perception and response investigations, and the great variety of situations that can arise in the world of transportation, this is not something that is likely to be resolved in the foreseeable future.

REFERENCE

OLSON, P.L. and SIVAK, M., 1986, Perception–response time to unexpected hazards. *Human Factors*, 28, 91–6.

Dynamic Situations

J.-M. Hoc

CNRS, University of Valenciennes, France

1. SUPERVISORY CONTROL

Dynamic situations are found in those work settings where the human operator partly controls a technical or physical process which has its own dynamics and which is most of the time also controlled by automatic machines (Hoc, J. M., Cacriabue, P. C. and Hollnagel, E. (eds), 1995). In such situations, the operator's tasks can vary from supervision (high automation and normal operation) to control (low automation and incident recovery) and are currently qualified as supervisory control (Sheridan 1992). Dynamic is used in contrast with static. A static situation does not change unless the operator acts on it (e.g. using a text editor). Various examples of dynamic situations can be found in industry (nuclear power plants, refineries, iron and steel industry, manufacturing, etc.), mobile driving (car, aircraft, or ship), traffic control (air or car traffic control), crisis management (fire-fighting or disaster management) or medicine (anesthesiology, but also patient management by general practitioners).

For a long time many studies have been devoted to industrial situations qualified by the term process control, where there has been a rapid shift from manual control to remote control bearing on automation. This shift was shown to have a direct implication on human–machine system reliability. Recent accidents in nuclear power plants or aviation reveal that, although high progress has been realized on the machine reliability side, the accident rate remains stable and can be attributed to the human–machine cognitive coupling (Hollnagel 1993).

Several authors have enlarged this process control category trying to characterize these various situations by their inherent complexity. For example, Woods (1988) describes four sources of complexity in dynamic situations. Problems are not stable (as in electronic troubleshooting, for example, where the system is out of work during diagnosis), but evolve along time (most often a dynamic system cannot be stopped during incident diagnosis and recovery). These systems are composed of numerous and interactive components (multiple incidents, causes, goals, tasks, and artificial and human agents). Their behavior is subjected to a large amount of uncertainty due to uncontrolled factors. They are subjected to various risks, and different levels of action must be considered to manage these risks. Static situations also can be complex to deal with and the management of a temporal structure is probably the main feature of the dynamic situations. In addition, the operators must adjust the dynamics of their own cognitive processes to the dynamics of the technical or physical process.

2. COGNITIVE DIMENSIONS

Most of the cognitive dimensions that are used to classify dynamic situations refer more or less to time. They are useful for making choices of appropriate observational or experimental methods in order to identify cognitive problems and find the best solutions. Five dimensions are selected here (a more complete account would need greater detail) that define situations taken as actual interactions between operators and tasks. The same task can result in very different types of situations when different kinds of operator are confronted with it. For example a task can be a problematic situation for a beginner and a routine situation for an experienced operator (expert, but in the sense of action-relevant knowledge).

2.1. Supervision Span

Supervision span at the same time concerns information and action. From the information point of view, the operator can have access to the process variables through a limited window that can be defined in temporal, causal, or spatial terms, sometimes isomorphic. A restricted information span can result in difficulty to anticipate and to make decisions at the right time. However, operator expertise plays a major role in determining the information span. A beginner can be unable to process a large information span due to having too high a level of complexity to manage. In contrast, an expert may reduce such complexity by using chunks or high-level schemas. From the action point of view, the supervision span can be defined as the set of variables accessible, more or less directly, to action. Restrictions can lead to an inability to act on the causes of the phenomena or prevent their occurrence, and also to restrict the action needed to deal with their consequences in the best way possible. The size of air traffic sectors is a typical example of the problems caused by a restriction of the supervision span. As the workload increases in a sector, it is possible to break it down into smaller ones — but there is an obvious limit. Due to the high speed of aircraft, too narrow a sector can result in an impossibility to anticipate conflicts and act in time.

2.2. Control Directness

The partial control inherent to dynamic situations implies that operator actions on the crucial process variables are very seldom direct. They trigger causal chains with intermediary effects on mediated variables before producing their results. At each node of such a chain, some factors can intervene and interact with the operator action without being fully controlled or predicted. That is why diagnosis is strongly intermixed with monitoring. Intermediary effects on intermediary variables are used as intermediary feedbacks to correct the action and to adjust it to the actual situation, which can be quite different from the anticipated situation. In addition, response latencies must be considered in the choice of action time. Causal chains and response latencies result in planning needs, integrating the operator actions and the dynamics of the process at the same time. Beginners are shown to act on the basis of present data with a risk of getting an effect at the wrong time whereas experienced operators are able to anticipate a future state and to act in order that the action effect occurs at the right time.

2.3. Process Information Accessibility

Many dynamic situations are opaque in the sense that the crucial variables used by operators to make decisions are not directly accessible. Often these variables must be calculated or assessed on the basis of surface variables, which can sometimes play the role of symptoms enabling operators to diagnose syndromes. Some effort has been devoted in control room computerization to make "visible" such hidden variables, for example, through "ecological interfaces" (see this entry). However, before being routinely used, such interfaces require deep knowledge to be understandable .

2.4. Process Continuity

A dynamic situation is continuous when the operators process trends, discontinuous (or discrete) if they process discrete states (e.g. discrete manufacturing). In the first case, the temporal functions representing the trends are interpreted in semantic terms (e.g. a particular phenomenon inferred from the curve shape) and action is decided to change the trends (e.g. to modify the phenomenon if it is undesirable). In the second case, the traditional state-transformation models are applied (networks where the nodes are the states and the arrows the transformation to go from a state to another). This distinction is related to the kind of representational system used by operators and has implications on the design of human–machine interfaces. However, discrete manufacturing, for example, opens another source of complexity when parallel operations must be considered when scheduling.

2.5. Process Speed

Process speed can be defined in relation to the sampling frequency (band width) necessary to avoid the risk of omitting to notice an important event or omitting to act at the right time. When this frequency is low, the process is slow (e.g. a frequency of 10 min is largely sufficient to supervise a blast furnace), rapid if not (e.g. fighter aircraft piloting requires an almost continual monitoring). Process speed introduces temporal constraints in the development of cognitive processes. Real-time planning is possible when the process is slow, whereas pre-planning is needed when the process is rapid. In the latter case, pre-planning implies risk anticipation to prepare appropriate responses to possible dangerous events

in order to use routines compatible with the process speed and the operator's skill.

3. REPRESENTATION AND PROCESSING SYSTEMS

Within many kinds of (dynamic or static) situation where machines are implied, operators have been shown to make use of mental models (Moray 1988) to manage them. These models have been proved to be strongly dependent upon action repertory and objectives (e.g. a motor mechanic has a representation of a car that is very different than that used by a common driver). They ensure a tight coupling between representation (objects, properties, including procedural properties) and processing, and can be called Representation and Processing Systems (RPS). In dynamic situations frequent shifts between different types of RPS have been demonstrated, which imply the design of different kinds of interfaces and of links between them. A Causal RPS enables operators to process causal (and temporal) relations between variables, especially in problem-solving situations where the supervised process can deviate considerably from its normal operation. In this case, the causality introduced by the designer can be invalid and a new and valid causality must be discovered to understand what is going on. A Functional RPS gives access to a break down of the process into goals and subgoals and can play a major role in planning and focusing on deficient functions. A Topographical RPS can guide topographical search towards the appropriate physical components upon which to act concretely or towards the deficient component to repair. A Transformational RPS is appropriate to support procedure discovery or application, especially in discrete situations, but also
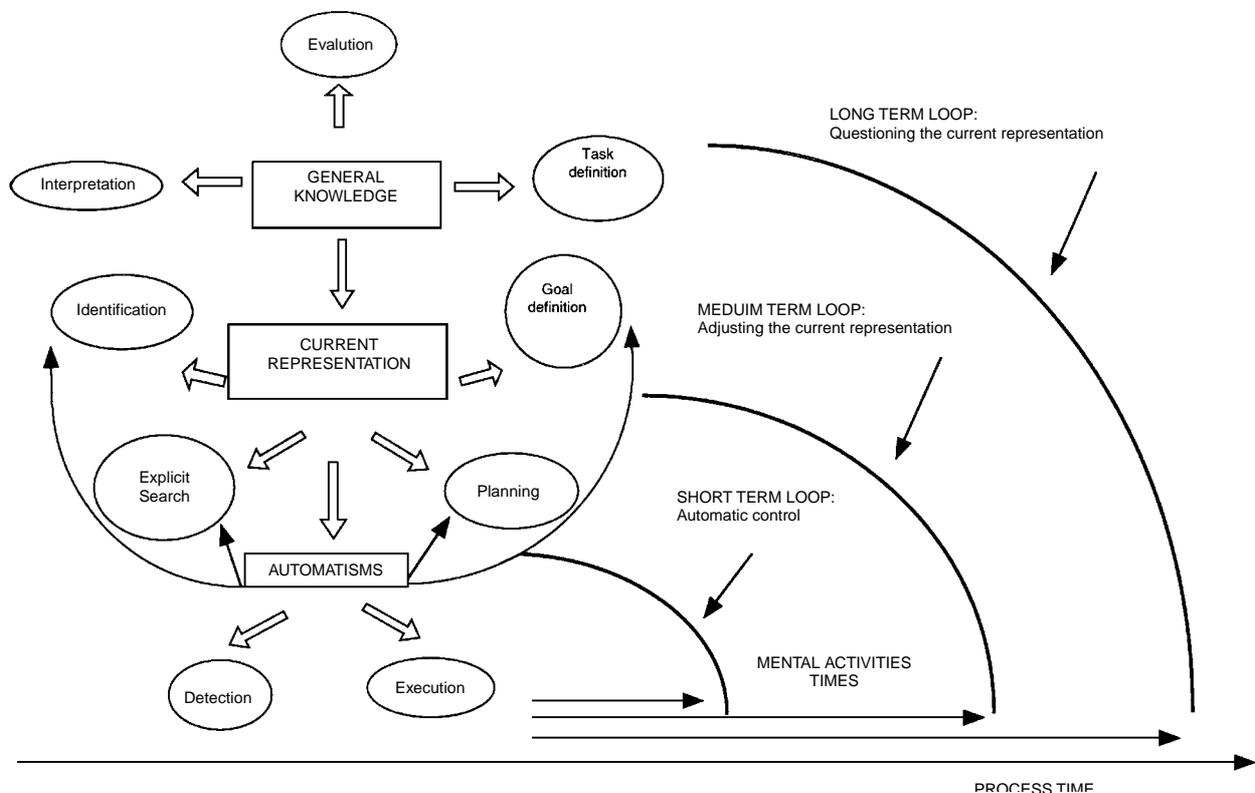
Figure 1.   The dynamic situation management architecture (after Hoc, Amalberti, and Boreham 1995).

Note: Rectangles represent cognitive control levels, ovals elementary activities, arrows influences.

in continuous ones (e.g. the extended use of procedures in nuclear power plants).

4. COGNITIVE CONTROL

After Rasmussen (1986), cognitive modeling of operators bears on a multi-processor conception of cognitive activities, roughly defined at three levels. Knowledge-based (KB) behavior relies on symbol (or concept) processing in problem-solving situations. Rule-based (RB) behavior processes signs requiring attention before triggering rules to be applied routinely. Skill-based (SB) behavior does not require attention and can be done automatically without suffering the well-known limitations in the attention channel. Rasmussen has proposed a model of diagnosis and decision-making of a sequential and procedural kind, from alarm detection to action implementation, including diagnosis, prognosis, and planning, except for some shortcuts permitted by the activity automatization (in experienced operators) from KB to RB, and SB behavior. More recently, Hoc, Amalberti, and Boreham (1995) have enlarged this model by a parallel cognitive architecture (see Figure 1) to deal with the necessary compromise between comprehension (diagnosis and prognosis) and action in dynamic situations. A certain level of comprehension is needed to act appropriately, but the priority to the short-term control of the situation does not always enable the operator to comprehend the situation fully before acting.

At the center of this architecture a Current Representation (CR) of the situation plays a major role in triggering different kinds of activity, on the basis of comprehension or incomprehension. This notion is close to that of Situation Awareness (Endsley 1995) — representation of the past, present, and future technical process evolution. But CR also integrates operator mental processes — plans and action decisions, data for risk and resource management (including internal risk and resource management by the means of metaknowledge — knowledge of one's own cognitive processes). CR can make use of general knowledge (KB level), can trigger different kinds of activity submitted to an inner determination (including RB level), and can supervise human automatisms (SB level).

In parallel, human automatisms can be executed to control the technical process without attentional control. However, SB behavior traces can activate some representations that will be processed at the CR level. These representations could be either confirmation or invalidation of the actual comprehension and play a role in error detection.

Because they are adaptive, cognitive processes need more or less frequent feedbacks to be effective. Roughly, three feedback temporal spans can be considered. A short-term feedback is necessary to the adaptation of human automatisms. This loop can be very efficient even if the comprehension of the situation does not correspond exactly to the current situation (e.g. the pilot will continue flying the aircraft even if some problem is not understood). Local action adjustments can be sufficient before solving the problem at the attentional level (CR). In parallel, the CR must be adjusted to comprehend the situation, but the feedbacks necessary to comprehend and to test the diagnosis/prognosis need the creation of a medium-term loop before reconfiguring the automatisms to be really adequate. Sometimes, different possible explanations of the problem or different problems are triggered that must be processed in a time-sharing way at the attentional level. Finally, RC can be deeply questioned. Adjustments are necessary, but not sufficient, to account for the new situation. In this case, a long-term loop is engaged that will produce its feedbacks at a late stage and reconfigure RC.

Such coupling/decoupling between the different levels of control will vary in relation to situation features. Very roughly, the coupling is at a maximum when the situation is well known and/or the process is slow. Conversely, the decoupling can be very deep when the situation is novel (a long time is needed to reach an appropriate representation and the technical process continue evolving in the meantime) and/or the process is rapid.

REFERENCES

ENDSLEY, M., 1995, Toward a theory of situation awareness in dynamic systems. *Human Factors,* 37, 32–64.

HOC, J.M., AMALBERTI, R., and BOREHAM, N., 1995, Human operator expertise in diagnosis, decision-making, and time management. In Hoc, Cacciabue, and Hollnagel (eds) *Expertise and Technology: Cognition and Human-Computer Cooperation* (Hillsdale, NJ: Lawrence Erlbaum), 19–42.

HOC, J.M., CACCIABUE, P.C., and HOLLNAGEL, E. (eds), 1995, *Expertise and Technology: Cognition and Human-Computer Cooperation* (Hillsdale, NJ: Lawrence Erlbaum).

HOLLNAGEL, E., 1993, *Human Reliability Analysis: Context and Control* (London: Academic Press).

MORAY, N., 1988, Intelligent aids, mental models, and the theory of machines. In Hollnagel, Mancini, and Woods (eds) *Cognitive Engineering in Complex Dynamic Worlds* (London: Academic Press), 165–75.

RASMUSSEN, J., 1986, *Information Processing and Human-Machine Interaction* (Amsterdam: North-Holland).

SHERIDAN, T.B., 1992, *Telerobotics, Automation, and Human Supervisory Control* (Cambridge, MA: MIT Press).

WOODS, D., 1988, Coping with complexity: the psychology of human behaviour in complex systems. In Goodstein, Andersen, and Olsen (eds) *Tasks, Errors, and Mental Models* (London: Taylor & Francis), 128–48.

Ergotoxicology: towards an Operational Prevention of Chemical Risk in the Working Environment

B. Mohammed–Brahim

Laboratoire Santé Travail Environnement, Université Victor Segalen
Bordeaux, France

1. PREVENTIVE ACTION IN INDUSTRIAL TOXICOLOGY

In his definition of industrial toxicology, Robert R. Lauwerys (1990) makes a distinction between *toxicity* and *risk* associated with industrial chemical hazards. Toxicity is defined as the inherent capacity of a substance to produce a harmful effect on the organism. Risk is the probability with which a toxic effect may occur in the conditions of use of a given substance. Curtis D. Klaasen (1986) uses the notion of *safety* defined as the probability that a toxic effect may be avoided in given conditions. Industrial toxicology, defined by Robert R. Lauwerys (1986) is therefore the science which, from the study of toxicity and determination of the risk, establishes measures of safety: "This objective [of industrial toxicology] can be attained if the conditions of exposure and the workers' tasks are defined in such a way that they do not cause an unacceptable health risk."

Yet, although toxicology has been a discipline founded on the discoveries of biology, on the one hand, and the development of industrial analysis, on the other — a fact which has led to the understanding of the metabolism and the mode of action of a large number of toxic substances — it has paradoxically been restrained in terms of condition of exposure to consideration of the immediate environment surrounding the workstation. Thus, there is a list of *environmental factors*, such as noise, light, climate, gravitation or irradiation (Lauwerys 1990), determining *exposure factors*, i.e. mode of administration, duration and frequency of exposure (Klaasen 1986). It has thus been stated that "much more data is needed to characterize the toxicity of a chemical substance than to characterize the risk associated with its use in a given situation" (Doul and Bruce 1986). Whenever explicit reference is made to working conditions, this concerns prescribed conditions: "The evaluation of risk takes into account the harmful effects on individuals . . . which may occur through the use of a substance at a prescribed dose and manner" (Klaasen 1986). In this case, all reference to work activity is avoided. This approach presupposes that the operator is passive, not active, in his own activity, and that he is reduced just to his physical and even biological component. There is an opposition between *real risk*, which is what toxicologists study, and the *risk perceived* by the worker, which may be only slightly or not even correlated with the former (Lauwerys 1986). Safety measures are therefore defined outside of the scope of man and with two possible approaches:

1. The admissible limits of exposure, defined as qualitative hygiene norms for protecting workers. These norms are respected by collective and individual action such as working in a closed environment, ventilated places, wearing masks and protective clothing.

On the one hand, these norms are the result of a decisional process in which the economic and social stakes are at least as important as scientific and technical knowledge. On the other, in a more mass-orientated or even Taylorian approach, they are imposed without taking the specific context or situation into account; the norm is applied to an asexuated human, who is biologically healthy, works eight hours a day, without interruption, five days a week, and is exposed to only one substance at a time. Of course, no situation corresponds to this fictional norm.

2. The medical, clinical and, if possible, biological monitoring of workers. Here, the worker is his own risk indicator. In fact, it is only when the clinical and biological indicators are significantly shifted from the "norm" that preventive action is called for. The occupational doctor therefore prescribes a temporary or permanent halt to the worker's activity according to the official requirements regarding his medical aptitude to continue. The industrial toxicologist then performs metrologic readings and recommends collective (ventilation, humidification, etc.) and individual (protective equipment, safety recommendations) action.

This process is what may be referred to as a "screening model" (figure 1) to prevent toxic risk. The potential risk associated with a prescribed task is in effect contained by the safety recommendations and the protective equipment (first screen). Their role is to maintain the level of exposure at an admissible limit (second screen) depending on the potential toxicity of the substance. Regulation is then achieved on the level of the individual, yet outside his own scope of action, in the way the doctor decides whether or not the worker is able to work in view of his organic status and the potential effect that the substance might have. Medical selection, therefore, constitutes the third and final component of this screen against risk.

2. ERGOTOXICOLOGIC ALTERNATIVE FOR OPERATIONAL PREVENTION

In the 1980s, Alain Wisner (1997) developed an approach called ergotoxicology which included the ergonomic analysis of work within the concept of occupational risk associated with exposure to chemical substances. In this approach, *factors linked to the job itself* and evidenced by ergonomic analysis are taken into consideration to the same extent as the factors of dispersion of the toxic effects produced by the same substance as described by
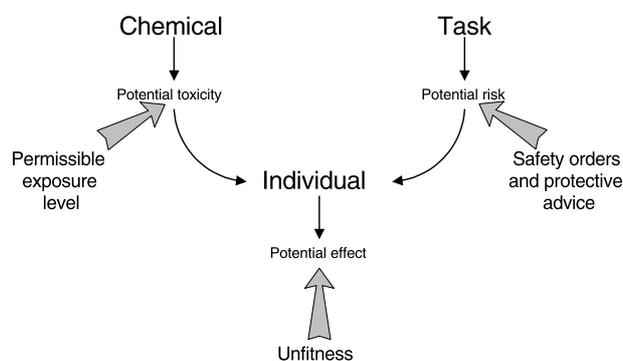
Figure 1. Occupational toxicology preventive model

toxicologists, i.e. factors such as the variability due to extrapolation from an animal model or inter-individual biologic variability. Ergotoxicology, therefore, is a major step forward in relation to the ambition of classical industrial toxicology, i.e. preventing alone, and efficiently, the toxic risk. When speaking of phytosanitary protection, Wisner (1997) adds that "ergotoxicology does not replace the study of biocides less dangerous for man" but that "it tends to prevent or reduce the existing risks". He implicitly recognizes that although toxicology remains the science of the study of poisons, prevention can only stand to gain from ergotoxicology.

Nevertheless, most of the latter developments have borne out the soundness of this cooperative approach in which work and the factors linked to it are considered not at the heart of the process itself, but rather in a complementary manner, just like the other factors already mentioned by the toxicologists. L. Sznelwar (1991) thus speaks of cooperation with the toxicologist to transpose the results of experimentation to the reality of the workplace. The process of prevention is therefore assisted by analysis of work, but in parallel to the toxicological approach according to the model under study.

Since then, many studies have concerned, at the same time, the ergonomic analysis of work, the metrology of the atmosphere, and/or biometrology and epidemiological investigation. With starting points as diverse as the evidencing of a relationship between the occurrence of health problems and the level of exposure or the working conditions (Khaleque *et al.* 1991) or the implementation of a new working process or the substitution of one toxic substance by another less toxic one (Jacobsen *et al.* 1991), these studies try to show beyond the visible activity any factors which might be linked to skill or experience (Bobjer 1991), to the representation and the perception of risk by the operators (Six *et al.* 1991). Other authors have developed techniques for instantaneous measurement coupled with video recording of activity in order to evidence variations in level of exposure with different operative modes in real time (Gressel *et al.* 1983; Heitbrink *et al.* 1983). Films have been used to train operators in chemical risk management (Rosen and Anderson 1989). Beyond the immediate determinants of the work situation, Sznelwar (1992), in an essay on exposure to pesticides in agriculture, discusses aspects relating to anthropotechnology.

We have also analyzed the work situation upstream, in particular, the stages of project design and consideration of the work of the preventers themselves in analyzing the work situation. Another of our research themes has been the wearing, pertinence, and efficacy of individual means of protection against chemical risk in relation to work activity, and the representations and perception of these that the workers may have.

3. ERGOTOXICOLOGY: A SIMPLE SEMANTIC CONSTRUCTION ?

The process in which there is the simultaneous but parallel involvement of various disciplines such as ergonomics, toxicology, and sometimes epidemiology, a characteristic of most of the above-mentioned studies, bears witness to this view of the question. Ergotoxicology is not just a matter of cooperation between the toxicologist and the ergonomist, nor just negotiation of a neck of the woods for the science of ergonomics within the traditional grounds of industrial toxicology. Ergotoxicology could be defined

as a process of analyzing work activity in the light of toxicological knowledge, and destined to evidence the determinants of chemical risk for the health of workers with a view to adopting preventive action in an operational perspective.

Ergotoxicology, therefore, is an alternative model with methods for understanding what is happening for a worker when he is more or less exposed to toxic risk owing to his professional activity. It is an approach which centers on the process of prevention within the activity of work.

Any scientific approach to man at work involves implicit considerations about models of man and the concepts of technology and work (Dejours 1995). Seen in terms of ergotoxicology, and since prevention is to be considered not only in terms of a dose–effect relationship but also as a work activity–response relationship, this notion should also apply to the prevention of toxic risk at work. We will not discuss here, however, the internal debate among ergonomists regarding the models of man and the relationship between work and health (Daniellou 1998).

Beyond the biological dimension underlying the dose–effect relationship, ergotoxicology involves cognitive, mental, and social dimensions in the occurrence of risk in the workplace, and therefore in its prevention. Thus the challenge is to highlight these dimensions in the understanding and prevention of chemical risk.

Since there are many ways in which a chemical substance is used during a transformation process, or when the worker is accomplishing several acts, the notion of exposure is a complex one and it cannot be understood in a deterministic linear manner according to the paradigm chemical substance — level of exposure — health risk. Ergonomic analysis evidences a number of individual factors of a biological, cognitive and mental nature, together with parameters external to the worker. Their interrelation is complex and evolutive, and will involve the material elements of the situation and the representations and perceptions that the worker has of them. Work activity, therefore, involves individual and collective regulations which become manifest in the avoidance strategies adopted by the worker. These regulations lead to a compromise favorable to the continuing state of good health of the worker, or may lead to delayed or chronic effects. They may even fail, causing an accident or illness, or even a renouncement, i.e. the worker's premature retreat from the workplace, accomplished with the worker's tacit complicity, and which corresponds to the *healthy worker effect* as described by epidemiologists.

The following model (figure 2), which places work at the center of the process, includes the norms and protective means recommended by industrial toxicologists within the overall framework of the work situation. It is more likely, in our opinion, to provide fruitful directions for prevention.

4. WHAT HAPPENS IN PRACTICE ?

4. 1. Design of Work Systems and Toxic Risk

We have examined the stages preceding the organization of work as a result of a request to analyze the toxic risk in a viticultural environment. The study centers on phytosanitary treatments and the risks involved in the removal of asbestos from a building.

The preoccupation of a regional health care mutual fund and the regional chamber of agriculture was to reduce the toxic risk

Individual determinants:
- biological: sex, age, health status...
- cognitive and psychic: motivation, training, experience...

External determinants:
technology, organization, norms, safety orders, protective advice, living conditions.

1 ⟶
Initial toxic(s) agent(s)

Transformation process

Polyvalence of worker

Ultimate toxic(s) agent(s)

Activity

Individual and collective regulations

Representations / Perceptions

Avoidance strategies

Failure:
renouncement
accident or illness.

Compromise:
- provisional: delayed or chronic effects
- favourable: maintenance of good health

Figure 2. Ergotoxicology preventive model

inherent in the use of pesticides. For this, the suggestion was to replace chemical treatment with what is known as "rational" treatment. Whereas in traditional pest control, treatment is performed systematically against all harmful agents and as blanket coverage over the whole vineyard according to a pre-established schedule, rational treatment involves selective applications according to forecast models for infection based on real-time diagnosis.

For this, we analyzed the organization of work and the treatment strategies over a year in two vineyards, one using traditional methods and the other the rational strategy. We used available data (timesheets, treatment programs, daily schedules), discussions with winegrowers, vineyard managers, foremen, and workers. Activity regarding vine protection (cropping techniques and pesticide applications) were assessed by observation.

These observations showed that rational treatment saved on the use of the most toxic substances. Nevertheless, analysis of the activity revealed a transformation in the organization of the work, with infrastructural and equipment modifications, a new know-how, more flexibility in the management of work and the staff with risk-taking both for the crop and the workers.

Without a rational approach to the new rational strategy, it was clear that the transformation itself was generating new constraints, with new moments at which the workers were exposed to levels of pesticides sometimes exceeding those to which they had previously been exposed.

The multiplication of the number of applications by staff who were not always qualified, and at moments which were not foreseeable, since they were precisely dependent upon natural conditions, meant that the number of occasions on which

exposure was possible increased. Moreover, the requirement of immediacy involved the mobilization of staff sometimes at a moment's notice. In any respect, chemical treatments were replacing manual prophylactic procedures, and were being conducted in conditions which did not respect the usual rules of efficacy for crops and safety for the workers (e.g. excess dose owing to slowing of tractor when blocked by heavy foliage still in place).

Management of the critical periods, in which the results of the model are at the limit of the admissible risk, represents a gamble taken exclusively by the vineyard manager with possibly disastrous consequences. We noted a case where the decision was taken not to treat: infestation was then suddenly triggered by a storm with continuous rain, so treatment was performed in a hurry with little protection against the leaching of the plants that the rain induced, which in turn represented a subsequent source of contamination for the workers.

It is clear, therefore, that the infrastructure, equipment, and procedures used may constitute constraints. Thus, in the vineyard which was using the rational strategy, an area was set aside for storing and preparing the pesticides. The reduction in quantity prepared and the higher frequency with which the preparations were mixed up quickly made the investment unsound. Consequently, the decision was taken to revert to the traditional mode of application in which the tractor driver both prepared and applied the chemical. Apart from the fact that he had to manipulate the products more, he also came more into contact with them: skin contact with froth from the mixture and the manual elimination of lumps, small-volume reservoirs requiring frequent refilling, etc.

Moreover, the products were stored for a longer period, with a concomitant phenomenon of overdosing at the moment of mixing; in effect, the operators felt that they should add a little extra to ensure the efficacy of a product whose expiry date would otherwise be attained.

In the other observation concerning the steps preceding the removal of asbestos from a site, we analyzed the process of the call for bids. As soon as the budget was made available, it was clear that the process of competition favored the choice of the company making the lowest estimate. Since the implementation of safety measures can represent between 60 and 80% of the total cost of this sort of job, it is easy to see where the savings are made. Moreover, the analysis showed the inadequacy at all stages of the consideration of the wide variability of worksites due to a great number of factors — state of the premises, type of surface asbestos is on, surface to be deflocculated and additional applications, immediate environment of the site, etc. Moreover, there was a lack of anticipation, and an ad hoc solving of problems as they arose, leading to a sense of improvisation undermining the collective and individual mobilization necessary to handle the omnipresent risks. This in turn led to delays with a concomitant lengthening of working time.

Whatever the sector of activity, this example shows that if organizational constraints are not taken into consideration, then the way in which the work is designed and managed can ride roughshod over any reference to exposure norms, the respect of safety recommendations, and the wearing of protective clothing, which are all part of the strategy of prevention.

4. 2. Protective Equipment against Toxic Risk

The issue of protective means against toxic risk may be examined in a number of ways.

4.2.1 Intrinsic quality of equipment

The supply of exterior air is strongly recommended for respiratory protection on sites from which asbestos is being removed. Yet there is still no set of rules applying to devices for producing air, even considering the latest revisions. And in our observations of asbestos defloculation procedures, we noted traces of oil on the protective masks. Levels noted were as great as three times the accepted threshold for divers. Such values were also found during sanding operations using respiratory protection by the supply of external air.

This is a typical example of what we term the screen model for the prevention of toxic risk: a filter is placed between the origin of the asbestos and the lungs of the worker, yet the possibility that there might be some other form of contamination (here, oil residues and carbon monoxide) was overlooked.

4.2.2 What happens in the work situation

In the same example, analysis of the efficacy of respiratory protection in the large number of asbestos defloculation sites now underway has shown the presence of fibers inside the masks worn by the workers, including those equipped with an external air source. Levels range from 0.01 to 7.8 f.cm^{-3}. While dry defloculation leads in all cases to a high level of airborne fibers in a confined zone, the use of wetting agents (water or other impregnating agent) does not notably reduce the level. Indeed, contrary to what might be expected, contamination levels in masks are independent of the degree of airborne contamination. Therefore, it is likely that the relationship between the degree of wetting used and the extent to which contamination is reduced is not a linear one.

On the other hand, contamination is strongly correlated with the following:

- the quality of the instruments used, which may either peel away whole layers of asbestos or serve only to disperse the particles by a ripping or powdering action. This parameter is also dependent on the skill of the worker in using the equipment or on the nature of the support.
- situations in which the worker has to temporarily unplug his external air source because the pipes have become entangled, or others in which special effort is required; such situations call for extra air to be breathed, resulting in an air depression within the mask allowing some fibers to slip through between mask and skin.
- the water jet used to clean the masks should be sufficiently strong to eliminate all traces of asbestos. This is rarely the case, and fibers may remain in the mask and be released at the subsequent moment of use.

4.3. Representations and Perception of Workers and Toxic Risk

Studies in ergotoxicology have shown how the worker not only attributes a meaning to his work but also to the risk to which he is exposed, how he identifies the people who are involved in the risk and the preventive means implemented, and how he interprets these factors through his own operative modes and safety behavior.

It is thus possible, through observation and speaking with the worker, to give substance to the risk that workers perceive. Avoidance strategies have been observed, such as working upwind to avoid pesticide clouds or shifting back when highly irritant sulfur is being prepared. Workers have even been seen to wet the ground where carbon-free copy paper is being unloaded, in order to "avoid the irritating effect of the dust" (in this case, formaldehyde is released but is neutralized by the presence of water). Workers are also known to describe odors as being "strong", "bad", or "sickening", and in a more colorful and perhaps even more discriminating way describe the "sharp" odor of chlorine or ozone, the "rotten hay" smell of phosgene or the "rotten eggs" whiff of hydrogen sulfide. All such odors give warning as to the level or imminence of the danger. Another example is the stench of "bitter almond in the mouth", a characteristic signal for those working with electrolysis and galvanization baths where levels of cyanohydric acid may suddenly rise.

Apart from the perception that workers may have of the presence of a given risk around them, the very representations that they have of them may differ:

- The danger may be perceived outside of the geographical area of risk. For example, a worker may describe how he "carries the danger with him" (e.g. marks on the skin like tattoos, stains or incrustations; bad breath if the product is excreted by the airways) or "takes the danger home to his loved ones" (fibers stuck on clothing worn home).
- The concomitant presence of other risks. The worker may suffer from more perceptible manifestations such as noise or a postural defect.

- The way the preventers behave or speak to the worker as if underestimating the risk, e.g. "I don't wear gloves when handling fungicides because my doctor at work says that they are less dangerous than the insecticides", or an overestimation of the type "if no preventer ever comes into the confined zone, then they can't be as sure of themselves as that!"
- The pertinence of the operative references: the perceived disproportion between the prescription of a few milligrams of insecticide for one cultivated hectare leads to a systematic overdosing by the workers; the trade names thought up by the manufacturers to signify the power of such insecticides (Apollo, Karaté K, etc.) do not offset this form of behavior, and are even more derided by the workers because of their sound.

The underestimation and the non-recognition of the interpretative system of the worker regarding the risk explain, at least in part, the limitations of preventive policies developed and implemented only through the know-how of the industrial toxicologist. In a recent publication, Stewart-Taylor and Cherrie (1998) describe the absence of relation between what they wrongly call the perception of the risk by the worker and the level of exposure to asbestos that they are subjected to on deflloculation sites. By re-assessing the items that they consider as important in explaining the perception that workers have of risk, it becomes clear that this perception is in fact a re-statement of risk as initially perceived by the preventers themselves. In effect, what these authors demonstrate substantiates our hypothesis, i.e. if it is not linked to the representational models of the worker, information on risk has little chance of being applied in the field of prevention.

4.4. Preventers and Toxic Risk

Irrespective of the undeniable impact of such information in improving working conditions, the analysis of activity may evidence practices which run counter to the avowed objective of the preventive measures.

The examples described above show that preventers frequently believe that they are acting upon virgin terrain which is simply to be transformed mechanically in order, in turn, to transform its behavior patterns. In this view, the worker is occulted, becoming a "reflexive agent", i.e. someone who not only develops his own representations of risk in an ongoing manner but also who never totally questions his own interpretative system, however much the preventer may insist on the degree of risk concerned.

The building of a relationship of confidence between employers, bosses and employees is based on a wager; a wager that is made in each interactive situation that may become very difficult to win whenever there are conflicts, which are often very costly for all concerned. Some situations leading to ineffective prevention are the direct result of the difficulties experienced by the preventers themselves:

- managing to keep up with ever-changing rules regarding safety;
- adopting a position between advising and coercing;
- protecting oneself against the risk, which limits ones own fine understanding of the worker's activity;
- hesitating between taking decisions in emergency situations and prudence in view of the fear of being sanctioned for inappropriate decisions.

By developing a risk analysis procedure which includes the action of risk preventers, ergotoxicology not only assesses that action but also examines the messages that it conveys.

4.5. Ergotoxicology and Risk Assessment

The hypothesis of a relationship between the determinants of activity, the regulations set up individually and collectively, and the level of expression of workers to toxic substances has been demonstrated during the search for a system of evaluating the retrospective exposure to pesticides of viticultural workers. Such a tool needed to be developed to demonstrate the relationship between the repeated contamination of workers by pesticides during treatment procedures in the vineyard and the long-term

Unwavering determinants
Division of work
Professional qualification
Machanization
Nature/Amount of treatment

Indirect indicators
Size of vineyard
Type of crop
Vine diseases

Uncertain factors
Individual protective equipment
Safety orders
State of health

Indirect indicators
Clothing
Years employed
Loading equipment
Product packaging

Conduct of pesticides treatment

Determinants of contamination
Quantity of product dispersed at upper airway
Respirable fraction
Quantity of product in contact with exposed skin
Quantity of product on clothing
Quantity of product crossing clothing to reach skin
Fraction of product in contact with skin and absorbed percutaneously
Temperature and humidity of air
Presence of organic solvents, emulsifiers and wetting agents in mixture
Part of body contaminated
State of body contaminated
Food
Nycthemeral cycle

Figure 3. Determinants of the pesticides risk for operator's health

Table 1. **Typical reconstructed activities of vineyard treatment workers and estimated exposure levels of pesticides**

Activity	Level of exposure *(mg/h)
Preparation with liquid/solid substances in open air	15
Pulverization with manual equipment	50
Application with pulverizer mounted on any type of low-vine tractor or high-vine straddler	1
Appllication with pulverizer mounted on intervine tractor for high vines	20
Application with pulverizer mounted on any type of tractor alternating between low and high vines	10
Etc.	

* Values obtained by extrapolation from documented meta-analyses

occurrence of neurologic symptoms. Unlike classical epidemiological approaches for estimating exposure, we used ergonomic work observations. This allowed us to identify the determinants of exposure defined as constant or repeated observables having an impact on the biological factors of contamination recognized in toxicology, and leading to a measurable level of exposure comparable with those described in the literature (figure 3). These observables were sought by a questionnaire given to workers at different phases of their professional life. The next stage was to classify typical homogeneous actions with respect to the factor of exposure. Each action was compared to a level of exposure established from a meta-analysis of published data (table 1). This tool was then applied to 1000 exposed or unexposed workers in order to evidence a relationship between a calculated level of exposure and the development of neuro-behavioral disturbances.

5. CONCLUSION

It would seem that by proposing a *modus operandi* allowing new hypotheses to emerge regarding the investigation and prevention of risks for workers, ergotoxicology is able to transform considerably the traditional notions of risk management in industrial environments.

The norms, safety recommendations, and individual protective equipment that represent the customary arsenal in the fight against risk are now showing their limitations. Neither the training of workers in safety measures nor official obligation to wear protective equipment have proved sufficient to transform risk behavior patterns or to make them be transformed. Preventers are wasting their time if they continue to speak in terms of "all you have to do is . . ." and call into question the impermeability of workers.

Processes which involve the full integration of the work components in an individual's activity, which offer a different model of man as both operator and actor in his work and in the construction of his health, are becoming more frequent in the analysis of a number of risks such as musculoskeletal disturbances of hypersolicitation. On the other hand, this is not the case where chemical risk is involved, partly because the normative approach is powerful in the representation of toxicologists and because ergonomists are reticent to tread a path which indeed seems complex.

REFERENCES

BOBJER, O., 1991, Job enlargement and improved operator competence to reduce RSI in the furniture industry. In Quénnec and Daniellou (eds) *Designing for Everyone* (New York: Taylor & Francis), pp. 960–2.

DANIELLOU, F., 1998, Evolutions de l'ergonomie francophone: théories, pratiques, et théories de la pratique. In Dessaigne and Gaillard (eds) *Des évolutions en ergonomie…* (Toulouse: Octares), pp. 37–54.

DEJOURS, C., 1995, *Le Facteur Humain*. (Paris: PUF).

DOUL, J. and BRUCE, M.C., 1986, Origin and scope of toxicology. In Casarett and Doull (eds) *Casarett and Doull's Toxicology, The Basic Science of Poisons* (New York: Macmillan), pp. 3–10.

GARRIGOU, A., MOHAMMED-BRAHIM, B., and DANIELLOU, F., 1998, Etude ergonomique sur les chantiers de déflocage d'amiante. *Rapport final. OPPBTP,* Paris.

GARRIGOU, A., MOHAMMED-BRAHIM, B., and DANIELLOU, F., 1999, Chantiers de déflocage. Au-delà du "risque amiante". *Travail et Changement*, 243, 6–8.

GARRIGOU, A., MOHAMMED-BRAHIM, B., DANIELLOU, F., DEJEAN, M., and FERRIÈRE, J.L., 1999, How do risk prevention specialists allow for operators' risk management strategies? *International Symposium. Training of occupational health and safety experts: issues at stake and future prospects* (Germany: Mayence).

GRESSEL, M.G., HEITBRINK, W.A., and JENSEN P.A., 1983, Video exposure monitoring: A means of studying sources of occupational air contaminant exposure. Part I: Video exposure monitoring techniques. *Appl. Occup. Environ. Hyg.*, 8(4), 334–8.

HEITBRINK, W.A., GRESSEL, M.G., COOPER, T.C., FISCBACH, T., O'BRIEN, D., and JENSEN PA., 1983, Video exposure monitoring: A means of studying sources of occupational air contaminant exposure. Part 2: Data interpretation. *Appl. Occup. Environ. Hyg.*, 8(4), 339–50.

HERY, M., POSSOZ, C., and KAUFFER, N., 1997, Expositions professionnelles des travailleurs employés sur les chantiers d'enlèvement de l'amiante. *Cahiers de notes documentaires, INRS*, 167(2), 217–24.

JACOBSEN, T., ANGELO, H., and KAMP A., 1991, Substitution of chloroform. A case study. In Quénnec and Daniellou (eds), *Designing for Everyone*, (New York: Taylor & Francis), pp. 463–5.

KHALEQUE, A., WADUD, N., and KHANAM, M., 1991, Industrial pollution and health hazards of workers. In Quénnec and Daniellou (eds), *Designing for Everyone*, (New York: Taylor & Francis), pp. 957–9.

KLAASSEN, C.D., 1986, Principles of toxicology. In Casarett and Doull (eds) *Casarett and Doull's Toxicology, The Basic Science of Poisons* (New York: Macmillan), pp. 11–32.

LAUWERYS, R., 1986, Occupational toxicology. In Casarett and Doull (eds) *Casarett and Doull's Toxicology, The Basic Science of Poisons* (New York: Macmillan), pp. 902–9.

LAUWERYS, R., 1990, Définition et but de la toxicologie industrielle. In R. Lauwerys (ed.) *Toxicologie industrielle et intoxications professionnelles* (Paris: Masson), pp. 3–5.

MOHAMMED-BRAHIM, B., 1996, Du point de vue du travail ou comment sulfater la vigne autrement. Approche ergotoxicologique du traitement phytosaniatire en viticulture. *Mémoire du DESS d'ergonmie, conception des systèmes de travail et santé,* Université Victor Segalen, Bordeaux.

MOHAMMED-BRAHIM, B., DANIELLOU, F., and BROCHARD, P., 1997, "Non pas que nous ne raisonnons pas, mais nous raisonnons autrement": à propos d'un regard de l'ergonomie sur la conduite "raisonnée" du

traitement pesticide en viticulture. *XXXIIème Congrès SELF*, Lyon, 107–14.

MOHAMMED-BRAHIM, B., DANIELLOU, F., and GARRIGOU, A., 1998, *Amiante. Chantiers de retrait secteur 2. Guide pour le médecin du travail. Bordeaux* (Bordeaux: AHI).

ROSÉN, G., and ANDERSSON, I.M., 1989, Video filming and pollution measurement as teaching aid in reducing exposure to airborne pllutants. *Ann. Occup. Hyg*, 33, 137–47.

SIX, F., and VAXEVANOGLOU, X., 1991, Work analysis on respiratory risks evaluation in coking plant In Quénnec and Daniellou, (eds), *Designing for Everyone* (New York: Taylor & Francis), pp. 952–3.

STEWART-TAYLOR, A.J., and CHERRIE, J.W., 1998, Does risk perception affect behaviour and exposure. A pilot study among asbestos workers. *Ann. Occup. Hyg.*, 42(8), 565–9.

SZNELWAR, L., 1992, Analyse ergonomique de l'exposition des travailleurs agricoles aux pesticides. Essai ergotoxicologique. *Thèse de Doctorat d'ergonomie,* Conservatoire national des arts et métiers, Paris.

SZNELWAR, L.I., 1991, L'exposition à des substances biocides utilisés comme pesticides en agriculture. Un essai ergotoxicologique. In Quénnec and Daniellou, (eds), *Designing for Everyone* (New York: Taylor & Francis), pp. 954–6.

WISNER, A., 1997, Ergotoxicologie dans les pays tropicaux. In A. Wisner (ed.) *Anthropotechnologie, vers un monde industriel pluricentrique* (Toulouse: Octares), pp. 179–89.

Fatigue and Driving

L. R. Hartley

Institute for Research in Safety and Transport, Murdoch University,
Western Australia 6150

1. INTRODUCTION

Over the past few years there has been renewed international interest in the impact of fatigue on transportation safety, although research dates back some 80 years. The reasons are several: a better understanding of the causes of fatigue; better information on the numbers of fatigue-related crashes; increasing globalization leading to longer trips; increasing deregulation leading to 24-h operations in many industries including transport; and increasing "just-in-time" delivery of products.

Despite the interest there is no complete agreement on a definition of fatigue. Unlike metal fatigue, human fatigue leaves few telltale signs in a crash. Its presence has to be inferred. It was not until 1947 that Bartley and Chute distinguished between the physiological consequences of prolonged activity, which they termed impairment, and the psychological consequences including poor performance and negative feelings, which they termed fatigue. More recently it has become clear that poor performance reflects the impact of fatigue on the effectiveness of task performance and not a decline in the efficiency of performance. Well before performance has declined there may have been an increased cost to the operator in maintaining good performance when fatigued. This process is recognized in the still widely held view that sufficient motivation will overcome fatigue; a position that is no longer tenable. With this consideration in mind some authors have emphasized that fatigue is a state where the individual declares him/herself unwilling to continue the task any longer, because of the increased cost to them of doing so and their declining performance. This has merit in drawing attention to fatigue as a graded condition of which the end point is sleep, and to its subjective nature. People with practical concerns may want a more operational definition such as "fatigue is a state of declining alertness which eventually ends in sleep". Such a definition has merit in drawing attention to the key deficits which occur such as the decrease in the speed with which decisions are made, including making control maneuvers in a vehicle. The definition suggests an envelope of performance capability that stretches as the individual suffers fatigue increasing both the mean and variability of the performance measure.

Vehicle drivers, both commercial and private, are engaged in a task requiring their continuous attention, they may spend many hours behind the wheel, the environment may be monotonous, they may have less sleep than they would wish and they may be driving at a time when they would normally be asleep. The US National Transportation Safety Board (1995) examined 107 single truck crashes where the driver survived and records were available for the previous 4 days. Of crashes, 58% was said to be due to fatigue and in 18% the driver was found to be asleep. In the fatigue crashes drivers had an inverted sleep-waking cycle, had driven at night with lack of sleep, had < 6 h of sleep or had fragmented sleep patterns. This article is concerned with the three common causes of fatigue: long hours of work, lack of sleep and the time of day, and the two consequences of driver impairment and crashes.

2. THE SIZE OF THE PROBLEM

Most is known about truck crashes and fatigue. In 1985 the American Automotive Association examined 221 truck crashes where the truck was towed away. Surviving drivers and families were interviewed and records tracked to determine working hours. Crashes were determined to be due to fatigue if the driver had worked > 16 h and if the characteristics of the crash were consistent with falling sleep such as running off the road or on to the wrong side of the road. It was estimated that fatigue was the primary cause in 40% and a contributory cause in 60% of crashes. One study concluded that in articulated vehicles the incidence of fatigue crashes ranges between 5 and 10% of all crashes, ~20–30% of injury crashes and ~25–35% of fatal crashes. Studies of car crashes where the crash had similar characteristics of running off the road or a head-on collision when not overtaking show much the same prevalence. A UK study revealed 20% of freeway crashes were attributed to fatigue by car drivers.

These figures show fatigue to be comparable with alcohol in its impact on safety. Alcohol is estimated to be involved in ~5% of all crashes, 20% of injury crashes and 40% of fatal crashes. A recent study demonstrated that 24 h without sleep produces comparable impairment to a blood alcohol concentration of 0.05%. A blood alcohol level just in excess of 0.05% doubles the risk of a crash. Despite the comparability between the crash risk of alcohol and fatigue, the former cause has received a great deal of attention the latter has not. This is undoubtedly due to past ignorance of the contribution of fatigue and the difficulties inherent in measuring and predicting fatigue. There is no breathalyzer for fatigue.

3. LONG HOURS OF WORK

There have been studies of long hours of driving a truck, driving performance and crashes. But there are few studies of the impact of long hours of non-driving work on the crashes and performance of private motorists. Most developed jurisdictions limit working hours for truck drivers presumably in the belief that this controls fatigue. But these limits vary widely across jurisdictions and there is scant evidence they are effective in controlling crash rates.

One of the first studies of the effect of hours of work on crashes found that there were about twice as many crashes during the second half as there were in the first half of the trip, irrespective of how long it was. This suggests length of driving hours does cause fatigue. Another study used a case control approach where details of > 300 trucks which crashed were matched with details of trucks from the same place and time 1 week later. The relative risk of a crash for drivers who did > 8 h driving was twice that of drivers who drove fewer hours. A European study found that accident risk is relatively low when drivers are on the road between 08:00 and 19:00 hours, and have completed < 11 h of work. However, these same drivers show approximately twice the risk of accident when driving between 20:00 and 07:00 hours. Drivers who have accumulated > 11 h of driving and non-driving work show twice the risk of accident between 08:00 and 19:00 compared with those who have worked < 11 h. For those drivers who have completed > 11 h of work, accident risk is doubled

again when driving between 20:00 and 07:00 hours as compared with daytime driving.

The most well-known on-road study of performance examined 12 truck and six bus drivers over 1 week under a range of schedules. Vehicle control, drowsiness, hazardous incidents, performance on tests and subjective ratings were all measured. When differences in the road geometry were taken into account vehicle control deteriorated after 8 h on regular schedules and after 5 h on irregular schedules. A more recent North American study using similar measures compared 10- and 13-h schedules but failed to find any differences in vehicle control between the schedules. However, the latter study has not controlled for road geometry. This study did find a major impact of time of day on vehicle control, drowsiness, subjective ratings and tests. All measures were worse when driving between midnight and dawn as compared to daytime.

So there is convincing evidence that long hours of work are an important cause of crashes and that poor vehicle control is a factor here. But there is also evidence that the timing of work, especially at night, is at least as an important a problem and drowsiness and vehicle control are the factors here.

4. TIMING OF WORK

The pattern of human behavior follows a 24-h cycle through sleep, awakening, increasing alertness during the daylight hours to preparation for sleep in the evening. A large number of biological functions demonstrate this pattern. For drivers the cycle of alertness is of most interest. A very large number of studies are now available that show alertness, especially vigilance, to be lowest during the hours we normally sleep, and the propensity to fall asleep is conversely greatest at this time. As when changing time zones, adaptation to a regular pattern of night work takes place. But no one has ever observed complete adaptation of the biological rhythms and alertness to a life of night-shift work; quite possibly there are too many external cues such as daylight to permit this to occur. With the exception of overnight delivery drivers few drivers consistently work nights. So it would not be expected that they would show much adaptation to night work but would have reduced alertness and increased drowsiness at night and the pattern of drivers' propensity to fall asleep show just that.

There are studies of crash risk at different times of day in which the number of crashes is adjusted to take account of the number of vehicles on the road. Truck crashes in which drivers were dozing were seven times more frequent between midnight and 08:00 hours than at other times of day and the highest risk was between 04:00 and 06:00 hours. A Swedish study found single vehicle truck crash risk increased by a factor of four between 03:00 and 05:00 hours and the increase in single vehicle car crashes was considerably greater. In a UK study of car drivers fatigue is reported as a factor in 36% of the crashes before dawn and declines to 4% of crashes during the morning, rising again through the afternoon and evening. A European study which showed a 4-fold increased risk of a truck crash when driving at night was described above. There is also evidence that successive periods of night driving pose a further increase in risk of a crash.

In the studies referred to under Long Hours of Work the physiological records of drivers' brain waves showed greater sleepiness when working at night, especially after midnight. And

video recordings taken of drivers' faces in a recent North American study showed drowsiness in 5% of the recordings and virtually all of these episodes occurred at night. Records of vehicle control also showed it deteriorated at night. The North American study concluded that night driving was a more important determinant of drowsiness and poor vehicle control than long hours of work alone.

5. TIMING AND DURATION OF SLEEP

As for other biological rhythms sleep is of the best quality and quantity if taken at night. Nightshift workers typically get 2 h less sleep than dayshift workers and some studies of night drivers have shown values as low as 4 h of sleep are obtained.

In the previously referred study of 107 single-vehicle truck crashes by the National Transportation Safety Board the most important factors in predicting night-time fatigue crashes were the duration of the most recent sleep period, the amount of sleep in the past 24 h and split sleep patterns. Truck drivers in fatigue crashes obtained an average of 5.5 h sleep in the last sleep period before the crash. This was 2.5 h less than the drivers involved in non-fatigue crashes. In split sleeps drivers typically work 4 h and sleep 4 h by sharing the driving. One study has found that this practice increases crash risk by a factor of 3. A study by the present author found drivers to have a 4-fold increase in hazardous incidents including crashes if they obtained < 6 h of sleep.

6. DRIVERS' REPORTS OF FATIGUE

Given that fatigue is as serious a safety problem as alcohol, how serious a problem do truck drivers think it is? In a study by the author 10% of truck drivers found fatigue to be at least "often" a problem for themselves. However, 36% thought fatigue to be at least "often" a problem for other transport drivers. This suggests a degree of over-optimism among drivers about their own capabilities as compared with every one else. In the National Transportation Board study of 107 single vehicle truck crashes in which 58% were attributed to fatigue ~80% of those drivers rated their last sleep as good or excellent. These data suggest that drivers underestimate their own state of fatigue and the impact of the causes of fatigue, despite some studies revealing that up to 20% of drivers will admit to falling asleep at the wheel in the past month.

Because it is unlikely that a technological device will be available in the near future to assess a driver's state of fatigue it is important to find out how accurate are drivers' self assessments of their fatigue, for this is the only information on which they can base their decision to cease driving. The research findings are not unequivocal. As reported in the recent North American study some drivers will drift in and out of microsleeps of many seconds. Unless these periods of sleep last minutes people are usually unaware of them occurring. However, people are aware of having to fight sleep before it over takes them. Perhaps the appropriate conclusion is that drivers will be aware they are fatigued and likely to fall asleep but do not know when this will occur and so continue driving. This problem has important liability implications.

7. CONCLUSIONS

Fatigue poses as great a safety problem for drivers as does alcohol. Crash risk starts to increase after as short a period as 6 h of driving and may be double after 11 h of driving or less. A deterioration

in driving performance may occur even earlier depending on the trip schedule. Dawn driving when the body rhythm of alertness is lowest increases crash risk by at least 4-fold and more for car drivers. Dawn driving performance is impaired and most people will demonstrate drowsiness and some drivers will move in and out of sleep at this time. Drivers who work at night will inevitably have poorer quality and quantity of sleep. Obtaining < 6 h of sleep increases risk by a factor of ~4 and splitting sleep periods imposes almost as great a risk. Most truck drivers do not recognize fatigue as a problem for them and many will be unaware of passing in and out of microsleeps.

The existing regulatory framework for controlling fatigue, termed Hours of Service, does not address the principle causes of fatigue, the timing of work and rest and the necessity to obtain adequate sleep.

REFERENCES

BARTLEY, S. H. and CHUTE, E. 1947, *Fatigue and Impairment in Man.* (New York: McGraw-Hill).

HARTLEY, L.R., 1995, *Fatigue and Driving* (London: Taylor & Francis).

HARTLEY, L.R., 1998, *Managing Fatigue in Transportation* (Oxford: Pergamon).

McDONALD, N., 1984, *Fatigue, Safety and the Truck Driver* (London: Taylor & Francis).

MACKIE, R.R. and MILLER, J.C., 1978, *Effects of Hours of Service Regularity of Schedules, and Cargo Loading on Truck and Bus Driver Fatigue.* Report No. HS-803 799 (Washington, DC: US Department of Transport).

NATIONAL TRANSPORTATION SAFETY BOARD, 1995, *Safety Study: Factors that Affect Fatigue in Heavy Truck Accidents.* Washington, DC, NTSB PB95-917001–SS-95/01.

Fire-fighting and Rescue Work in Emergency Situations and Ergonomics

A. Punakallio

Department of Physiology, Finnish Institute of Occupational Health, Laajaniityntie, Vantaa, Finland

S. Lusa-Moser

Department of Physiology, Finnish Institute of Occupational Health, Topeliuksenkatu, Helsinki, Finland

V. Louhevaara

Department of Physiology, University of Kuopio, Finland

1. INTRODUCTION

According to Louhevaara *et al.* (1997), ergonomics is defined as a multidisciplinary field of science based on physiology, psychology, sociology, and the application of technical sciences. It considers human capacities, needs, and limitations in the interaction of technical and organizational work systems. The integrated knowledge of ergonomics is used to develop work contents and environments through the use of job design and redesign measures for promoting health, work ability, safety and professional competence as well as the productivity and quality of work.

Fire-fighting and rescue work in emergency situations involves special ergonomic problems. Without taking into consideration work equipment and methods, it is almost impossible to influence the exposure of fire-fighting and rescue work by ergonomics. Therefore, physical work capacity is a crucial factor in a firefighter's work, and it should be sufficient for the physical stress factors of various operative tasks regardless of age. However, the exposure, safety issues, and individual work capacity in fire-fighting and rescue work can be influenced by measures related to the organization of work and management, as shown in figure 1.

2. EXPOSURE CAUSED BY ENVIRONMENTAL AND PHYSICAL STRESS FACTORS

Emergency situations are usually occasional, unpredictable, stressful, and dangerous. In these situations a firefighter must act determinedly, effectively, and safely. Every firefighter, regardless of age, needs to operate in various emergency situations at least a few times a year. Emergency situations take place under a wide range of thermal conditions because of variations in personal protective equipment, season and climate, and the intensity of the environmental heat and muscular work. Firefighters are exposed to many kinds of injuries and burns. According to a Finnish questionnaire study (n=846) of Lusa-Moser *et al.* (1997), 12% of the firefighters who responded had experienced local skin damage and 6–28% of the firefighters had repeatedly had extensive skin flushing in different body parts after the fire-fighting events.

The increased risk of injuries in fire-fighting and rescue work is mainly due to the use of heavy machines and equipment, dangerous entrances and restricted spaces for moving, extreme heat stress, and rapidly changing environmental conditions. The mean annual frequency of injuries which caused at least three
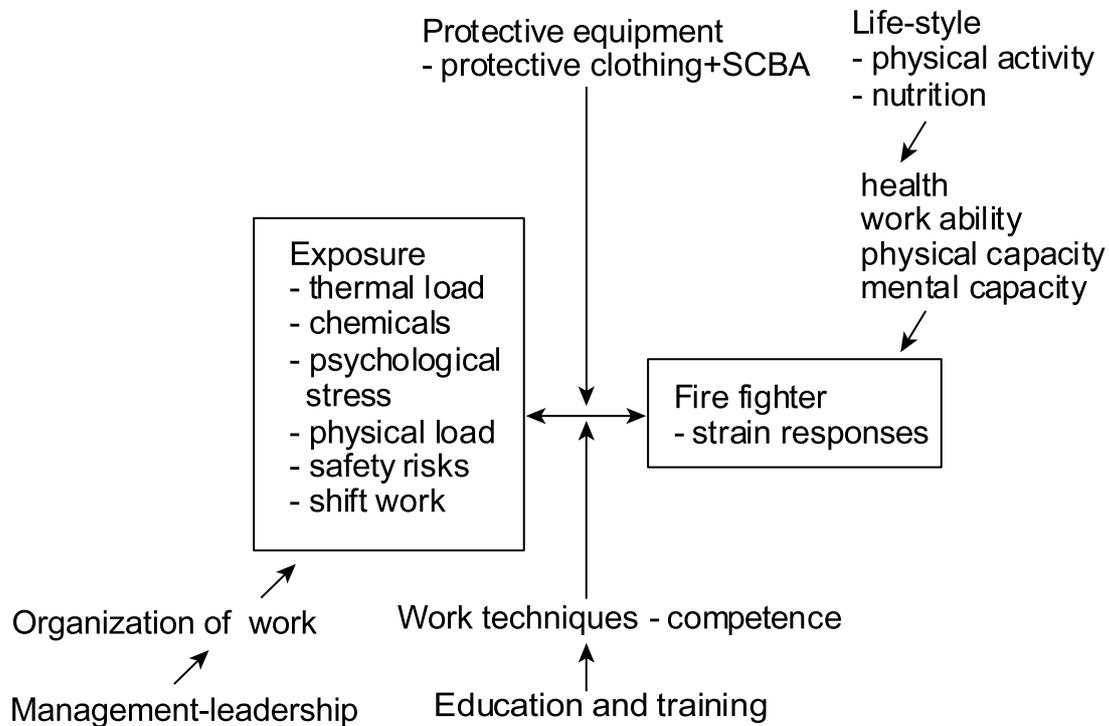
Figure 1. In emergency situations of fire-fighting and rescue work the exposure and individual strain responses to the exposure can be affected by various ergonomic measures for improving protective equipment, organization of work, professional competence and work methods

days sick of leave in 3000 professional operational Finnish firefighters was 388. The main events resulting in injuries were slipping and tripping, overstrain or an accident involving human motion, and stepping on objects or being struck by objects. According to Reichelt and Conrad (1995), in the USA the incidence of work-related injuries in the fire service was reported to be 4.7 times higher than in the private industry when almost half of the firefighters (41%) reported injuries.

Firefighters need also efficient protection against combustion gas containing a variable mixture of compounds. The number of different chemicals and substances detected in fires is over 800. Their toxicity varies greatly and depends mainly on the type of fuel, the heat of the fire, and the amount of oxygen available for combustion.

In a building fire, a firefighter may have to climb stairs while wearing personal protective equipment and carrying tools weighting up to 50 kg. After climbing and clearing the entrance, a firefighter should be able to help a victim or resist the propulsive force of a water spout. In rescue operations associated with traffic accidents, a firefighter must often use hydraulic tools weighting 15–20 kg in poor working positions. Lifting and carrying tasks while using heavy personal protective equipment demands good aerobic and muscular capacity as well as motor skills. In the study of Lusa (1991), the opening of a ceiling with a motor saw was simulated. In the task, the compressive force in the low back was found to be so high that the risk for back injuries increased considerably according to the norms of the National Institute of Occupational Health and Safety in the USA. Dynamic stress can be assessed with the measurement of absolute oxygen consumption (l/min) during work. Louhevaara *et al.* (1985) reported a mean oxygen consumption of 2.1–2.8 l/min in various fire-fighting tasks and different thermal conditions. The cardiorespiratory strain varies according to an individual's cardiorespiratory and muscular fitness, body composition, as well as motor and occupational skills.

Firefighters are also exposed to psychological stress induced by unpredictable and dangerous situations, and to stress caused by their role as help providers. In addition to their own safety, firefighters are responsible for the safety of their co-workers and others involved in an event. Shift and night work markedly emphasize the effects of different environmental stress factors such as lightning and ambient temperature. A firefighter may also experience considerable stress due to the intermittent nature of the work. According to the report of Glazner (1996), a firefighter is more likely to be injured at mealtimes or during a night shift. Due to the shift-work nature of fire-fighting and rescue work, both disruption of eating patterns and fatigue increase the risk of work-related injuries.

3. PERSONAL FIRE PROTECTIVE EQUIPMENT

Personal fire protective equipment, including protective clothing and a self-contained breathing apparatus (SCBA) are necessary protection against combustion gases and heat in most emergency situations. Already in a thermoneutral environment, the use of personal fire protective equipment weighting 25 kg increases cardiorespiratory strain by 20–30% at submaximal work levels, and also increases thermal strain significantly. In the study of Lusa *et al.* (1993), two walking tests were performed on a treadmill, and experienced firefighters served as subjects. First,

Table 1. Heart rate, oxygen consumption, rectal temperature and blood lactate at the end of the treadmill exercise without and with the protective equipment. The values given are means ± SD

Variable	Without equipment	Protective equipment
Heart rate (beats·min⁻¹)	129 ± 4	169 ± 4***
Oxygen consumption (1·min⁻¹)	2.2 ± 0.1	3.0 ± 0.1***
Oxygen consumption (ml·min⁻¹·Kg⁻¹)	26.5 ± 0.6	36.4 ± 1.0***
Rectal temperature (℃)	37.6 ± 0.1	37.9 ± 0.1*
Blood lactate (mmm01·1⁻¹)	1.1 ± 0.2	2.5 ± 0.7

*$p < 0.05$, ***$p < 0.001$

the subjects wore sportswear and then, after the rest period, repeated the test with fire protective equipment. Table 1 shows that in the test with the protective equipment, heart rate, oxygen consumption, rectal temperature and blood lactate were higher compared to the control test with sportswear. In addition, the use of the SCBA hampers breathing, and may cause incomplete gas exchange (Lusa 1994).

The use of the SCBA also sets special demands on motor skills, affecting the postural control, for example, by changing the center of gravity. The face mask of the SCBA also limits the visual field and this may also have an effect on postural control. The SCBA is usually used in difficult environmental conditions. Working in smoky or dark places requires that other body control systems (e.g. the kinesthetic sense, the sense of orientation) have to be more alert because vision cannot be used maximally. Moving with the personal fire protective equipment in fire-fighting and rescue situations demands good motor skills, particularly when moving on roofs or climbing up ladders. Punakallio *et al.* (1999) reported that firefighters, regardless of age, made more errors and were slower in the functional dynamic balance test while wearing the fire protective equipment than when wearing sportswear. The youngest group of firefighters was, on average, quicker and more accurate in the test with the fire protective equipment than the oldest group was in sportswear. The good physical capacity of the young firefighters seemed to compensate for the disadvantages caused by the equipment to the motor skills.

The currently accepted multilayer fire protective suit designed to fulfill the European standard provides more effective heat isolation than the conventional fire suit. It protects firefighters effectively from powerful heat radiation, although, on the other hand, it makes heat transfer more difficult when the thick layer limits adequate body cooling through evaporation; the suit is also heavy (up to 3 kg). It is not possible to increase the weight of the fire protective equipment because of the obvious risk of both cardiorespiratory and thermal overstrain. A reduction in the weight of the protective equipment is problematic without decreasing safety and the protective properties too much. It is, however, possible to lighten the equipment by replacing the present steel air containers by containers made of lighter fiberglass materials. These containers are about 9 kg lighter than the traditional steel containers; however, they are also more expensive. The lighter containers would be useful in fire-fighting and rescue work because of decreased cardiorespiratory and musculoskeletal strain, and better control of balance.

4. WORK TECHNIQUE, COMPETENCE, AND TRAINING

In fire-fighting and rescue work, the range of duties has widened markedly in recent years; an example is the sharp increase in the daily transport of patients. This underlines the importance of continuous education and training. Good work methods also contribute to safer work, reducing, for instance, the risk of injuries due to physical overstrain. Adequate competence and skills can also compensate for a lack of physical capacity with respect to job demands.

In the questionnaire study of Lusa-Moser *et al.* (1997), two-thirds of the firefighters (64%) considered that they had inadequate training and guidance for some of their tasks. In particular, the older firefighters felt that they had insufficient training and guidance to enable them to perform the whole variety of fire-fighting and rescue tasks. This lack was also illustrated by the fact that older respondents claimed more frequently than the young firefighters that they encountered too difficult work situations. This may be due to a number of reasons. Extremely stressful and difficult situations that have occurred in the past affect an individual's experimental world, and negative memories may augment the mental burden in new situations if the past situations have not been dealt with thoroughly — for example, by means of defusing and debriefing. On the other hand, the decrease in physical capacity related to aging may also affect experiences resulting from extreme situations. The responsibility for fire-fighting and rescue tasks for which older firefighters in particular may have had inadequate training can increase the strain as well. These facts also underlie the importance of continuous education and training in fire-fighting and rescue work.

5. MANAGEMENT — LEADERSHIP

The enhancement of management was considered very important by all firefighters according to Lusa-Moser *et al.* (1997). One-quarter of the firefighters who responded (26%) get, when needed, little support and help from their superiors; the rest of the respondents felt that they got support at least to some extent (figure 2). Furthermore, almost one-half of the respondents (45%) considered that they got very little or no feed-back on their success at work. Three out of four firefighters (75%) evaluated that they could influence much, or to some extent, matters which concerned them at the fire department. Older firefighters had significantly more possibilities to influence work matters than did the younger ones. However, older firefighters felt a lack of opportunity to advance more often than did the younger firefighters. About one-half of all the respondents (45%) thought that they had only few opportunities to advance in their work. Almost everyone (95%) felt that they had at least fairly good possibilities to influence and use their knowledge and skills at work. Moreover, the firefighters having good psychological resources were more likely to feel that they could utilize their own talents and abilities at work, compared to those having moderate or poor resources.
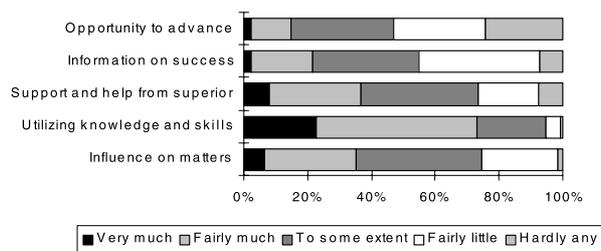
Figure 2. The factors associated with a firefighter's well-being at work and work role (percentage of respondents)

According to Lusa-Moser *et al.* (1997) nearly three-quarters of the respondents (70%) considered that the division of labor was correct and equal at fire departments, while 18% considered that it was unfair. Within their own work crews, the firefighters felt that co-operation was good, despite differing attitudes towards work. There have been attempts to solve negative attitudes towards work that hamper the progress of work by discussing and consulting (69%), by ordering and resorting to authority (24%), while the remainder of the respondents thought that there had been no attempts to deal with negative attitudes. The psychosocial atmosphere within the work shift was considered comfortable by about one-half of the firefighters (46%), and a minority (16%) thought that the atmosphere was supportive. However, 29% felt that the atmosphere was prejudiced, while 8% felt that the atmosphere was tense and competitive.

In the Finnish study of Härmä *et al.* (1991), firefighters were able to sleep, on average, six hours during a 24-hour work shift and only a few of them slept under two hours. The quality of sleep was a little worse at work than at home, but the structure of the sleep did not differ from that registered on days off. As a matter of fact, the sleep during the work shift contained even more phases of deep sleep than that at home, which is important for the recovery of alertness. It was concluded that recovery in the studied 24-hour shifts was sufficient for the firefighters. However, if the number of alarms increases, the proportion of firefighters who cannot get sufficient sleep during a work shift increases. It was recommended that night shifts be added, for example, temporally using the rest pause for sleeping.

6. PHYSICAL CAPACITY AND HEALTH

Accidents and fires can occur at any time and place, and under a variety of circumstances. There are almost no standards (norms) to influence the ergonomics of work or the work environment. Despite improvements in professional competence, work methods, and equipment, a firefighter's work ability and safety depend largely on physical capacity. Finnish multidisciplinary studies have led to the standardization of health examinations and assessments of the physical capacity of firefighters. The minimum levels of the cardiorespiratory and muscular fitness of a firefighter who needs to carry out smoke-diving tasks have also been defined. The maximal oxygen consumption of an operative firefighter should be at least 3 l/min or 36 ml/min/kg. At these levels, there is a certain margin in the maximal oxygen consumption of a firefighter to tackle unpredictable situations without obvious risk for overstrain, disorders or illnesses.

According to the laboratory study of Punakallio *et al.* (1999), about one-half of 50- to 54-year-old Finnish firefighters did not

reach the minimum requirements of maximal oxygen consumption. The proportion of unfit firefighters increased significantly even in the 40–44-years age group as compared with the firefighters aged 30–34 years. According to the Work Ability Index, most of the firefighters had good subjective work ability (Punakallio *et al.* 1999). However, the prevalence of poor work ability increased with age. About one-fifth of the firefighters aged over 49 years (19%) rated their work ability as poor. The corresponding proportion for poor work ability among municipal male workers in Finland is 15%. The work ability of firefighters in relation to their physical job demands was perceived to decrease more steeply than the work ability related to psychological job demands. In addition to age, factors related to work and lifestyle were associated with good work ability based on the Work Ability Index. These factors explain 53% of the variability of the Work Ability Index shown in table 2.

7. WORK DEVELOPING ASPECTS IN THE FUTURE

The average age of firefighters is increasing rapidly: nowadays the average age of professional firefighters in many fire departments in Finland is over 40 years. The proportion of firefighters whose health and physical capacity are not sufficient for the job demands is increasing continuously due to the recent change in the retirement age. Nowadays, every firefighter should be able to perform smoke-diving tasks up to the age of 65 years. In larger fire departments, unfit firefighters can be transferred to physically less demanding tasks. This is not possible in smaller fire departments where only a few firefighters are on duty and where each one should be able to perform all tasks. Therefore, it should be advisable to plan, from the beginning of each firefighter's career, alternative career paths which demand less physical capacity than that required by, for instance, smoke-diving tasks. The valuable working skills and knowledge collected during the early part of the career should be utilized later in other fire-fighting and rescue tasks. Possible alternatives could be, for example, prevention of accidents, and education and counseling. With the use of this new strategy, the remaining work ability could be supported to prevent early retirement due to work disablement. The proportion of disabled firefighters is already increasing in the 40- to 49-year-old age group of firefighters, and especially in those older than 50 years. According to Lusa-Moser *et al.* (1997) there are several important factors which support coping at work (figure 3).

In the emergency situations of fire-fighting and rescue work,

Table 2. The factors associated with the Work Ability Index. The model explained 53% of the variability of the index

Factor	Connection
Age	–
Smoking	–
Daytime tiredness	–
Psychological stress	–
Physical exercises	+
Life satisfaction	+
Meaning of work	+
Perceived stress tolerance	+

All the factors were statistically significant, at least at the p<0.05 level
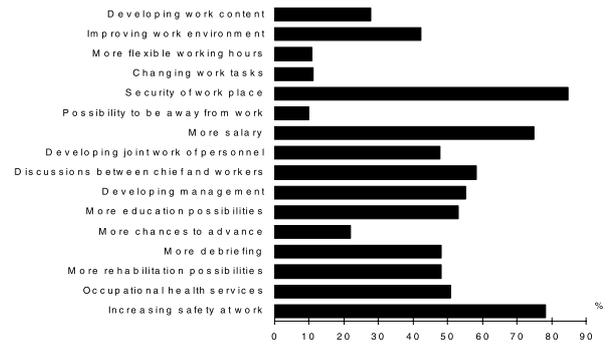
Figure 3. The most important factors affecting coping at work in fire-fighting and rescue work (percentage of respondents)

ergonomics is primarily needed for developing work equipment and methods, as well as for the organization of work.

REFERENCES

GLAZNER, L.K., 1996, Factors related to injury of shiftworking firefighters in the Northeastern United States. *Safety Science*, 21, 255–63.

HÄRMÄ, M., LUSA, S., LOUHEVAARA, V., and NYGÅRD, C.-H., 1991, The sleep and recovery of firefighters in a 24-hour shift schedule. *International Symposium of "Shiftwork and job demands". Extended Abstracts,* 71–3.

LOUHEVAARA, V., HOPSU, L., HUTH, E., KRUGER, D., SOGAARD, K., and NIELSEN, J., 1997, Ergonomics. In Kruger, Louhevaara, Nielsen, and Scheider (eds) *Risk Assessment and Preventive Strategies in Cleaning Work* (Hamburg: Werkstattberichte, Wissenschaft + Technik, Wb 13).

LOUHEVAARA, V., TUOMI, T., SMOLANDER, J., KORHONEN, O., TOSSAVAINEN, A., and JAAKKOLA, J., 1985, Cardiorespiratory strain in jobs that require respiratory protection. *International Archives of Occupational and Environmental Health*, 55, 195–206.

LUSA, S., 1994, Job demands and assessment of the physical work capacity of firefighters. *Studies in Sport, Physical Education and Health 33* (Jyväskylä: University of Jyväskylä).

LUSA, S., LOUHEVAARA, V., SMOLANDER, J., KINNUNEN, K., KORHONEN, O., and SOUKAINEN, J., 1991, Biomechanical evaluation of heavy tool-handling in two age groups of firemen. *Ergonomics*, 34, 1429–32.

LUSA, S., LOUHEVAARA, V., SMOLANDER, J., POHJONEN, T., UUSIMÄKI, H., and KORHONEN, O., 1993, Thermal effects of fire-protective equipment during job-related exercise protocol. *Safe Journal,* 23, 36–9.

LUSA-MOSER, S., PUNAKALLIO, A., LOUHEVAARA, V., VIIKARI-JUNTURA, E., ILMARINEN, R., OLLILA, J., KORHONEN, O., LINDQVIST-VIRKAMÄKI, S., and LUUKKONEN, R., 1997, Health and functional capacity of firefighters in different age groups. Part I: Questionnaire study — the association of life habits, health and mental factors with perceived work ability. *Respect for Aging – A Programme to Promote the Health, Work Ability and Well-being of Ageing Workers 30* (Helsinki: Finnish Institute of Occupational Health and Finnish Work Environment Fund) (in Finnish).

PUNAKALLIO, A., LUSA-MOSER, S., LOUHEVAARA, V., VIIKARI-JUNTURA, E., ILMARINEN, R., OLLILA, J., KORHONEN, O., and LUUKKONEN, R., 1999, Health, physical and mental capacity of firefighters in different age groups. In Ilmarinen and Louhevaara (eds) *Finnage - Respect for the Ageing: Action Programme to Promote Health, Work Ability and Well-being of Ageing Workers in 1990-1996,* People

and work research reports 26 (Helsinki: Finnish Institute of Occupational Health), pp. 117–27.

REICHELT, P.A. and CONRAD, K.M., 1995, Musculoskeletal injury: ergonomics and physical fitness in firefighters. *Occupational Medicine: State of Art Reviews*, 10(4) (Philadelphia: Hanley & Belfus), pp. 735–46.

Human Control of the Behavioral Environment

T. J. Smith

Human Factors Research Laboratory, Division of Kinesiology, University of Minnesota, 1900 University Avenue SE, Minneapolis, MN 55455, USA

1. INTRODUCTION

Work is the most ubiquitous form of human activity on the planet. In the strict physical scientific sense of the term (application of energy to drive chemical reactions (internal work) or to move body effectors (external work)), human engagement in work is synonymous with life itself, beginning at the moment of conception and terminating only with death. In the behavioral sense, all conscious human activity represents work directed at some sort of goal or another. In the occupational sense, all humans routinely engage in purposeful task or job work activity throughout most of their lives. In the evolutionary sense, since emergence of the species work activity has served both as the engine of the human condition and as the means of human self-selection.

Encyclopedic treatments of the term "work" typically focus only on its physical scientific meaning, ignoring the much broader human perspective. Dictionary definitions (the term has relatively recent Old English origins dating back less than 1500 years) hint at its broader implications — Webster provides over 30 definitions of "work" used as a transitive or intransitive verb, and a dozen definitions of its use as a noun. However, it is with the historical perspective that the true human meaning of the term begins to emerge. All of the so-called prehistorical and historical ages of human development and civilization — Stone, Copper, Bronze, Iron, Classical, Medieval, Renaissance, Industrial, Information — refer explicitly or implicitly to distinctive modes and expression of human work activity.

Work serves also as a primary focus of human cultural, religious, intellectual, political and scientific thought. The emergence of human civilization essentially rested upon a series of transformations in how work was organized. The first words of the Bible — "And God created heaven and earth" — refer to work of the Almighty. Human work was an object of reverence for the ancient Hebrews, a Judeo-Christian tradition that persists to this day. The organization of work — in warfare, commerce, cultural and religious expression, and society — has represented a primary focus of the state from its earliest origins. Political thought and philosophy revolve around the best approach to such organization. Certain forms of government such as communism exalt work as the centerpiece of their design. The evolution of technology reflects the application of increasingly sophisticated artifacts to aid human work performance. All academic, intellectual and scientific disciplines concerned with human performance — from art to athletics, from business to engineering — are concerned in one way or another with how human work is organized and expressed. And of course the scientific focus of this encyclopedia — ergonomics/human factors — is concerned with all dimensions and conditions of human work. Ergonomics is defined as, "laws and customs of work."

Given the extremely broad scope of the topic, it would be impractical to consider here all of the possible human implications of the term "work." Sections therefore touch upon scientific theories of work, landmarks in the evolution of work, general conclusions about laws and customs of work, and the future of work.

2. THEORIES OF WORK

What is the purpose of human work? In light of the universal significance of work performance for defining, organizing and guiding the human condition, it is remarkable that no generally accepted theory of human work has yet been promulgated. Scientific concepts of work essentially have not been developed much beyond conventional views of labor as a means of production and source of wealth (Smith 1776), as a limited aspect of technology (Marx 1867), as an exploited competitor in class struggle (Marx and Engels 1848), or as essentially a mechanistic component of work process subject to "scientific management" (Taylor 1911). A variety of sources can be cited (Galbraith 1958, McGregor 1960, Smith 1962, 1965, Horowitz 1998) to support the conclusion that all of these viewpoints have been largely discredited as they apply to motivational behavior in the performance of work.

It is equally remarkable that despite its putative role as *the* major scientific discipline concerned with human behavior at work, Ergonomics/Human Factors science has been largely silent in addressing this question. For example, no theoretic treatment of work is to be found in handbooks concerned with Ergonomics/Human Factors (Salvendy 1982, 1987). Moreover, work receives short shrift as an object of formal academic focus by Ergonomics/Human Factors science. Among 79 North American graduate programs in Ergonomics/Human Factors catalogued by the US Human Factors and Ergonomics Society in 1997, only one (Department of Work Environment, Lowell University, Lowell, MA) has the term "work" in the title.

The perspective offered here is that purposeful work represents the principal behavioral activity by means of which humans: (1) self-control their interaction with their environment; (2) self-define themselves as individuals; (3) create and human factor new modes and dimensions of work technology and design; and (4) have self-selectively guided their evolutionary trajectory as a species (Smith 1965). That is, with its goal of human factoring the ecosystem to achieve environmental control, human performance of work has a mutually adaptive influence both on the nature of the performer as well as on the nature of work itself.

The major assumptions of this behavioral cybernetic theory (Smith and Smith 1987) of human work are summarized in Figure 1. The theory views work as a closed-loop process in which the worker employs various behavioral strategies to control sensory feedback from the physical, social and organizational design attributes of the work environment. In so doing, the worker self-organizes and self-defines his/her individual biosocial identity as a distinct member of society, both psychologically and physiologically.

In addition to individual self-definition, another major feedback influence of work behavior assumed by the model in Figure 1 is on the nature of work design itself. That is, interaction with the work environment prompts consideration on the part of the individual worker and the organizational work system as
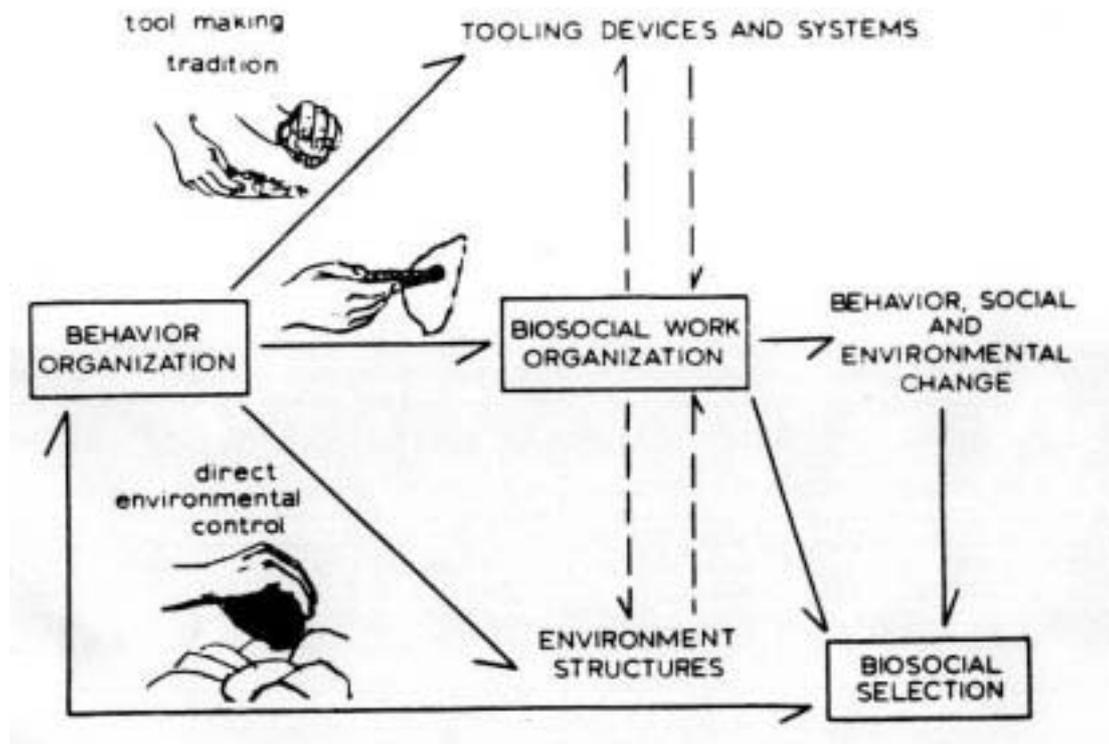
Figure 1. Behavioral cybernetic theory of human work. Throughout evolution, humans have endeavored to control their behavioral environment through the use of work to fabricate tools and other types of technology, and to organize and structure their society. Work activity of this sort tends to catalyze behavioral, social, technological and environmental change, which in turn feeds back to influence the primary features of behavioral organization of the individual.

to how work design features can be modified and human factored to facilitate work performance. In this manner, work itself has a positive feedback effect on cognitive creation of new modes and dimensions of work design and work technology.

Finally, the model in Figure 1 assumes that the behavioral feedback effects of work represent the principal mechanism of human self-selection in evolution. Figure 2 further elaborates upon this concept. The premise in Figure 2 is that the evolutionary emergence of all of the distinctive biological features of species *homo* — upright posture, relatively hairless body, apposed thumb and fingers mediating manipulative skill, verbal communication, distinctive modifications in neurohumoral regulation; and specialization in brain structure and function — was influenced and guided in a major way as a self-selective process through the planning, organization and execution of work, and through the fabrication of tools and other artifacts to facilitate the performance of work.

A number of lines of evidence can be cited to support the behavioral cybernetic models of work depicted in Figures 1 and 2 (Smith 1962, 1965). Various observations from behavioral science indicate that both the distinctive personalities as well as the cognitive and motor behavioral proficiencies of different individuals are influenced in a major way by their daily work activity throughout life. Indeed, the motivational behavior of work preoccupies human attention and time to a far greater extent than other sources of motivation such as eating and drinking, aggression, activity–rest cycles, sexual behavior, or relaxation. These behavioral feedback effects of work augment whatever contributions innate biological differences may make to defining

individual differences, and we may speculate that such effects date back to the dawn of the species.

A second line of evidence is that human evolutionary origins and development are closely associated with the fabrication and use of tools and other artifacts for the performance of work, a process that has shown continuous progression since the dawn of the species. That is, most of the behavioral and technological functions of modern machines employed to facilitate work performance — e.g. those of shaping, striking, cutting, smashing, forming, lifting, or turning — are derivative and can be traced back to human use of hand tools in remote prehistory for similar purposes (Oakley 1957).

Thirdly, information gathered from present-day primitive tribal and village groups suggests that transitional processes in the social, cultural and technological development of human societies and civilizations are related more prominently to distinctive adaptive changes in the organization and implementation of work than to those in other behavioral factors such as changes in patterns of plant or animal husbandry, consumption, etc. (Woolley 1946, Wolf 1954). Indeed, the latter depend upon the former. The relatively permanent nature of tools (some stone tools have been dated back over 2 million years), machines and other technological artifacts of work suggests a dominant role of behavioral feedback mechanisms of work activity and technology in defining both the direction and pattern of human development.

Finally, using parallels from present-day observations of human behavior, it can be argued that all of the mechanisms of natural selection invoked by classical evolutionary theory to
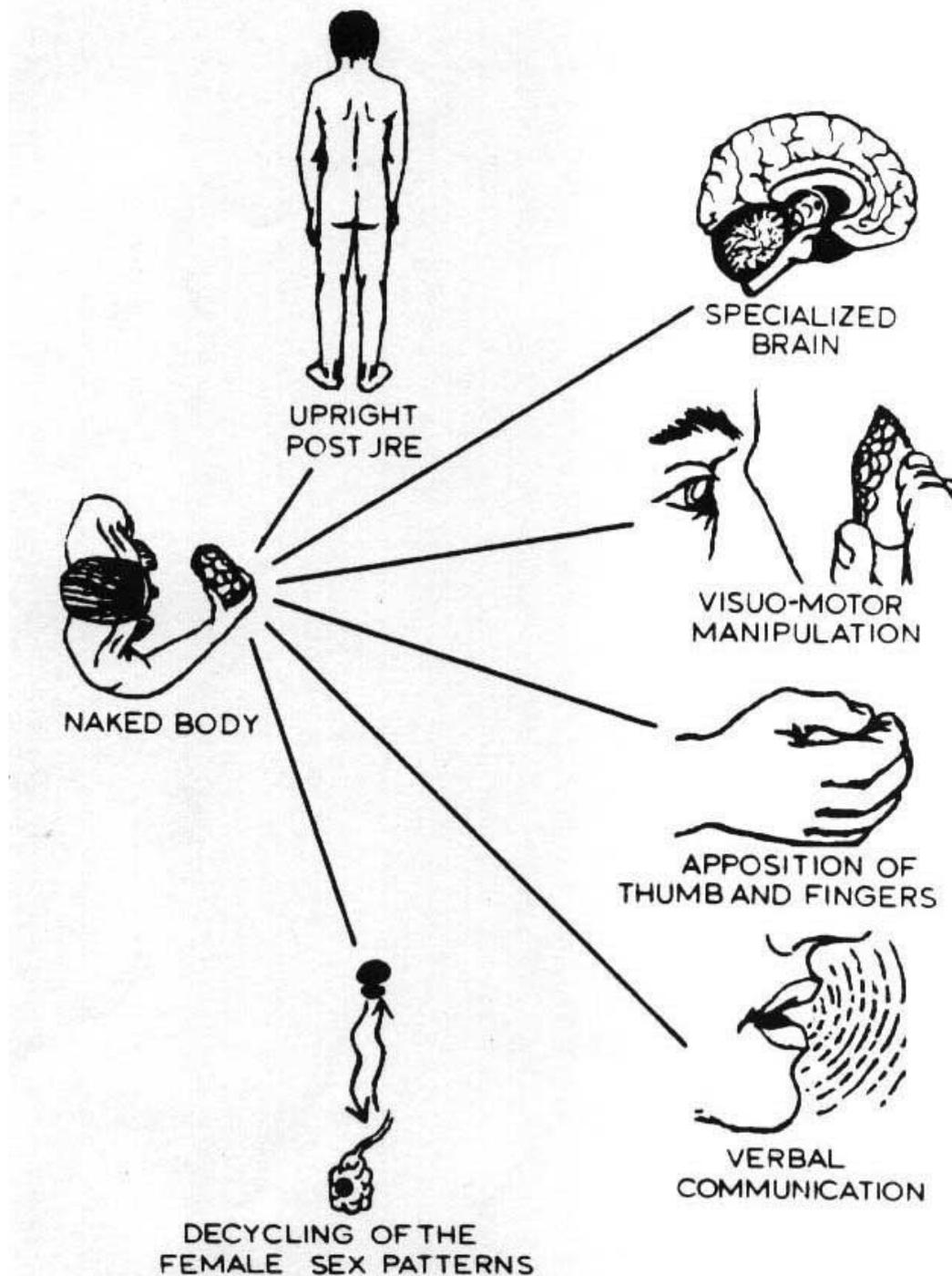
Figure 2. Work theory of feedback selection. Through the use of work to control the behavioral environment, humans tend to self-select anatomical and physiological attributes of behavioral organization and personality that tend to enhance such control, as a consequence of differential survival and reproductive success. Work behavior and natural selection thereby become coupled in a feedback manner.

explain the evolution of the modern humans — transformation of social and family structures, exploitation of new sources of nutrition, advantages in intra- and interspecies competition, dispersion into new ecosystems, heightened motor behavioral capabilities, improved exploitation of external sources of environmental energy — are directly linked in a behavioral feedback manner to advances in the organization, technology and conduct of work to human factor and control the environment and to thereby facilitate differential reproductive success. From this perspective, human evolution may be viewed as a self-selective process largely influenced and guided by work behavior.

What is distinctive about human work? No one who has observed a bee or ant colony can doubt that non-human species are entirely capable of intense, highly organized, and socially coherent patterns of work. What then distinguishes the special nature and outcomes of human work? Suggested answers to this question include human capacity for verbal communication and

cognition, factors that also are implicated in addressing the related question of "humanness."

An undeniably unique feature of human work performance, that both complements and underlies its other distinct attributes, relates to the use of work to track and thereby control time (Smith 1965). Human capabilities in the control of time encompass perception and comprehension of the past, the future, time continuity, time persistence, historical time, relative time (early, late, then, before, etc.), time interruption, time marking and recording, temporal change, clock time, space–time interaction, time prediction, and space–time relativity.

Humans rely upon these capabilities to control spatio-temporal feedback from prevailing conditions in the present and past, and to thereby feedforward control their actions into the future. No other species displays comparable sophistication in temporal awareness and control. However, to attain such control, humans have had to create work-derived artifacts for marking and tracking time, because innate comprehension of time apparently differs little among humans and other animals. For example, languages of extant aboriginal cultures contain no words for temporal concepts such as "early," "late," "then," "now," "past," "'future," "history," etc. One of the key events in the evolution of modern humans therefore was the augmentation of space-structured behavior (common to humans and many other animals) with capabilities for time-structured behavior through fabrication of artifacts for telling time. Marshack (1964) offers evidence that such artifacts date back some 30 000–40 000 years, anticipating by thousands of years the massive Neolithic and

Bronze Age megaliths and temples erected in Egypt, on the Salisbury Plain, and elsewhere for tracking time.

Through fabrication of such artifacts, human control of work and of time have become inextricably integrated. From a cybernetic perspective therefore (Figure 1), one of the key self-selective influences of work behavior in human evolution has been the use of work to human factor technological and social means for tracking time, thereby allowing projection of work-derived artifacts, knowledge and skill into the future to benefit generations yet unborn.

3. LANDMARKS IN THE EVOLUTION OF HUMAN WORK

Landmarks in the evolutionary progression of human work can be characterized in terms of the use of work to guide the emergence of new forms of societal organization (Figure 3), of technology (Figure 4), or of modes of symbolic communication (Figure 5). The heavy lines in the figures suggest that the development and elaboration of each of these manifestations of work has been more or less continuous and progressive throughout evolution (Smith 1965), in patterns that largely parallel the non-linear growth of human population itself. However, at certain critical periods in prehistory and history, innovative patterns of adaptive organization of work emerged to mediate the creation of new types of architecture and modes of societal integration, new tools, machines and other technologies, and new methods of symbolic communication.

In particular, Figure 3 indicates that forms of societal
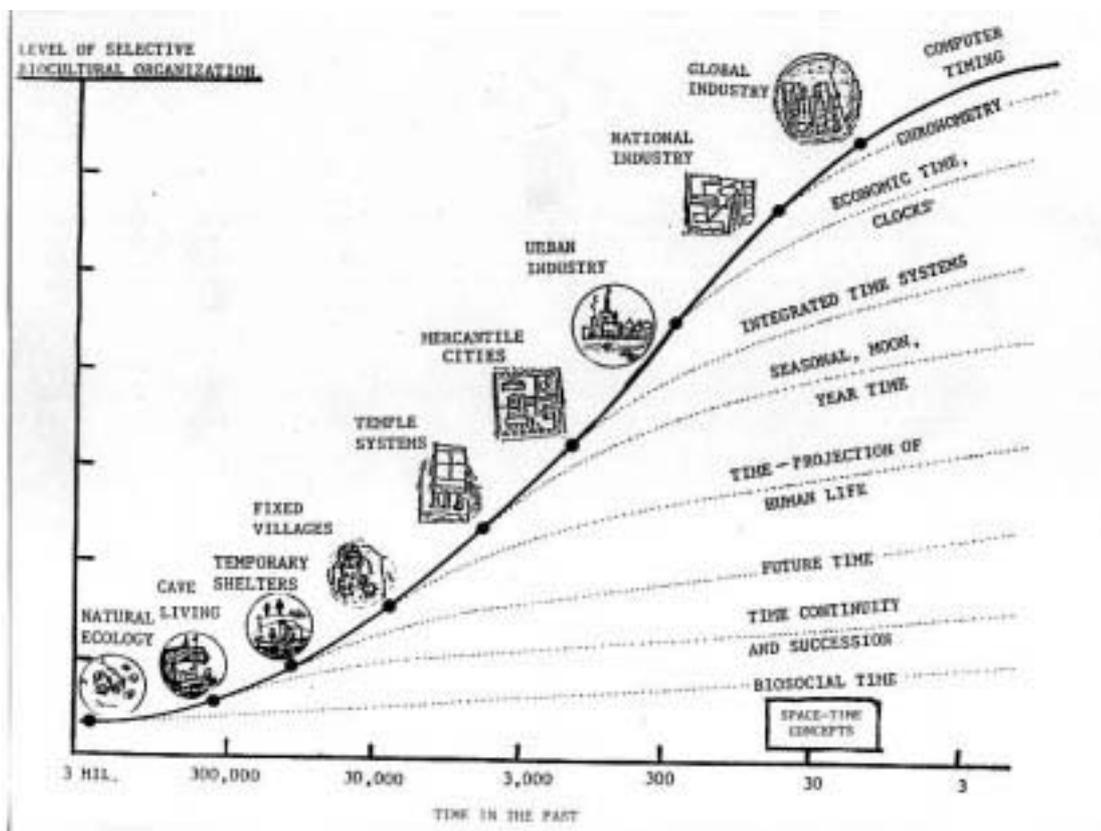
Figure 3. Progressive emergence in evolution of innovative forms of societal organization, feedback coupled to advances in work behavior and in time perception and control. As new forms of organization emerge, old forms persist and are conserved.
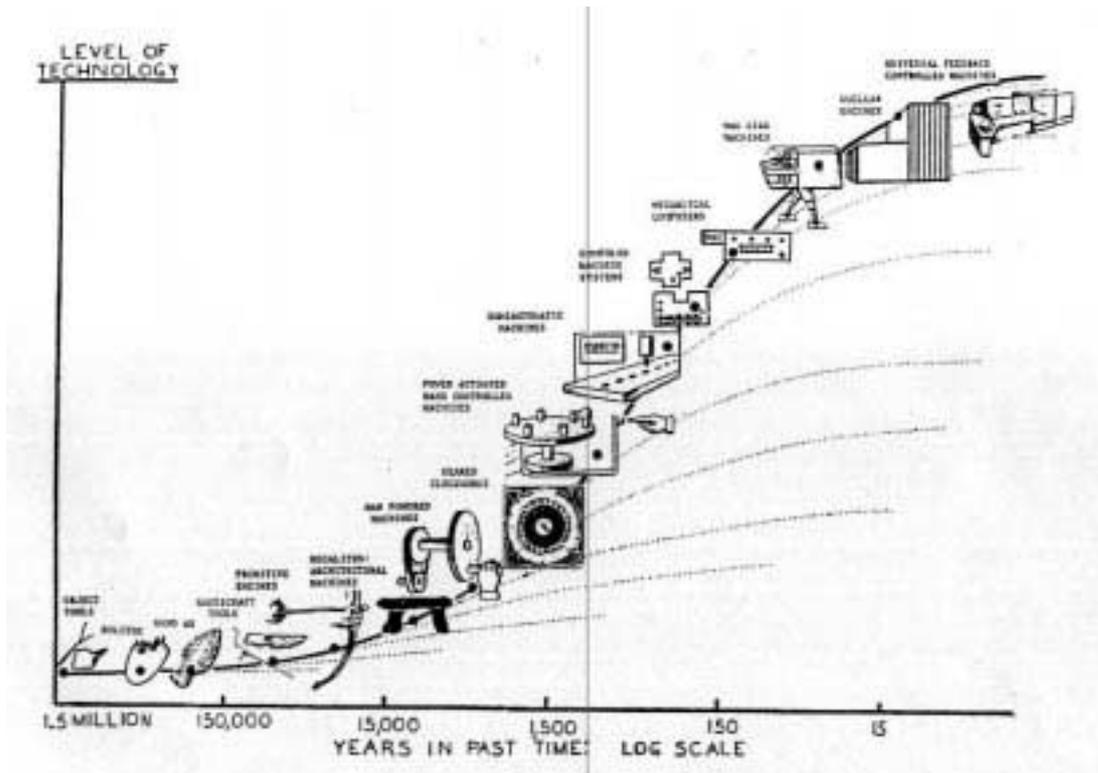
Figure 4. Progressive emergence in evolution of innovative forms of technology, feedback coupled to advances in work behavior. As new forms of technology emerge, old forms persist and are conserved.
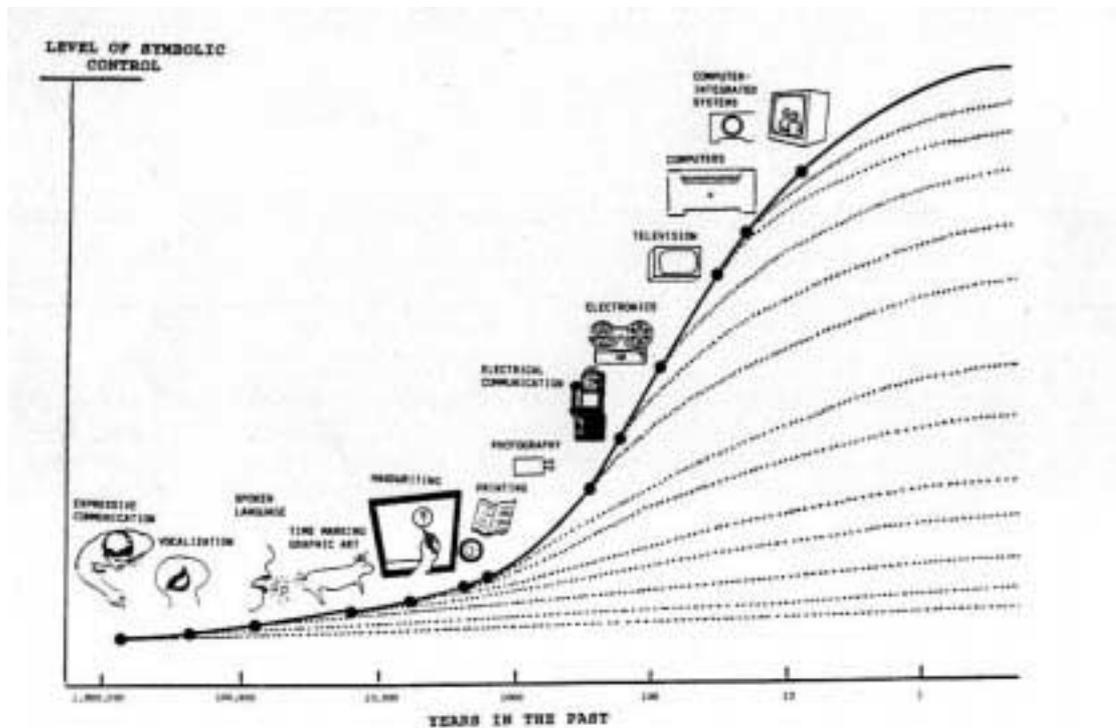
Figure 5. Progressive emergence in evolution of innovative modes of symbolic communication, feedback coupled to advances in work behavior. As new forms of communication emerge, old forms persist and are conserved.

organization have progressed from primitive shelters to fixed village and temple systems, to ramified mercantile and industrial centers, and finally to the national and global commercial systems of today. The figure also suggests that the emergence of each of these new levels of societal organizational sophistication has been accompanied and guided, in a feedback manner, by the emergence of new behavioral sophistication in the use of work to control time. Figure 4 indicates that technological innovation has progressed from primitive hand tools to megalithic temples used as architectural time-keeping machines, to human-powered and thence to human-controlled machines powered by environmental sources of energy, and finally to the semi-automated, automated and cybernetic technologic systems of today. Finally, Figure 5 indicates that modes of symbolic communication have progressed from spoken language and non-verbal communication to graphic expression, handwriting and printing, to various modes of electronic communication, and finally to the computer-mediated communication systems of today.

Figures 3–5 are meant to suggest that the innovative advances depicted are both integrated and cumulative. That is, work-related elaboration of new forms of societal organization, technology and symbolic communication has occurred in an interdependent manner. In addition, all of the innovations indicated in the figures, plus related skills, patterns and systems of work on which they depend, can be found today. Legacies in work innovation accumulate because of the unique human capacity for conserving knowledge and information about new technology and work methods over time.

Thus, many of the so-called "modern" features of work in fact have ancient origins. For example, many observers agree that a key prehistoric landmark in the organization of work was the establishment of relatively fixed settlements, facilitated in a feedback manner by innovations in agricultural and animal husbandry work. A likely behavioral feedback effect of this trend was establishment of more elaborate systems of community decision-makers, today termed administrators, managers, authorities, bureaucrats and government. This was necessary to ensure that the work of residents directed at expanding the scope of community control over critical environmental conditions — food supply, security, design and fabrication of technology, commerce etc. — was effectively organized and managed. Division of labor certainly exists in mobile hunting and gathering societies and also is apparent in non-human species. However, emergence of larger populations in more static settlements dramatically reduced the ability of a given resident to effectively control all key environmental conditions on an individual basis. Hence, the need for better defined systems of organizers and managers to organize and guide community work processes.

One of the earliest historical records of this 'macro-ergonomic' approach to work organization is that of the ancient Sumerians starting ~7000 years ago. These peoples — precursors of the Hebrews, Babylonians, Greeks and Romans — elaborated over the next three millennia most if not all of the basic concepts of work organization that we observe and practice today (Kramer 1963). These include writing and phonetic notation, laws of work and commerce, ethical principles and mechanisms of justice, organization of crafts, professions and trades, rules of fair play, concepts of freedom, and valuation of qualities of ambition and success, all hallmarks of current work practices. Anticipating

subsequent Greek and Roman developments by some 2000 years, Sumer was the first ancient civilization effectively to combine and organize labor, crafts, professions, and civil and religious authority as an integrated societal system of work.

As another example, it is likely that methods of mass production, based on what we today call factory technology and work, originated with Greek, Persian and Asian civilizations to deal with the demands of providing for burgeoning populations and of equipping large armies numbering in the hundreds of thousands (Smith 1965). For example, the largest naval battle in recorded history between the Greeks and Persians over 2000 years ago clearly points to a well-defined system of work for fabricating and equipping thousands of ships. A third example is the use of massive hydraulic, geared and levered machines for mining work in northern Europe in the fifteenth and sixteenth centuries, documented by Agricola in 1556, which anticipated by some three centuries the application of similar machines to mechanize work during the industrial revolution.

What factors prompted the emergence of successive innovation in forms of societal organization, technology and symbolic communication, depicted in Figures 3–5. Answers to this question remain a matter of speculation. Diamond (1997) points to favorable environmental circumstances. He postulates that Eurasian success in expeditions of conquest over other peoples may be traced ultimately to availability of native plant and animal species suitable for domestication in the Fertile Crescent area. This factor, coupled with suitable climatic conditions and lack of major geographic barriers, allowed early dispersal of domesticated crops and livestock from the Mediterranean basin to the Asian subcontinent and perhaps beyond, thereby facilitating early transformation of hunting and gathering modes of work in these areas to activities and systems of work associated with agriculture, mining and development of settled communities. A self-selective impetus for the innovation of new technologies plus approaches to work organization was thereby generated to address associated needs of large, established communities, such as better communication (writing), better weapons for defense or conquest (mining), and better coordination and integration of work on a societal basis (government and religion). Early success in work innovation thus is assumed to anticipate later success in subjugating other peoples. More broadly, we may postulate that there are reciprocal feedback relationships (Figure 1) between the availability of favorable environmental conditions readily controlled through work, the development of more stable and organizationally sophisticated settlements and societies, and the innovation of new modes and forms of work technologies and practices, a pattern that undoubtedly has prevailed throughout human evolution.

4. GENERAL LAWS AND CUSTOMS OF WORK

Laws and customs of work — the definition of "ergonomics" — refer to work design features and characteristics that conform to the capabilities and limitations of the worker. As evidenced in this encyclopedia and in diverse Ergonomics/Human Factors handbooks, textbooks and publications, there are literally hundreds of laws and customs of work that have been promulgated, in the form of guidelines, recommendations and standards. Customarily, these take the form of specifications pertaining to the design of any and all components of the work

system: tools, machines, equipment, work stations, software or hardware interfaces, material handling, work layout, production flow, work environment, jobs, or organizational systems. The underlying premise of such specifications is that modifications of work design to conform more closely to the behavioral abilities and skills of the worker will enhance work performance through improvements in productivity, quality, safety, health and/or efficiency of the workforce and the work system.

This encyclopedia is largely concerned with the topic of laws and customs of work. The discussion here therefore confines itself to the following series of general conclusions that can be drawn about "laws and customs" pertaining to the interaction of work design and work performance.

1. *Work performance is highly context specific*. This means that the behavior of the worker is specialized to a substantial degree in relation to the design context of the workplace and the work environment. In the domain of cognitive behavior, empirical evidence dating back ~150 years indicates that task design factors, as opposed to innate ability or learning factors, account for the preponderance of performance variability observed across a broad range of problem-solving, perceptual, and psychomotor tasks (Smith *et al.* 1984). Jones (1966) concludes that when it comes to actual cognitive performance in the workplace, as much as 90% of such variability may be attributable to task design. Evidence regarding the most robust "laws" in Ergonomics/Human Factors and psychological science — those of Hick and Fitts — shows that almost all of the variance in reaction or movement time is attributable to design of the task, namely number of choices or target size and separation respectively. Essentially all Ergonomics/Human Factors guidelines and standards pertaining to musculoskeletal performance at work reference work capacity and injury risk to design factors — weight and height of lift, task repetition and force requirements, etc. In the realm of work physiology, guidelines regarding both physical work capacity and fatigue susceptibility are referenced to task design factors such as work load and task duration. Recommendations pertaining to scheduling effects on work performance reference design factors such as when the shift is scheduled, length of shift, shift start and stop times etc. One corollary of this conclusion therefore is that "laws" of work cannot and should not be generalized across all work design contexts — instead they must be customized for different work environments and types of work. Another corollary is that insight into sources of variability in work performance for a particular type of work should be based on observations of that performance in the work design context under consideration.

2. *Work motivation also is context specific*. This means that there is no universal "law" that explains why people want to work. Although evidence on this point is less comprehensive, suggestive support is provided by observations that among a range of factors implicated in work motivation — related to opportunities that work provides for gainful return, for social interaction, for purposeful activity, for pride in accomplishment, for self-fulfillment etc. — no single factors or set of factors uniformly applies to all workers and work sectors. For slavery — a class of work quite prevalent for thousands of years until recent times — fear and intimidation were key motivators. For early humans, plus the millions of workers living in poverty today, simple survival through acquisition of food and avoidance of harm likely represents the most important motivational influence for work. The only generalization that appears to apply is that human motivations for work and for self-control of the environment essentially are identical (see Section 2).

3. *Effects of new work design on work performance are not necessarily predictable*. This conclusion follows from point 1. If work performance is context specific, new work designs (of tools, machines, interfaces, jobs, etc.) will evoke new patterns of variability in work performance which cannot necessarily be predicted from patterns associated with existing designs. To deal with this issue, there is growing acceptance of iterative design testing, in which some measure of usability (user acceptance, productivity, quality, safety risk, etc.) of a prototype design is assessed, and the design modified based on the results. This process is repeated until an acceptable level of performance with the new design is achieved.

4. *Operational validity of work performance models rests upon their recognition of context specificity*. Some models of work performance, such as ecological or various lifting and fatigue models, are based on the premise of performance–design interaction, and therefore build context specificity in performance directly into the framework of the model. Others, such as information processing, mental workload, human reliability or time–motion assessment and prediction models, disregard such interaction in their formulation. Context specificity in work performance is a matter of empirical record. The operational validity, and consequent predictive power, of models that ignore this record may therefore be questioned, relative to models that are context sensitive.

5. *Organize work as closed-loop system*. Cybernetic or closed-loop control of behavior represents an inherent biological feature of work performance at the level of the individual (Smith and Smith 1987). The same cannot be said of organizational behavior. Instead, closed-loop linkages between work effects (outcomes) and work investment (inputs) have to be built into the design — the macro-ergonomics — of work organizations. There is ample observational and economic evidence to show that if yield and investment of a work system (the ratio of which defines work productivity) are not closely coupled, the system ultimately is destined to flounder if not fail. The need to close the loop between the effort and effects of work undoubtedly was recognized by early hominid work groups, was manifest in ancient Sumerian and Egyptian work systems, and represents an organizational design imperative for successful complex work systems of today (Kramer 1963, Juran 1964). Closed-loop control of work may be implemented for a broad range of organizational design alternatives, from authoritarian to unsupervised. However, macro-ergonomic observations on organizational design and management point to strategies commonly associated with success in achieving this goal. One is clearly defined decision-making authority — every control system needs a controller. A second is the participatory approach — workers represent a good resource for tracking both the effort and effects of work. A third is to customize input and output measures of

work system performance to the particular type of system under consideration. A fourth is realistic accommodation of the inherent variability of a control system, particularly at the organizational level. Although this represents a key feature of modern quality management systems, the same cannot be said of other management domains.

6. *Human error and design error often are synonymous.* The invocation of "human error" as the underlying cause of work-related accidents and injuries is not uncommon in the Ergonomics/Human Factors and safety fields, a proclivity originating with Heinrich ~70 years ago (Heinrich 1931). From the perspective of point 1 however we might expect that hazards in the workplace, manifest as work design flaws, should be a major contributor to errors in work performance. Indeed, a variety of field observations support the general conclusion that design factors, often management-related, are implicated in about half of all work-related accidents and injuries (Smith 1999). Before invoking pure human error therefore, the possible contribution of flaws in physical and/or organizational work design to errors in work performance should be carefully scrutinized.

7. *Ergonomic standards addressing upper limb musculoskeletal injury risk should accommodate individual differences and be referenced to work design.* This conclusion is based on three considerations. First, motor behavioral theory tells us that there are essentially an infinite number of alternative trajectories that can be employed for guiding movements of the hands and arms from one point in space to another. This is because the total degrees of freedom in the combined muscle–joint systems of these effectors (7) exceeds that of the spatial environment in which movement occurs (6). Second, motor behavioral research tells us that the motor control system is inherently imperfect. This motor control variability ensures that no particular arm–hand movement is ever repeated with precisely the same spatio-temporal pattern by a given individual. Collectively, these two phenomena underlie the observation that upper limb musculoskeletal performance is highly individualized and context specific (point 1), such that in the same work design context, different individuals may perform differently, and the same individual may perform differently at different times. The idea of "general" ergonomic standards for upper limb musculoskeletal performance that are presumed to apply for all workers across all work design contexts therefore is scientifically untenable. Instead, such standards should be customized for different classes of workers (categorized by age, gender, anthropometry, etc.) and for different types of work.

8. *To modify work performance, modify work design.* As a cybernetic, self-regulated process, work behavior cannot be directly controlled by an outside force or agent. However, variability in work performance can be influenced by modifications to the design of the work environment in which the performance occurs (point 1). Therefore, modifications to work design (encompassing the design of training programs) offer the most realistic, and often the most cost-effective, avenue for modifying work performance. This conclusion represents the *raison d'être* for Ergonomics/Human Factors science.

5. FUTURE OF WORK

From a biological perspective, the basic behavioral and physiological attributes and capabilities that humans bring to the performance of work likely have remained fairly consistent over previous millennia, and are unlikely to change much in the millennium to come. Some historical support for this view is available. For example, the ergonomic guideline introduced by the ancient Greeks over 2000 years ago that packs carried by Hoplite soldiers be limited to 70 lb is retained by armies of today.

What has changed is the amount of knowledge and information that has accumulated about the organization and conduct of work, and the number of technological and communication innovations developed to facilitate the work process (Figures 3–5). However, given that variability in work performance is context specific, the nature of the impact of new designs of technology and work systems on work performance is not entirely predictable (Section 4, points 1 and 3). Hancock (1997) suggests that given the numbing effect of much computer-based work on job satisfaction, more avenues for job enjoyment and opportunities for leisure must be introduced to sustain worker acceptance of this type of work. Drucker (1999) addresses the possible impact of the newest major technological innovation in work design — the internet — on the future of work, and expresses uncertainty as to what the outcomes will be. In support of such indecisiveness, he points out that the industrial revolution, which also involved major transformations in work design, had many unintended and unanticipated consequences (Section 4, point 3) for the performance of work.

Nevertheless, at least four predictions for the future trajectory of work can be offered with a reasonable degree of assurance, all of which already are somewhat in evidence. The first is that the persistence of new work designs will rest upon the degree to which they augment work productivity, albeit with some lag built in. For example, many observers attribute recent productivity gains to the wholesale computerization of work, an effect that has taken some three decades to emerge. Work designs that do not benefit work productivity eventually will be discarded.

The second and third predictions are that work will become more disseminated on a regional basis, but more integrated on a global basis. Advances in distributed communication such as the internet will facilitate both of these transformations. The need for maintaining production facilities in discrete, physically defined locations, for purposes such as manufacturing, mining or agriculture, should not change much in the foreseeable future, although the degree of remote control of such work likely will steadily increase. However, for many other types of work across the entire commercial sphere, from education to finance to retail, concepts of the "job" and of "going to work" almost certainly will undergo fundamental modifications.

Finally, the decline in population growth and progressive aging of the workforce in most western societies, coupled with an unabated demand for workers, suggests that respect for work capabilities of older workers will undergo a resurgence, and that the concept of retirement as an expected and accepted concomitant of aging may be fundamentally revised.

REFERENCES

AGRICOLA, G., 1556, *de re Metallica*, trans. from the first Latin ed. by H.C. Hoover and L.H. Hoover (London: Mining Magazine, 1912).

DIAMOND, J., 1997, *Guns, Germs, and Steel: The Fates of Human Societies* (New York: W.W. Norton).

DRUCKER, P.F., 1999, Beyond the information revolution, *Atlantic Monthly*, 284), 47–59.

GALBRAITH, J.K., 1958, *The Affluent Society* (Boston: Houghton Mifflin).

GOULD, J.D., 1990, How to design usable systems. In M. Helander (ed.), *Handbook of Human-Interaction* (Amsterdam: Elsevier), 757–89.

HANCOCK, P.A., 1997, On the future of work, *Ergonomics in Design*, 5, 25–9.

HEINRICH, H.W., 1931, *Industrial Accident Prevention: A Scientific Approach* (New York: McGraw-Hill).

HOROWITZ, D., 1998, *Marx's Manifesto: 150 Years of Evil* (Los Angeles: Center for the Study of Popular Culture).

HUMAN FACTORS AND ERGONOMICS SOCIETY, 1997, *Directory of Human Factors/Ergonomics Graduate Programs in the United States and Canada* (Santa Monica: Human Factors and Ergonomics Society).

JONES, M.B., 1966, Individual differences. In E.A. Bilodeau (ed.), *Acquisition of Skill* (New York: Academic Press), 109–46.

JURAN, J.M., 1964, *Managerial Breakthrough: A New Concept of the Manager's Job* (New York: McGraw-Hill).

KRAMER, S.N., 1963, *The Sumerians* (Chicago: University of Chicago Press).

MARSHACK, A., 1964, Lunar notation on Upper Paleolithic remains, *Science*, 146, 743–6.

MARX, K., 1867, *Das Kapital*.

MARX, K. and ENGELS, F., 1848, *Communist Manifesto*.

McGREGOR, D., 1960, *The Human Side of Enterprise* (New York: McGraw-Hill).

OAKLEY, K.P., 1957, *Man the Tool-Maker* (Chicago: University of Chicago Press).

SALVENDY, G. (ed.), 1982, *Handbook of Industrial Engineering* (New York: Wiley).

SALVENDY, G. (ed.), 1987, *Handbook of Human Factors* (New York: Wiley).

SMITH, A., 1776, *Inquiry into the Nature and Causes of the Wealth of Nations*.

SMITH, K.U., 1962, *Work Theory and Economic Behavior*. Indiana Business Paper no. 5 (Bloomington: Indiana University Foundation for Economic and Business Studies).

SMITH, K.U., 1965, *Behavior Organization and Work: A New Approach to Industrial Behavioral Science*, revd (Madison: College Printing & Typing).

SMITH, T.J., 1999, Macroergonomics of hazard management. In H.W. Hendrick and B. Kleiner (eds), *Macroergonomics* (Mahwah: Lawrence Erlbaum). [accepted]

SMITH, T.J., HENNING, R.H. and SMITH, K.U., 1994, Sources of performance variability. In G. Salvendy and W. Karwowski (eds), *Design of Work and Development of Personnel in Advanced Manufacturing* (New York: Wiley), 273–330.

SMITH, T.J. and SMITH, K.U., 1987, Feedback-control mechanisms of human behavior. In G. Salvendy (ed.), *Handbook of Human Factors* (New York: Wiley), 251–93.

TAYLOR, F.W., 1911, *The Principles of Scientific Management* (New York: Harper & Bros).

WOLF, W., 1954, *Die Welt die Aegypter* (Stuttgart: Gustav Kilpper).

WOOLEY, L., 1946, *Ur: The First Phases* (London: Penguin).

Human Error

J. Reason

Department of Psychology, University of Manchester, Manchester M13 9PL, UK

1. INTRODUCTION

Error is defined here as the deviation of human performance from some intended, desired or ideal standard. Such deviations may have bad outcomes, but they can also be inconsequential or even benign – as in trial-and-error learning. Errors are not intrinsically bad, though their consequences and the conditions provoking them may be. Systematic error forms arise from highly adaptive cognitive processes. It is the circumstances of their occurrence rather than the underlying mechanisms that largely determine the nature of the outcome.

Before the mid-1970s errors were rarely studied for their own sake, although there were some notable exceptions. Two developments then occurred to make error a legitimate object of scientific study. The first was theoretical. The mid-1970s saw the reappearance of Bartlett's schema concept in several different guises (Reason 1990). Common to these new approaches was the revival of Bartlett's (1932) insight that knowledge of all kinds was stored in long-term memory as active, organized structures. A key feature of these new schema theories was the idea that knowledge structures (schemata) contain informational slots or variables. If concurrent mental processing or information from the outside world fails to provide inputs to these slots, they take on default or stereotypical values. Such a concept proved central to the understanding of a recurrent characteristic of many different error types, namely that they take forms that are more familiar and more conventional than those which were intended or appropriate at that time. In short, many errors appear as strong habit intrusions.

The second development was the growing concern over the terrible cost of human error in hazardous technologies, particularly those in which a relatively few individuals exercised centralized control over complex, well-defended systems (Hollnagel 1993; Kirwan 1994). Three overlapping eras of safety concern can be identified: the *technical failure era*, the *human failure era* and the *system failure era*. The 1950s and 1960s saw technical failures as the primary threat (e.g. failure of engines or the structural integrity of aircraft fuselages and pressure vessels). In the 1970s and 1980s, human unsafe acts were recognized as being the single largest contributor to both major and minor accidents. Between the late 1980s and the present time, however, there has been a marked shift of emphasis towards a wider acknowledgement of the situational, managerial, organizational and regulatory contributions to system breakdown. We are now in the system failure era in which people at the "sharp end" of a hazardous technology (e.g. pilots, control room operators, ships' officers, etc.) are seen more as the inheritors of "organizational accidents" than as their principal instigators.

2. VARIETIES OF HUMAN ERROR

There is no one taxonomy of human error – nor should there be, since different classificatory schemes are created to serve different purposes (Norman 1981). In what follows, errors are classified according to the type of deviation involved. The various classes of error are produced by different psychological processes, vary widely in their ease of detection and recovery, occur at different levels of the system and require different remedial measures. They are described below in order of increasing complexity.

2.1. Trips and Stumbles

This deviation is between an internal representation of the upright (generated by the vestibular system and cerebellum) and a sudden, unintentional, off-vertical body posture. The nervous system is hard-wired to detect such discrepancies and immediately initiates reflex postural corrections without the intervention of any conscious decision-making. Detection and recovery is automatic.

2.2. Slips and Lapses

These are the unintended deviations of action from current intention. They occur at the level of execution and are commonly termed absent-minded actions. Actions-not-as-planned take a variety of forms depending upon which cognitive process is primarily implicated. In general, the term "slips" will be reserved for those discrepant actions involving the deployment of an inappropriate level of attention (this may be either too much or too little) to the monitoring of actions or perceptions. "Lapses," on the other hand, are commonly associated with memory failures, either forgetting (in the case of intentions – a very common lapse) or the inability to retrieve some known memory item (e.g. failing to achieve a timely recall of the name of a familiar person). Another distinction is that action slips are publicly observable, while lapses tend to be more private events – though their consequences may be evident to others. Since the intention associated with the deviant behavior (i.e. the standard against which such discrepancies are identified) is generally well known to the actor, the detection of slips is usually fairly rapid, though memory lapses may have to wait upon a particular occasion before they are noticed (e.g. the failure to post a letter may only be realized when one has returned home to find the unposted letter).

Three conditions appear to be necessary for the occurrence of an absent-minded slip of action: (1) the performance of an habitual sequence of actions in familiar surroundings, (2) the occurrence of some change, either in the usual plan of action or in its circumstances and (3) a marked degree of "attentional capture" associated with internal preoccupation or with external distraction. Slips are thus a penalty for the ability to automatize our recurrent actions. Though such slips may occasionally have tragic consequences, they are a relatively small price to pay for the enormous benefits of performing the routines of everyday life without having to make moment-to-moment demands upon the limited resources of conscious decision-making. Continuous "present-mindedness" would be a far worse fate.

2.3. Mistakes

Mistakes occur at the higher level of formulating intentions, specifying goals and planning the means to achieve them. The deviation here is not between action and intention, but between the selected course of action and some more adequate means of attaining the immediate and longer-term goals. Since the future state of the world is uncertain and the ideal path towards a given

Table 1. Relationship between rule-related situations and types of mistake.

Type of situation	Type of mistake
Good rules	Mistaken circumvention (misvention) bad outcome successful outcome
Bad rules	Strong-but-wrong rule applied Inelegant or ill-advised rule applied Mistaken compliance (mispliance) Malicious compliance (malpliance)
No rules	Mistaken improvisation

goal is rarely known (even after the event, it can be a matter of debate), mistakes are extremely hard to detect and recover by their perpetrators.

In the world of hazardous systems, these plans of action are rarely made from scratch. They are generally guided and constrained by rules, regulations and procedures – subsumed here under the general term "rules." Every organization faces the challenge of limiting the enormous variability of human behavior to that which is safe, productive and efficient. Each does so by a mixture of external and internal controls, both of which involve rules. External controls rely heavily – but not exclusively – on prescriptive procedures. Internal controls exist primarily within the head of the individual and are acquired through training and experience. The factors that shape the balance between external and internal controls have been discussed at length elsewhere (Perrow 1967); for our present purposes, the term "rules" will be used to cover both types of control.

For any given situation, the rules (either on paper or in the head) may take one of three forms: (1) *good rules* that are appropriate for the situation, (2) *bad rules* that are either inappropriate for the situation or are inelegant and/or ill-advised and (3) *no rules* are available. The latter state corresponds to what Rasmussen (1982, 1986) has termed the "knowledge-based level" of performance. Mistakes can be conveniently classified according to which of these three situations prevail. The classification is summarized in table 1.

2.3.1. Mistaken circumventions

The operators of the Chernobyl nuclear power plant clearly committed misventions when they persisted with the electrical testing at a dangerously low power regime and then proceeded to disable successive plant safety systems. But most violations of safe operating procedures tend to go unpunished and may become routine as an easier or more convenient way of working. Such rewarded but unwise violations create the breeding ground for more dangerous misventions. The violation of good rules has at least two adverse effects: it makes the commission of subsequent errors more likely, and it increases the probability that these subsequent errors will have bad consequences.

2.3.2. "Strong-but-wrong" mistakes

These involve the use of a normally good rule in circumstances for which it is no longer applicable. The mistake lies in failing to spot the counter-indications. A simple example would be braking suddenly to avoid a pedestrian when the road is icy. The problems

are that the rule has demonstrated itself to be appropriate in most circumstance (hence it is "strong") and the counter-indications are often hard to detect (Reason 1990).

2.3.3. Application of inelegant or ill-advised rules

Many people acquire "bad" rules during the course of performing some task. The application of such rules may go uncorrected for long periods. Indeed, they may even appear to be serviceable. The problem is that they are inelegant (costing unnecessary effort or resources) or, under certain conditions, ill-advised (having dangerous consequences). A technician involved in re-wiring a signal box committed a tragic example of this type of mistake. Instead of removing the old wires, he bent them back. Subsequently, the old wire made a connection with the terminal, causing the signal failure that led to the Clapham Junction rail disaster of 1988.

2.3.4. Mistaken compliance

Procedure writers are fallible. The rules they create may prove incorrect, unworkable, unintelligible, ambiguous or unavailable when needed. Such rules can also be rendered wholly inappropriate by an unforeseen change of circumstance. On the *Piper Alpha* oil and gas platform, for example, the procedures directed personnel to assemble in the accommodation in the event of an emergency. But this area was directly downwind of the flames on the evening of 6 July 1988 when an explosion occurred onboard the platform. Most of those who died followed the emergency procedures (mispliance). Most of the survivors jumped into the water (successful violation).

2.3.5. Malicious compliance

It is unlikely that all malpliances are mistakes, but some of them certainly are. A malpliance is committed when the workforce deliberately adheres to procedures that are known to be bad or unworkable to thwart the management.

2.3.6. Mistaken improvizations

In well-established work domains most of the likely situations have either been anticipated or encountered previously and are thus covered by standard operating procedures. On rare occasions, however, entirely novel circumstances occur for which no rules are available. The blow-out of the oxygen tank on *Apollo 13* was one such case; the loss of all three hydraulic systems on a United Airlines DC10 aircraft over Iowa was another. In both instances the performance of the crews was truly inspired. More usually, however, people do not perform well in situations for which they lack both training and procedures. Mistakes are inevitable under such circumstances. The evidence suggests that when the people on the spot are both highly experienced and extremely talented, the probability of their improvizations having a satisfactory outcome is ~0.5, the chance level (Reason 1997). It could be argued that the term "error" is inappropriate in wholly novel situations, since it is only through trial-and-error learning that those involved can iterate towards an effective outcome. Nonetheless, the category of "mistaken improvizations" will be retained here both for logical completeness and to emphasize the difficulty of making hard-and-fast distinctions between error and correct performance.

3. ACTIVE FAILURES AND LATENT CONDITIONS

So far, human error from a largely person-centered perspective has been considered. The difficulty with this approach is that it encourages those in authority to treat errors as a moral issue and to focus their remedial efforts on "deviant" individuals rather than on the system at large. One of the basic rules of error management is that one cannot change the human condition, but one can change the conditions under which people work. The current trend in human factors research is to view errors and violations from a system-centered perspective. And it is from this standpoint that one arrives at the final distinction in this chapter, that between active failures and latent conditions.

3.1. Active Failures

These are the unsafe acts committed by those at the immediate human–system interface. Such errors can have an immediate impact upon the safety of the system, though their effects may be relatively short-lived. Not so long ago these proximal errors tended to be the main focus of accident inquiries. In recent years, however, the scope of such investigations has extended both "upstream" of the front-line operators and backwards in time.

3.2. Latent Conditions

These are the product of decisions made by designers, builders and top-level managers. Such decisions may later prove to be mistaken, but they need not be. Nearly all such strategic decisions, good or bad, have the potential for seeding latent conditions into the system. Latent conditions (or organizational pathogens) have two kinds of impact: they can translate into error-provoking conditions in the workplace (e.g. time pressure, inadequate tools and equipment, insufficient training, etc.) and they can also open up long-lasting gaps or weaknesses in the system's defenses (e.g. unworkable procedures, weak containment, inoperable engineered safety features, etc.). These conditions may lie dormant for many years before they combine with local triggering events to create a breakdown of the system's defenses. They are called "conditions" rather than "failures" because they are present in all systems and do not differentiate between an organization that has just had an accident and one that has not. While their specific forms may not be made manifest until after a bad event, they are rooted in generic organizational processes such as designing, building, managing and maintaining. Although it may not be possible to prevent the introduction of latent conditions into the system, it has been demonstrated that their likely hiding places can be identified and remedied before they contribute to an adverse event (Reason 1997).

4. ERROR MANAGEMENT

Errors and violations at the "sharp end" may be likened to mosquitoes. It is possible to swat or spray them one by one, but they keep on coming. The only effective remedy is to drain the swamps in which they breed. The swamps, in this case, are the latent error-provoking and accident-producing conditions present in all technical systems.

REFERENCES

BARTLETT, F.C., 1932, *Remembering: A Study in Experimental and Social Psychology* (Cambridge: Cambridge University Press).

HOLLNAGEL, E., 1993, *Human Reliability Analysis: Context and Control* (London: Academic Press).

KIRWAN, B., 1994, *A Guide to Practical Human Reliability Assessment* (London: Taylor & Francis).

NORMAN, D.A., 1981, Categorization of action slips. *Psychological Review*, 88, 1–15.

PERROW, C., 1967, A framework for the comparative analysis of organizations. *American Sociological Review*, 32, 194–208.

RASMUSSEN, J., 1982, Human errors: a taxonomy for describing human malfunctions in industrial installations. *Journal of Occupational Accidents*, 4, 311–35.

RASMUSSEN, J., 1986, *Information Processing and Human–Machine Interaction* (Amsterdam: North-Holland).

REASON, J., 1990, *Human Error* (New York: Cambridge University Press).

REASON, J., 1997, *Managing the Risks of Organizational Accidents* (Aldershot: Ashgate).

Human Reliability Analysis

E. Hollnagel

Graduate School of Human-Machine Interaction, Linköpings Universitet,
S-581 83 Linköping, Sweden

1. INTRODUCTION

The term "human reliability" is usually defined as the probability that a person will correctly perform some system-required activity during a given time period (if time is a limiting factor) without performing any extraneous activity that can degrade the system. The historical background for the development of the set of methods that are commonly referred to as Human Reliability Analysis (HRA) was the need to describe incorrect human actions in the context of Probabilistic Risk Assessment (PRA) or Probabilistic Safety Analysis (PSA). The premises for HRA were, and are, therefore that it must function within the constraints defined by PRA/PSA and specifically that it can produce the human action probabilities that are needed by the PRA/PSA. The accident sequence that is analyzed by a PRA/PSA is typically represented as an event tree (see Figure 1). A node in the sequence of events that may lead to the accident represents a specific function, task, or activity that can have two different outcomes, usually denoted *success* and *failure*. A node can either represent the function of a technical system or component, or the interaction between an operator and the process. For example, if the analysis considers the sequence of events that are part of landing an aircraft, the event "*timely extraction of flaps*", which is an action that must be taken by the pilot, is represented by a node. From the perspective of the PRA/PSA there is a need to know whether it is likely that an event will succeed or fail, and further to determine the probability of failure in order to calculate the combined probability that a specific outcome or end state will occur. If the node represents the function of a mechanical or electronic component, the failure probability can, in principle, be calculated based on engineering knowledge alone. If the node represents the interaction between an operator and the process, engineering knowledge must be supplemented by a way of calculating the probability that the human, as a "component", will fail. Historically, the role of HRA has been to provide the
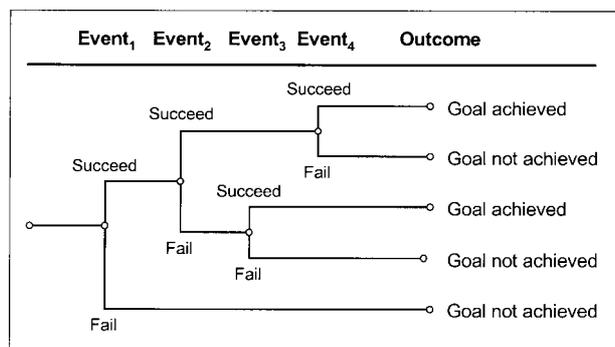
foundation for calculating this probability. The sought for value has traditionally been called a human error probability (HEP), but as the following will show this is both a misunderstood and misleading term.

2. HUMAN ERROR PROBABILITIES AND PERFORMANCE SHAPING FACTORS

The practice of HRA goes back to the early 1960s, but the majority of HRA methods were developed in the middle of the 1980s — mainly as a consequence of the concern caused by the accident in 1979 at the nuclear power plant at Three Mile Island. Partly due to the conditions under which it was developed, HRA methods from the beginning used procedures similar to those employed in conventional reliability analysis. The main difference was that human task activities were substituted for equipment failures and that modifications were made to account for the greater variability and interdependence of human performance as compared with that of equipment. The traditional approach is first to determine the HEP for a node, either by using established tables, human reliability models, or expert judgement. The characterization of human failure modes is usually very simple, for instance in terms of "error of omission" and "errors of commission". Since human actions clearly do not take place in a vacuum, a second step is to account for the influence of possible Performance Shaping Factors (PSF) such as task characteristics, aspects of the physical environment, work time characteristics, etc. This influence is expressed as a numerical factor that is used to modify the basic HEP. The resulting formula for calculating the probability of a specific erroneous action (P_{EA}) is shown below:

$$P_{EA} = HEP_{EA} * \sum_{k=1}^{N} PSF_k * W_k + C$$

From the standpoint of behavioral sciences, this formula makes two fundamental assumptions. First, that the probability of failure can be determined for specific types of action independently of any context. Second, that the effects of the context are additive, which is the same as saying that the various performance conditions (such as interface quality, stress, level of training, task complexity, etc.) do not influence one another. Neither of these assumptions are reasonable, and either alone constitutes a grave deficiency of the approach. Quite apart from these principal objections, HRA methods in practice turned out not to be sufficiently effective and the need for substantial improvements was gradually realized. In 1990, this was expressed as a criticism against HRA on the following points (Dougherty 1990):

- Existing empirical data are insufficient to support quantitative predictions of human performance in complex systems. This problem had actually been recognized by HRA practitioners since the early 1960s. As alternatives to empirical data, HRA often relied on either expert judgement or data from simulator studies.
- Expert judgements can be used *in lieu* of empirical data, but there is a lack of agreement in the use of expert judgement methods. The methods neither have satisfactory between-expert consistency, nor produce accurate predictions.
- Data from simulator studies can be used instead of empirical data, but the calibration to real life situations is inadequate.

Figure 1. A simplified event tree representation.

The veracity and validity of simulator data have not yet been convincingly demonstrated.

- The accuracy of predictions from HRA methods is debatable and generally unproved, particularly for non-routine tasks. Benchmark studies usually produce divergent results, for many different reasons.
- The psychological realism in most HRA methods is inadequate, and the assumptions about human behavior are often highly questionable from a psychological point of view.
- The treatment of important Performance Shaping Factors is inadequate, there is too little emphasis on PSFs relating to management, organization, culture, etc.

Considering the historical basis for HRA methods, it is reasonable to suggest that one improvement should be a better integration of psychological theory with HRA models and methods. In particular, it is strongly felt by many with a behavioral science background that quantification should await an improved theoretical foundation.

3. CONTEXT AND COGNITION IN HRA

The above-mentioned criticism caused an intensive debate within the community of HRA theoreticians and practitioners, and pointed to two problems that need to be addressed by HRA methods: the problem of context and that of cognition. Technical systems that depend on human–machine interaction to accomplish their function, such as nuclear power plants, are typically tightly coupled and have complex interactions. Plant operators and plant components should therefore be seen as interacting parts of an overall system that responds to upset conditions. The actions of operators are not simply responses to external events, but are governed by their beliefs as to the current state of the plant. Since operators make use of their knowledge and experience, their beliefs at any given point in time are influenced by the past sequence of events and by their earlier trains of thought. In addition, operators rarely work alone but are part of a team — especially during abnormal conditions. Altogether this means that human performance takes place in a context which consists both of the actual working conditions and the operator's perception or understanding of them. It also means that the operator's actions are a result of cognition and beliefs, rather than simple responses to events in the environment, and that the beliefs may be shaped — and shared — by the group.

The problem of *context* is a noticeable feature of other current theories, such as the so-called multi-threaded failure models that describe how accidents in systems may evolve (Reason 1990). Events are described as determined by a combination of psychological, technological, and oganizational or environmental factors. The issue of organizational reliability has recently received specific attention by the PRA/PSA community, and proposals for an organizational reliability analysis have been made. On the level of human performance, which is the focus of HRA, context is important because human action always is embedded in a context. Given a little thought this is obvious, but the preferred mode of representation that the analyses use — the event tree or operator action tree — is prone to be misleading, since it represents actions without a context. One unfortunate consequence of that has been the preoccupation with the HEP, and the oversimplified concept of "human error". The consequence of acknowledging the importance of the context is that HRA should not attempt to analyze actions separately, but instead treat them as parts of a whole. Similarly, "human error" should be seen as the way in which erroneous actions can manifest themselves in a specific context, rather than as a distinct and well-defined category.

In the search for a way of describing and understanding the failure of human actions, several classes of models have been used. In brief, HRA methods seem to include one of the following types of operator model:

- Behavioral, or human factors, models that focus on simple manifestations (error modes). The error modes are usually described in terms of omissions, commissions, and extraneous actions, and the methods aim at deriving the probability that a specific manifestation will occur. Since causal models are either very simple or non-existent, the theoretical basis for predicting performance failures is inadequate. Behavioral models are therefore also weak in accounting for the influence of context.
- Information processing models that focus on internal "mechanisms" — for example, decision-making or reasoning. The methods aim at explaining the flow of causes and effects through the models. Causal models are therefore often complex, but with limited predictive power, and little concern for quantification. Error types typically refer to the cause as much as the manifestation (e.g. slips, lapses, mistakes, violations), or to the malfunctioning of a hypothetical information processing function. Context is not considered explicitly, at most in terms of the input to the operator, and information-processing models are better suited for retrospective analysis than for predictions.
- Cognitive models that focus on the relation between error modes and causes, where the latter refer to the sociotechnical environment as a whole. Unlike information-processing models, cognitive models are simple and context is explicitly represented. Cognition is the reason why performance is efficient — and why it is sometimes limited — and the operator is seen as not only responding to events but also as acting in anticipation of future developments. Cognitive models are well suited for both predictions and retrospective analyses. There are, however, only a few HRA approaches that pay more than lip service to this type of model.

The problem of *cognition* can be illustrated by the popular notion of "cognitive error". The problem facing HRA analysts was that human cognition undoubtedly affected human action, but that it did not fit easily into any of the established classification schemes or the information-processing models. One solution was simply to declare that any failure of a human activity represented in an event tree, such as diagnosis, was a "cognitive error". However, that did not solve the problem of where to put the category in the existing schemes. A genuine solution is, of course, to realize that human cognition is a cause rather than an effect, and that "cognitive error" therefore is not a new category or error mode. Instead, all actions — whether erroneous or not — are determined by cognition, and the trusted categories of "error of omission" and "error of commission" are therefore as cognitive as anything else. Furthermore, the cognitive viewpoint implies that unwanted consequences are due to a mismatch between cognition and context, rather than to specific "cognitive error mechanisms".

4. PRINCIPLES OF A CONTEMPORARY HRA

The current dilemma of HRA stems from its uneasy position between PRA/PSA and information processing psychology. As argued above, HRA has inherited most of its concepts and methods from the practice of PRA/PSA. This means, that HRA uses a fixed event representation based on a predefined sequence of steps, that the methods are a mixture of qualitative and quantitative techniques, and that the underlying models of operator actions and behavior are either probabilistic or simplified information processing models. Yet information-processing psychology is characterized by almost the opposite approach. The representation is typically information flow diagrams of internal functions, rather than binary trees of external events; the methods are generally qualitative/descriptive, and the models are deterministic (because an information-processing model basically is a deterministic device). HRA practitioners have therefore had great problems in using the concepts from information-processing psychology as a basis for generating action failure probabilities; a proper adaptation would practically require a complete renewal of HRA. Even worse, information-processing models are almost exclusively directed at retrospective analysis, and excel in providing detailed explanations of how something can be explained in terms of internal mental mechanisms. HRA, on the other hand, must by necessity look forward and try to make predictions.

Since neither classical HRA nor information-processing psychology is easily reconciled with the need to account for the interaction between context and cognition, it is difficult to find a solution to the current predicament. The main problem is that the cocneptual foundation for HRA methods either is misleading or is partly missing. The solution can therefore not be found by piecemeal improvements of the HRA methods. In particular, attempts to rely on information processing psychology as a way out are doomed to fail because such models are retrospective, rather than predictive.

The solution to the current problems in HRA is not just to develop a new approach to quantification, but to develop a new approach to performance prediction that clearly distinguishes between the qualitative and quantitative parts. The purpose of *qualitative* performance prediction is to find out which events are likely to occur, in particular what the possible outcomes are. The purpose of *quantitative* performance prediction is to find out how probable it is that a specific event will occur, using the standard expression of probability as a number between 0 and 1. Qualitative performance prediction thus generates a set of outcomes that represent the result of various event developments. The validity of the set depends on the assumptions on which the analysis is based, in particular the detailed descriptions of the process, the operator(s), and the interaction. If the assumptions are accepted as reasonable, the set of outcomes will by itself provide a good indication of the reliability of the system, and whether unwanted outcomes can occur at all. That may be sufficient in the first instance. It may only be necessary to proceed to a quantitative performance prediction if significant unwanted consequences are part of the set of possible outcomes.

5. QUALITATIVE PERFORMANCE PREDICTION

The qualitative performance prediction must be based on a consistent classification scheme that describes human actions in terms of functions, causes, dependencies, etc., as well as a systematic way or method of using the classificaiton scheme. A consistent classification scheme is essential if assignments of erroneous actions are to be justified on psychological grounds. A clear method is necessary because the analysis will otherwise become limited by inconsistencies when the classification scheme is applied by different investigators, or even by the same investigator working on diferent occasions. It can further be argued that a classification scheme must refer to a set of concepts for the domain in question, i.e., a set of supporting principles, specifically a viable model of cognition at work. The model will guide the definition of specific system failures from a consideration of the characteristics of human cognition relative to the context in which the behavior occurs. The overall approach of a contemporary HRA should be something along the following lines:

1. *Application analysis*. It is necessary first to analyze the application and the context. This may in particular involve a task analysis, where the tasks to be considered can be derived from teh PRA/PSA as well as from other sources. The application analysis must furthermore consider the organization and the technical system, rather than just the operator and the control tasks.
2. *Context description*. The context must be systematically described in terms of aspects that are common across siutations. If insufficient information is available it may be necessary to make assumptions based on general experience, particularly about aspects of the organization, in order to complete the characterization.
3. Specification of target events. The target events for the human actions/performance can be specified in several ways. One obvious source is the PRA/PSA, since the event trees define the minimum set of events that must be considered. Another is the outcome of the application and task analyses. A task analysis will, in particular, go into more detail than the PRA/PSA event tree, and may thereby suggest events or conditions that should be analyzed further.
4. *Qualitative performance analysis*. The qualitative performance analysis uses the classification scheme, as modified by the context, to describe the possible effects or outcomes for a specific initiating event. The qualitative performance analysis, properly moderated by the context, may in itself provide useful results, for instance by showing whether there will be many or few unwanted outcomes.
5. *Quantitative performance prediction*. The last step is the quantitative performance prediction, which is discussed in detail below.

6. QUANTITATIVE PERFORMANCE PREDICTION

The quantification of the probabilities is, of course, the *sine qua non* for the use of HRA in PRA/PSA. The quantification has always been a thorny issue, and will likely remain so for many years to come. Some behavioral scientists have argued — at times very forcefully — that quantification in principle is impossible. More to the point, however, is whether quantification is really necessary, and how it can be done when required. From a historical

perspective, the need for quantification may, indeed, be seen as an artifact of how PRA/PSA is carried out.

At present, the consensus among experts in the field is that quantification should not be attempted unless a solid qualitative basis or description has first been established. If the qualitative performance analysis identifies potentially critical tasks or actions, and if the failure modes can be identified, then it is perhaps not necessary to quantify beyond a conservative estimate. In other words, the search for specific HEPs for specific actions may not always be necessary. To the extent that quantification is required, the qualitative analysis may at least be useful in identifying possible dependencies between actions. The description of the context may also serve as a basis for defining ways of preventing or reducing specific types of erroneous actions through barriers or recovery.

Traditionally, quantification involves finding the probability that a specific action may go wrong and then modifying that by the aggregated effect of a more or less systematic set of PSFs. An alternative approach is to begin the prediction by identifying the context and the common performance conditions (Hollnagel 1998). Following that, each target event is analyzed *in the given context*. This means that the analysis is shaped by the defined context and that the list of possible causes of a failure contains those that are the likely given the available information. Once the set of possible causes has been identified, the need will arise to quantify the probability that the target event fails. Since, however, the target event is associated with a set of context specific error modes, it is reasonable to estimate the probability on that basis — rather than treating the target event in isolation.

Initially, it may be necessary to use expert judgments as a source of the probability estimates. But keeping the criticisms of HRA in mind, it is crucial that the expert judgemnts are calibrated as well as possible. Meanwhile, every effort should be made to collect sufficient empirical data. Note, however, that empirical data should not be sought for separate actions but for types of actions. When the observations are made, in real life or in simulators, both actions and context should be recorded, and a statistical analysis should be used to disentangle them. This may, eventually, provide a reliable set of empirical data to supplement or replace expert judgements. The principles of data collection are thus fundamentally different from a traditional HRA. Furthermore, data collection can be guided by the classification scheme, since this provides a consistent and comprehensive principles according to which the data can be organized and interpreted. That by itself will make the whole exercise more manageable.

REFERENCES

DOUGHERTY, E.M. Jr., 1990, Human reliability analysis: Where shouldst thou turn? *Reliability Engineering and System Safety*, 29(3), 283–99.

HOLLNAGEL, E., 1998, *Cognitive reliability and error analysis method* (Oxford: Elsevier Science).

REASON, J.T., 1990, *Human error* (Cambridge, UK: Cambridge University Press).

Learning and Forgetting

J.R. Buck

Department of Industrial Engineering, University of Iowa, IA 52242, USA

1. INTRODUCTION

Learning is a phenomenon of people where human performance *improves with experience*. As tasks are repeated elements of the task are better remembered, cues are more clearly detected, skills are sharpened, transitions between successive tasks are smoothed, eye-hand coordination is more tightly coupled, and relationships between task elements are discovered. The aggregation of these and other personnel improvements will offer faster performance times, fewer errors, less effort, and there is often a better disposition that results. Several historical studies on individual person learning include Barnes and Amrine (1942), Knowles and Bell (1950), Glover (1966), Hancock and Foulke (1966), and Hancock (1967).

The detection of learning follows the observation of these changes in performance due *solely* to experience. Clearly, changes in the methods of performing a task, automating jobs, imparting information about the job, training, new incentive systems, and many other things can cause performance changes other than just learning. Thus, detection involves the identification of an improvement trend as a function of more experience. It also involves the elimination of other explanations for this improvement. Like a theory, *learning* can never be proved; it can only be disproved. After detecting learning, its measurement involves fitting mathematical models, called *learning curves*, to describe performance data changes with experience. This description involves the selection of the most accurately fitted model and the determination of the parameters so that improvement can be predicted.

Some of the same causes that contribute to an individual person's improvement with experience are similar to the causes of improvement by crews, teams, departments, companies, or even industries with experience. As a result, similar mathematical models are often fit to organizational performance changes. However, the term *progress curves* is more often applied to cases involving assembly lines, crews, teams, departments, and other smaller groups of people. (Professor S. Konz (1990) champions this terminology without necessarily distinguishing between individual person learning and the learning by small groups of people.) Whereas the term *experience curves* is often applied to larger organizational groups such as companies and industries (see Conley (1970) or Hax and Majluf (1982) for studies on experience curves over industrial sectors). The principal distinction between these different types of improvement curves is that between-person activities (e.g. coordination) occur as well as within-person learning. In the case of progress curves, there are important effects due to numerous engineering changes. Experience curves also embody scientific and technological improvements as well as progressive engineering changes and individual-person learning. Regardless of the term used, the same models of learning are frequently applied. In our discussions below, only learning and progress curves will be considered and the distinction will be individual person or group changes in performance with experience.

1.1 Some Applications of Learning Curves

Learning and progress curves serve numerous important purposes. One of the more obvious applications is costing out production operations. Since part of the cost of production is direct labor and person-hours of direct labor change with learning over the production run, any substantial changes in performance over a production run will decrease the amount of direct labor. A number of US governmental activities require learning curve pricing of services and products. Accordingly, labor costs decrease with greater learning so that longer production runs exhibit greater reduction in direct labor hours.

Another application of learning or progress curves is method evaluation. When there are alternative methods for performing various tasks, the best method may depend upon learning. Some methods may produce consistently better performance on one criterion but worse on another. In those siutations one has to make a trade-off between those criteria. More often one method has an advantage for shorter production runs and another method for greater production runs. In those cases one needs to find the break-even run size for those methods and establish a decision rule for method selection based on run sizes.

A third application of learning curves is in production control where specific designs of production operations are determined. Engineers of the production control function need to identify bottleneck operations and then solve the problems causing those bottlenecks. Alternatively, they may set up parallel workstations for part of the production operations. In order to identify bottlenecks and evaluate potential solutions, engineers need models of performance changes for each operation over the production run.

Quality control is a fourth application of learning and progress curves. That functional oganization is concerned with the causes of defective product units. Since some causes of defectives are due to human errors during production and the magnitude and number of those errors change with experience, learning and progress curves are needed by quality control engineers to predict these sources of defectives and to help reduce those and other causes of defectives.

1.2 Modeling Human Learning Behavior

Modeling human behavior is complex but interesting and modeling human learning behavior is more so. As a consequence you might ask, "Why then go to the effort?" The answer is simple, to obtain quantitative estimates of performance with more or less experience.

Typically, the first time a task is ever performed, both performance times and errors are the greatest. Namely, speeds and accuracies are quite low. When the task is performed repeatedly, people think of ways to improve and their behavior follows their thinking. Accordingly, learning models show decreasing performance times as a function of more task repetitions. That chance is precisely what one would expect for performance time *averages* but not necessarily actual data changes. People do not exhibit constant performance time decreases as most models predict. That is one of the typical simplifications.

All of the commonly used models vary as the rate of performance time decreases with experience through the specification of model parameters. However, some models show constantly decreasing performance time which appears to

eventually approach zero cycle times, a ridiculous situation. Other models show performance time decreasing to a preset level but that case requires the module to predetermine the amount of learning in addition to the rate of learning. However, those more realistic models still show constantly decreasing performance times down to the limit. On the other hand, anecdotes of learning phenomena frequently show various learning plateaus of performance times along a learning function but no known models provide this realism. As with other models, learning models are simplified descriptions of reality.

1.3 Why Use a Model of Learning?

With learning data, why does one need a model? There are a number of reasons:

- Learning data may be situation specific but models are more general and may be used in other situations that appear similar.
- It is difficult to extract prediction from data. It is much easier with a model that is fit to various products, especially within the same product family.
- One would expect slower learning rates for products with more complex manufacturing requirements (e.g. more metal to cut, higher tolerances to maintain, or small and complex assembly pieces). Models where the parameters vary with those features in expected ways provide face validity of the model that in turn helps in model acceptance.
- While models may simplify a situation, even to the extent of over-simplification, many people can use the simplified basis as a first-order description.

2. PERFORMANCE CRITERIA AND EXPERIENCE UNITS

Performance *time* is the most common criterion used in industry today. Production cycles are the most commonly used variable for denoting experience. If t_i is the performance time on the ith. cycle, a learning curve should predict t_i as a function of n cycles. Since learning implies improvement with experience, one would expect $t_i t_i + 1$ and $t_i - t_i + 1 \geq 0$ for the average case of i = 1, 2, . . ., n cycles.

An associated criterion to the time on the ith. cycle is the cumulative average performance time on the ith. cycle or A_i. Cumulative AVERAGE times consists of the sum of all performance times up to and including the nth. cycle divided by N or:

$$A_n = \frac{\sum_{i=1}^{n} t_i}{n} \tag{1}$$

Note that for n = 1, $A_1 = t_1$. With learning, t_i tends to decrease as i increases and so does A_n but at a slower rate than t_i decreases. This effect can be shown by the first-forward difference of A_n which is:

$$\Delta A_n = A_n + 1 - A_n$$

$$\Delta A_n = \frac{\sum_{i=1}^{n+1} t_i}{n+1} - \frac{\sum_{i=1}^{n} t_i}{n} = \frac{\sum_{i=1}^{n} t_i + t_{n+1}}{n+1} - \frac{\sum_{i=1}^{n} t_i}{n(n+1)}$$

$$= \frac{\sum_{i=1}^{n} t_i + n t_{n+1} - (n+1)\sum_{i=1}^{n} t_i}{n(n+1)} \tag{2}$$

$$\Delta A_N = \frac{\sum_{i=1}^{N+1} t_i}{N+1} - \frac{\sum_{i=1}^{N} t_i}{N} = \frac{\sum_{i=1}^{N} t_i + t_{N+1}}{N+1} - \frac{\sum_{i=1}^{N} t_i}{N}$$

$$\Delta A_N = \left[\frac{N}{N+1} - \frac{N+1}{N+1} \right] A_N + \frac{t_{N+1}}{N+1} = \frac{t_{N+1} - A_N}{N+1} \tag{3}$$

So long as t_{n+1} is less than A_n, ΔA_n is negative and the cumulative average time continues to decrease. Another criterion frequently of interest in ergonomics is *accuracy*. This criterion can be measured in the number of correctly performed sequential tasks. However, it is usually easier to measure the inverse of accuracy as errors in operations. In a single operation case, observations over a production sequence are a series of zeros and ones, corresponding to no error and an error. When more than a single operation is performed, say **M** operations, each cycle of production will consist of e_i equal to 0, meaning no error, or 1, 2, . . ., M errors for each cycle. A more understandable practice is to define e_i as the fraction of the possible errors where the observed number of errors is divided by the **M** possible errors at an operation. In this way e_i is 0, some proper fraction, or 1. It also follows that a learning curve could be fitted to the series of e_i values over the n observations of sequential units of production or to the cumulative average errors. If learning is present, then one would expect to see a general decrease in **ei** with increases in i = 1, 2, . . . , n and also the cumulative average errors would similarly decrease but with a rate lag compared to the serial errors.

3. SOME FUNCTIONAL LEARNING MODELS

Models for learning originated as elementary mathematical functions of serial data that fit the data trend. This restriction to elementary functions tends to make the model fitting easier. Also elementary functions tend to have fewer parameters, usually making fitting easier, and they tend to be more robust but there is no guarantee of such properties with simplicity.

3.1 Powerform Learning Models

A psychologist by the name of Snoddy (1926) originally reported the powerform function as fitting an individual's learning. (This is the earliest reference to learning models known to the author.) Snoddy identified the powerform function as a model of individual learning. This function is:

$$t_n = t_1 n^{-c} \tag{4}$$

where **c** is the *learning rate parameter*. Later Wright, working for an aircraft manufacturing company in Witchata, Kansas, noticed that the direct labor in building successive aircraft reduced by a constant fraction with each doubling of the number of aircraft produced. This characteristic is a property of the powerform model that he identified. Wright used this model as a progress function in the aircraft manufacturing industry. A property of the powerform model is that t_{2n}/t_n ratios are constant for all **n**. A numerical example of this series in the form of $5n^{-0.322}$ is:

n =	1	2	3	4	5	6	8	10	100	1000
tn =	5.0	4.0	3.5	3.2	3.0	2.8	2.6	2.4	1.13	0.541
t_n/t_{2n} =	4/5 = 0.8	3.2/4.0	2.8/3.5	2.6/3.2	2.4/3.0	–	–	–	–	–

It is clear from this example that t_n tends toward zero as n increases and so the asymptote of this model is 0. The aircraft

manufacturing industry assumes that asymptote is equal to t_{1000} by agreement. The example above also demonstrates that the ratio $t_{2n}/t_n = 0.8$ for all n. That result is caused by the parameter c and other values of c give different ratios as:

c =   0.074   0.153   0.234   0.322   0.415   0.515   0.621   0.737

tn/t2n = 0.95   0.90   0.85   0.80   0.75   0.70   0.65   0.60

However, the parameter c can be computed for any ratio using the equation:

$$C = - \frac{\log \left(\frac{t_{2n}}{t_n} \right)}{\log 2} \tag{5}$$

For completeness it should be stated that A_n be substituted for t_n when this model is used to describe the cumulative average time. Not that when there is a good fir of T_n to a powerform equation, An does not behave as a powerform series and conversely.

There have been numerous reports of model-based learning reported in the literature. (The few references cases given here come from Konz (1983) and they show Conway and Shultz (1959), Glover (1966), and Hancock (1967).) A few of these studies are summarized in Table 3 showing typical values of the powerform parameters. It is icommon practive in many companies to assume that a similar job has a similar learning curve.

A variant of the powerform model was devised at Stanford University and so it was known as the Stanford model. This model is:

$$t_n = k (d + n)^{-c} \tag{6}$$

where the only changes from the classical powerform is the parameter k replacing t_1 and the additional parameter d in the last factor. The parameter d was added to slow down the early serial values. Parameter k makes $t_1 = k (d + 1)^{-c}$. Note that the traditional powerform $d = 0$ and $k = t_1$. Table 4 describes the Stanford model for $d = 0, 0.1, 0.5$, and 1.0 in order to demonstrate how it differs from the traditional powerform using the same data as above. Table 4 also reports ratios of t_{2n}/t_n for those illustrations and the parameter values for parameter c and k. It is clear that the larger the value of d, the quicker those ratios decrease.

3.2   A Discrete Exponential Learning Curve

An alternative model to the powerform is an exponential model and a discrete serial form is:

$$t_n = t_1 \, \alpha^{n-1} \, [t_1 - t^*] + t^* \tag{7}$$

where α is the *learning rate* parameter and it is restricted to $0 < \alpha < 1$. Pegels (1969) first identified this model of learning and he noted its natural asymptote at t^* in contrast to the zero of the powerform model. Later Buck, Tanchoco, and Sweet (1976) showed that basis of this model is the constant-parameter forward-difference model or:

$$t_2 = \alpha \, t_1 + \beta \tag{8}$$

That is, the next value in the series is α times the current value plus β. It further follows that:

$$t_3 = \alpha \, t_2 + \beta = \alpha[\alpha \, t_1 + \beta] + \beta = \alpha^2 t_2 + \alpha\beta + \beta$$
$$t_4 = \alpha \, t_3 + \beta = \alpha[\alpha^2 t_1 + \beta] + \beta = \alpha^3 t_2 + \alpha^2\beta + \alpha\beta + \beta$$

Table 1. Some powerform learning examples in the literature

Improvement	Number of cycles	Task and authors
68%	1,000	Test and adjust – R. Conway and A. Schultz (1950) *Journal of Industrial Engineering*, 10, 39–53
73%	11,000	Electronic assembly – as above
79%	1,000	Service time – IBM electronics
80%		Man-hours per airframe – Wright in WWII
85%	2,000	Labor hours per gun barrel on a boring machine – F. Andress (1954) *Harvard Business Rev.*, 32, 87–97
86%	12,300	Cost/model-T – Ford (1910)
88%	–	Machine labor/product unit – W. Hirsch (1952) *Int. Journal of Production Research*, 8, 293–305
90%	–	Bench inspection – G. Nadler and W. Smith (1963) *Int. Journal of Production Research*, 2, 115–135

Table 2. Examples of the Stanford model for alternative values of parameter d

d	n	1	2	3	4	5	6	. . . 8
0	$t_n =$	5.0	4.0	3.5	3.2	3.0	2.8	2.6
c = 0.322	$t_{2n}/t_n =$	0.8	0.8	0.8	0.8	0.8	0.8	0.8
0.1	$t_n =$	5.0	4.0	3.497	3.175	2.945	2.769	2.504
c = .345	$t_{2n}/t_n =$	0.8	0.794	0.792	0.789	–	–	–
k = 5.167								
0.5	$t_n =$	5.0	4.0	3.453	3.094	2.834	2.635	2.344
c = 0.437	$t_{2n}/t_n =$	0.8	0.774	0.763	0.758	–	–	–
k = 5.969								
1.0	$tn =$	5.0	4.0	3.414	3.020	2.731	2.509	2.185
c = 0.550	$t_{2n}/t_n =$	0.8	0.755	0.735	0.724	–	–	–
k = 7.322								

More generally,

$$t_n = \alpha^n t_1 + \beta[\alpha^{n-1} + \alpha^{n-2} + \ldots + \alpha + 1]$$

$$t_N = \alpha^n t_1 + \beta \frac{1 - \alpha^n}{1 - \alpha} \qquad (9)$$

By defining

$$t^* = \frac{\beta}{1 - \alpha},$$

equation (9) is equivalent to equation (8). The next serial value in this exponential series after t_n is t_{n+1} and it is descrtibed as:

$$\Delta t_n = t_{n+1} - t_n = \alpha t_n + \beta - t_n = (\alpha - 1) t + \beta \qquad (10)$$

Equation 10 shows that the change with the next serial value is a constant $(\alpha - 1)$ times the last value plus the constant β and that corresponds perfectly with equation (9). It should be stated that when this exponential series is converted to logarithmic form, the function is S-shaped. (Both Jordan (1965) and Muth (1986) observed S-shaped logarithmic functions in practice.)

It should be stated that when $\alpha = 1$, this series becomes linear with the rate of change equal to β.

An advantage of this exponential learning curve is that cumulative sums T_n, consisting of $t_1 + t_2 + \ldots + t_{n-1} + t_n$, can be easily computed for finding cumulative averages. The basic formula for this is:

$$T_n = \sum_{n-1}^{n} t_1 = nt^* + (t_1 - t^*) \frac{1 - \alpha^N}{1 - \alpha} \qquad (11)$$

A series starting with 5 times units as t_1, with $\alpha = 0.5$, $\beta = 1,5$ yields a cumulative sum of 21.9325 the same as: 5 + 4 + 3.5 + 3.25 + 3.125 + 3.0625 = 21.9375. Also not in equation (11) that the first term on the right-hand side is the steady-state time values for thee six units after all learning is complete and the second term in that equation is the transitory time or the learning effect. Since the steady state value is $t^* = 3.0$, the transitory values are t1 - 3 = 2 + 1 + 0.5 + 0.25 + 0.125 + 0.0625 = 3.9375 which is 17.52% of the total 21.9325 and the steady-state is a little over 82%. However, if this same series increases to n = 25, the cumulative serial sum is almost 79.0. The steady-state consists of 25 3 3 = 75, which is 95% of the 79.0. Hence, the learning part amounts to only 5% of this fast-acting series. A slower-acting series is described as an identical sseries with a = 0.9. The steady state component of this slower series consists of 66% of six serial values and 80% of 25, hence there is considerably more learning effect.

A variation of this discrete exponential learning curve model is the continuous version, which was developed by Bevis, Finnicat, and Towill (1970) and extended by Hutching and Towill (1975) who applied it to performance accuracy. This is

$$t_n = t_1 + [t_1 - t^*]e^{-bt} \qquad (12)$$

where the exponent t is continuous running time while the tasks is being performed and b is the rate parameter. In one version of these authors proposed the paramter g as the time duration until learning is half completed or the time savings are $(t_1 - t^*)/2$ and parameter b = 1/g, which is similar to a half-life value.

4. FITTING LEARNING CURVES TO EMPIRICAL DATA

Besides the fact that different functional forms of learning curves yield different serial shapes and possess different properties, those different functional forms are also fit differently to data. Part of the reason is that those functional forms allow different mathematical operations and provide different mathematical conveniences because of their properties.

4.1 Fitting the Powerform Model

Fitting this model to data is simplified by the fact that this function is *linear* when converted to the logarithmic form. That is, converting equation (4) to logarithmic form results in:

$$\log tN = \log t1 - C \log N \qquad (13)$$

When both N and t_N are changed to their logarithmic values of any logarithmic base, the plots are *linear* functions of log N.

The data from the previous powerform example are changed to logarithmic values and plotted in Figure 1 where a straight-line plot is evident. A simple linear regression can be made on these logarithmically transformed data where the negative slope of the regression provides a direct estimate of parameter **c**. Since this plot was made using data that were generated by a powerform equation, the fit is perfect, including that of t_1.

In regard to de Jong's variation of the powerform model, an identical data fitting procedure can be used if the incompressibility parameter **M** can be estimated first. Once **M** is estimated, subtract the product of **M** and t_1 from every value of t_i and continue as before. That operation will give an estimate of the compressible portion of de Jong's model; one only needs to add back the constant product of **M** and t_1. However, it is not altogether clear how one could estimate the fraction **M** except by recognizing ultimate performance of similar tasks, or perhaps using laboratory studies of a similar task.
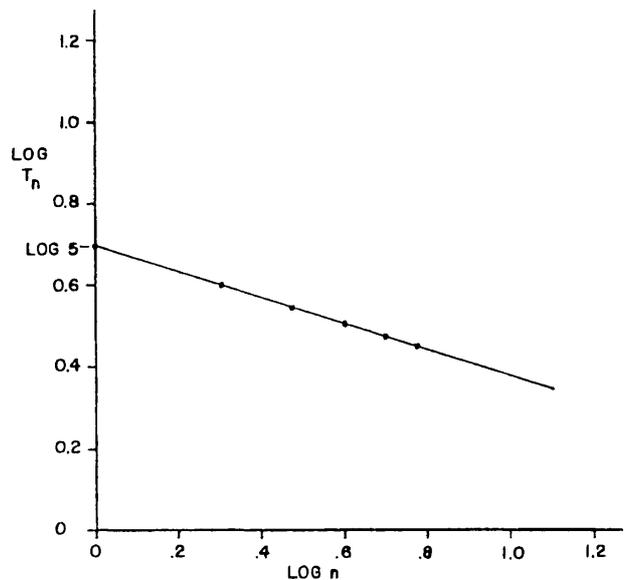
Figure 1. A log-log graph of the powerform example with log T_n as a function of log N as used for fitting this learning curve

4.2 Fitting the Discrete Exponential Model

This action required empirical data and a different approach. It was noted during the discussion of this model that the first-forward differences Δt_N is a *linear* function of the performance magnitude t_n as shown in equation (10). Accordingly, the best linear fit provides estimates of parameters α and β. A linear regression of Δt_N, as a function of t_N, has the slope of $(\alpha - 1)$. The ordinate intercept of that regression is β and the abscissa intercept is t^* (the asymptote). To illustrate, consider the previous numeral example above where:

n =	1	2	3	4	5	6
t_n =	5.0	4.0	3.5	3.25	3.125	3.0625
Δt_n =	-1.0	-0.5	-0.25	-0.125	-0.0625	—

A regression of Δt_N as a function of t_N yields a slope of -0.5, an ordinate intercept of 1.5, and an abscissa intercept of 3. When the slope value is equated to $(\alpha - 1)$ and solved, then it is seen that $\alpha = 0.5$. Figure 2 shows a plot of these regression data.

The above example of fitting the discrete exponential model to data resulted in finding parameters α and β. Parameter t^* is a direct result of those parameter values. With the values of these parameters substituted in equation (7), only the value of t_1 is needed and every value of the series can be predicted. One could use the first data point for t_1, but if there is any error in that value, the entire series of predictions is wrong. A better procedure is to expand the above procedure with the least-squared-error estimate of **t1** based on many data points, not just the first one. It is easier to compute first the mythical value of t'_0 as:

$$t'_0 = \frac{\alpha^2 - 1}{\alpha^{2n} - 1} \sum_{i=0}^{n-1} \alpha^i t_i + \frac{(\alpha^{n-1} - 1)(\alpha^n - 1)}{(\alpha^{2n} - 1)} \qquad (14)$$

and then to convert back to t'_1 use the basic relationship:

$$\mathbf{t'_1} = \alpha\, t'_0 + \beta \qquad (15)$$

Equation (14) provides a least square error estimate of fit so that the whole known series contributes to the initial value for estimation. It there is error added to each of the points over the learning sequence, by finding an estimate of t_0 from each of several points, the average value of can go into equation (14) to help stabilize the resulting t'_1 in equation (15).
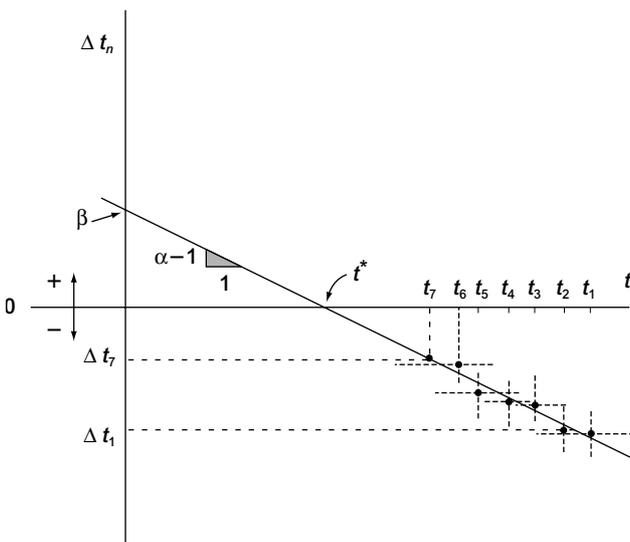
4.3 A Numerical Example of Learning

Consider the numerical example immediately above where the series went from 5 to 4 to 3.5, and so forth. Suppose that all we had were these first three series values. Thus, N = 3 for estimation purposes. The mythical t_0 value is 7 as $0.5(7) + 1.5 = 5 = t_1$. When is $\alpha = 0.5$ and is $\beta = 1.5$, a series starting at 7 goes exactly through those points 5, 4, and 3.5. Note the calculations below for n = 2, 3, and 4, respectively, using equation (14) and the actual data. The first equation below considers only the mythical series value and $t_1 = 5$. The second equation includes $t_2 = 4$ as well. Both equations yield the same result.

$$t'_0 = \frac{.5^2 - 1}{.5^4 - 1}\, [7 + .5(5)] + 3\, \frac{.5(.5^{2-1} - 1)(.5^2 - 1)}{(.5^4 - 1)} = 7.6 - 0.6 = 7.0$$

$$t'_0 = \frac{.5^2 - 1}{.5^6 - 1}\, [7 + .5(5) + .25(4)] + 3\, \frac{.5(.5^{3-1} - 1)(.5^3 - 1)}{(.5^6 - 1)} = 8,0 - 1.0 = 7$$

$$t'_1 = 0.5(7.0) + 1.5 = 5.0$$

If there was noise present and we would expect to estimate the same α and β, but the noise could change the whole series by changing the first few serial values.

Figure 3 shows some interesting points about fitting this discrete exponential model to empirical data. When α is very close to unity, the slope of the regression is nearly flat. If a large amount of learning takes place between the first and second observation and the slope is approaching flatness (i.e. near zero slope), the two intercepts can become negative, leading to very unreasonable implications for a learning curve. Examples of this were reported by Buck and Cheng (1992) who performed a series of experiments. For this reason, one needs to constantly check reasonableness of models. Another problem with the discrete exponential model occurs when people try to over-compensate

Figure 2. Fitting a discrete exponential model to data
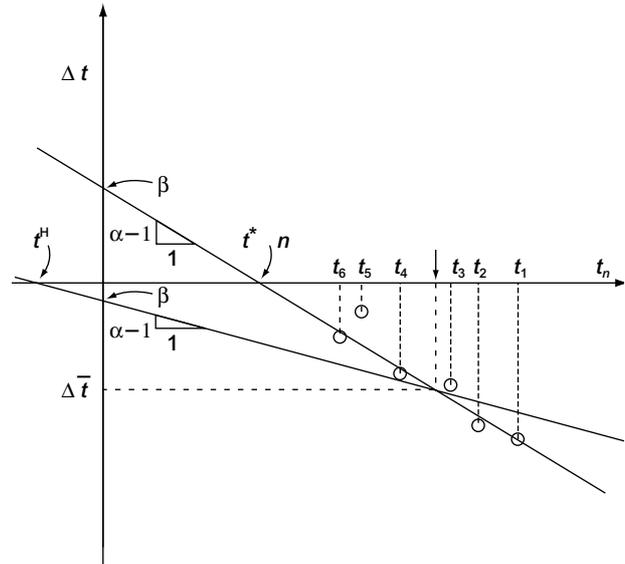
Figure 3. Possible errors in the fitting of the discrete exponential model to data

for immediate feedback. In that case, performance time may alternate between successive cycles. With strong data alternations, α can become negative to reflective this form of behavior. The discrete exponential learning curve is more pliable than the powerform model so that the exponential will reflect those oscillations when they dominate the learning effect.

Learch and Buck (1975) demonstrated this alternation phenomena in an elementary learning situation.

5. COMPARING ALTERNATIVES FOR LEARNABILITY

One of the important features of many ergonomic designs is the capability of a user being able to use the design proficiently at the earliest possible time. As a consequence, new designs can be tested in usability. Fitting a learning curve to the performance changes of naive users provides the basis for assessing learnability. In terms of the powerform model, the learning rate is specified by the parameter **c**. The discrete exponential model describes learning in two ways; parameter a which shows the learning rate and the difference between t_1 and t^* which describes the amount of learning expected. Unfortunately, the expected amount of learning is not available directly from the powerform model but a similar estimate can be obtained by computing t_{1000} and subtracting it from t_1. A design with a greater learning rate and a small amount of learning but with a low t^* is a preferred design on the basis of learnability.

6. COGNITIVE CONTENT AND THE LEARNING RATE

Some recent studies from Arzi and Shatub (1996) and Dar-El and Vollichman (1996) have shown two interesting results. The first is that tasks with higher cognitive content tend to have lower learning rates. An immediate implication is that quicker learning (or lower performance times) occur sooner when the cognitive content is stripped from the task. This implication speaks immediately to the economics of learning which follows. The second finding of importance is that after forgetting and a transitional restarting people return quickly to the former learning curve which they terminated at the previous stopping of practice. Hence, forgetting is, to a certain extent, unlearning. This observation describes the effects of repeated practice.

7. ECONOMICS OF LEARNING

When there is improvement in performance times due to learning, there is a shorter production duration to produce **n** items. Hence, the direct labor cost is smaller with greater learning. In fact, as an individual completes more of the production, the time per unit decreases. At the same time, those future economic advantages are discounted further in time and so some of the effects are less. However, the learning curve directly states the changes in labor time as the person improves and the Department

of Defense often requires learning curve adjustments on manufacturing contracts.

8. FINAL REMARKS

Learning is the result of practice and training. The focus here is on the description of the effect using well-known functions that decrease with time or experience. Those models are classical in describing performance changes with experience but none are ordained. That is, there is no true learning curve, because there is no fully accepted theory on behavioral learning. Hence, the best learning curve is the one which best fits available data and one really needs to describe learning with available data for fitting the models. The models formalize the description of the effect and they allow one to project what future effects are likely to be.

REFERENCES

ARZI, Y. and SHATUB, A., 1996, Learning and forgetting in high cognitive and low cognitive tasks, Proceedings of the First International Conference (ICAE '96), Istanbul, Turkey, A. F. Ozok and G. Salvendy (eds), *Advance in Applied Ergonomics*: West Lafayette, IN: US Publishing, pp. 672–75.

BUCK, J. R. and CHENG, S.W.J., 1993, Instructions and feedback effects on speed and accuracy with different learning curve functions, *IIE Transactions*, 25(6), 34–47.

BUCK, J. R., TANCHOCO, J. M. A. and SWEET, A. L., 1976, Parameter estimation methods for discrete exponential learning curves, *AIIE Transactions*, 8(2), 184–94.

CONWAY, R. and SCHULZ, A., 1959, The manufacturing process function, *Journal of Industrial Engiengring*, 10, 39–53.

DAR-EL, E. and VOLLICHMAN, R., 1996, Speed vs Accuracy under continuous and intermittent learning, Proceedings of the First International Conference (ICAE '96), Istanbul, Turkey, A. F. Ozok and G. Salvendy (eds), *Advance in Applied Ergonomics*: West Lafayette, IN: US Publishing, pp. 689–94.

DE JONG, J.R., 1957, The effect of increasing skill on cycle time and its consequences for time standards, *Ergonomics*, 1, 51–60.

GLOVER, J., 1966, Manufacturing process functions II: selection of trainees and control of their progress, *International Journal of Production Research*, 8, 293–305.

HUTCHING, B. and TOWILL, D.R., 1975, An error analysis of the time constraint learning curve model, *International Journal of Production Research*, 13(2), 105–35.

JORDAN, R., 1965, How to use the learning curve, *Materials Management Institute*, Boston, MA.

PEGELS, C.C., 1969, On startup of learning curves: an expanded view, *AIIE Transactions*, 1(3), 216–22.

SALE, D.R., 1978, The effect of alternate periods of learning and forgetting on economic manufacturing quantity, *AIIE Transactions*, 10(3) September.

SNODDY, G.S., 1926, Learning and Stability, *Journal of Applied Psychology*, 10, 1–36.

TOWILL, D.R. and KALOO, U., 1978, Productivity drift in extended learning curves, *OMEGA*, 6(4), 295–304.

WRIGHT, T.P., 1936, Factors affecting the cost of airplanes, J*ournal of Aeronautical Science*, 3, 122–8.

Lifting Techniques

R. Burgess-Limerick

Department of Human Movement Studies, University of Queensland, Brisbane 4072 Australia

1. INTRODUCTION

Compensation claims for occupational injuries attributed to lifting have typically represented a significant proportion of claims (~20%), and the majority of these (typically ~70%) are for injuries to the back. This paper discusses the postures and patterns of movement used to lift low-lying objects and the implications of these postures and movements for back injuries in particular.

Current knowledge about the mechanisms of back injuries and the biomechanical consequences of different lifting techniques are described; and attention is directed toward the postures and patterns of movement normally employed to lift low-lying objects. The paper concludes by presenting recommendations for preventing back injuries caused by lifting.

Starting from upright standing, lifting a low-lying object placed in front of the person involves a cycle of movement beginning with flexion of the knees, hips and lumbar vertebral spine (and dorsi flexion of the ankles). The load is then grasped and raised against gravity by extension of the knee, hips and lumbar vertebral spine (and plantar flexion of the ankles).

Lifting technique has typically been defined in terms of the posture adopted just before the load is lifted (at the start of the extension phase). It has commonly been proposed that the postures adopted to lift loads from a low level may be characterized in terms of two extremes: one, described as a stooped posture, is where the knee joints are almost fully extended and the hip joints and vertebral column are flexed to reach the load; the second, described as a full squat, is where the knee joints are fully flexed and the trunk is held as vertical as possible. It has become a matter of dogma that the latter posture is the "correct" manner of lifting. However, on the basis of current knowledge of injury mechanisms and lifting biomechanics, it is clear that lifting from a full squat posture is not an appropriate recommendation. It is also evident that a self-selected lifting technique typically involves adopting a posture at the start of the extension phase which is intermediate between full squat and stooped postures, and that the pattern of movement normally employed reduces muscular effort.

2. MECHANISMS OF INJURY DURING LIFTING

During lifting, large extensor moments about the joints of the lumbar vertebral column are produced by the paravertebral musculature to overcome the flexor moment caused by the weight of the upper body and load. Injury to musculo-ligamentous structures commonly occurs as a direct consequence of the high forces involved.

These high forces also result in large compressive and shear forces acting between each pair of vertebra. Unless the lumbar spine is in a posture of extreme flexion, the mechanism of failure in compression is failure of the endplates of the vertebral bodies and the underlying trabeculae as the nucleus pulposus bulges upward and downward (Adams and Dolan 1995). The magnitude of compressive forces experienced during a single lift is unlikely to cause endplate failure, and an injury of this type caused by lifting is more likely to be cumulative.

Cumulative damage to the vertebral endplates may occur in a number of ways. Microdamage to vertebral endplates is likely during heavy lifting, and injury may arise if the microdamage accumulates more rapidly than can be repaired. Repeated compressive loading will also reduce the failure tolerance of the tissues, resulting in injury if repeated loading continues (McGill 1997). Damage may also be additive in that prolonged exposure to other sources of loading, and especially whole-body vibration, may render the vertebral bodies vulnerable to injury during lifting.

Lifting from postures involving extreme lumbar vertebral flexion has the potential to contribute to injury. Extreme lumbar vertebral flexion is characterized by absence of electromyographical activity in erector spinae (e.g. McGill and Kippers 1994). In this situation the anterior moment caused by the weight of the upper body and the load is balanced by an extensor moment created by tension in the paravertebral ligaments, interspinous ligaments, posterior fibers of annulus fibrosus and passive elements of the musculotendinous tissues. The first tissues to be injured in this situation are the interspinous ligaments (Adams and Dolan 1995). Disruption of the posterior fibers of annulus may follow if extreme lumbar flexion is combined with compression and lateral bending or torsion.

If damage to the posterior annulus progresses, seepage of the nucleus pulposus through the annulus may result (an intervertebral disc prolapse). While intervertebral disc prolapse only accounts for small proportion of claims for back injuries (5–10%), the injury frequently results in chronic back pain and accounts for a considerably larger proportion of claims costs.

Compressive load alone will not cause intervertebral disc prolapse, and damage to the intervertebral disc is unlikely to occur as a consequence of one-time loading (although this is possible if high compressive load is placed on the spine while hyperflexed and laterally bent) (Adams and Dolan 1995). Injury to the intervertebral disc is more likely the consequence of an accumulation of microdamage due to repeated compressive and torsional loading applied while the lumbar spine is extremely flexed.

Anterior shear forces are also very high when loads are lifted from a posture of extreme lumbar flexion and this represents a risk of injury. However, the orientation of the fibers of the erector spinae muscles (in particular, the pars lumborum fibers of longissimus thoracis and iliocostalis lumborum) is such that when tension develops in these muscles a posterior shear force is created on the superior vertebrae that counteracts the anterior shear created by the weight of the upper body and load (McGill 1997). The erector spinae are active unless lumbar flexion is extreme, and consequently the anterior shear forces are reduced in postures that do not involve extreme lumbar flexion.

Prolonged exposure to static postures involving extreme lumbar vertebral flexion will also cause the tissues to creep (the ligaments do not return to their resting length immediately upon unloading). The consequence may be a temporary loss of stability after the period of sustained extreme lumbar flexion which may lead to a higher likelihood of injury in subsequent loading in any posture (McGill 1997). The abdominal muscles normally contribute to stability of the spine, and failure to contact these muscles appropriately may also increase the risk of injury.

3. THE "FULL SQUAT" RECOMMENDATION

According to Brown (1973), the recommendation that lifting should be carried out from a "full squat" posture has been promulgated since the 1930s. Many researchers have subsequently noted that this recommendation is unjustified (e.g. Whitney 1958, Brown 1973, NIOSH 1981). The authors of the otherwise influential NIOSH Work Practices Guide for Manual Lifting (1981) observed that the full squat posture reduced stability (the heels are inevitably lifted from the ground and the knees are in an unstable "loose packed" posture when maximally flexed) leading to the possibility of injury due to unexpected perturbations; and that the technique increased the distance of wide loads from the spine (increasing the load moment and consequently the resulting extensor moment and compressive forces). It was concluded that the squat lift recommendation was based on simplistic mechanical logic that failed to take dynamic loading on the back and the knees into account.

The proponents of the full squat technique suggested that the stresses on the vertebrae are better distributed with the lumbar spine in a lordotic posture. In fact, lordosis poses several disadvantages relative to postures of partial flexion (Adams and Dolan 1995), including increased loading of apophyseal joints and increased compression of the posterior annulus. Lordosis has been advocated to reduce hydrostatic pressure in the nucleus, but this only indicates that the load has been shifted to the annulus and apophyseal joints.

An additional pragmatic problem with the full squat recommendation is that it cannot be utilized in many situations. Maximal knee flexion has the consequence of lengthening the quadriceps beyond their optimal length leading to decreased knee extensor strength. The result is that lifting capacity is reduced: submaximal loads require greater muscular effort leading to more rapid onset of muscular fatigue; maximal loads cannot be lifted at all.

From the discussion of injury mechanisms it is evident that the only appropriate recommendation regarding posture of the lumbar spine is to avoid extreme lumbar vertebral flexion, and trunk rotation and lateral flexion. There is no basis for avoiding postures involving moderate lumbar vertebral flexion.

4. SELF–SELECTED LIFTING TECHNIQUE

The traditionally recommended full squat posture is seldom, if ever, spontaneously adopted in the absence of specific instruction. Investigations of self-selected lifting technique have revealed that the postures typically adopted to lift low-lying loads are intermediate between full squat and stoop extremes, and might be termed semi-squat (e.g. Burgess-Limerick et al. 1995, Burgess-Limerick and Abernethy 1997). Lifting a low-lying load from a semi-squat posture typically involves ~45° of lumbar vertebral flexion, that is, ~75% of the normal range of movement. In conjunction with the absence of an electromyographical silent period in erector spinae, this suggests that the passive structures of the back are not substantially stretched during lifting from this posture. Some people in some circumstances adopt stooped postures involving greater lumbar flexion, although typically this occurs when the load is relatively light.

An adequate description of lifting technique requires consideration of the pattern of interjoint coordination as well as the posture adopted at the start of the lift. The posture adopted at the start of extension influences the pattern of subsequent interjoint coordination by determining the range of movement available at each joint. The semi-squat posture most commonly adopted at the start of extension allows a pattern of interjoint coordination that appears functional.

The coordination of self-selected lifting involves contemporaneous movement of the lower limb and trunk joints, that is, the joints flex and extend at the same time rather than sequentially (as is sometimes modeled). However, the joints are not perfectly synchronized: a consistent pattern of deviation from synchronous coordination is commonly observed. Knee extension typically occurs more rapidly earlier in the lifting movement relative to extension of the hip, and the onset of rapid lumbar vertebral extension is delayed substantially after the start of the lift. The moderate lumbar flexion observed lengthens the erector spinae relative to its length in normal standing, and the delay before rapid lumbar vertebral extension delays rapid shortening of the erector spinae. Estimation of the length changes of the biarticular hamstring muscles has revealed that they are also relatively lengthened at the start of the extension phase, and that the pattern of coordination between knee and hip joints also has the consequence of delaying rapid shortening of the hamstrings.

Muscles are stronger when lengthened, and when not shortening rapidly, and thus this pattern of coordination increases the strength of the hamstrings and erector spinae early in the extension phase when the acceleration of the load is greatest by both lengthening the muscles, and delaying their rapid shortening. Delaying shortening of the hamstrings has the additional functional consequence of allowing the monoarticular knee extensors, paradoxically, to contribute to hip extension through a tendinous action of the hamstrings. The pattern of coordination observed thus reduces the muscular effort required to perform the task, and the pattern of interjoint coordination is exaggerated with increased load mass.

A different pattern of coordination between hip and knee occurs when a stooped posture is adopted at the start of extension. The large range of hip flexion and small range of knee flexion involved results in the hamstrings being lengthened further than if a semi-squat posture were adopted. A stooped posture has the advantage of lowering the center of gravity of the upper body less than a semi-squat posture and thus less work is done in lifting the upper body during each lift. However, during lifting from a stooped posture the hamstrings must immediately shorten rapidly because the knee is unable to extend rapidly. This counteracts to some extent any strength advantage that might accrue as a consequence of the increased hamstring length and prevents the monoarticular quadriceps from contributing to hip extension.

5. RECOMMENDATIONS FOR PREVENTING BACK INJURIES DUE TO LIFTING

Training people to perform lifting in safer ways has been consistently proposed as a means of reducing the risk of injury; however, research evaluating the effectiveness of lifting training programs involving the uninjured worker has generally failed to find any evidence of persistent modification in lifting technique (Pheasant 1986). If it can be assumed that muscular fatigue contributes to injuries suffered as a consequence of lifting, then a technique that reduces muscular effort may be preferred. Rather than prescribing a single "best" technique that is not likely to be

appropriate in all situations, it may be preferable to provide education in general lifting guidelines and to use exploratory learning techniques (Newell 1991) to assist lifters to discover individually appropriate postures and patterns of movement.

General lifting guidelines that can be justified on the basis of current knowledge include: wherever possible, remove exposure to manual lifting by providing mechanical aids; and if manual lifting must be undertaken, one should:

- Reduce the load mass.
- Raise the initial height of the load.
- Keep the load close.
- Adopt a posture at the start of the lift that involves a moderate range of motion at the knee, hip and vertebral column.
- Avoid lifting from a posture of extreme lumbar vertebral flexion.
- Avoid trunk rotation while lifting.
- Avoid lateral trunk flexion while lifting.
- Avoid lifting after prolonged periods of extreme lumbar vertebral flexion.
- Avoid high acceleration of the load (lift smoothly).

The risk of injury to the back caused by lifting can also be reduced by reducing exposure to whole-body vibration and by strengthening the bones, ligaments and muscles by appropriate exercise.

ACKNOWLEDGEMENT

REFERENCES

ADAMS, M. A. and DOLAN, P., 1995, Recent advances in lumbar spinal mechanics and their clinical significance. *Clinical Biomechanics*, 10, 3–19.

BROWN, J.R., 1973, Lifting as an industrial hazard. *American Industrial Hygiene Association Journal*, 34, 292–7.

BURGESS-LIMERICK, R. and ABERNETHY, B., 1997, Toward a quantitative definition of manual lifting postures. *Human Factors,* 39, 141–8.

BURGESS-LIMERICK, R., ABERNETHY, B., NEAL, R. J. and KIPPERS, V., 1995, Self-selected manual lifting technique: functional consequences of the interjoint coordination. *Human Factors*, 37, 395–411.

McGILL, S. M., 1997, The biomechanics of low back injury: implications on current practice in industry and the clinic. *Journal of Biomechanics*, 30, 465–75.

McGILL, S.M. and KIPPERS, V., 1994, Transfer of loads between lumbar tissues during the flexion-relaxation phenomenon. *Spine*, 19, 2190–6.

NATIONAL INSTITUTE FOR OCCUPATIONAL SAFETY AND HEALTH, 1981, *Work Practices Guide for Manual Lifting. No. 81–122* (Cincinnati: US Public Health Service).

NEWELL, K.M., 1991, Motor skill acquisition. *Annual Reviews of Psychology*, 42, 213–37.

PHEASANT, S. T., 1986, B*odyspace: anthropometry, ergonomics and design* (London: Taylor & Francis).

WHITNEY, R. J., 1958, The strength of the lifting action in man. *Ergonomics*, 1, 101–28.

Loads on the Lumbar Spine during Dynamic Work

W.S. Marras

Biodynamics Laboratory, Institute for Ergonomics, The Ohio State University, 1971 Neil Avenue, Columbus, OH 43210, USA

1. WHY ASSESS LOADS ON THE SPINE?

Occupationally related low back disorders (LBD) are typically the leading cause of lost work days as well as the most costly occupational health problem facing industry today. Statistics gathered from industry have made it clear that occupationally related LBD risk is associated with manual materials handling (MMH) tasks such as lifting lowering, pushing, pulling, carrying and performing tasks in non-neutral postures. Until recently tools were available that could biomechanically analyze jobs for workers performing tasks in static postures. This greatly reduced the applicability of the models to realistic work tasks observed in industrial settings. This chapter will show how biomechanical analyses can be performed to assess spine loading during realistic, dynamic task activity of the spine.

Industrial studies have identified relationships between the load imposed upon the spine and the risk of low back disorders. The more load imposed upon the spine, the greater the risk of suffering a LBD. Thus, it is important to be able to assess the magnitude of the load imposed upon the spine if we are to develop an appreciation for the risk imposed upon a worker though the performance of a particular job task. Hence, estimates of spinal loads provide a benchmark or measure of the percentage of workers who will be at risk of suffering a LBD due to the job.

According to biomechanical logic, loads imposed upon the spine are compared to the tolerances of spinal structures to assess the risk of injury in a certain job. Figure 1a shows the magnitude of a load imposed upon the spine during several cycles of work. As the work task is repeated, the load imposed upon the spine increases and decreases in a cyclical pattern. This figure also shows the relationship of this loading to the tolerance level of the spine. As long as the magnitude of the load does not exceed the tolerance of the spine, one would not expect an injury to occur due to the work. Realistically, the tolerance of the spinal structures varies between individuals. This tolerance level can be viewed as a *distribution* if tolerances among the populations of workers. In addition, the loading imposed upon the spine can vary from cycle to cycle even if the same task is repeated. Finally, both the loading imposed upon the spine as well as the tolerance can vary over time. Figure 1b shows how biomechanical logic can account for cumulative trauma through a gradual decrease in the spine tolerance though out the work period. Thus, by considering the variability in loading and comparing this to the variability in spine tolerance over time, one could estimate the percentage of workers that would be expected to be affected by a job design characteristics.

Loading of the spine can occur in three dimensions. Figure 2 shows the three types of forces or loading that can occur on the
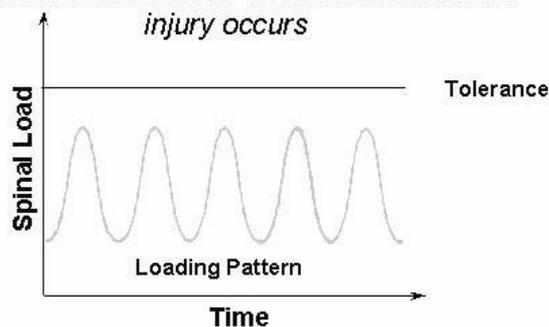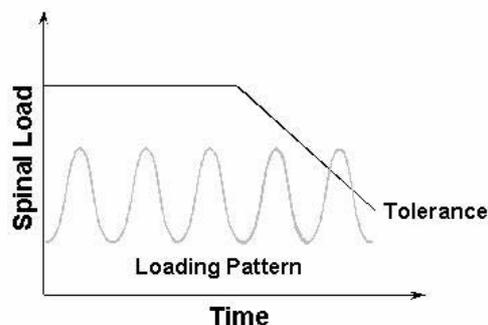
Figure 1. (a) Load–tolerance relationship assumption used for biomechanical assessments; (b) adjustment of load–tolerance relationship for cumulative trauma.
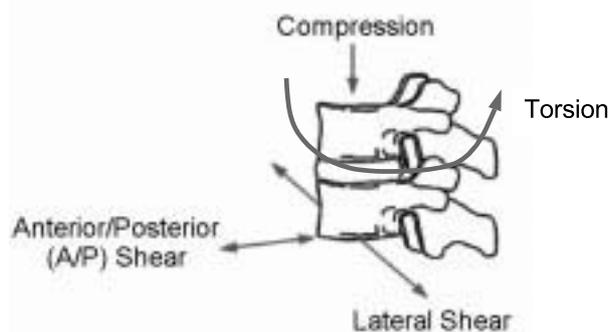
Figure 2. Forces that can affect spine loading.

spine. Loading can occur in compression, shear or torsion. In actual MMH situations, the spine is loaded simultaneously in multiple dimensions.

2. HOW DOES SPINE LOADING AFFECT SPINE DEGENERATION?

Figure 3 shows the sequence of events that occurs during work related degeneration of the spine. This sequence represents one

of the major pathways believed to occur for low back disorder. As indicated in the figure, excessive loading, generated from both within and outside the body (internal and external forces), cause microfracturing of the vertebral end plates. These end plates serve as a transport system for nutrient delivery to the disc fibers. If this loading becomes excessive and exceeds the end plate tolerance, a microfracture occurs. This microfracture is typically painless since few pain receptors reside within the disc. As healing occurs, scar tissue develops over the microfracture. Since scar tissue is thicker and denser than normal tissue, this scar tissue interferes with nutrient delivery to the disc fibers. This loss of nutrient results in atrophy to the disc fibers and weakens the disc structure. This process represents the beginning of cumulative trauma to the spine and can result in disc protrusions, disc herniation, and instability of the spinal system. It has been commonly accepted that compressive loads on the end plate of 3400 N represent the level at which vertebral end plate microfractures begin to occur. Loadings of 6400 N are expected to affect 50% of people < 40 years (NIOSH 1981).

Tolerances have also been estimated for shear loading of the spine. These are expected to occur between 750 and 100 N.

3. HOW ARE SPINAL LOADS ASSESSED?

3.1. Internal and External Loading

To evaluate the load imposed upon the spine, one must consider the mechanism by which forces are transferred to the spine. Two types of forces typically are typically imposed on the spine. External forces represent those forces due to the effect of gravity acting on the object being moved as well as the worker's body. As shown in Figure 4, external loads represent the mass of the object lifted. They can also represent the forces generated by the force of gravity acting on the workers arms and torso. The second type of force imposed upon the body are internal forces. Internal forces are those forces imposed on the spine due to the reactions of the body to the external forces. Forces generated by muscles as well as passive forces in the connective tissue represent internal forces. However, as suggested in Figure 4, the magnitude of the internal forces typically are much larger than the external forces since they must operate at a mechanical disadvantage.

Thus, the key to estimating accurately internal loads is to accurately account for the internal loads that are needed to support the external loads. The sum of the internal and external forces occurring in three-dimensional space define spinal loading. However, a major limitation in controlling the incidence of occupationally related LBD has been the inability *accurately* to assess the loading due to the internal forces acting within the torso and imposing loads on the lumbar spine. For spine loading estimates to be accurate, they must be able to assess internal forces in the torso under realistic work conditions that often involve whole-body free-dynamic MMH conditions.

3.2. Single Equivalent Muscles

Historically, researchers have attempted to estimate the activity of the internal forces generated by the trunk muscle through several methods. Early approaches assumed that a single equivalent muscle force in the back could represent the internal muscle forces. Stick figures with a single equivalent extensor muscle were used to asses the contribution of the trunk muscles to spine loading. However, latter research showed that such models could not accurately represent risk and that assuming a single equivalent muscle was a gross simplification of the biomechanical system. These approaches also were unable to account for how muscle behavior responded to realistic trunk motion.

3.3. Multiple Muscle Systems

These early single equivalent muscle attempts made it clear that a multiple muscle system representation was need to describe the activity of the trunk's musculoskeletal system. However, this type of representation further complicated the issue of resolving muscle forces within the torso. Over a dozen muscles can support external forces imposed on the trunk during a MMH task. But only three external forces and three external moments can be monitored outside the body (external forces). Thus, this results in a statically indeterminate situation since there are far more unknowns (internal muscle forces) than knowns (external forces)
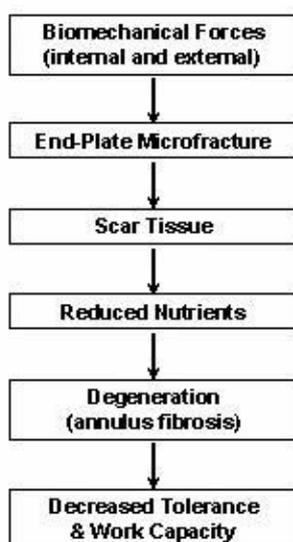
Figure 3. Sequence of events occurring on the spine during degenerative process.
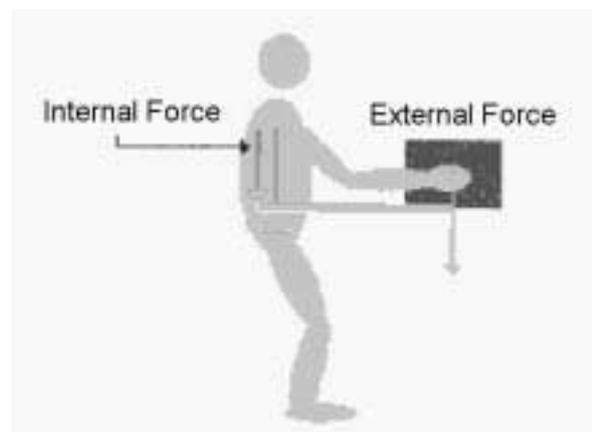
Figure 4. Internal and external forces acting on the spine.

and it becomes impossible to determine which muscle support the external loads.

Several approaches have been employed to estimate the contribution of the loads in the multiple muscle system. First, assumptions were made regarding which muscles would be active during a task and which ones would be silent. This reduced the size of the problem and permitted one to solve for the internal muscle forces and, therefore, estimate spinal loads. Unfortunately, laboratory monitoring of the muscle activities during simulated MMH rarely indicated that these assumptions were realistic. It was often the case that more muscles were active than assumed.

Second, optimization and neural network algorithms were used to assess the contribution of the internal muscle forces. In the case of optimization, no objective functions could be identified that resulted in a realistic muscle activity descriptions. Optimization often worked under static prolonged loading conditions but failed to predict the coactive nature of the trunk musculature. Neural networks used historical records of muscle activities to predict muscle how muscle would behave during lifting tasks. These networks were often used to classify muscle usage patterns but were not able to describe the range of responses between workers. In addition, they were unable to adapt to new lifting situations. Hence, these predictions tool failed to realistically estimate the response of the trunk muscles to dynamic real-world lifting conditions.

Finally, biologically assisted models were used to estimate the forces generated within the trunk muscle during a lift or MMH activity. Instead of attempting to predict which muscles were active or inactive in response to an external loading condition, biologically assisted models monitor the biological output from many muscles to directly assess which muscles are active in response to an external load. Electromyography (EMG) is often used to monitor the muscles under these circumstances (McGill and Norman 1986). EMG is a measure of the electrical depolarization of a muscle during work. The EMG signal can be calibrated to represent the force generated by a given muscle during a MMH condition.

4. OSU EMG–ASSISTED FREE–DYNAMIC BIOMECHANICAL MODEL

Over the past 18 years such an EMG-assisted biomechanical model has been under development in the Biodynamics Laboratory at the Ohio State University (Marras and Granata 1997). This model monitors the electrical activity of the 10 major trunk muscles used to support external loads and sums these forces in three dimensions to predict compression, shear and torsion loading on the spine. The model is responsive to dynamic trunk motion and can monitor complex MMH motions of the trunk.

The logic behind this model is represented graphically in Figure 5. The model assumes that we can pass one imaginary transverse plane though the thorax and another imaginary transverse plane through the pelvis. According to the laws of physics, only muscles that pass through both of these planes would be capable of imposing loads on the lumbar spine. We directly monitor every muscle that passes through both of these two planes except for two deep muscles. One of these muscle is not active during lifting and the other would require needle electrodes for assessment. Because of its proximity to the spine,
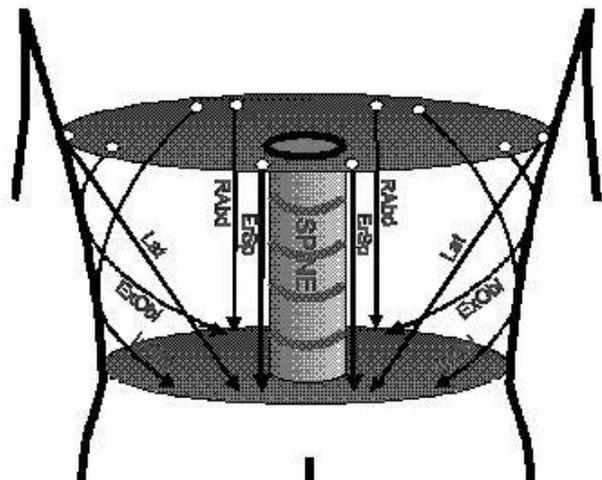
Figure 5. Ohio State University (OSU) EMG-assisted spine model torso.

this muscle contributes little to spine loading, therefore, it is not monitored in the model. The maximum physiologic cross-sectional area of a muscle dictates maximum force generation capacity of a muscle. Our recent studies (Marras *et al.* 1999a, b) have documented these cross-sectional areas of the male and female trunk using MRI. Our MRI studies have also documented the lines of actions of the trunk muscles for males and females and have allowed us to accurately model the geometric locations of these internal forces within the trunk.

The EMG activity of the 10 trunk muscles is adjusted for muscle velocity, muscle length, maximum muscle activity, and the size of the muscle. Once these adjustments are made, the continuous internal force generated by these muscles during the task of interest is estimated. The resultant muscle force vector is multiplied by its distance from the spine to assess its contribution to spine loading. These forces are then summed in the vertical and horizontal directions to predict compression and shear.

The ultimate proof of whether a model is robust enough to assess loading in three-dimensional space comes from validation studies. While it is unethical to directly monitor loads on the human spine using instrumentation, we can monitor model performance using indirect validation measures. Three validation measures are used to realistically assess model performance. In these validations, the external loads imposed about the torso are monitored using a force plate. Our model also *predicts* these external loads via the EMG activity. A comparison of these measures indicates the degree of agreement between the measured and predicted external load. Tests have indicated that the model has performed very well in over 1600 trials. A recent study (Marras *et al.* 1999a, b) has also explored model variability associated with the repetition of a task compared to independent evaluations on different days. Model performance was found to be excellent both within and between days.

5. USE OF THE MODEL

Figure 6 shows a worker instrumented with EMG electrodes so that loads on the spine can be assessed as he performs a work
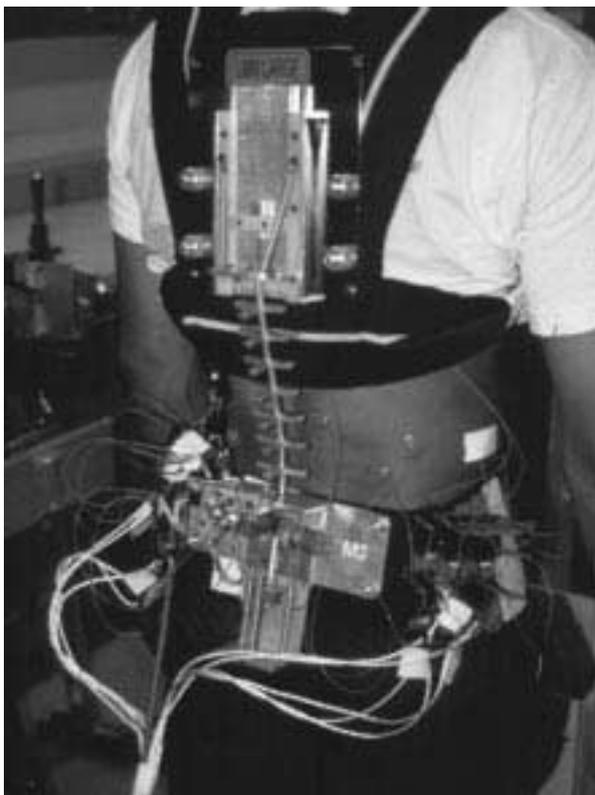
Figure 6. A subject instrumented with electrodes and the lumbar motion monitor (LMM). Instrumentation is used to run the EMG-assisted model.

task. The worker has also been fitted with a lumbar motion monitor (LMM). This device tracks the motion of the torso with respect to the pelvis. In this manner, the muscle lengths can be calculated and muscle velocities can be measures. This information is used to adjust the EMG signal so that muscle force can be represented.

The model has been developed so that it can be represented in a "windows" environment. Figure 7 shows an example of this program. The model includes a video of the worker performing the task along with numerous biomechanical performance measures. Windows may be selected that show the instantaneous kinematic, kinetic, and muscle activities indexed along with the video of the task. In this manner, the model can show how the spinal loads develop through an activity, which internal forces are responsible for the load development, how the position and activities of the worker are associated with the loading, as well as model performance measures. Dozens of windows are available thoroughly to evaluate factors that contribute to spine loading during task performance. In this manner, it is possible to determine which portions of a task place the worker at risk and which portions of the task are considered low risk.

6. MODEL USE EXAMPLE

In this manner, the model can also be used to evaluate the risk associated with a particular task. The risk associated with alternative task designs can also be assessed so that the risks or benefits of a job design alternative can be objectively assessed. In addition, since the model is sensitive to individual differences between workers, the model could be used to assess how risk
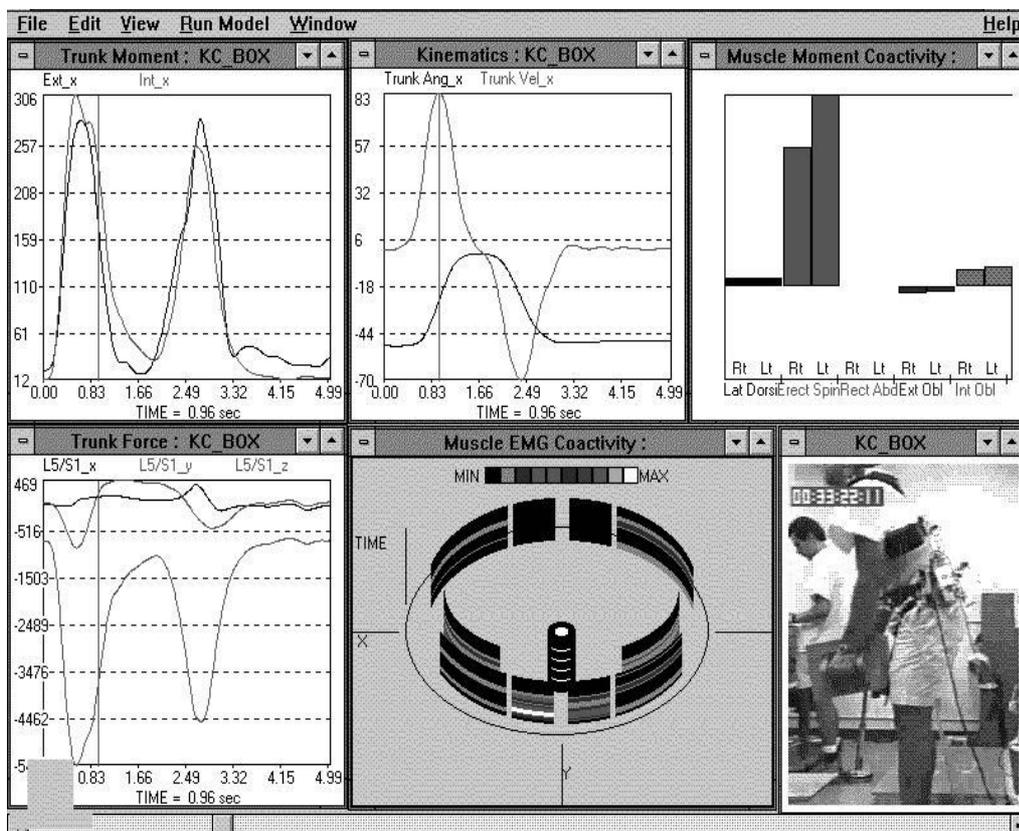
Figure 7. Windows environment adapted for the OSU EMG-assisted model.

Maximum Compression Force
60 lb. Box - No Handles
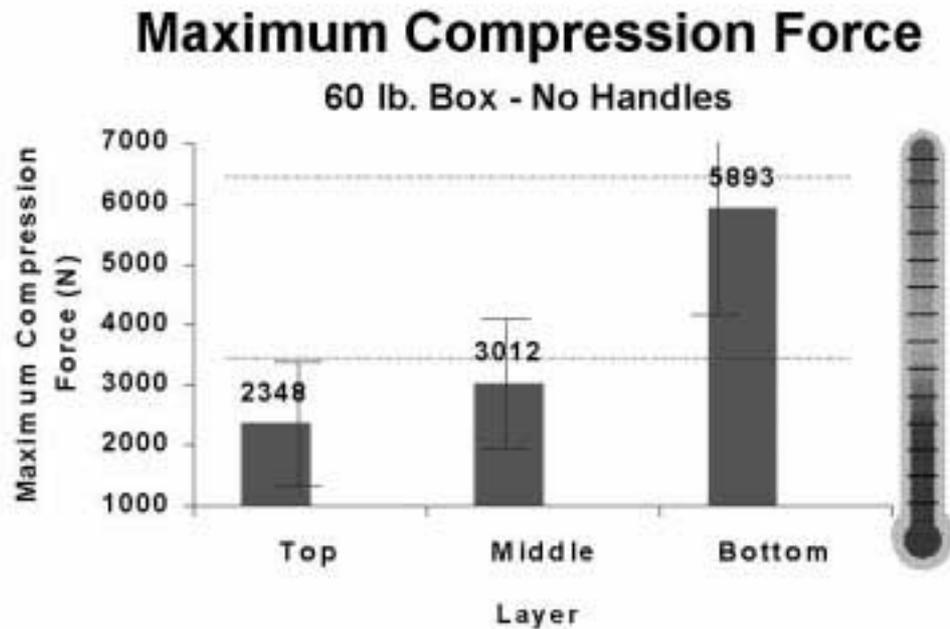
Figure 8. Maximum spine compression as a function of lifting from different vertical regions of a pallet; an example of the results obtainable from the OSU EMG-assisted model.

varies for a given task between workers or between various work techniques.

Figure 8 shows how this model was used to explore how spine loading (compression) changes as loads are lifted from different regions of a pallet in a distribution center environment. For this analysis, the vertical regions of a pallet have been subdivided as top, middle or bottom levels or layers. This figure shows the magnitude of compression compared to two tolerance benchmarks mentioned earlier. In addition, to aid interpretation a risk "gauge" has been included to the right of this figure. This figure indicates that most of the risk in lifting 60 pound boxes from a pallet occurs when lifting from the bottom layer where were average maximum spine compression approaches 6400 N. Hence, the threshold of vertebral end plate microfracture tolerance would be exceeded for many people. This information also makes it possible to estimate the percentage of workers that would be at risk of these microfractures so that the magnitude of risk can be assessed. Thus, we can see, though this quantitative analysis, that some sort of lifting assistance would be beneficial but this would only be necessary at the bottom layer of the pallet.

7. CONCLUSIONS

This chapter has shown how quantitative evaluations of the workplace can be used as a tool to evaluate spine loadings

associated with tasks that comprise a given job. The benefit of such analyses is that they permit one to quantitatively evaluate 'how exposure to a work task factor is too much.' Through such an accurate and complete assessment of work it is possible to significantly reduce the risk of LBD associated with the physical demands of work tasks.

REFERENCES

MARRAS, W.S. and GRANATA, K.P., 1997, The development of an EMG-assisted model to assess spine loading during whole-body free-dynamic lifting, *Journal of Electromyography and Kinesiology*, 7, 259–68.

MARRAS, W.S., GRANATA, K.P. and DAVIS, K.G., 1999a, Variability in spine loading model performance, *Clinical Biomechanics*, 14, 505–14.

MARRAS, W.S., JORGENSEN, M.J., GRANATA, K.P. and WIAND, B., 1999b, Size and prediction of female and male spine loading trunk muscles derived from MRI, *Clinical Biomechanics* (submitted).

McGILL, S.M. and NORMAN, R.W., 1986, Partitioning of the L4–5 dynamic moment into disc, ligamentous, and muscular components during lifting, *Spine*, 11, 666–78.

NATIONAL INSTITUTE FOR OCCUPATIONAL SAFETY AND HEALTH (NIOSH), 1981, *Work Practices Guide for Manual Lifting*. Technical Report, DHHS(NIOSH) Publication no. 81-122.

Manual Materials Handling in Unusual Postures

J. L. Smith

Texas Tech University, Lubbock, TX, USA

Early research in manual materials handling examined physiological, biomechanical and psychophysical capabilities of workers in standard lifting postures, which were usually described as two-handed, symmetric, sagittal plane lifting (or lowering). In 1981, NIOSH published *Work Practices Guide for Manual Lifting* which established recommended manual materials handling limits for such "standard postures". In 1991, revisions to the guide were published in which angle of twist was one of the components added to the recommended weight limit calculation (Waters *et al.* 1994). However, more unusual postures have received little attention. Part of the obvious reason is that, by definition, unusual postures imply that the situation is unique and hence trying to establish norms for unique postures has limited value.

In an effort to provide input data for a computerized CAD model that included a maintenance person, research efforts at Texas Tech University established capacity data for male and female subjects performing a variety of manual materials handling activities in unusual (or non-standard) postures. The capacity data were presented in terms of predictive models as well as the distribution of capacity data in Smith *et al.* (1989, 1992). The psychophysical approach was utilized to generate capacity data for the non-standard postures of interest. About 100 subjects were used for each task (50 males, 50 females, aged 18–25, selected to be representative of the US Air Force population in terms of height and weight characteristics). Each subject was asked to determine his or her individual capacity (without overexertion or lifting in an unsafe manner). In addition, each subject performed a variety of static and dynamic strength tests. Models were generated to predict handling capacity from one or more of the strength tests. A dynamic test, the incremental 6-foot lift, was the best predictor of handling capacity.

Examples of non-standard postures examined for the US Air Force study involved lifting and lowering while assuming a stooping, squatting, kneeling or sitting posture. Such postures represented those found while handling materials in restricted environments, such as aircraft cargo bays. Other unusual postures included lying face down and lifting objects through restricted openings, lying face up and lifting, and lying on the side of the body while performing a variety of activities. More common examples of non-standard tasks might include handling objects with only one hand or arm, holding an object in place with one or both hands, and carrying an object while crawling.

For the ergonomist needing capacity information for workers required to handle materials in non-standard postures, the best approach would be to simulate those activities in an environment where the testing would not interfere with production. Workers could be brought to the test area, given instructions for the test and asked to perform the required tests. Strength testing should also be conducted if capacity models are desired. To minimize data contamination and to reduce the risks of injury, subjects should be isolated during testing, not provided with feedback and be told to not discuss the test with their fellow workers.

REFERENCES

AYOUB, M. M. and MITAL, A., 1989, *Manual Materials Handling* (London: Taylor & Francis).

MITAL, A., NICHOLSON, A.S. and AYOUB, M.M., 1997, *A Guide to Manual Materials Handling* (2nd ed.) (London: Taylor & Francis).

NIOSH, 1981, *Work Practices Guide for Manual Lifting.* DHHS (NIOSH) Publication no. 81-122 (Cincinnati: US Department of Health and Human Services).

SMITH, J.L., AYOUB, M.M. and MCDANIEL, J.W., 1992, Manual materials handling capabilities in non-standard postures. *Ergonomics*, 35, 807-31.

SMITH, J.L., AYOUB, M.M., SELAN, J.L., KIM, H.K. and CHEN, H.C., 1989, Manual materials handling in unusual postures. In A. Mital (ed.), *Advances in Industrial Ergonomics and Safety* (London: Taylor & Francis).

WATERS, T.R., PUTZ-ANDERSON, V. and GARG, A., 1994, *Applications Manual for the Revised NIOSH Lifting Equation* (Cincinnati: US Department of Health and Human Services, Public Health Service).

Manual Materials Handling: Multi-person Lifting

M. A. Sharp† and V. J. Rice‡

†US Army Research Institute of Environmental Medicine, Natick, MA01760 USA

‡US Army Academy of Health Sciences, San Antonio, TX78234 USA

1. INTRODUCTION

Manual materials handling (MMH) of very heavy or bulky objects often necessitates the use of multiple-person teams. Examples of these tasks are loading large rolls of material onto or off of machines, transferring medical patients from a bed to a wheelchair, carrying injured persons on stretchers and moving furniture. More than 50% of lifting tasks and 49% of the lifting and carrying tasks performed by US Army soldiers involve teamwork in groups ranging from two to eight persons. For Army tasks, the most commonly occurring team size is two-person, followed by four- and then three-person teams. This is likely true for industrial team MMH tasks as well.

2. BACKGROUND

The majority of basic research examining team-lifting performance has concentrated on either a single maximum lift or a maximum acceptable weight of load (MAWL) for repetitive lifting. In applied research, patient transfers in multiple person teams is the task that has received the most attention.

2.1. Single Maximum Lift

The one repetition maximum (1RM) isometric (no movement) and isokinetic (controlled speed of movement) strength of teams of two and three men and two and three women and the 1RM dynamic lifting strength of two-, three- and four-person teams of men, women and mixed-gender have been examined. The team 1RM lifting strength is less than the sum of the individual 1RM lifting strengths by 10–40% for all team sizes and lifting modes examined. The difference between the sum of individual lifting strengths and team lifting strength varies with the team gender and the specific lifting task. To examine the relationship between the sum of the individual lifting strengths of team members and team-lifting strength, the percentage of the sum of individual lifting strengths relative to the team strength (% sum) is calculated: % sum = (team strength/sum individual strengths) x 100. For

Table 1. Team strength data.

	Team Size	Men	Women	Mixed-Gender
Dynamic[1]	2-person	89.8	91.9	79.7
	3-person	85.0	90.9	78.5
	4-person	86.0	90.3	83.7
Isometric[2]	2-person	94.1	79.1	
	3-person	88.6	87.0	
Isokinetic[2]	2-person	66.5	70.5	
	3-person	60.3	72.8	

[1] Sharp et al., 1997, [2] Karwowski and Mital, 1986; and Karwowski and Pongpatanasuegsa, 1988.

dynamic 1RM lifting strength, % sum was significantly greater for single gender teams (87% for men, 91% for women) than for mixed-gender teams (80%) when lifting in teams of two, three or four. Table 1 lists the % sum for three modes of lifting by gender and team size. Research shows teams of men lift heavier loads than mixed-gender teams, while teams of women lift lighter loads than mixed-gender teams.

2.2. Maximum Acceptable Weight of Lift (MAWL)

In addition to maximum lifting strength, the load a team finds acceptable for repetitive manual materials handling tasks is important. The load acceptable to 95% of the working population has been used to set limits for safe materials handling loads. MAWL is defined as the load a person is willing to work with under a given set of task conditions. For example, an individual may be asked to determine the MAWL for an 8 h work day when lifting a box from the floor to a 70 cm high table at the rate of three times per min. The person is given a box that is either too heavy or too light and asked to perform the defined task for 20 min. The individual then adds or subtracts weight from the box until it is judged acceptable. The individual adjusts the load so that s/he does not become overtired, overheated or out of breath. MAWL has been determined for repetitive lifting and lifting and carrying in two-person teams only. No information is available for larger teams. When determining the team MAWL, both team members must agree on the load adjustments and final load.

Unlike the 1RM for team lifting, the MAWL for repetitive lifting or lifting and carrying in teams tends to be equal to or greater than the sum of individual MAWL (% sum = 98 – 140%). When roughly matched for strength, team members work at the same exercise intensity (%O$_{2max}$) and perceive the repetitive MMH to be easier when working in pairs than when performing the same task alone. This is true despite the fact they may be lifting > 100% of the sum of their individual MAWL. This reduced perception of effort does not occur for the weaker individual of a pair with large differences in individual strength, such as mixed-gender teams. In this case, the lower strength person has a slightly higher heart rate, works at a slightly greater exercise intensity and perceives the repetitive task to be more difficult in a team, than when working alone. Because they are working harder, the weaker individual will fatigue faster and may be at an increased risk for injury during team MMH tasks.

2.3. Comparison of 1RM versus MAWL for Team MMH

While the team MAWL is equal to or greater than the sum of the individual MAWL, this is not true of 1RM team lifts. Task differences may be the source of these differing results. The 1RM team lift is a one time maximum effort involving little decision making. The team is either physically able or unable to lift the load. In contrast, the MAWL is a repetitive, submaximal self-selected load. The individual or team must estimate the load they are willing to handle for an extended period. Because there is a subjective element to MAWL selection, personality interactions may influence the team MAWL while not affecting the 1RM lift. One study found Type A individuals work at a higher percentage of their aerobic capacity, and is able to determine their MAWL in a shorter period than Type B individuals. A Type A individual

might be more insistent about exercising at a higher intensity by selecting a heavier load. The higher MAWL for teams than for the sum of individuals making up the team might be an expression of competitive behavior or a higher level of arousal. For social reasons individuals may be more motivated to select heavier loads when working in teams than when working alone.

Variation in lifting technique may also be a source of differences in the % sum obtained during a 1RM versus a MAWL. It is easier to standardize lifting technique during a single maximal lift (1RM) than during repetitive submaximal lifting or lifting and carrying (MAWL). As muscle groups fatigue during a repetitive lifting task, a typical strategy used to continue performing the task is to change lifting technique to allow for physiological recovery. A change in technique would not be necessary during a single maximum lift, as adequate rest is provided between lifting attempts. During 1RM testing, changes in lifting technique are often used as a criterion for test termination.

2.4. Biomechanics of Team MMH

Few studies have been conducted to compare the biomechanics of individual and team lifting. Differences in lifting technique produce differences in the spinal compression and lateral shear forces for one- versus two-person lifting. Spinal compression and moments about the sagittal plane are reduced for two-person symmetrical lifts compared to individual symmetrical lifts when the load is equated on a per-person basis. When lifting asymmetrically, however, two-person lifts result in significantly higher lateral shear forces and moments than one-person lifts. These biomechanical differences are likely due to differences in lifting technique. Changes in pelvis and trunk positioning and limitations in the ability to position one's body with respect to the load occur during two-person lifting and change the biomechanics of the lift.

Biomechanical analyses of patient handling in two-person teams have found that the maximum limits for spinal compression are often exceeded. This is due to the awkward lifting position (reaching, twisting) and the unequal distribution of the patient's body weight between the two handlers.

2.5. Prediction of Performance on Team MMH Tasks

Three prediction equations have been developed to estimate maximum lifting strength or maximum acceptable weight of lift in two-person teams. In two of three equations the lifting strength of the weaker individual was a good predictor of team lifting strength. The third equation found the lifting strength of the team member with the higher body weight (presumably the stronger individual) was a good predictor of the team load lifted. When all equations were applied to the same sample, the equations based on the lifting strength of the weaker individual were more accurate. This would lead to the conclusion that the weaker individual limits the load that can be lifted in a two-person team. These attempts were all limited by a small number of observations and only one study included mixed-gender teams.

3. CONCLUSIONS/RECOMMENDATIONS

- The 1RM for dynamic two-person team lifting is 10–20% lower than the sum of individual 1RM lifts, but little further decrease is found with the addition of one or two more people. If a recommended load for an individual performing a task has been determined, the %sum (Table 1) can be used to estimate the maximum load for two to four persons lifting as a team. For example, if a safe load for one person is 25 kg, a two-person team performing a similar lift should not attempt a load that is more than 80–90% of two times the individual load (25 x 2 = 50 x 0.8 – 0.9 = 40–45 kg). No further reduction in the per-person load is needed for three- or four-person lifting tasks, so an additional 25 kg can be added for each additional person up to four. This recommendation is based on optimum conditions where there is adequate team coordination, space and handholds.
- Team lifting and carrying MAWL tends to be equal to or greater than the sum of the individual MAWL for the same task. Therefore, doubling the individual MAWL provides a reasonable estimate of the load two-person teams will find acceptable for a repetitive MMH task.
- If at all possible, lifting teams should be roughly matched for strength and height. When a large strength discrepancy exits between two persons performing a repetitive team-lifting task, the weaker individual works at a higher relative intensity, predisposing that person to early fatigue and possibly injury.
- In two person team lifting (1RM and MAWL), the lower strength individual limits the load lifted.

REFERENCES

FOX, R.R., 1982, A psychophysical study of bi-manual lifting. Unpublished Masters thesis, Texas Technical University, Lubbock, p. 1.

JOHNSON, S.L. and LEWIS, D.M., 1989, A psychophysical study of two-person manual materials handling tasks. In *Proceedings of the Human Factors Society 33rd Annual Meeting*, 651–3.

KARWOWSKI, W., 1988, Maximum load lifting capacity of males and females in teamwork. In *Proceedings of the Human Factors Society 32nd Annual Meeting*, 680–2.

KARWOWSKI, W. and MITAL, A., 1986, Isometric and isokinetic testing of lifting strength of males in teamwork. *Ergonomics*, 29, 869–78.

KARWOWSKI, W. and PONGPATANASUEGSA, N., 1988, Testing of isometric and isokinetic lifting strengths of untrained females in teamwork. *Ergonomics*, 31, 291–301.

SHARP, M., RICE, V., NINDL, B. and WILLIAMSON, T., 1997, Effects of team size on the maximum weight bar lifting strength of military personnel. *Human Factors*, 39, 481–8.

SELAN, J.L., HALCOMB, C.G. and AYOUB, M.M., 1987, Psychological variables effecting lifting capacity: a review of three studies. *Trends in Ergonomics/Human Factors IV*, 987–93.

ULIN, S.S., CHAFFIN, D.B., PATELLOS, C.L., BLITZ, S.G., EMERICK, C.A., LUNDY, F. and MISHER, L., 1997, A biomechanical analysis of methods used for transferring totally dependent patients. *SCI Nursing*, 14, 19–27.

Manual Work and Back Stress in Industrially Developing Countries

J. Charteris and P.A. Scott

Department of Human Kinetics and Ergonomics, Rhodes University, Grahamstown, South Africa

INTRODUCTION

While occupation-related back stress among manual workers is a global problem (Pope *et al.* 1991) it is arguably true that ergonomists of the industrially developed world, with expertise in manual materials handling situations typical of those societies, would be both appalled by, and ill-equipped to deal with, the sorts of abuses of the human spine which are so typically encountered in IDC contexts (figures 1a and b).

The issues raised in this chapter summarize the authors' experience of conditions in several countries of the Southern African Development Community (SADC: Angola, Botswana, Democratic Republic of Congo, Lesotho, Malawi, Mauritius, Mozambique, Namibia, Seychelles, South Africa, Swaziland, Tanzania, Zambia, Zimbabwe). Mindful of the sterling efforts being made in many of these countries where historical circumstances often militate against rapid improvement of working conditions, we offer these insights in a generalized form to typify SADC conditions rather than to dramatize worst-case scenarios in any specific country. Our familiarity with conditions in Far Eastern and Latin American countries, though less extensive, suggests strongly that the Southern African model is adequately representative of issues pertaining in IDCs in general.

1. COMMON DENOMINATORS IN THE LABOR-INTENSIVE ETHOS OF MANUAL WORK IN IDCS

Infrastructural deficiencies bring to the fore several issues that need to be considered in appreciating the situation to be addressed. Manual materials handling stresses in general and lift-related back stress in particular in IDC contexts are hampered by the following issues:

1.1. Unreliability of Existing Databases

While the immediate causes are biomechanical and can be dealt with at a micro-ergonomic level, this issue must also be addressed at the epidemiological level. Even in industrially advanced countries epidemiological research in occupational back stress is fraught with methodological problems of definition, classification, and diagnosis. These pall to insignificance against the problem faced in IDCs where record-keeping is somewhere between unreliable and non-existent. With defective records to go by (whether in the form of insurance, workers' compensation, hospital, occupational health clinic or factory records), there is little hope in IDCs of getting a good handle on the nature and extent of the back-stress problem. People die of aids-related diseases, not of occupational back stress; consequently poor resources and more urgent priorities deflect attention away from often insidious, cumulative manual work stresses, and this fulfills the old adage that if you're not looking for something you tend not to find it. In societies where the workforce is exploited, there is sometimes overt, and always indirect, pressure on workers to suffer in silence rather than place their jobs at risk by complaining. Under these conditions a strained back is less likely to be reported than a bleeding hand because evidence for the latter is manifest, while that for the former could be countered by an accusation of malingering.

In IDCs Workers' Compensation data are poor indicators of the real problem of occupational lift-related back stress because they typically take no account of short-term absences which are not recorded in time-loss reports. Nor do they take cognizance of the vast biomass of the self-employed in the informal sector,

(a)

(b)

Figure 1. Practical examples of loads handled in Africa
(a) An 80kg concrete block maneuvered by a 62kg manual laborer
(b) African women carry head loads of up to 80% body weight and 150% body length

of injuries exacerbated by but not caused by occupation, or of non-reported work-related incidents. Often Workers' Compensation injury statistics are produced several years in arrears, and when improvements are made it is in the direction of speeding up the process of compilation rather than improving the quality of the disclosures. In some SADC countries accident reports are listed by gross anatomical regions. Thus all injuries to the "trunk", however caused, are lumped together, making it impossible to determine spinal-stress incidence statistics in meaningful form.

1.2. Willful or Ignorance-based Misrepresentation of Evidence

In defense of Workers' Compensation databases worldwide, it must be noted that they are not designed specifically to serve purposes such as are the focus of this article; internationally they are recognized as lacking validity on diagnosis of occupational back stress (Abenhaim and Suissa 1987). Even good records contribute little of value to solving the problem. "Poor correlations between pathology and symptoms, and between symptoms and function continue to frustrate patients and their practitioners, particularly in determining work readiness" (Hazard *et al*. 1992).

As a result, what is difficult in medically sophisticated and worker-oriented societies of the industrially developed world becomes a morass of misunderstanding and apathy born of ignorance in the medically unsophisticated, worker-exploited societies of many industrially developing countries. Not surprisingly, therefore, medical writers in the IDCs can argue, as Simon (1992) and du Toit (1992) have done, that in their countries occupational back injuries are few and far between and that back-related disability is ubiquitous in industrially developed countries in part because it is engendered by medical overcare, and that in less industrialized countries less coddling results in less disability.

Contradicting these views in the worldwide experience of a high rate of back injuries among construction workers and miners the consensus of authorities is that: "The incidence of low back pain ranges from 60 to 80% of the adult population worldwide" (Pope *et al*. 1991).

1.3. Absence in IDCs of Comprehensive National (let alone regional international) Strategies to Address the Ill Effects of Occupation-related Back Stress

Why should occupational back stress reduction receive attention at the national level in industrially developing countries? The answer is as simple as it is important — to promote worker well-being and to enhance national productivity. Among the results of occupational back disorders, other than the obvious pain and disability of the sufferer, are socio-economic losses in the form of compensation payments, absenteeism, replacement, training of new recruits, continuity disruption, administrative busywork, reduced/lost earnings, reduced efficiency and increased risk of re-injury in those who suffer in silence, to mention but a few. It is generally accepted that measurable costs in this connection are but a small proportion of the total. Thus, when relatively small sums are reported as reflective of a nation's compensation costs in respect of work-related back stress, these are the superficial pay-out costs reflecting only the tip of the iceberg.

Work-safe Australia, whose national campaign to reduce back injuries at the workplace has been effectively presented for some years, is convinced, as we are, that good design/redesign of manual handling operations, in tandem with effective educational programs for all concerned, are effective preventive measures.

There is a need in IDCs to see the link between productivity and occupational back stress, particularly because the ergonomic implications of productivity are not well appreciated, leading to a failure to comprehend the extent to which an ergonomic orientation can assist in reversing the malaise. Typically in IDCs the per capita GDP has for decades remained static. Where population growth rates of over 2.6% p.a. are common and employment rates increase around 0.6% p.a., effectively this means that the sub-Sahara African population would double every 27 years while the number of people employed would only double every 117 years — the need for jobs increasing 4.3 times faster than the ability to provide them. Seen against this backdrop, poor work practices, whether in terms of inefficient production or low levels of safety, begin to spiral out of control.

2. PARTICIPATORY ERGONOMICS

Education is sorely lacking in IDC work practice; when attention is given it is often in the form of training in the pejorative sense of requiring compliance rather than inculcating understanding and a sense of personal responsibility. Even in industrially developed countries there is an obsession with absolute load-level specification, with complete disregard to individual variability and the fact that, for the same remuneration, a less well-suited manual worker is being exposed to significantly greater stresses than a physically better equipped counterpart. The immanent causative factors contributing to the high incidence of work-related back stress in IDCs and developed countries alike are the same (see table 1):

- suboptimal positions while working
- labor-intensive manual materials handling
- suboptimal work—rest ratios
- whole-body vibration exposure

In the case of the first three above, the common denominator of primary concern is, of course, the task-inherent risk factor of physical load in terms not only of the mass being moved but its location relative to the center of mass of the mover, and the manner in which it is being accelerated (Charteris and Scott 1989).

Occupational back stress can be alleviated in IDCs using what we call "no-cost or low-cost intervention strategies" that are situationally feasible improvements on current practices and which, above all, are supported by the people doing the work. This requires that "First World" criteria are disregarded, but not the sentiments behind them. "Exert less; produce more" is an attainable ethos, but it requires interventions by ergonomists who are also completely familiar with local geophysical, economic and socio-cultural conditions.

3. DISPELLING THE MYTH THAT THE BACK-STRESS PROBLEM BEGINS AND ENDS WITH SETTING OF MAXIMAL PERMISSIBLE LOAD-LIMITS

In IDC contexts what is needed are easily implementable risk-identification tools and simple risk-rating techniques, such as liftRISK proposed by Charteris and Scott, whose effect is to

Table 1. Results of a 10-year survey by the authors of Risk Factors for low-back stress in a wide range of Southern African Industries/Employment sectors.

TYPICAL INDUSTRY	SELECTED TASKS	REPRESENTATION		WORST FACTOR(S)	COMMENTS
		TASK DESCRIPTION	RISK RATINGS		
SUPERMARKET/ WAREHOUSING	BULK STORES	SHELVING CARTONS – WAIST LEVEL – LOW SHELVES – HIGH SHELVES	LOW MEDIUM/HIGH HIGH	LIFT-RATE STOOP STRETCH	HIGH WORK FOR SHORT PERIODS GREAT TASK DIVERSITY
BUILDING/ CONSTRUCTION	BRICK STACKING	CONVEYOR TO PALLET PALLET TO CONVEYOR PALLET TO PALLET	MEDIUM MEDIUM MEDIUM/HIGH	REACH REACH; RATE RATE; STOOP; REACH	HIGH WORK RATE
MINING	WETFLAP CATHODES ORE-CRUSHER	LIFTING COPPER SHEETS TO BEAT MAINTENANCE REPLACEMENT OF HEAVY BARS	HIGH HIGH	MASS; STOOP MASS	HAZARDOUS CONDITIONS EXCESSIVE HEAT
FISHERIES	UNLOAD TRAWLERS FOOD PROCESSOR FILLETING LINE STACKING BOXES LOADING TRUCK	STACKING CONTAINERS ON QUAY LOADING COMPRESSOR LOADING CONVEYOR PALLETISING CONVEYOR TO FLAT BED	MEDIUM LOW LOW LOW HIGH	STOOP; REACH REACH STOOP RATE; REACH; STOOP RATE; REACH; STOOP	CASUAL LABOUR (UNTRAINED) AWKWARD CONTAINERS PAIRS SELF-PACED INTENSE
HEALTH-SERVICE	PATIENT HANDLING	ONTO; OFF BED	LOW/HIGH	REACH; TWIST; STOOP	HIGHLY VARIABLE; LOW PREDICTABILITY
ROAD-WORKS	PAVEMENT LAYING	CONCRETE RODS (60KG) PAVING SLABS (80KG) CEMENT CURBING (96KG)	HIGH HIGH HIGH	MASS; STOOP MASS; STOOP MASS; STOOP	PAIRS, POOR UNDERFOOT SURFACES
CONSTRUCTION SITE	BRICK LAYING ROOFING	SELF-PACED BRICK HANDLING PLACING ROOF TRUSSES	LOW LOW	STOOP; STRETCH STRETCH	AS IN BRICK STACKING BUT SLOWER PACE HAZARDOUS CONDITIONS
MOTOR INDUSTRY	ASSEMBLY LINE	WINDSCREENS ENGINE MOUNTING SEAT INSTALLING BATTERY FITTING TYRE LIFTING	MODERATE HIGH MODERATE MODERATE/HIGH MODERATE	STRETCH MASS; STOOP; REACH STOOP; REACH MASS; REACH; STOOP REACH; STRETCH	AWKWARD POSTURES
REFRIGERATION	FROZEN FOOD HANDLING	LIFTING BAGS; BINNING	MODERATE/HIGH	MASS; STOOP	TEMPERATURE AWKWARD CONTAINERS
SOFT-DRINKS BREWERIES WINERIES	CRATE STACKING	CONVEYOR TO PALLETS	MODERATE	RATE; MASS; STOOP	BELT-PACED, SUBOPTIMAL
AGRICULTURE	HARVESTING MARKET SUPPLY	GRAPE-BASKET LOADING DE-TRUCKING PALLETISING	MODERATE/HIGH MODERATE MODERATE	STOOP; STRETCH; MASS STRETCH; MASS STOOP; REACH	HIGH INTENSITY AWKWARD POSTURE

sensitize users to the need to address excessive task-inherent risk factors in realistic ways. Attention should be placed less on legislating maximal permissible loads to be lifted – in situations where, in any case, enforcement simply does not occur – and more on encouragement designed to reduce excessive stooping, reaching and stretching, and excessive rates of lifting. These, in the IDC model presented here, account for 34%, 26%, 11%, and 10% of the back-stress factors encountered in situ, while excessive mass per se was implicated in less than 20% of instances.

In IDCs in the foreseeable future issues of overpopulation, underemployment, and lack of education (particularly health-related education) transcend all other problems. They underlie even such basics as access to clean water, electricity, and housing. Against this backdrop, any implementation by isolated modern companies of work conditions typical of the developed countries would be as societally ineffectual as restricting aids education to the affluent. True, such practices serve as exemplars, but where social conditions militate against the following of these examples less idealized but feasiably attainable schemes should be followed. The best way to ensure sustainable growth of an ergonomic ethos in IDCs is via simple, feasible, and unremitting education of workers to act as co-responsible agents in their own work-related safety and well-being, while persisting with encouragement to employers to implement no-cost/low-cost changes in work practices. These small changes in working conditions, while not ideal, are manifestly implementable and do constitute improvements on current practices. Emulating the best practices of the developed

countries, while superficially laudable, is not sustainable in the IDC context where labor-intensive practices will prevail as long as the only maarketable resources on which the bulk of the populace can draw are in terms of semi-skilled physical work capability.

REFERENCES

ABENHAIM, L.L. and SIUSSA, S., 1987, Importance and economic burden of occupational back pain: a study of 2500 cases representative of Quebec. *Journal of Occupational Medicine*, **29**, 670–4.

CHARTERIS, J. and SCOTT, P.A., 1989, Risk identification in manual materials handling operations: a prototype expert system which reveals task-operator mismatch. *Ergonomics SA*, **2**(2), 73–85.

Du TOIT, J.G.A., 1992, The "injured-on-duty" back. *Continuing Medical Education*, **10**(7), 1257–63.

HAZARD, R.G., REEVES, V., and FENWICK, J.W., 1992, Lifting capacity: indices of subject effort. *Spine*, **17**(19), 1065–70.

POPE, M.H., ANDERSSON, G.B.J., FRYMOYER, J.W., and CHAFFIN, D.B., 1991, *Occupational low back pain: assessment treatment and prevention* (St Louis: Mosby Year Book).

SIMON, M.J., 1992, Low back pain: how much of a disability? *Continuing Medical Education*, **10**(7), 1185–91.

Mental Fatigue and Related Phenomena

T. Marek

Institute of Management, Jagiellonian University, Krakow, Poland

1. INTRODUCTION

The term mental fatigue is used with different meanings. Mental fatigue is considered as a decrease in physical and cognitive performance. It is considered also as a condition affecting the mental processes, including feelings of fatigue. Ergonomists lay stress on consequences of fatigue. It is work decrement. Very often mental fatigue is viewed in a context of activation theory. According to this workers perform best at some optimum level of activation. If a worker is overloaded, he/she suffers from an exhaustive form of fatigue. In the case of insufficient demands the (load) worker suffers from a type of fatigue known as boredom or monotony.

2. FATIGUE SENSATION

One of the most predominant mental fatigue factors is sensation. In a tired condition workers feel tired or dull in bodily parts, are inhibited in doing work and are forced to give up. The sensation is not unpleasant until a worker can stop and rest. It is harmful when rest is impossible. Generally, sensation consists of three factors: bodily drowsiness sensation; weakened cognitive processes and motivation sensation; and physical complaints. As Grandjean and Kogi (1971:18) pointed out, "sensations of fatigue have a protective function similar to those of hunger and thirst. They force us to avoid further stress, and allow recovery to take place." The sensation of mental fatigue is a biological signal for recovery.

3. NEURAL ASPECTS OF MENTAL FATIGUE

From the cognitive neuroscience point of view mental fatigue is a decrement in the functioning of brain centers due to long-lasting work. According to the classical knowledge (Grandjean and Kogi 1971) it is state of the central nervous system induced by prolonged activity and controlled by antagonistic brain stem systems which are responsible for arousal and inhibition. Changes in arousal level in plus and/or in minus was always considered as one of the crucial symptoms of mental fatigue.

According to current knowledge the relations between mental fatigue and arousal are much more complex. There are several arousal systems (e.g. Robbins and Everitt 1996) based on the four neurotransmitter systems: noradrenergic, serotonergic, dopamine and cholinergic. These are responsible for different forms of arousal and so mental and cognitive functions. For example, dopamine systems can be identified with a process of activation that modulates the speed and vigor of responding as well as motivational states. Noradrenergic and cholinergic projections to cortex have much more subtle roles in modulating processes of attention (e.g. action selection, discrimination, and concentration of attention, as well as inflation on some aspects of memory). The functional balance between all these systems in particular work conditions is the basic factor for effective action. The balance is established due to particular work demands and

individual resources, and differs from action to action. Mental fatigue may start from the inability of the worker to establish and keep appropriate functional balance between the above mentioned systems.

The second crucial symptom of mental fatigue is attention impairment. There are three attentional networks: orienting network; executive network, and alerting network (e.g. Posner 1996). Each network subserves the different attentional processes. In the case of the orienting network, the parietal lobe disengages attention, then the midbrain areas act to shift the attention index and finally the pulvinar engages attention in a new location. The executive network consists of two parts: the part responsible for action awareness and the part that allows one to control the action. Finally, the alerting network plays a crucial role in vigilance to the task. While this network is active, the executive network is inhibited. There are many connections between arousal and attentional network systems. Some of them are well recognized but some are not. Networks of arousal and attentional systems and their functioning are crucial for mental fatigue phenomenon. Dependent on external conditions and work demands, an appropriate model of action is established for both types of networks. Changes which appear due to long lasting work in some parameters of this model cause the mental fatigue. Because of a large variety of parameters there are different patterns of mental fatigue symptoms.

Each mental action seems to be accompanied by neurochemical changes. These changes, especially related to serotonergic deactivating mechanisms, are related to mental fatigue (Grandjean and Kogi 1971, Ursin 1986). Owing to these changes, "something is built up which can only be removed by rest and/or sleep." (Ursin 1986: 62).

4. MENTAL EFFORT AND FATIGUE AND HUMAN BRAIN METABOLISM

Traditionally, two types of mental effort are distinguished (Mulder 1986, Marek and Fafrowicz 1993): one related to the difficulty of the task, processing complexity, and the second to the control of the state and adjusting parameters in case these are suboptimal. The first type of effort creates the essence of mental effort, which is related to task and environmental demands. The second type reflects adaptation and compensatory effort, which follows all deficits and suboptimal mental states of the subject.

Many authors point out that a high level of mental effort caused by a high level of task variability and complexity are crucial in mental fatigue.

There are two essential questions. The first is whether mental effort is related to human brain metabolism. The second is whether mental fatigue is a consequence of a depletion of the brain metabolic energy.

It is well known that energy is required for many functions of neurons. First, it is needed to synthesize neurotransmitters that are essential for synaptic transmission. Second, demands related to the molecular pumps of the neuron membrane. The task of the membrane is to maintain normal intracellular levels of potassium and sodium. While computational transaction occurs within the neuron, the demands for restorative action of the membrane pumps appear. The neurons consume more oxygen and glucose when processing demands are higher. The metabolic rate for glucose is a sensitive indicator of the local brain energetic

demands related to given task-specific demands. The metabolic rate for glucose is measured in humans by positron emission tomography (PET) (Roland *et al.* 1996) and functional nuclear magnetic resonance (fNMR) (Petegrew 1991). It is well documented now that the magnitude of glucose utilization by the given part of the brain or cortical regions is a function of stimulus or task complexity. The processing effort (mental effort) is correlated with increased glucose utilization (e.g. Beatty 1986, Warburton 1986). However, there is no evidence that energetical resource limitations could occur in the brain due to even the most intense mental effort. Thus, mental fatigue is not a consequence of a depletion of the brain metabolic energy. It seems to be the process that is switched on to prevent a depletion of metabolic energy at the active processing sites in the brain.

The crucial aspect for the process is the second type of mental effort that is related to the resource control (compensatory control). Mental fatigue as Mulder (1986: 190) pointed out, "may be best characterized as the inability to continue to exert executive resource control." It creates a preventive mechanism against the depletion of metabolic energy in the brain.

REFERENCES

BEATTY, J., 1986, Computation, control, and energetics: a biological perspective. In Hockey, Coles and Gaillard (eds), *Adaptation to Stress and Task Demands: Energetical Aspects of Human Information Processing* (Dordrecht: Martinus Nijhoff), 43–52.

GRANDJEAN, E. and KOGI, K., 1971, Introductory remarks. In Hashimoto, Kogi and Grandjean (eds), *Methodology in Human Fatigue Assessment* (London: Taylor & Francis), 17–30.

MAREK, T. and FAFROWICZ, M., 1993, Changes in attentional disengagement process under repetitive visual discrete tracking task measured by oculographical index. In Marras, Karwowski, Smith and Pacholski (eds), *The Ergonomics of Manual Work* (London: Taylor & Francis), 399–402.

MULDER, G., 1986, The concept and measurement of mental effort. In Hockey, Coles and Gaillard (eds), *Adaptation to Stress and Task Demands: Energetical Aspects of Human Information Processing* (Dordrecht: Martinus Nijhoff), 175–98.

PETEGREW, J.W., 1991, Nuclear magnetic resonance in neuroscience research. In Boller and Grafman (eds), *Handbook of Neuropsychology*, vol. 5 (Amsterdam: Elsevier), 39–56.

POSNER, M.J., 1996, Attention in cognitive neuroscience: an overview. In Gazzaniga (ed.), *The Cognitive Neurosciences* (Cambridge, MA: MIT Press), 615–24.

ROBBINS, T.W. and EVERITT, B.J., 1996, Arousal systems and attention. In Gazzaniga (ed.), *The Cognitive Neurosciences* (Cambridge, MA: MIT Press), 703–20.

ROLAND, P.E., KAWASHIMA, R., GULYAS, B., and O'SULLIVAN, B., 1996, Positron emission tomography in cognitive neuroscience: methodological constraints, strategies, and examples from learning and memory. In Gazzaniga (ed.), *The Cognitive Neurosciences* (Cambridge, MA: MIT Press), 781–8.

URSIN, H., 1986, Energetics and self regulation of activation. In Hockey, Coles and Gillard (eds), *Adaptation to Stress and Task Demands: Energetical Aspects of Human Information Processing* (Dordrecht: Martinus Nijhoff), 53–70.

WARBURTON, D.M., 1986, A state model for mental effort. In Hockey, Coles and Gillard (eds), *Adaptation to Stress and Task Demands: Energetical Aspects of Human Information Processing* (Dordrecht: Martinus Nijhoff), 217–30.

Mental Models

J. R. Wilson

Institute for Occupational Ergonomics, University of Nottingham, Nottingham NG7 2RD, UK

1. INTRODUCTION

Consider some typical situations investigated by ergonomists. First, an operator at a remote control task who must carry out a job from a vantage point not used before, trying to get oriented and to guide the system through its critical operations in the most reliable, safe, effective manner. Second, a factory worker at an industrial machining center who must try to understand why parts she is producing are showing hairline cracks, and what combination of the many parameters available on the machine would solve the problem. Third, a team responsible for scheduling a transport system from a control center, coping with a variety of visual and auditory on-line communications and also using their knowledge of operating procedures. Fourth, a maintenance engineer of a continuous process plant trying to integrate information from system diagnostics with data from historic print out and what he can see from visual inspection.

In all these cases the people concerned will form some kind of internal, or mental, representation of the system they are working with; this representation can be termed a *mental model*.

In fact, in most circumstances people will construct and employ several mental models, for instance of the behavior of the physical variables in the systems (electrical, chemical or mechanical), or of the structure and form of the system, or even of the rules governing the operation of the system. These mental models may vary in their degree of abstraction and may be formed from observation of the system itself, from knowledge of operating, emergency or maintenance procedures, from instructions and training, or even from experience of other similar systems worked with in the past.

The argument for the existence of mental models may seem eminently reasonable. What, though, does it mean — in theory and in practice — to ergonomists and human factors specialists?

2. WHAT ARE MENTAL MODELS?

The notion of mental models — that people at work hold in their minds a representation of the systems with which they are working, and upon which they draw to assist their understanding and operation of those systems — is an extremely attractive one in ergonomics. The idea is that, if we can find a way to identify and represent the mental models we expect to be held by operators in a particular situation, and if we can communicate these models successfully to designers, then systems would be designed with interfaces which far better match the needs and expectations of the operators. Indeed, to "know the user's mental model" is often an early guideline in advisory texts on design for human–computer interaction. This begs the question: "what is a mental model?" There is no agreement about this, but accounts from ergonomics (as against cognitive psychology) taken together indicate that mental models:

- are internal (mental) representations of objects, actions, situations or people;
- are built on experience and observation, of both the world in general and the particular entity of interest;
- are simulations that are run in mind to produce qualitative inferences;
- underpin our understanding of a system;
- allow one to describe, predict and explain behavior of a system;
- constitute topography, structure, function, operation of the system;
- contain spatial, causal and contingency relations;
- may contain grouped schemata;
- are instantiated each time they are required, and, therefore, vary over time;
- are parsimonious, being developed only to the minimum extent required for a particular purpose; and
- are incomplete, unstable, multiple and non-unique.

3. MENTAL MODELS IN ERGONOMICS

For nearly 60 years the notion that people form mental models of their environment has been a topic of psychological enquiry, starting with the work of Craik (1943). And ergonomics researchers of human–machine systems have long regarded the concept as being almost self-evident (e.g. the collection of Edwards and Lees 1974). In other domains there are parallel notions, such as naive theories, device models, mental maps and domain models. In the past few years researchers in human–computer interaction have picked up the notion, attempting to define user models as well as models of the user (e.g. Allen 1997). During all this time, developments of the mental model notion within ergonomics and within cognitive psychology (e.g. Johnson-Laird 1983) appear to have followed independent tracks, at least until recently. This is perhaps caused by the role of psychologists purely to explain behavior but the need of ergonomics to utilize such explanations to improve design.

In ergonomics theories and concepts are sought, from
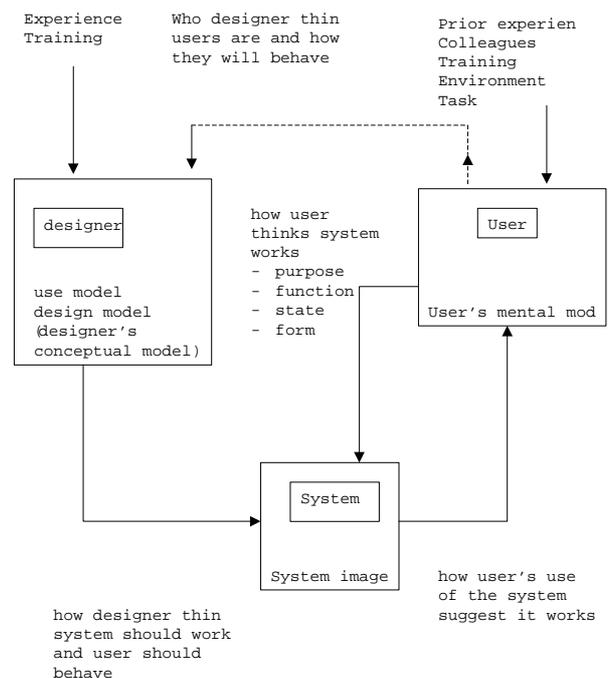
Figure 1. Use of mental models in design.

cognitive psychology and elsewhere, to adapt and apply them to the specification and implementation of systems in offices, services, factories, transport, the home and leisure. If we can predict or understand the mental model that a new system user might hold about that system and its domain, or what model might be built by interaction with the system, then we can improve interface design, training, operating procedures and so on. In Norman's (1983) view, by understanding the potential users' mental model, and by adapting their own conceptual model accordingly, designers might develop a system "image" that better matches, sustains and helps develop an appropriate user mental model. Figure 1 expands from these ideas, and defines — optimistically — the potential role of understanding the user mental model in design.

4. USER MODELS AND MODELS OF THE USER

As well as confusion over the different ways in which "mental model" is employed in cognitive psychology and in ergonomics, there is often also confusion over whether we are talking of *models of a person* or *models held by a person*. In the first instance we can produce a model of a person, of some of their basic characteristics or of their performance or behavior. This is what is often known as a *user model* when applied within human–machine system design or human–computer interaction; examples from the former are HOS and from the latter GOMS and TAG. In contrast to these user models, we have a user's or a persons' model, for which the term mental model is usually used, where we are talking of the model that the person holds in their mind of any entity or concept. Of course any user model may itself contain components of the anticipated mental model held by a user, but we must distinguish clearly between the two.

Adding to the confusion, the mental model held by the user, in any form, must be identified and reported to be of value in design, and, therefore, must itself actually be the model held by an experimenter or researcher of the model held by the user. At some point the investigator must have examined user performance in practice or else experimented in controlled circumstances to provide data from which they constructed a representation of the user mental model. By extension, we can talk of the designer's model of the users' mental model, the model of the user embedded in training and instructions or, going to greater extremes, the designer's model of system requirements based upon a researcher model of the potential users' mental model of the likely system design!

5. FORMS OF MENTAL MODEL

It may be practically useful to accept two distinct forms of mental model (Wilson and Rutherford 1989). First is a *Conceptual Mental Model*, a non-formal, non-computational representation of users' knowledge and how this is employed in task performance (perhaps as a "walk through" simulation). Such mental models are incomplete, unstable and may often be multiple; for example, the models instantiated by an operator to assist in fault diagnosis at a flexible manufacturing cell will vary not only in form each time but also in type since they may variously be of functional, spatial or causal relationships. Certainly there will be more than one mental model formed; these will overlap in content and use, and will have gaps or missing data depending upon the operator's experience and training.

This notion of a conceptual mental model is useful to ergonomists within user-centered design, task analysis, training specification and job design. It is also potentially a stage on the way to identifying the second form, a *Formal Mental Model*; this (with reference to Johnson-Laird 1983) being formal, computational and practically useful if it can be implemented. Such different forms of mental model are also implied by Rasmussen (1986: 140–8) and Nielson (1990: 155–6).

The majority of ergonomics literature appears implicitly to refer to the first form, the conceptual mental model. It is regarded almost as axiomatic that people form some kind of internal representation that constitutes their topographical, structural and functional understanding of a physical system and which allows them to describe, explain, understand and predict system behavior. It is easiest to conceive such a model as comprising a system simulation that can be reconstituted and run to derive or confirm understanding. Borrowing from Payne (1988), users' conceptual models may be seen as embodying a weak commitment to the notion, postponing the general questions of *how* people represent and use knowledge in favor of understanding *what* knowledge they represent and how they use it in making inferences in specific domains. Two stronger commitments proposed by Payne are that mental models may be run to produce qualitative inferences readily and that they rely on experiential knowledge. Within this view it is possible to see included such varied ideas as a mental map, an internal device model, a mental simulation, a naive (physics) theory, or a domain model (Gentner and Stevens 1983).

There is room for both types of mental model — the conceptual mental model and the formal mental model — in human factors. A conceptual model has seemingly great face validity. It is useful in itself as an encouragement and focus for user-centered design, task analyses, skill-based manufacturing and so on. The conceptual model also is an early stage in the process of identifying and specifying a computational or formal mental model. The utility of the mental model, if it can be specified computationally, is immense. This is a big "if," however, and there are many other serious questions, about what mental models are, whether the theoretical notion is necessary or if existing theories of knowledge representation and inference suffice.

6. IDENTIFICATION OF MENTAL MODELS

Given that it seems worth at least making the effort to try to identify specific mental models of systems held by potential users, how might this be done? Unfortunately, current understanding does not allow simple identification of a "best" method and researchers use many different methods to collect basic data and then to derive or interpret a "mental model." Sometimes there are doubts about the quality of such exercises. Where mental models are induced from subjects' responses, the route the investigator takes from the information elicited to the hypothesized mental model is often unreported. Experimenters also may employ intuitive rather than formal procedures to suggest possible mental models, since little is known about the creative and inductive processes underlying model generation.

One danger in exploratory studies of mental models is that preconceptions about the form and content of likely mental models may only be confirmed as a result of the choices made about relevant data, methods and analyses. The most appropriate

strategy to identify mental models may consist of two distinct stages (Rutherford and Wilson 1991), first suggesting a mental model on the basis of initial data and the experimenter's domain knowledge, and second, testing the validity of the suggested mental model, probably through controlled experimentation. Most mental models research appears to be at the first, inductive, stage, which is one reason why a variety of approaches to mental model identification and description are observed.

The methods that might be used in the process of mental model identification are many and varied, covering much of ergonomics methodology. Classical laboratory experiments, quasi-experiments in the field or observation of actual behavior can all be employed, with data collected from direct recording, interviews, questionnaires, think-aloud verbal protocols, and (two person) constructive interaction. Problem-solving experiments, in which participants perform "how to" and "what if" tasks, can be carried out, in which verbal, written, and graphical protocols are taken as well as performance measures, and other possibilities are techniques from the better documented world of expert knowledge elicitation.

7. REPRESENTATIONS OF MENTAL MODELS

Since there is a large variety of methods that can be used to identify mental models there is also a large number of forms in which they can be represented. A simple form, if of restricted value, is a textual account, either following performance observation or on the basis of experimental findings. A more valuable text form will be a natural or reduced transcription of the verbal protocols collected from one individual or two or more in a constructive interaction experiment. Commonly found in brief descriptions of the notion are pictorial representations (akin to the idea of "pictures in the mind"), and often these are used to illustrate the model people might hold of the action of physical variables such as electricity, gravity or the solar system. Hierarchical tree diagrams might be used as the representational form of systems of knowledge, showing levels and inter-connectedness of entities. More dynamic mental models might be represented by flow charts, and this form might be incorporated into a pictorial representation to produce a mental model of, say, a continuous process plant with its vessels, pipes, valves, levels and flows all included. If a more computational account is produced, this might be on the basis of logical syllogisms or mathematical notation.

8. LIMITATIONS AND CRITICISMS

Current thinking on mental models highlights a number of issues to do with limitations for the utility and usability of the notion of mental models. These include:

- The limitations in usefulness of the notion itself, particularly if mental models are incomplete and multiple. Moray's (1990) Lattice Theory was developed as one proposed way round this dilemma, but there have been critics following attempts to use this theory.
- Is the notion even needed? Mental models may be accounted for perfectly adequately within the human information processing model, particularly as regards LTM/WM (Bainbridge 1991). The relationship of mental models to schemata, frames and scripts also remains to be considered. Furthermore, one can distinguish between the notion of

mental models and the notion of interface compatibility to produce a population stereotype?

There is a paradox of inconsistencies within mental models; the mistakes people make are one of the most interesting things we can understand about their potential performance, and prominent among these mistakes are inconsistencies. However, the methods used to identify and represent mental models may well lead to operators or subjects recognizing their own inconsistencies, and accounting for them before providing the information on which the mental model is built by the researcher.

One of the main problems with mental models is to do with the methods by which we identify and represent them. All those reported in the literature have some weaknesses and potential biases. In particular, the decisions made about which method and measures to use to extract information to construct the mental model may imply that the researcher will have already made some (conscious or unconscious) decision about the form this mental model will take, therefore biasing the representational form they use. There appears to be little or no way around this dilemma, but there is certainly need for more explicit accounts of how the domain provides data which are then represented and interpreted into a mental model.

If individuals can be said to hold mental models then it seems reasonable to think that teams do also. The problems of instability and non-uniqueness, and of the methods used to identify mental models for individuals, will be just as great, if not greater, for teams. There is also an interesting question about whether a high performing team will, in fact, closely share a mental model or models; alternatively, the very existence of a number of different mental models among the team may mean that they can deal with all situations (familiar or not) and less easily get locked into the first and easy diagnoses due to information availability and cognitive lock up biases.

It may be that the use and value of mental models within human factors is higher for explanation (i.e. a fire fighting tool) than for prediction (as a design tool).

9. CONCLUSIONS

Mental models have become a concept much employed within ergonomics, a growth of interest, however, not paralleled by any greater clarity in the issues surrounding the notion! If it is truly to be useful and usable within ergonomics we need to be clear about relevant methods of data collection and analysis, procedures for conversion of data into a mental model, and the specification, test and refinement of that model.

Some fundamental questions must be answered to do with what a mental model might be, how it might be identified and represented and how it might subsequently be applied in systems analysis and design. Fundamentally, are we sure whether we are talking about anything other than "operator/user knowledge" when the term is applied within the fields of human–computer interaction and human–machine systems?

The notion of mental models will hopefully turn out not to be a sidetrack of arcane academic interest. It can allow the production of useable tools in design as well as enhancing understanding of the need for user-centered design. In the end, however, the evocative nature of the mental model concept, its tendency to be meaningful to designers, and the very process of elicitation required to represent such models may well be more

important than worrying about the exact form and content of mental models.

REFERENCES

ALLEN, R.B., 1997, Mental models and user models. In M. Helander, T. Landauer, P. Brabhu (eds) *Handbook of Human–Computer Interaction* (Oxford: Elsevier), 49–63.

BAINBRIDGE, L., 1991, Mental models in cognitive skills; the example of industrial process operation. In Y. Rogers, A. Rutherford and P. Bibby (eds) *Models in the Mind: Perspectives, Theory and Application* (London: Academic Press).

CRAIK, K.J.W., 1943, *The Nature of Explanation* (Cambridge: Cambridge University Press).

EDWARDS, E. and LEES, F.P. (eds), 1974, *The Human Operator in Process Control* (London: Taylor & Francis).

GENTNER, D. and STEVENS, A.L., 1983, *Mental Models* (London: Lawrence Erlbaum Associates).

JOHNSON-LAIRD, P.N., 1983, *Mental Models* (Cambridge: Cambridge University Press).

MORAY, N., 1989, A lattice theory of approach to the structure of mental models. *Philosophical Transactions of the Royal Society of London, B*, 327, 577–83.

NIELSEN, J., 1990, A meta-model for interacting with computers. *Interacting with Computers*, 2, 147–60.

NORMAN, D.A., 1983, Some observations on mental models. In D. Gentner and A.L. Stevens (eds) *Mental Models* (Hillsdale: Erlbaum), 7–14.

PAYNE, S.J., 1988, Methods and mental models in theories of cognitive skill. In J. Self (ed.). *Artificial Intelligence & Human Learning* (London: Chapman & Hall).

RASMUSSEN, J., 1986, *Information Processing and Human–Machine Interaction* (Amsterdam: North-Holland).

RUTHERFORD, A. and WILSON, J.R., 1991, Searching for the mental model in human–machine systems. In Y. Rogers, A. Rutherford and P. Bibby (eds) *Models in the Mind: Perspectives, Theory and Application* (London: Academic Press).

WILSON, J.R. and RUTHERFORD, A., 1989, Mental models: theory and application in human factors. *Human Factors*, 31, 617–34.

Mental Workability and an Increasing Life Span

R. Cremer

TNO Work & Employment, PO Box 78, 2130 AS Hoofddorp, The Netherlands

1. INTRODUCTION

Satisfactory job performance depends both on the characteristics of the individual employee and the condition of the work environment. An important characteristic of employees is their information-processing capacity. A concept which is highly related to mental work capacity. In a broad sense, processing capacity or mental resources have been conceived as the composite of both structural and energetic components of information processing. Structural processing components have traditionally been identified with specific information-processing operations (Sternberg 1969) or a limited capacity working memory system. In addition, energetic components have been associated with some form of mental energy, effort or attention (Kahneman 1973).

Generally it is expected that there is a general decline in the capacity of the information-processing system with advancing age. Results of various studies demonstrate that with respect to speed measures, the complexity effect, that is the greater slowing of old subjects in more complex than in simple RT tasks, is a robust phenomenon (Cerella 1991). Moreover it is concluded in stress and aging studies that the combined effects of task complexity and stress conditions yield a complex pattern of performance measures, which could not be explained simply on the basis of simple resource-depletion models of aging. Related behavioral problems may occur and become more obvious with aging in particular types of work activities. In discussing the relevance for work it is debatable how results from laboratory studies can be interpreted for real-world activities such as work.

2. MODERN WORK, MENTAL DEMANDS AND EFFECTS OF COGNITIVE AGING

The ongoing introduction of new technologies is expected to bring about increasingly more changes in our daily life and work activities. In modern work mental demands have an increasing significance. First automation and later emphasis on information and communication processing (ICT) made many types of work more abstract and complex. Workers have sometimes difficulty in recognizing where they are involved in the chain between input and output in production and services. Most dominant in modern work is the abundant use of information suppliers and processors.

Many human operators or employees are expected to process an increasing amount of information. In their work, high demands on the senses require shift regulation in relevance but also switches between modalities either between senses or between cognitive processes. Malfunction in selecting (problems in attention with advancing age) can then easily overburden the human system. Combined effects of slowing and poor selection mechanisms may then lead to the experience of time pressure or inadequate control over the abundance of sensory input.

Furthermore it can be concluded that the more knowledge-intensive functions are the more mental effort is required for effective performance. Knowledge handling is a so-called "fluid ability," which is particularly vulnerable with advancing age. The counterweights of such abilities are the "crystallized" abilities. Relations between workability and these abilities will be explained below.

2.1. Weaknesses and Strengths in Information Processing with Advancing Age

In following the Piagetion stages it is established that there are early periods of rapid development in childhood, a learning and experiencing phase which is prolonged in young adulthood and only in adolescence signs of decrease of function become manifest. In this cycle of change disuse of cognitive function with advancing age is often confused with biological effects of aging. Whenever knowledge is gained or abilities are learned, not using them for a while usually means forgetting resulting in reduction of efficiency and automation. However with normal aging (as opposed to pathological aging) regular use of knowledge is a guarantee for preservation of knowledge, abilities or skills. Whereas effects of unhealthy aging (pathology or disease) may interact with the normal aging process. For example, a brain-injured person may show early signs of cognitive aging. On the other hand mechanisms such as compensation, deliberate use of strategy and coping (putting more effort into a job) may, to a large extent and to an advanced age, be useful in maintaining an adequate style of performance.

Largely, mental workability is composed of the following factors: (1) processes required for information handling (or "computational" mechanisms, (2) the "carriers" for such processing: memory and attention (or cognitive "resources") and (3) the subjective interpretation (emotions) and finally the preparedness to act (motivation and effort or the "allocation of resources").

The information processes are related to: input (sensory and sensory-perceptual), central cognitive (problem solving and deciding) and output (motor preparation and motor execution). Other dimensions are the energetic resources: e.g. input of information arouses the system, and effort can strengthen central processes, whereas activation is typical for output related processes. Furthermore the effectiveness of these combined processes depend on speed, duration, alteration and interference during processing. Speed of information processing (efficiency of transmission of signals in the nervous system) is often thought to be related to cognitive aging. When older nervous systems require more time (are slowed), an additional risk is that of "lost" information traces. For example, loosing time in one process may stop the continuous flow in other parallel or simultaneously ongoing processes. Finally, sensory input (e.g. environmental noise, or social pressure) may be interfering more easily in old than young adults. Such interference may disturb attention and memory functions, a situation, which may be particularly distractive for older adults in effortful processing of new or complex information. All the changes referred to above occur over the lifespan. However, they do not only concern functional loss: structural decline but also gains: life experiences and skill development takes place when growing older.

When referring to work and age, at least two aspects are important but not always highlighted: the accumulation of effects

of mental load over time and positive changes during the lifespan. The relation between these factors is that age related change may cause a general reduction of workability leading to an increased vulnerability for stresses at work. Age-related change as main variable will be explained below in terms of positive and negative outcome.

There is substantial evidence to suggest that the decline in mental functions is most conspicuous for "fluid" abilities, in contrast to "crystallized" abilities that are much more resistant with age (Horn 1982). Crystallized abilities are associated with verbal, acquired or specialized skills. These functions depend strongly on experience, information that has been stored by a learning process. Since performance is positively related to experience, crystallized abilities may improve with age owing to gradual accumulation of experience and knowledge. Fluid abilities refer to cognitive functions such as speed of decision making, short-term memory and attention, which mainly play a role during the acquisition and transformation of information. The impact of the decline in fluid cognitive abilities with advancing age in our society is substantial, especially if we consider the increased life expectancy of elderly persons, and the many changes and often increasing demands in terms of information and communication technology.

This functional approach of (work) capacities accounts for the dynamics of changes over the lifespan. At a young age knowledge is gained about facts and procedures, in adulthood the emphasis is on the application of established knowledge, skills and experience, and the integration of new and existing knowledge. This latter style of functioning makes information-processing skills automated, less time consuming and less effortful. The use of this condensed knowledge (skills and abilities) is a useful starting point for development (learning) at later age than application of cognitive abilities that require effortful processing. Also in discussions on work, health and aging of workers, arguments are usually restricted to the phase in which changes become visible in terms of deterioration of function, both physically and mentally. In a broader approach, not just decline but also growth should be incorporated when speaking about changes over the lifespan.

It is hardly possible to map the pattern of change over the lifespan, however a general principle is that the lifespan and a professional career take place with continuous change (growth and decay) of effectiveness and potential capacities. In other words the lifespan can be considered as a continuum of development from adulthood into senescence. A variety of life events may affect the life cycle positively or negatively. Personal development, education and a job career are among such events. Also general health and the onset and progression of the aging process are major determinants.

From many classical textbooks on aging it may be concluded that in laboratory settings major effects of aging are measured. However laboratory conditions and experiments have usually little relevance for work, and research in work conditions not always have relevance for functioning in daily life. These conclusions have to do with the validity of aging studies. In daily life an individual may choose what to do (ignore tasks and activities which are difficult and complex, instead activities (e.g. hobbies) are chosen which are over-learned and highly experienced. In this respect aging at work takes place somewhere in between. Performance of elderly workers is not as strenuous and valued as in laboratory studies. Still work is not a natural situation, specific work demands determine the constraints for an aging employee. So it is unavoidable that aging at work also means confrontation with limitations of performance. Adaptation of working conditions for example in terms of cognitive ergonomics may help to solve some of the problems that aging workers encounter.

2.2. Adaptation at Work to Age-related Change

In considering adaptations at work a useful framework is the information processing chain (see the explanation above). The processes are sensitive for aging and affect practical functioning. However, ergonomic adaptations at the workplace can compensate for some of the shortcomings (Table 1).

Apart from the focus on computational processes, one can design adaptations on the level of the energetics that supports the processes. Such effects can be achieved by reducing: demands on processing speed, time-on-task, switches between modalities and processes and interfering sources.

3. DISCUSSION

Over time changes take place both in the nature of work and in attitudes towards an aging workforce. Modern work (e.g. information and communication technology) increases mental

Table 1. Information processing, ageing and adaptation: some basic principles and examples.

Information processing	Practical effect	Adaptation of the work environment
Sensory	clarity of information uptake: signal-to-noise ratio; distinction relevant/irrelevant information; visual clarity, simplicity of information	amplify signals, reduce noise; improve clarity of visual and auditory signals;
Sensory-perceptual	sensory storage capacity, processing capacity; logic and uniformity of presented information	simplicity and uniformity of information; reduce disturbance
Cognition	memory and attention functions (working memory capacity), capacity for reasoning, problem solving	reduce complexity; apply individual expertise and knowledge; apply familiar concepts; improve logical relations
Motor preparation	stimulus/response relation; preparation for action	apply direct stimulus-response relations; apply logical and compatible choices
Motor execution	response speed, accuracy of response	apply enforcement of movement (electronic) aids in order to simplify choice

demands of work. On the other hand the general attitude towards older workers becomes friendlier and more tolerant. Lifelong learning at work can be applied as a successful strategy against changes in the nature of work and individual changes over the lifespan. Continuous upgrading of knowledge, the "hardening" of mental workability during the active lifespan means the "survival" of the cognitive system. This strategy is the opposite of disuse, the process of (forgetting) to use acquired knowledge, abilities and skills.

As was referred to in the introduction, the relevance of laboratory findings for real-world activities should be taken into consideration. In his 1982 book on "adult cognition," Salthouse gave an impressive overview of studies indicating age-related functional change. In a final summary chapter categories were defined, however no statements about relative rate of decline of magnitude of the age trend were given. In his book the following list in this specific order was given about age-related decline in abilities: reasoning and decision making, memory, spatial abilities, perceptual-motor and cognitive speed and sensory factors. Moreover these five categories of specific processes are proportionally more vulnerable to overall task complexity (Salthouse 1982). All these findings are results from laboratory studies. An important question that was raised in his book dealt with the observation why these findings are not so relevant in the behaviors of everyday life. Some explanations were particularly relevant for the concept of workability with advancing age. It was suggested that the laboratory tasks may be irrelevant to the skills of daily living. Also that most of the activities of daily living are only minimally demanding, suggesting that most people will have enough in reserve to handle nearly all normal situations. Finally it is argued that more dramatic age differences have not been noted in, for example, elderly workers because tasks have been highly practised.

Arguments about the ecological validity of laboratory findings in cognitive aging studies may have changed in the past decades. For example, vulnerability in cognitive speed, memory, perceptual load and generally increasing complexity in work demands may place older worker in a weaker position than in earlier years. Also more field studies are conducted in recent years. In a book specifically devoted to "work and aging" (Snel and Cremer 1994) a fair number of studies is devoted to real work situations providing evidence for age-related decline in real work conditions. This promising development in applied research will help in finding better and proactive solutions for an extended working life of aging workers in good health.

REFERENCES

CERELLA, J., 1991, Age effects may be global not local: comments on Fisk and Rogers (1991). *Journal of Experimental Psychology: General*, 120, 215–33.

HORN, J.L., 1982, The theory of fluid and crystallised intelligence in relation to concepts of cognitive psychology and aging in adulthood. In F.I.M. Craik and A. Trehub (eds) *Aging and Cognitive Processes. Advances in the Study of Communication and Affect*, vol. 8 (New York: Plenum), 238–78.

KAHNEMAN, D, 1973, *Attention and Effort* (Englewood Cliffs: Prentice-Hall).

SALTHOUSE, T.A., 1982, *Adult cognition: An Experimental Psychology of Human Aging* (New York: Springer).

SALTHOUSE, T.A., 1996, The processing-speed theory of adult age differences in cognition. *Psychological Review*, 103, 403–28.

SNEL, R. and CREMER, R., 1994, *Work and Aging: A European Prospective* (London: Taylor & Francis).

STERNBERG, S., 1996, The discovery of processing stages: extensions of Donders method. *Acta Psychologica*, 30, 276–315.

Mental Workload

P.S. Tsang

Department of Psychology, Wright State University, Dayton, Ohio45435, USA

1. INTRODUCTION

Growing complexity and increasingly automated features of modern human-machine systems are presenting operators with fewer physical demands and greater cognitive demands. Unlike physical demands, cognitive or mental demands are not directly observable. The concept of mental workload is used to benchmark the mental demands of complex systems.

There are several broad objectives for assessing mental workload. The first is prediction. To be able to anticipate the workload level of a new system is valuable for system designs. Workload prediction is also useful when there is a modification to the system configuration (e.g., the incorporation of automated components), the operational procedure, or the personnel. The second objective is evaluation. The most common application is comparing the workload of alternative systems.

Workload evaluation is also used to determine the changing demands across different phases or conditions of a task or mission. The third objective is diagnosis. The appropriate workload assessment can potentially isolate the troublesome spot and identify the precise nature of the demand. The objective of the workload assessment is one main determinant in selecting the appropriate workload metrics.

1.1. Mental Workload Fundamentals

O'Donnell and Eggemeier (1986) defined mental workload as "that portion of the operator's limited capacity actually required to perform a particular task." All mental processing requires some resources. The greater the task demand, the greater the processing resources needed to maintain performance at a certain level. If task demands exceed available resources, performance falters. So having some indications of the expended resources would be invaluable for ensuring optimal performance.

Gopher and Donchin (1986) added: "Mental workload is clearly an attribute of the information processing and control systems that mediate between stimuli, rules, and responses. Mental workload is an attribute of the person-task loop, and the effects of workload on human performance can therefore be examined only in relation to a model of human information processing."

A common model is the multiple resource model (Wickens 1992). This model states there are specific attentional resources for different types of cognitive processing. Attentional resources are defined along three dimensions:

- *Stages of processing*: perceptual/central processing and response processing require separate resources.
- *Processing codes*: spatial processing and verbal processing require separate resources.
- *Input/output modalities*: visual and auditory processing require different resources; manual and speech responses require different resources.

When performing multiple tasks, the higher the similarity in the resource demands among the task components, the more severe the resource competition and the lower the level of performance. With extremely difficult tasks, it would be necessary to select one of the tasks to perform or else face performance degradation. Whether any additional tasks can be accommodated depends on whether there are resources not needed by the high priority task.

2. WORKLOAD METRICS

2.1. Types of Workload Measure

There are four major types of measure: subjective ratings, operator performance, psychophysiological measures, and analytic methods. Subjective metrics elicit the operator's perception of the mental demands of the system. Performance metrics are part of the system performance of interest. Psychophysiological metrics detect changes in the operator's body that are related to task demands. Analytic methods are modeling efforts. Typically, subject matter experts (SMEs) are asked to provide workload estimates in a structured way that is based on a specific workload model and detailed task/mission analysis. Each of the workload measures is thought to provide valuable information. But they all have different strengths and weaknesses.

2.2. Properties of Workload Measures

Workload measures can be characterized by a number of properties: sensitivity, diagnosticity, intrusiveness, validity, reliability, ease of use, and operator acceptance. These are important considerations in selecting an appropriate workload assessment technique.

Sensitivity refers to how well a metric detects changes in mental workload. Often the degree of sensitivity depends on the level of the workload. For example, performance measures are more sensitive at higher levels where the limit of the operator capacity is being reached and performance deficits are expected.

Diagnosticity is the extent to which a metric can reveal the precise nature of the workload. The multiple resource model provides a framework for determining the aspect of processing that is demanding. For example, if performance of a secondary task that mostly uses response processing resources deteriorates, then the output of the primary task should be examined for a potential workload problem. On the other hand, if excessive visual workload is indicated, then the system displays need to be examined.

Intrusiveness is an important consideration. If the workload assessment itself interferes with task performance or adds to the workload, then the workload measure is contaminated. For example, subjective reports that are required during the task performance can take up mental resources that should be allotted to the primary task.

Construct validity considers whether the metric is measuring mental workload. Although important, it is difficult to verify because the workload concept is complex and difficult to define operationally. Two other validity criteria should be considered. Concurrent validity is when different workload metrics correlate in a systematic fashion. For example, subjective rating tends to increase as performance deteriorates. Predictive validity is an especially important consideration for the analytic methods because SME workload estimates are analytical rather than experiential. They should be correlated with operational measures for validation.

Reliability describes the replicability of the workload measures

under similar conditions. Measures that fluctuate independently of task demands have no predictive value and are therefore not useful. One common test is the test/retest correlation. Under similar testing conditions, the metrics obtained at time t_A should correlate well with those obtained at time t_B.

With respect to ease of use, subjective measures are the easiest to administer. Since most operational systems do not have built-in performance recording, the system under evaluation must be specially modified for this purpose. Technical knowledge is required to properly analyze the mental requirements of primary and secondary tasks. Psychophysiological techniques require specialized equipment and expertise in data collection and analysis. Analytic methods require the adoption of a specific workload model and an intimate knowledge of the system to be evaluated.

Operator acceptance is crucial for a successful workload assessment. It is important to explain the nature of the assessment to the operators. Anonymity must be assured in order to gain and maintain cooperation.

2.3. Strengths and Weaknesses of the Workload Metrics

2.3.1. Subjective measures

Operators are typically asked to provide a number on a rating scale to represent the level of mental effort required to accomplish a task. One distinction enjoyed by subjective measures is their use as projective measures. Although direct performance or physiological data cannot be obtained from a system yet to be realized, operators (usually SMEs) could provide workload estimates based on a detailed description of the functionality of the new system. There are several common approaches to subjective workload assessment:

- A rating scale asks for a single unidimensional rating for the overall workload level or it asks for ratings on multiple dimensions.
- The ratings are obtained immediately after task performance or they are obtained retrospectively, having experienced all the task conditions.
- The ratings are absolute or they are relative. An absolute rating is based on the task of interest by itself. For relative ratings, operators compare the task of interest to a single standard or to multiple task conditions (the redundant method).

Descriptions of the more common subjective metrics can be found in O'Donnell and Eggemeier (1986).

Subjective metrics have high operator acceptance since operators can express their opinions. And because they are easy to use, they are the most commonly chosen of all metrics. The measurement units are not task dependent, so subjective metrics can be used to compare different tasks. The metrics are also more sensitive to information represented well in consciousness (such as working memory demands) than to demands that are not well represented (such as response execution processing demands). Multidimensional measures provide diagnostic information. Subjective measures appear to be reliable and have concurrent validity with performance measures. Ratings that are not obtained during task completion are nonintrusive.

One potential drawback is a susceptibility to memory problems if the ratings are made considerably after task performance.

Subjective estimates may also be susceptible to operator bias, past experience, and ego. It is advisable to explain to the operators the importance of responding honestly.

2.3.2. Performance measures

The primary task method assesses the actual operator or system performance. Deteriorating or erratic performance may indicate an unacceptable workload level. The secondary task method can be used when primary task performance is unavailable due to a lack of performance recording in the operational system or a paucity of performance indicators in a highly automated system. Also performance measures are insensitive to very low workload levels since a stable level of performance could be maintained by increased effort. Adding a secondary task would increase the demand to a level where performance measures would be sensitive.

The secondary task is performed concurrently with the primary task, and it is important to know the precise nature of its demands. Primary and secondary tasks should compete for the same processing resources such that changes in the primary task demand would result in changes in the secondary task performance. The degree of interference is used for inferring the workload level. Laboratory tasks such as memory and mental arithmetic tasks are often used as secondary tasks. Alternatively, a naturally occurring part of the overall task can be used as an embedded secondary task. For example, changing demands of the primary flight task would be expected to affect the response latency of the secondary task of responding to radio messages.

The primary task metric is the most direct measure. Except for extremely low workload conditions, performance is sensitive to a variety of task demands. Performance measures are reliable to the extent that performance has stabilized at the time of assessment. But primary task performance is not always available. When primary performance is available, it is specific to the system evaluated and the workload results may have limited generality. Furthermore, different units of measurement for different task performances could pose a scaling problem when comparisons are needed or when aggregating different performance measures.

The pattern of interference observed with secondary tasks of known demand characteristics could pinpoint the type of processing resources demanded by the primary task. This diagnostic information is invaluable for optimizing system performance. Considerable background knowledge is needed to properly conduct a secondary task assessment. And the addition of an extraneous task has another potential drawback: not only might it add to the workload, but it may also fundamentally change the processing of the primary task. It could be intrusive and it could misrepresent the true level of the workload. The embedded secondary task technique is proposed to circumvent this difficulty. Further ideas about the performance measures can be found in Tsang and Wilson (1997).

2.3.3. Psychophysiological measures

Two general classes of psychophysiological measures function interactively: central nervous system measures (CNS) and peripheral nervous system measures. CNS measures include electroencephalographic (EEG) activity, event-related brain potentials (ERPs), magnetic activity of the brain, and measures of brain metabolism such as positron emission tomography. One

component of the peripheral nervous system that is responsive to workload manipulations is the autonomic nervous system (ANS). Measures of ANS activity include cardiovascular measures, measures of pupil diameter, and respiratory measures. Kramer (1991) provides a comprehensive review of the psychophysiological metrics.

A distinction of the psychophysiological measures is that not only are they indicators of mental workload, they are also indicators of vital bodily functions. They are affected by a variety of environmental artifacts as well as the physical state of the operators. Elaborate controls and baseline assessments are therefore required for many of the psychophysiological measures. In addition, they are generally quite complex, consisting of multiple components. Components of the physiological signals that are indicative of mental demands could be small compared to those reflective of bodily functions. Sophisticated data analyses are often needed to extract the relevant information. If the ultimate interest of the workload assessment is system performance then it is important to establish the link between physiological measures and system performance.

A notable strength of the psychophysiological metrics is their nonintrusivenss on task performance. Most physiological measures are obtained by attaching small electrodes to the operator; the detected potentials are then amplified and recorded for later processing. Physiological measures do not require the operators to perform any additional tasks or to deliberately generate subjective workload estimates. However, some physiological techniques are equipment-intensive and may have environmental as well as mobility restrictions. Fortunately, technological advances have reduced many of these problems. Most physiological measures are available on a continuous basis. They are therefore most suited for assessing moment-to-moment or phasic workload changes and for assessing highly automated systems that provide few observable performance indicators. Selectively sensitive physiological metrics (e.g., ERP) can provide valuable diagnostic information.

2.3.4. Analytic methods

Analytic methods rely on mathematical, engineering, and psychological modeling of the workload context of interest. Most are designed to capture the workload of dynamic complex multi-task environments. For example, PROCRU (procedure-oriented crew model) was used to examine the approach-to-landing phase of commercial airflights and W/INDEX (workload index) was used to examine the design of the advanced tactical fighter. Their development is largely driven by the inordinate costs of implementing a system that eventually produces an unacceptable level of workload. Workload prediction is the primary goal of analytic methods. McMillan *et al.* (1989) provide an orientation to the analytic methods.

One distinct advantage of the analytic methods is that each parameter and assumption of the model must be made explicit. This serves several important functions: (a) it fosters careful consideration of the relevant input and output variables in the model, (b) it provides specific predictions that could be tested empirically, and (c) it makes it easier to communicate the findings

Analytic methods are used less often than the other methods, and most have been developed only recently. Current research suggests that analytic methods which assume multiple resources or multiple channels of processing account for more performance variance than the time line method (which assumes serial, single-channel processing). Models which take resource conflicts into consideration—concurrent tasks with similar types of task demand would impose greater workload than concurrent tasks with dissimilar demands—account for more performance variance than models which do not. This line of research helps to identify the parameters and assumptions needed for effective application of analytic methods.

2.4. Multiple Workload Measures

2.4.1. Benefits of multiple workload measures

Multiple workload measures are likely to be needed to obtain a complete picture of the mental workload incurred in a system. This is because mental workload is multidimensional and most real-world tasks are complex and involve multiple task components. Some workload metrics (e.g., pupil diameter) only reveal the global level of the overall demand whereas others are specific in the nature of the information they can provide (e.g., secondary task performance). Global information may not be useful for identifying individual troublesome components. Selectively sensitive metrics may fail to reveal other relevant aspects of the workload altogether. Using multiple measures based on a careful analysis of their strengths and weaknesses is recommended whenever feasible.

The selection of workload metrics should be based on the assessment's objectives, the technical constraints, the available expertise, and the properties of the different metrics. Kantowitz (1992) advocates that the selection of workload metrics should be based on theory. Kantowitz suggests that trying to interpret data without the guidance of a theory is like assembling bricks randomly when constructing a building without a blueprint. It is important to combine a theoretical understanding of human information processing with a theoretical understanding of psychometric measurements. Information processing dictates what parameters should be measured; psychometrics suggests how to measure them.

2.4.2. Dissociations among workload measures

When different workload measures provide seemingly contradictory information, they are said to dissociate. Given that mental workload is multidimensional, and different workload metrics are differentially sensitive to the different workload dimensions, some dissociations among workload measures are to be expected. Dissociations between performance and subjective measures have received the most extensive investigation. Vidulich (1988) provides a summary of these findings. The understanding of the dissociations can be very helpful in evaluating the nature of the system's workload.

3. CONCLUSION

An array of workload metrics is now available for operational use. Performed properly, workload assessment can be invaluable. This does not mean that all questions have been answered successfully. Further theoretical development is needed to better integrate the large body of existing findings and to explain the intricate relationships among the different workload metrics.

As automation escalates and managerial responsibilities multiply, there is a growing concern for operators to be able to

maintain an awareness of their whole situation. In very much the same way that the concept of workload arose out of practical needs, the concept of situation awareness has grown from a need to understand how human operators manage dynamic and complex systems (Adams *et al.* 1991). But according to researchers, situation awareness and mental workload are not interchangeable concepts. The utility of one does not diminish the utility of the other.

REFERENCES

ADAMS, M.J., TENNEY, Y.J., and PEW, R.W., 1991, *State of the Art Report: Strategic Workload and the Cognitive Management of Advanced Multi-Task Systems*, BBN Report 7650, Crew Systems Ergonomics Information Analysis Center, Wright-Patterson Air Force Base, Ohio.

GOPHER, D. and DONCHIN, E., 1986, Workload—an examination of the concept. In Boff, Kaufman, and Thomas (eds) *Handbook of Perception and Human Performance*, Vol. II, *Cognitive Processes and Performance* (New York: John Wiley), Ch. 41.

KANTOWITZ, B.H., 1992, Selecting measures for human factors research. *Human Factors*, 34, 387–98.

McMILLAN, G.R., BEEVIS, D., SALAS, E., STRUB, M.H., SUTTON, R., and VAN BREDA, L. (eds), 1989, *Applications of Human Performance Models to System Design* (New York: Plenum).

KRAMER, A., 1991, Physiological metrics of mental workload: a review of recent progress. In Damos (ed.) *Multiple Task Performance* (London: Taylor & Francis), pp. 279–328.

O'DONNELL, R.D and EGGEMEIER, F.T., 1986, Workload assessment methodology. In Boff, Kaufman, and Thomas (eds) *Handbook of Perception and Human Performance*, Vol. II, *Cognitive Processes and Performance* (New York: John Wiley), Ch. 42.

TSANG, P.S. and WILSON, G.F., 1997, Mental workload. In *Handbook of Human Factors and Ergonomics* (2nd ed.) G. Salvendy (eds.) (Wiley, New York), pp. 417–49.

VIDULICH, M.A., 1988, The cognitive psychology of subjective mental workload. In Hancock, and Meshkati (eds) *Human Mental Workload* (New York: North-Holland), pp. 219–29.

WICKENS, C.D., 1992, *Engineering Psychology and Human Performance* (2nd ed.) (New York: Harper Collins).

Mental Workload Measurement

G. F. Wilson and F. T. Eggemeier*

Crew System Interface Division, Human Effectiveness Directorate, Air Force
Research Laboratory, 2255 H Street, Wright-Patterson AFB,
OH 45504-7022, USA and *Department of Psychology, University of
Dayton, 300 College Park, Dayton, OH 45469-1430, USA

1. INTRODUCTION

Mental workload is a multidimensional construct, and refers to the ability of the operator to meet the information processing demands imposed by a task or system. It is generally assumed that workload is related to operator performance such that low to moderate levels of workload are associated with acceptable levels of operator performance. High levels of workload or information processing demand that cannot be effectively accommodated by the operator are associated with degraded levels of operator performance.

A major purpose of mental workload measurement is evaluation of the workload levels imposed by a task or system with the objective of identifying and eliminating workload-related performance decrements. Although numerous approaches have been used to measure mental workload, most procedures can be classified as belonging to one of four major categories: (1) performance-based, (2) subjective, (3) physiological, and (4) analytic. Each category is described below.

2. PERFORMANCE–BASED MEASURES

Performance-based measures derive an index of operator workload from some aspect of the operator's capability to perform a task or a system function. Two major categories of performance-based measures are primary task measures and secondary task methodology.

2.1. Primary Task Measures

Primary task measures assess the capability of the operator to perform the task or system function of primary or principal interest (e.g., pilot the aircraft, control the system). Speed and accuracy measures are commonly employed, and so are assessments of multiple aspects of operator performance (e.g., ability to maintain proper altitude, speed, and glide slope on final approach in an aircraft). The use of primary task measures is based on the hypothetical workload-performance relationship noted above. This relationship assumes that as workload increases beyond the information processing limits of the operator, degradations in primary task performance will result from the inability of the operator to deal with the processing demands of the task or system.

Because they represent a direct assessment of the operator's ability to perform required tasks or functions, primary task measures should always be included in assessments of operator workload. However, although primary task measures can be considered to represent sensitive indexes of workload when an information overload exists, these measures can be insensitive to variations in workload at levels that do not actually exceed the information processing capability of the operator. At low or moderate levels of workload, it is assumed that the operator can compensate for some increases in information processing demand and maintain acceptable levels of primary task performance. Because of this insensitivity, it is typically recommended that potentially more sensitive workload assessment techniques (e.g., physiological, secondary task, subjective) should also be employed in a comprehensive assessment of operator workload. Several reviews of the workload assessment literature (e.g., O'Donnell and Eggemeier 1986, Eggemeier and Wilson 1991, Tsang and Wilson 1997) provide examples of the successful use of primary task measures and describe possible sensitivity limits in more detail.

2.2. Secondary Task Methodology

Secondary task measures assess the capability of the operator to perform the primary task or function of interest concurrently with an additional or secondary task. In the typical application of the secondary task measurement paradigm, the operator is instructed to perform the primary task or function at unimpaired single-task levels, and to permit secondary task performance to decline if degradations in performance cannot be avoided under the concurrent primary/secondary task condition.

Assuming no degradation of primary task performance under the concurrent task condition, secondary task performance provides an index of any excess information processing capacity that is afforded by the primary task. High levels of secondary task performance would be associated with low primary task workload or excess processing capacity, whereas low levels of secondary task performance would be indicative of high workload and little or no excess capacity. By requiring concurrent performance of two tasks, the secondary task paradigm seeks to ensure that total task demands will exceed operator information processing capacity, thereby avoiding the potential insensitivity of primary task measures when used alone.

The secondary task paradigm has been widely applied to measure operator workload. A number of reviews (e.g., O'Donnell and Eggemeier 1986, Eggemeier and Wilson 1991, Wierwille and Eggemeier 1993, Tsang and Wilson 1997) provide examples of the successful use of this approach in laboratory and applications environments, as well as references to guidelines (e.g., matching the processing demands of the primary and secondary tasks) that should be considered when implementing this type of procedure.

Some possible difficulties with the use of this approach in applied environments include the potential for the secondary task to be associated with degradations of primary task performance (so-called primary task intrusion), lack of operator acceptance due to perceived artificiality of some secondary tasks, and the instrumentation necessary to present and record the performance associated with the secondary task.

In order to overcome these potential difficulties, some researchers have suggested the use of so-called embedded secondary tasks which are functions (e.g., radio communications in the flight environment) that would normally be performed by an operator with a lower priority than the primary task within the course of system operation. Because they are part of the normal system procedures, embedded secondary tasks can be expected to reduce potential intrusion and operator acceptance problems.

3. SUBJECTIVE MEASURES

A second major approach to assessment of workload is the use of subjective reports to assess the amount of demand or processing load imposed on the operator by performance of a task or system function. Application of this type of procedure typically involves having the operator provide some subjective assessment of the workload associated with a task or system function following its performance. Subjective procedures have been very frequently employed to assess operator workload, and the rating scale is the most commonly used subjective report procedure.

Within the past 20 years, a number of subjective rating scales have been specifically developed to assess workload and have been extensively evaluated with respect to their ability to provide sensitive indexes of subjective workload in both laboratory and applications settings. These scales include the subjective workload assessment technique (SWAT), the NASA task load index (TLX), and a modified version of the Cooper-Harper scale (MCH).

Several reviews (e.g., O'Donnell and Eggemeier 1986, Eggemeier and Wilson 1991, Wierwille and Eggemeier 1993, Tsang and Wilson 1997) provide additional information pertaining to the nature and successful application of these and other subjective rating procedures, as well as discussions of some methodological factors (e.g., potential effects of context on ratings) that should be considered in their use. In general, subjective rating scales have proven to be sensitive measures of workload that minimize the risk of primary task intrusion, enjoy a high degree of operator acceptance, and require minimal instrumentation. As a consequence, this type of approach has proven extremely popular, particularly in applications environments.

A number of the most frequently used scales (e.g., SWAT, NASA-TLX) are multidimensional. Multidimensional scales afford not only an overall estimate of the amount of subjective workload associated with task performance, but ratings of individual dimensions of subjective workload (e.g., time pressure, mental effort) as well. Information pertaining to individual dimensions of subjective workload is of potential use in identifying or diagnosing the sources of high subjective workload in task performance.

4. PSYCHOPHYSIOLOGICAL MEASURES

A third approach to mental workload assessment is the use of physiological signals from the operator's body. Changes in cognitive activity, including mental workload, are associated with changes in several measures of body function. Common measures are heart rate, blink rate, respiration rate, and brain wave activity. Measures of body fluids, such as salivary cortisol, have also been used. Typically, electrophysiological measures are implemented which require the application of electrodes to the body surface and equipment to amplify and record the signals. Further data reduction is usually required, such as deriving the heart rate from the electrocardiographic signals. However, systems are available that automatically carry out this data reduction. Because the psychophysiological signals are continuously available, they can be used to derive uninterrupted records of operator state. This is often helpful when unexpected events occur since it is possible to go back and examine the data around the unexpected event.

For the most part, psychophysiological data collection does not intrude upon the operator's primary task. Operators readily adapt to wearing the electrodes and forget that they are in place. Small, lightweight, portable recording devices can be worn by operators for ambulatory recording on the job. Wilson and Eggemeier (1991), Kramer (1991) and Tsang and Wilson (1997) provide reviews on the use of psychophysiological measures to monitor mental workload.

4.1. Heart Rate

Heart rate is probably the most widely used of the psychophysiological measures for mental workload assessment. Much of the research has been conducted in aviation settings. In general, increases in heart rate are found with increased mental demands of the task. For example, heart rates associated with landing an aircraft are higher than those recorded during lower-workload cruise periods. Even differences in the gradient of approach to a landing are associated with differences in heart rate; steeper gradients produce higher heart rates than less steep approaches.

The pilot in command of a two-pilot cockpit exhibits a higher heart rate than the nonflying pilot. The higher heart rate will switch back and forth as pilot and copilot change roles during a flight. Heart rate recordings can be used as a debriefing tool during test and evaluation trials to identify periods of high workload. These periods of higher heart rate can be used to focus debriefing discussions. Heart rate measures have been used to assess the workloads associated with telemarketing and with driving automobiles, race boats, and buses.

In general, heart rate is a sensitive measure of mental workload but may not be diagnostic with regard to the source of the workload. For example, both takeoff and landing segments of flight show roughly equivalent increased heart rates, making it impractical to discriminate between them. Assessing the variability of the rhythm of the cardiac cycle has also been proposed as a measure of mental workload. The hypothesis is that higher levels of task demand cause a decrease in the magnitude of the cardiac rhythm variability. However, this is based on laboratory results using simple, single-task paradigms and application to real-world settings does not seem to support this notion.

4.2. Eye Blinks

Visual information is critical in the operation of most systems. Eye blinks effectively render us blind for a few tenths of a second. Operators decrease their blink rate in high-workload situations having a large visual component. This has the positive effect of permitting the operator to miss less of the crucial visual input required for task performance. Blinks are typically measured electrophysiologically or with the use of small cameras. Blinks are identified and their rate, length of closure, and amplitude are measured.

High-workload conditions such as automobile driving in congested areas, landing an aircraft, and seeking a target are associated with lowered blink rates and possibly shorter duration of eyelid closure. The pattern of blinking can also be used to monitor visual demands of a task. Slow, periodic blinking patterns are related to times of visual search, with inhibition of blinking when there is extremely high visual demand.

There are also situations where blink rates may not decrease in proportion to the visual demands of a task. For example, dur-

ing certain types of aircraft landing the pilots may shift their gaze from inside the cockpit to outside. Because there is a tendency to blink when the gaze is changed, the blink rate may be higher than expected. To interpret the data correctly, it is important to know the situation. Eye blinks are both sensitive and diagnostic: they are sensitive to different levels of workload and, generally speaking, they are related only to the visual demands of a task.

4.3. Electrical Brain Activity

Brain wave activity, the electroencephalogram (EEG), is also useful for monitoring mental workload. There are two ways to analyze EEG data. One decomposes the complex EEG waveform into its constituent frequency bands and quantifies the energy in each band. The standard EEG bands are typically used: delta, theta, alpha, beta, and gamma. Increased workload appears to be associated with decreased activity in the alpha band but increased activity in the theta band. Beta and gamma band activity may also increase under conditions of higher mental demand.

In situations where discrete stimuli are presented during a task, it is possible to derive event-related potentials (ERPs). These small brain signals are time-locked to the processing of information and require several repetitions of the stimuli in order to average out the background EEG noise. A large positive-going component of the ERP has been the focus of this research, the so-called P300. Most of this research has been accomplished in the laboratory but some has been performed in simulators and in flight. The use of ERPs is still experimental in nature.

Both forms of brain wave activity are prone to artifact contamination, biological and nonbiological. The biological artifacts include blink, muscle, speech, and movement contamination. Electrical interference from power sources is the main nonbiological artifact. However, reliable methods for reducing and correcting these artifacts have been developed and they increase the utility of EEG measures.

5. ANALYTICAL METHODS

Analytical methods are typically used during early design phases of system development. Since the system does not exist, the workload estimates are predictive in nature. These methods are valuable in that they permit the estimation of workload before the system is built, hopefully avoiding later costly modifications after the system is in the field. Since the system does not exist, subject matter experts are often used to evaluate the demands the system will place upon the operator. These techniques depend upon models of the human operator and how interactions between the operator and the system result in mental workload

in the operator. Model development is valuable because it requires a specific statement of the relationship between the operator and the relevant factors.

The simplest approach is time line analysis. The procedure is to estimate the time to perform each required task then estimate the total time available to complete all the required tasks. Since most situations are time limited, it is possible to see whether the total job can be accommodated in the available time. One problem with this approach is it assumes that only one thing can be accomplished at a time. Recent techniques allow for simultaneous processing of multiple tasks and include task analysis/workload (TAWL), time line analysis and prediction (TLAP) and workload index (W/INDEX). Tsang and Wilson (1997) review these methods and give examples of their application.

Since these techniques are fairly new, they have not been extensively tested and compared to one another. However, as human-operated systems become more and more complex, analytical procedures should enjoy greater use. The complex nature of new systems places high demands upon the human operator, and the retrofit costs are very high. Therefore, inclusion of workload considerations early in the design process can help to reduce total system and operational costs.

REFERENCES

EGGEMEIER, F.T. and WILSON, G.F., 1991, Performance-based and subjective assessment of workload in multi-task environments. In Damos (ed.), *Multiple-task Performance* (London: Taylor & Francis), pp. 217–78.

KRAMER, A.F., 1991, Physiological metrics of mental workload: a review of recent progress. In Damos (ed.) *Multiple-task Performance* (London: Taylor & Francis), pp. 279–328.

O'DONNELL, R.D and EGGEMEIER, F.T., 1986, Workload assessment methodology. In Boff, Kaufman, and Thomas (eds) *Handbook of Perception and Human Performance*, Vol. II, *Cognitive Processes and Performance* (New York: Wiley), pp. 42–1 to 42–49.

TSANG, P. and WILSON, G.F., 1997, Mental workload measurement and analysis. In Salvendy (ed.) *Handbook of Human Factors and Ergonomics* (New York: Wiley), pp. 417–49.

WIERWILLE, W.W. and EGGEMEIER, F.T., 1993, Recommendations for mental workload measurement in a test and evaluation environment. *Human Factors*, 35, 263–82.

WILSON, G.F and EGGEMEIER, F.T., 1991, Physiological measures of workload in multi-task environments. In Damos (ed.) *Multiple-task Performance* (London: Taylor & Francis), pp. 329–60.

Mental Workload: Theory, Measurement and Application

M. S. Young and N. A. Stanton

Department of Psychology, University of Southampton, Southampton
SO17 1BJ, UK

1. ABSTRACT

Mental workload is a pervasive concept throughout the ergonomics and human factors literature and it represents a topic of increasing importance. As modern technology in many working environments imposes more cognitive demands upon operators than physical demands, the understanding of how mental workload impinges on performance is critical. Practitioners should therefore know how to assess mental workload accurately and apply the results appropriately. This chapter provides a general overview of the current state of affairs regarding the understanding, measurement and application of mental workload in the design of complex systems.

2. WHAT IS MENTAL WORKLOAD?

There is no universally accepted definition of mental workload, although an analogy is often drawn with physical load. In this sense, mental workload can be broadly comprised of two components: stress (task demands) and strain (the resulting impact upon the individual).

Attentional resource theories form a useful basis for describing mental demand. These theories assume that individuals possess a finite attentional capacity which may be allocated to one or more tasks. Essentially, mental workload represents the proportion of resources available to meet the task demands. If demands begin to exceed capacity, the skilled operator either adjusts their strategy to compensate, or else performance degrades. Such a view makes clear predictions about mental workload in any given situation and observations of performance provide simple measures of mental workload.

Investing resources in a task is a voluntary and effortful process to meet demands, so performance can be maintained at the cost of individual strain or vice versa. This raises the notion of optimal state – the strain of underload or overload is caused by a mismatch between demands and capabilities. Excessive load can affect selective attention, leading to narrowed or inefficient sampling. Resource models of workload have their drawbacks. For instance, they do not consider nonattentional factors, such as experience.

An alternative perspective, then, takes into account the level of operator skill and the extent to which cognitive processing is automatic. Automatic processing is associated with expert performance and is characteristically fast, unconscious and free from attentional resource constraints. The converse is controlled processing and these two elements lie on a continuum. From this view, mental workload in real world tasks is determined by the balance of automatic and controlled processing involved. This is consistent with the attentional resources approach, as automatic

processing releases attentional resources for other tasks, with a resulting decrease in mental workload.

Thus we see that mental workload is a multi-dimensional concept, determined by characteristics of the task (e.g. demands, performance), or of the operator (e.g. skill, attention; Leplat, 1978). Collating the information obtained so far enables us to propose an operational definition of mental workload: "The mental workload of a task represents the level of attentional resources required to meet both objective and subjective performance criteria, which may be mediated by task demands, external support and past experience." In this definition, the level of attentional resources is assumed to have a finite capacity, beyond which any further increases in demand are manifest in performance degradation. Performance criteria can be imposed by external authorities, or may represent the internal goals of the individual. Examples of task demands are time pressure, or complexity. Support may be in the form of peer assistance or technological aids. Finally, past experience can influence mental workload via changes in skill or knowledge.

Although this definition is not exhaustive, it does represent a consensus of the research. Furthermore, it can be used to predict and explain changes in mental workload as a result of variations in demand, training, or other mediating variables.

There is an abundance of knowledge about what can increase mental workload. For instance, an increase in perceived work required to achieve a goal and decreased time available to complete a task, can both increase subjective workload. Furthermore, performance and workload are to some extent negatively correlated, such that performance failure is associated with higher perceived workload.

On the other hand, there seems to be less certainty about the factors which can lower mental workload. Performance feedback and reducing the number of decision options available, have both been associated with decreases in workload.

Some of these factors that affect mental workload may seem intuitively obvious, however this is not always the case. The particular problem of automation has attracted a lot of attention in recent years. It is paradoxical that simple automation can both reduce and increase mental workload. For instance, it has been observed that glass cockpits in commercial aircraft have relieved workload in some areas, such as reduced display clutter and more automated flight procedures. However, the same systems have increased workload in other areas, such as more decision options
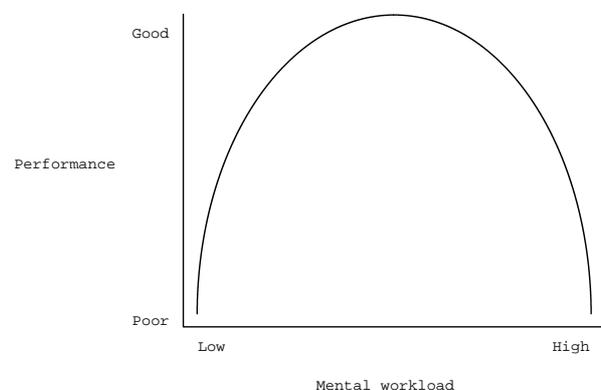
Figure 1. Relationship between mental workload and performance.

in a given situation and confusion about the operating mode. There is now a strong feeling that mental underload can be equally detrimental to performance as mental overload. Indeed, the current opinion is that there is an optimum level of mental workload which is associated with best performance (figure 1).

Therefore, simply reducing mental workload may not actually improve performance. To be able to apply this knowledge to the design of systems, it is first necessary to measure mental workload.

3. HOW IS MENTAL WORKLOAD MEASURED?

The measurement of mental workload is as diverse a topic as its theoretical counterpart, with many techniques available. Indeed, researchers in applied domains tend to favor the use of a battery of measures to assess workload, rather than any one measure. This use of multi-dimensional devices to assess workload seems sensible given our previous understanding of workload as a multi-dimensional concept. The main categories of workload measures are: primary task performance; secondary task performance; physiological measures; and subjective rating (see Meshkati, Hancock and Rahimi, 1990, for more detail on these techniques).

3.1. Primary and Secondary Task Measures

Performance measures on primary and secondary tasks are widely used in workload assessment. The basic premise is that a task with higher workload will be more difficult, resulting in degraded performance compared with a low workload task. Of course, though, following from attentional resource theories, an increase in difficulty (workload) may not lead to performance deficits if the increase is still within the capacity of the operator. Thus a secondary task, designed to compete for the same resources as the primary task, can be used as a measure of spare attentional capacity. In the secondary task technique, participants are instructed to maintain consistent performance on the primary task and to attempt the secondary task as and when their primary task demands allow them to. Differences in workload between primary tasks are then reflected in performance on the secondary task.

Although these measures have proven useful in many operational environments, there are mixed opinions as to their application. Secondary tasks have been criticized for only being effective in assessing gross changes in difficulty and are therefore considered quite insensitive. In cases where multiple primary task measures are accessible, these can be more sensitive than a single secondary task. Moreover, researchers using secondary task measures must be aware that the secondary task itself can be intrusive, particularly if the primary task imposes low workload, (Wierwille and Gutmann, 1978). However, a simplistic secondary task can be advantageous when primary task measurement may be unfeasible or uneconomical. Furthermore, secondary tasks have been demonstrated to assess different aspects of mental workload. Where the primary task is effective in measuring long periods of workload and performance, the secondary task is useful for quantifying short periods of workload and spare mental capacity. Finally, the use of primary and secondary task measures can be effective in discriminating individual differences in attentional resources, a particular problem in mental workload research.

Researchers wishing to use a secondary task are well advised to adopt discrete stimuli which occupy the same attentional resource pools as the primary task. For example, if the primary task is driving, this involves visual input, spatial processing and a manual response. Therefore, the secondary task should have the same characteristics. This ensures that the technique really is measuring spare capacity and not an alternative resource pool. Of course, this does raise the problem of intrusiveness, in which case one might consider an embedded secondary task. This is ostensibly a part of the primary task, however is not intrinsic to overall performance (e.g. providing readings from an instrument panel while driving).

3.2. Physiological Measures

Physiological measures are many and varied. For instance, various researchers have used respiration, heart rate, heart rate variability, electrodermal response, eye movements and pupillary responses as indices of mental effort. These measures offer advantages such as continuous monitoring of data, greater sensitivity and that they do not interfere with primary task performance (Fairclough, 1993). While many physiological measures have been reliably associated with mental effort, there are a number of disadvantages involved with this method of measuring workload. First, these measures are confounded easily, as they tend to be hypersensitive to extraneous noise from sources such as muscle movements and the rhythmicity of activity in the central and autonomous nervous systems. Also, there is a certain amount of physical obtrusiveness involved in using such bulky equipment. Therefore, it is generally recommended that physiological measures are only applied if unobtrusive and reliable and in conjunction with other measures of workload.

3.3. Subjective Measures

Many authors claim that the use of subjective reports may well be the only index of "true" mental workload. Subjective mental workload scores have been related to perceived difficulty and to the presence of automation. Again, such measures materialize in many forms, however these can be categorized as uni- and multi-dimensional. Uni-dimensional measures, such as the Cooper–Harper scale (1969), tend to be simpler to apply and analyze, but only offer a general workload score (Cooper and Harper, 1969). In contrast, multi-dimensional measures provide some diagnosticity towards the sources of mental workload, but are often more complex in their procedures. For example, the NASA Task Load Index (TLX; Hart and Staveland, 1988) and the Subjective Workload Assessment Technique (SWAT; Reid and Nygren, 1988) require participants to rate subjective workload on several scales. These scales combine into an overall workload score, therefore with the advantage of providing a simple rating of mental workload, in addition to the extra information available from the subscales.

Criticisms of subjective techniques are primarily concerned with the metacognitive abilities of the operator. That is, given the fact that the measures are necessarily administered post-task, one might question their reliability, particularly for long task durations. A recent attempt to develop an instantaneous (i.e. "on-line") measure of subjective mental workload met with limited success. Unfortunately, while it correlated with other measures of workload, it was disruptive to the primary task. It is recommended that researchers use established subjective mental workload measures, which have proven reliability across applications.

4. WHAT CAN DESIGNERS DO TO COPE WITH OPERATOR MENTAL WORKLOAD?

This section uses the information reviewed so far to examine what constitutes good and bad design from the perspective of mental workload.

Designers of complex systems need to recognize that operators are not inherently flawed and so should not attempt to simply remove as many tasks as possible from them (for instance, by implementing automation). This can have the effect of creating mental underload and leaves the operator with an incoherent set of tasks that remain unautomated. Humans obviously bring a valuable contribution to the task, or else they would not be there at all. Rather than task automation *per se*, the goal should be to design tasks and use available technology to exploit the unique skills and flexibility of human operators.

Ergonomics and human factors researchers generally agree that, while physical workload should be kept to a minimum, mental workload should be optimized. Human performance is at its best when task demands are harmonized to the capacities available. Although workload optimization is not easily achieved, a number of options have been discussed in the literature.

One possibility is the use of adaptive interfaces (e.g. Hancock and Verwey, 1997). These are essentially intelligent systems that can detect periods of elevated workload and assist the operator by presenting relevant information in a prioritized fashion. Adaptive systems typically use physiological measures to monitor workload and regulate the automation appropriately. For instance, a system in a car could integrate electrodes in the steering wheel to monitor electrodermal response. This could then be interpreted in terms of mental workload and used as input for the adaptive interface.

A related approach is to adopt dynamic allocation of function. Again, this monitors the task situation for peaks of workload and, in such cases, relieves the operator of some elements of the task. These tasks are returned to the human operator when demand returns to a more manageable level. The theory behind this is to maintain workload at a consistent level, rather than cycling through peaks and troughs which may degrade performance. Dynamic task allocation holds potential for improving performance at high levels of demand. However, questions remain concerning the criteria for and methods of, transferring control between the human and machine. Nonetheless, such human–system cooperation is a forward step, as performance does improve with technological assistance.

Finally, in some cases, mental workload can be reduced without affecting performance. Technological decision aids can relieve workload with no effect on task understanding and redundant task information can actually improve understanding and performance while decreasing workload.

In summary, it is possible to use technological interventions while keeping the human at the center of the design. An example already being tested is the electronic copilot, an adaptive system to assist pilots of military high-speed aircraft. The essential features when designing automated systems are feedback, assistance and optimization. Feedback is necessary to maintain communication between the human and the machine. Assistance for operators is a preferable alternative to relieving them of their tasks outright. People are surprisingly good at a lot of tasks, only sometimes they do need a little help. Finally and perhaps most importantly given the discussions here, mental workload optimization is crucial to maintaining effective task performance. Such optimization inevitably involves a balancing act between demands and resources of both task and operator. This view, which essentially captures the spirit of ergonomics, offers a way forward for designers concerned with the problem of managing mental workload in technological systems.

REFERENCES

COOPER, G. E. and HARPER, R. P., 1969, *The Use of Pilot Rating in the Evaluation of Aircraft Handling.* Report no. ASD–TR–76–19 (Moffett Field: NASA).

FAIRCLOUGH, S., 1993, Psychophysiological measures of workload and stress. In A. M. PARKES and S. FRANZEN (eds) *Driving Future Vehicles* (London: Taylor & Francis), pp. 390–9.

HANCOCK, P.A. and VERWEY, W.B., 1997, Fatigue, workload and adaptive driver systems. *Accident Analysis and Prevention*, 29, 495–506.

HART, S.G. and STAVELAND, L.E., 1988, Development of NASA-TLX (Task Load Index): results of empirical and theoretical research. In P. A. Hancock and N. Meshkati (*eds*) *Human Mental Workload* (Amsterdam: North-Holland), pp. 139–83.

LEPLAT, J., 1978, Factors determining workload. *Ergonomics*, 21, 143–9.

MESHKATI, N., HANCOCK, P.A. and RAHIMI, M., 1990, Techniques in mental workload assessment. In J. R. Wilson and E. N. Corlett (eds) *Evaluation of Human Work: A Practical Ergonomics Methodology* (London: Taylor & Francis), pp. 605–27.

REID, G.B. and NYGREN, T.E., 1988, The subjective workload assessment technique: a scaling procedure for measuring mental workload. In P. A. Hancock and N. Meshkati (eds) *Human Mental Workload* (Amsterdam: North-Holland), pp. 185–218.

WIERWILLE, W.W. and GUTMANN, J.C., 1978, Comparison of primary and secondary task measures as a function of simulated vehicle dynamics and driving conditions. *Human Factors*, 20, 233–44.

Models of Human Performance for Application to Military Simulation

R. W. Pew, PhD

BBN Technologies, Cambridge, MA 02138, USA

1. INTRODUCTION

Modeling and simulation provide tools that are being widely used to support system concept development and evaluation, engineering design, training, operations concept and procedure development, and system test and evaluation. (See R. Laughery section on human performance modeling for a general discussion of how human performance models are created and used.) This section will discuss some special classes of models and simulations that are particularly applicable to large-scale system design, such as military system development and military or civilian air traffic control. How human performance models play a role in these kinds of simulations is of particular interest.

The military makes extensive use of modeling and simulation in many of the same ways it is used in research and development and in business and industry; however, there are two categories of simulation that are more specific to the military. These are wargaming, sometimes called constructive simulations, and distributed interactive simulation, sometimes referred to as distributed virtual simulation.

Wargaming refers to simulations that represent the actions of friendly and enemy forces and predict battle outcomes. Originally they were played out by hand on a large game-board representing the military terrain. Today, in the most sophisticated versions, the actions of multi-service forces are represented in computer simulations, in which models of individual military units attack and defend according to a plan, engage the enemy, move across the battle area, while the equipment and personnel losses on each side are calculated using probabilistic models. The US Army JANUS model and the Navy Simulation System (NSS) model are examples of this approach. The models do not necessarily run in real time (they may run faster or slower than real time) and they typically run to completion once a simulation has been initiated. Often multiple runs are executed in order to obtain statistically reliable probabilistic outcomes.

Usually, in constructive models, the human participants, if they are represented at all, are programmed to follow the plan rigidly and there is no provision for representation of individual decision-making or variability in human performance, other than probabilistic models of detection of the enemy and of kill-probability. However, as the introduction of information technology changes, the nature of military engagements from primarily force on force in which the greatest firepower wins, to greater emphasis on information warfare and the importance of effective command and control, there is a recognized need to add representations of human decision-making and human fallibility in the execution of plans, such as the effects of limited training, fatigue, workload, or battlefield stress to these constructive simulations. This is where human performance models come into play.

Distributed interactive simulation (DIS) refers to a relatively new approach to simulation in which simulation is used to create a battlefield environment where human participants execute the engagements just as they would in a real battle. The earliest realization of this approach, developed in the early 1980s, was called SIMNET. In SIMNET multiple mock-ups of the inside of Army tanks were created. Each included computer-simulated workplaces for all of the four crewmembers of a real M-1 tank. Each tank's battlefield environment was simulated using computer graphics. When the commander of a particular tank looked out of his viewport, he saw the terrain, significant landmarks, and the other friendly and enemy tanks within visible range in their own battlefield positions just as he would in a real battle. This was accomplished by connecting all the tank simulators in a computer network and periodically broadcasting data on each tank's position, orientation and significant events, such as firing a round or being fired upon. Every other tank simulator within visual range would process those data and recalculate the state of its own visual environment accordingly. Initially, approximately 80 such tank simulators participated in exercises, some representing friendly and some enemy forces. Later helicopter and Air Force fighter simulations were added to the mix so that the tank forces could be augmented by close air support. Because the data communication can take place over a wide-area computer network, these units do not even have to be geographically co-located. In one demonstration the friendly tank forces were located at Fort Knox, KY, the enemy forces in Cambridge, MA, and the helicopter units at Fort Rucker, AL.

The current state-of-the-art in fielded distributed interactive simulation is represented by the Close Combat Tactical Trainer (CCTT). It is a family of virtual simulations currently being deployed by the Army. It is organized into unit formations of battalion task forces equipped with M1 tank simulators, M2 Bradley infantry fighting vehicle simulators and supported by AH64 attack helicopter simulators. It supports brigade-level operations with high task fidelity down to the crew and squad levels.

Originally distributed interactive simulation (DIS) was developed for training applications, but it has been so successful that it is now forecast to be used for concept development training, doctrine and tactics development, mission rehearsal and in system evaluation. Although it began as an Army project, now all the services are developing this kind of simulation for their own missions and there are R&D level explorations of Joint and Combined Arms training capabilities. Joint training refers to the participation of multiple services. Combined Arms training includes the forces of multiple countries, such as the NATO forces. There is now also considerable interest in combining distributed interactive simulations with constructive simulations so that some aspects of a mission are carried out without human intervention and others rely on actual officer and enlisted participation. For this to work requires temporal synchronization of the two kinds of representation since the DIS components must, by definition, operate in real time.

2. THE ROLE FOR HUMAN PERFORMANCE MODELS

The need for modeling of human performance arises in a different

way in DIS. In order to reduce the number of individuals required to participate in training exercises, the concept of a Semi-Automated Force (SAF) was invented. The notion was that a single participant could control the action of a platoon of enemy tanks so that each unit would not have to be operated by a full crew. Using simple concepts of artificial intelligence, the platoon could be programmed to transit in single file, transition to firing positions, and engage the friendly forces under the control of a unit commander who only had to dial in the objective, the way points, and the type of formation. This level of automation has been achieved, but now there is great interest in extending the concept to more fully automated forces, to include both enemy and collateral friendly forces, and to extend the concept to individual soldiers (referred to as dismounted infantry) on the one hand and to aircraft support missions on the other. To accomplish this requires models that reflect the behavior of individual soldiers, vehicle crews, unit command, control and communication, planning, decision-making, battle execution and, ultimately, that reflect the behavior of larger organizations. Based on current capabilities for modeling human performance, these are tall orders indeed.

It is easy to underestimate just how challenging this is. At the dismounted infantry level, human soldiers take account of what they hear, see, and feel on the battlefield. They use their memory of the results of visual and auditory scanning to produce an integrated view of what is happening around them, sometimes referred to as situation awareness. They make decisions about where to go and what to do next on the basis of this awareness together with what their team leaders tell them in person or over the various communication channels. Models of behavior at this level of complexity are just emerging from the laboratory. Command and control involves real-time planning, replanning, and monitoring of the execution of plans. A commander issues guidance to his or her staff; the staff defines the mission and decomposes it into specified and implied tasks; they identify resources, including personnel, equipment, and supplies, enumerate the logistics to get them to the battlefield and specify the timing of the operational tasks. Models of these kinds of human activity exist as isolated capabilities, but no one has achieved even an approximate model of human behavior integrated at this level of organizational complexity.

Most models of individual or small-team behavior are currently built to represent task execution. Task network models, described in another section of this encyclopedia, simulate the execution of tasks, but typically only produce the resultant measures of performance rather than the task-execution behavior itself. To pick a military example, if the simulation assessed the movement-to-contact of an armored platoon, the outcome of the model might be the time to complete the action, the attrition in forces and the level of supplies used. In order for models to participate in an interactive simulation they must also produce, as outputs, a trace of the physical movement of the battle entities, the individual units that are lost and the final positions of each unit, and they must generate these outputs in real time.

Most of the models used today in DIS are production-rule programs that implement if-then behavioral task-level rules. These programs test to see if a pre-specified set of conditions is present, and if it is, then trigger a sequence of actions. The firing of each subsequent rule is dependent on new sets of conditions. For example, just hypothetically, a firing routine might be specified as: If an enemy target is in visual range and if it is within the field of view, then with probability, p, it is detected. If a target is detected and it is identified as an enemy tank and it is moving toward the friendly positions, assign it a targeting priority of q. If such a tank is identified and no other target has been detected having a higher priority, then execute the firing sequence. The computer architecture that is most commonly used currently for implementing these rule sets in terms of a "finite-state machine" is called ModSAF, standing for Modular Semi-Automated Forces. See the website, http://www.mystech.com/-smithr/elecsim94/modsaf/modsaf.txt for a more detailed description of ModSAF.

At the research and development level, the state-of-the-art is represented by the Intelligent Forces Models created using the Soar architecture to represent the combat behavior of fixed- and rotary-wing pilots in combat and reconnaissance missions. The FWA-Soar simulated a fleet of up to 500 fixed-wing aircraft in a context that includes such elements as enemy aircraft, surface-to-air missiles, and enemy radar installations. It was used as the primary source of air missions for STOW-97 (Synthetic Theater of War-97) a demonstration of the capability to represent a full theater of war synthetically. See http://ai.eecs.umich.edu/ifor/stow-review.html for a fuller description of FWA-Soar.

3. THE PATH TO IMPROVED HUMAN PERFORMANCE MODELS

In 1998 the National Research Council of the US National Academy of Sciences published a book-length report, *Modeling Human and Organizational Behavior: Application to Military Simulations*, that summarizes the state-of-the-art with respect to this kind of human-performance modeling. The report recommends that, in order to achieve the kind of realism goals desired by the military sponsors, much greater attention must be paid to the psychological content of the models which are used for these purposes.

It suggests five broad areas of experimental psychology research and modeling that are especially relevant to improved human behavior representation. The areas of *attention and multitasking* refer to the human ability to focus on one thing at a time but to have multiple threads of activity going on simultaneously. *Memory and learning* are fundamental to functioning in an interactive environment. *Decision-making* is a central component of military command and control. *Situation awareness* refers to sensing and interpreting the state of the operational environment. And finally, *planning* is concerned with generating detailed preparations for future actions and anticipated events.

The report summarizes ten specific architectures that have been developed specifically for building integrative human performance models having the kind of scope necessary to be applied to military simulation problems of the kind addressed here. The Soar modeling environment, referenced above is an example such an architecture.

4. MODELING BEHAVIORAL STRESS

There is also considerable interest in modeling the effects of various kinds of stressors on human performance. A stressor, sometimes called a moderator variable, is any environmental or

internal variable that changes the way behavior is carried out. For example, when soldiers are subjected to excessive workload, they may do all the tasks required of them incompletely, they may skip some tasks, or they may just queue up pending tasks and accomplish them when they have more time. The models should reflect these different effects of workload pressure. When a soldier must continue to work without enough sleep, it is likely that there will be momentary lapses of attention during which changes in the environment will not be noticed and these lapses will effect the resulting behavior. There is also interest in accounting for emotional stresses of the battlefield, but we are only beginning to understand how they will effect behavior and therefore how to model them.

5. A SUGGESTED METHODOLOGY

Experience has shown that building models of human performance is extremely challenging and requires interdisciplinary teams. Computer scientists contribute the modeling expertise, behavioral scientists contribute knowledge of the psychological literature and the understanding of the vagaries of human performance, and military subject-matter experts contribute knowledge of military doctrine and tactics for specific application domains. Because human behavior is so context-dependent, success at modeling is, at this point, most successful in limited, well-defined domains, such as executing specific types of maneuvers on well-documented terrain following specifically prescribed doctrine.

Although behavioral theory can play an important role in good models, the process should begin with intensive knowledge acquisition to define and understand all the activities and tasks that are required of the entities being modeled. Then the team needs to examine the data requirements associated with those activities and tasks. What data are used when carrying out each task and what are their sources? What outputs are produced and where do they go? How long does each task take? What are the possible mistakes and errors that are likely to occur and how do they come about?

To develop models with psychological content also requires going deeper into the processes and preparing what is sometimes called a cognitive task analysis. This kind of analysis investigates the perceptual cues that are available to support situation awareness and decision making required in carrying out each task, as well as the memory requirements and movement control associated with successful performance. Only after these kinds of information are available is it possible to begin defining the models themselves.

The most practical approach to actually building the models is to create successively more detailed approximations to the final desired outcome. One starts with the simplest basic structure of a task, and then iteratively increases the fidelity with which the task is iteratively carried out until the desired level of realism is achieved. At each stage of development the models should be compared to real human performance in order to understand the model's limitations and to focus the effort at the next stage. When the model is ready for application it should be thoroughly validated by running formal comparisons between the behavior of the model and the behavior of humans accomplishing the same activities and tasks. This is one of the most challenging phases of model development and is the one that is most often neglected. There are already too many human performance models being used that are inaccurate and poorly understood because of failures to carry out adequate validation studies.

6. SUMMARY

Modeling human performance is of growing interest in the human factors and ergonomics community in general, and there are many potential applications in the military services. Current capabilities to achieve executable models that simulate the responses of humans, rather than simply producing performance measures as outcomes, are very limited, both in scope and in robustness in the face of modest changes in context. It will be important to incorporate greater understanding of psychological theory and to undertake model development with the participation of interdisciplinary team members for further progress to be achieved.

7. FURTHER READING

The reader interested in greater detail of military human performance models is referred to Pew, R.W. and Mavor, A.S. (eds), 1998, *Modeling Human and Organizational Behavior: Application to Military Simulations*, (Washington DC: National Academy Press) from which this section was derived.

Monitoring an Operator's Psychophysiological Fitness for Work

A. Y. Burov

Center of Ecology, Ergonomics and Safety, ELAN Ltd, PO Box 3, 04214, Kyiv, Ukraine

1. INTRODUCTION

In modern times manual work has been more and more replaced by mental one. In this respect it is critical not only to design operators' work, but also to re-organize the particular operator's working process itself. This reorganization is based on comparing the general professional requirements with an operator's individual characteristics (physical, psychophysiological, social) as well as his or her capacity for adaptation. Adaptation of individual characteristics and general requirements affects the dynamics of an operator's work functional structure during his or her lifetime. An ergonomic approach to improving human interaction with the work tools and means, as well as with the environment, is to provide mutual adaptation of a human working functional structure and the work environment. The mechanisms of this adaptation may differ depending on the conditions, location and time of the operators' work, i.e. they depend not only on the interaction "human–machine–environment", but also on the dynamics of this interaction throughout the whole work cycle.

The analysis of safety and efficiency of enterprises with complex technological cycles shows that a human (an operator) is the least reliable link in the human–technology system. His or her fitness for work depends on four groups of factors:

1. Professional expertise.
2. Psychophysiological professionally important qualities (PPIQ).
3. Functional state.
4. Parameters of the environment.

These factors determine the output operator's parameters that must meet work condition requirements both in values and over time. Only these parameters of human fitness for work could ensure the proper efficiency level of the human–technology system. The efficiency of mental work to a great extent depends on a person's ability to manage varying information flows. Given the rate of information growth and the span of human life, it becomes essential to maintain human psychophysiological condition at the proper level in case it is impossible or unreasonable to replace him or her by another person.

The objective was to develop the methodology and principles of the psychophysiological monitoring of the operators' fitness for work and to develop a psychophysiological monitoring system to assess the psychophysiological state of the operators, to predict their efficiency.

2. METHODOLOGICAL BASIS OF OPERATORS' FITNESS FOR WORK

While reviewing the role of a human operator as an element of the human–technological system, human factor or ergonomics approaches to the evaluation of the operator's fitness for work have typically assumed that the functional characteristics of an operator (e.g. expertise, information processing capacity, etc.) are invariant. Presumably, these are not influenced by such variables as operators' mood swings, sleep history, work time expenditure, etc.

Some PPIQ could change and reduce below the proper level. Both theoretical analysis and the analysis of actual data show that the range of psychophysiological parameters may vary greatly within several days, hours and shorter time intervals for both exogenous and endogenous reasons, which may result in the decreasing of an operator's professional fitness level below the critical one. A pre-shift medical control is able to reveal only rough, considerable deviations in a human condition ("permission check"), but it does not reveal such functional deviations at an early state due to insignificant changes in amplitudes and the impossibility of predicting stable durability parameters over time.

Therefore, it is evident that the tests and techniques used so far do not reflect physiological maintenance of the current activity. At the same time, the vegetative maintenance of mental activity is connected with the operation of the whole hierarchy of regulatory physiological systems and at times become apparent due to significant shifts in some physiological parameters.

In this respect, there seems to be a need to use objective physiological parameters as evaluative criteria of a functional state and mental fitness for work. However, the research into these parameters is quite hard to conduct in real-life conditions. Moreover, when it is necessary to project human fitness for work, most physiological parameters cannot ensure the required precision because of their short-time forecasting basis. While the requirements for professional expertise could be defined for various kinds of manual and mental work, the evaluation of functional condition over time is much harder to make, for it depends not only on the psychophysiological parameters, but also on the changes in an operator's state, especially during different periods (year-by-year variations, day-by-day variations, work-schedule and shift variations, etc.).

The physiological mechanisms of the adaptation of an operator's activity functional system to the industrial environment may vary depending on the conditions, location and time of human activity. Hence, ergonomics may contribute to improving human working conditions if it takes into account not only macro-ergonomic requirements (spatial aspect of human interaction with the industrial environment), but also ergodynamic ones (temporal aspect). Direct or indirect measurement of psychophysiological changes makes it possible to review a functional state of the operator structure (or a moment-by-moment condition of an operator), which helps to predict an operator's individual fitness and reliability for effective work.

3. PSYCHOPHYSIOLOGICAL BACKGROUND

Selecting appropriate physiological parameters to define an operator's functional state is of utmost importance. Our data clearly show that the dynamics of mental activity, especially its biorhythmic structure (Burov 1986, 1991), plays a leading role in the analysis of professional fitness and the maintenance of the operators' functional state. When mental activity as the primary output of the operator psychophysiological condition is examined,

it becomes evident that we are able to predict relationships between the parameters of mental activity and those of physiological maintenance (Alajalova *et al.* 1975). Prior researches (Burov 1986) indicate that this relationship can be most clearly seen in rhythmical characteristics of these processes rather than in stable baselines or a tonic level of responding. Rhythmic characteristics stand for temporal rhythms such as the well-known circadian rhythm and estrous cycle rhythm, as well as other rhythms that are less frequently referred to.

A breakdown in these rhythms reflects the strains associated with adjusting to demanding activities and stressors. Hence, they may be used to evaluate a psychophysiological regulation of activity and its impact on operational efficiency, and possibly to predict an ability to function effectively. Our research is aimed at developing psychophysiological monitoring systems to assess human psychophysiological states and to predict an operator's functional efficiency. We have started to evaluate its validity during the performance of monotonous tasks under laboratory conditions as well as in applied environments (Burov 1986). The laboratory findings clearly support the idea that variations in the relationship between oscillatory components of task performance measures and physiological measures provide a useful index of operators' efficiency ($R = 0.85–0.93$, $p < 0.01$). Applied research employing the system in fossil and hydroelectric power plants as well as that of dispatchers has evaluated the validity of the predictions concerning a moment-by-moment efficiency of an operator by correlating the prediction indices with the expert rating of an operator performance as well as with operational output measures (e.g. fuel usage) during a 1-year period among 73 power plant operators. The findings show that the approach is useful in predicting operators' efficiency.

4. PRINCIPLES OF DEVELOPING THE SYSTEM OF MONITORING OF OPERATORS' FITNESS FOR WORK

The approach is based on the following principles (Burov and Chetvernya 1996):

- Test techniques used to evaluate psychophysiological parameters of an operator's fitness for work should be aimed at evaluating not the absolute levels, but the time structure of task performance.
- Information concerning the human condition should characterize him or her as a complex system.
- An adequate adaptive model should describe parameters of psychophysiological maintenance and their sets.
- Information models should describe regional, professional, group and personal characteristics of operators under survey.
- Psychophysiological test procedure control should be exercised with the data collected in the system database with adaptive algorithm.

The simulation of an operator's work in laboratory conditions enables a psychological test to be developed that is most adequate for investigating operators' work types. The test is an inspection procedure as well as information selection about the human condition and his or her fitness for work: the reliability, rate and rhythm of performance. These parameters reflect the pressure of regulative systems and physiological "cost" of activity, on the one hand, and the efficiency of the professional operators' work, on the other hand (Burov 1986, 1991).

A psychophysiological check of an operator's ability to process information is performed as the same type of cognitive tasks performance under computer control. According to the performance of all tasks, parameters of rate, dynamics and biorhythmical structure are calculated. Among these numbers, a subset of parameters that are the most informative ones in relation to human professional durability are chosen.

Our research shows that formally to describe a human work efficiency in conditions of simulating an operator's activity, the statistical models in classes of multiple regression models of 4–6 degrees with volumes of training samples of 24 supervisions are the most exact and reliable ones.

5. APPLICATION

Despite all the possible measures to make a professional selection, from 40 to 60% of accidents arise because of personnel problems. According to our data, natural fluctuations of the human functional state account for failures in cases which do not enable an operator to realize his/her qualifications completely. The approach discussed is based on using an individual, i.e. inherent in each particular person, temporal organization of mental activity in systems for a pre-shift (intrashift) check of operators' fitness for work. Thus the prediction model is not fixed: instead of a specific model, the algorithm of constructing adaptive models is used.

The system to project operators' efficiency is supposed to open up the following opportunities:

- To collect indirect information on the human psychophysiological state.
- To investigate the dynamics of the human psychophysiological state during his or her professional career.
- To give human information which could be used to make predictions of his/her professional efficiency and improve it in a certain way.

In practical terms, the use of such systems ensures psychophysiological support during the whole operator's professional life. Generally, such systems are employed to provide an organizational support (task performance, projection, recommendations), information support (condition evaluation and projection of fitness for work) and technical support (software). Information on a human's psychophysiological conditions should be preserved during his or her lifetime, so that psychophysiological parameters do not remain constant and exceed the critical limits for different people in various moments (stages).

An adequate psychophysiological model of an operator's work allows for 3–4 min of work to evaluate a human functional state and to predict his or her professional endurance for the coming work shift. Such a model of activity is used in systems of operators' pre-shift checks at energy enterprises that provide a prediction accuracy of 85–90 %.

REFERENCES

ALAJALOVA, N.A., LEONOVA, N.A. and RUSALOV, V.M. 1975, About parity of features of aim formation at accomplishment of automated mental operations and of super slow potentials of brain. *Human Physiology*, 5.

BUROV, A. 1986, Valuation of functional state of operators' on parameters of mental serviceability. *Human Physiology*, 2.

BUROV, A. 1989, Biorhythmic aspects of the duty efficiency of generating units operators. *Energetics and Electrification*, 1, 27.

BUROV, A. 1991, Psychophysiological correlates of operators fitness for work at power plants. *Energetics and Electrification*, 2.

BUROV, A. and CHETVERNYA, YU. 1996, Methodological principles of psychophysiological forecasting of operators' reliability. In *Advances in Occupational Ergonomics and Safety I. Proceedings of the XIth Annual International Occupational Ergonomics and Safety Conference*, 8–11 July 1996, Zurich. International Society for Occupational Ergonomics and Safety, Ohio, vol. 2.

Noise: Measuring, Evaluation, and Rating in Ergonomics

H. Strasser and H. Irle

Ergonomics Division, University of Siegen, Paul-Bonatz-Str. 9–11, 57068 Siegen, Germany

1. INTRODUCTION

Both ergonomists and practitioners responsible for occupational health and safety in a company normally use and appreciate indices of workload and environmental exposures presented in the simplest possible figures and numbers. Therefore, in traditional standards, rules and safety regulations, the physical environment is normally rated in 8-h-based mean values via connecting intensity and duration of stress by means of a multiplication, i.e. a mutual settlement of high load within a short exposure time and a low stress height within a longer lasting exposure. This principle is well-based on the experience that a low workload can be tolerated for a longer duration than a high workload. But does this confirm the hypothesis that equal energy or dose, or equal demanded output, also involves equal short or long-term human responses? It will be shown that standards and conventional guidelines for occupational health and safety are more closely related to physics than to physiology. Yet, to really protect man at work, ergonomics must be much more concerned with physiological costs of work and environmental stress than with physical principles of equal energetic dose.

2. GRADUAL ASSESSMENT OF INTENSITY, FREQUENCY AND EXPOSURE TIME OF NOISE

As shown in the upper row of Figure 1, the intensity of sound events has always been quantified in decibels by the sound pressure level in a logarithmic scale. Of course, that is a pragmatic scale because a tremendous span of > 6 decimal powers of sound pressure and even 12 decimal powers of sound intensity can be condensed into easily manageable values of only 3 digits (e.g. 0–120 dB).

However, scientists and practitioners nowadays still have to work with this scale, despite the somewhat paradoxical fact that the psychophysical basic law of Weber–Fechner has meanwhile proven to be incorrect for acoustic stimuli. Although the formula for the sound pressure level is due to Weber–Fechner's law as outlined in the upper part of Figure 2, the resulting logarithmic scale is not in accordance with human sensation. For instance, a 3 dB difference means that two intensities vary by the factor 2, and a variation of 10 or 20 dB, e.g. from 80 or 90 to 100 dB, is associated with the 10- or even 100-fold in energy; a 10 dB increase or decrease results in a doubling or halving of the loudness for the human sensation, however.

Also, a level of 100 dB can never be assessed as the double value of 50 dB. Therefore, instead of the incompatible logarithmic scale, a scale of loudness with linear units in Sone due to sensation derived from Stevens's law of power should be used (see middle part of Figure 2). The above-mentioned increase of 10 dB, e.g. from 90 to 100 dB, is then represented by the doubling of 32 to 64 Sone.

If we were to make our money transactions utilizing the traditional logarithmic scale, we would, no doubt, handle the decibels a little bit more cautiously than is sometimes done in practice. Provided that: 0 dB corresponds with $1, 30 dB would be equivalent to $1000 and 60 dB would mean that we would already be millionaires. But also trillions or quadrillions (Figure 3) in national debt expressed in the small figures 120 or 150 dB
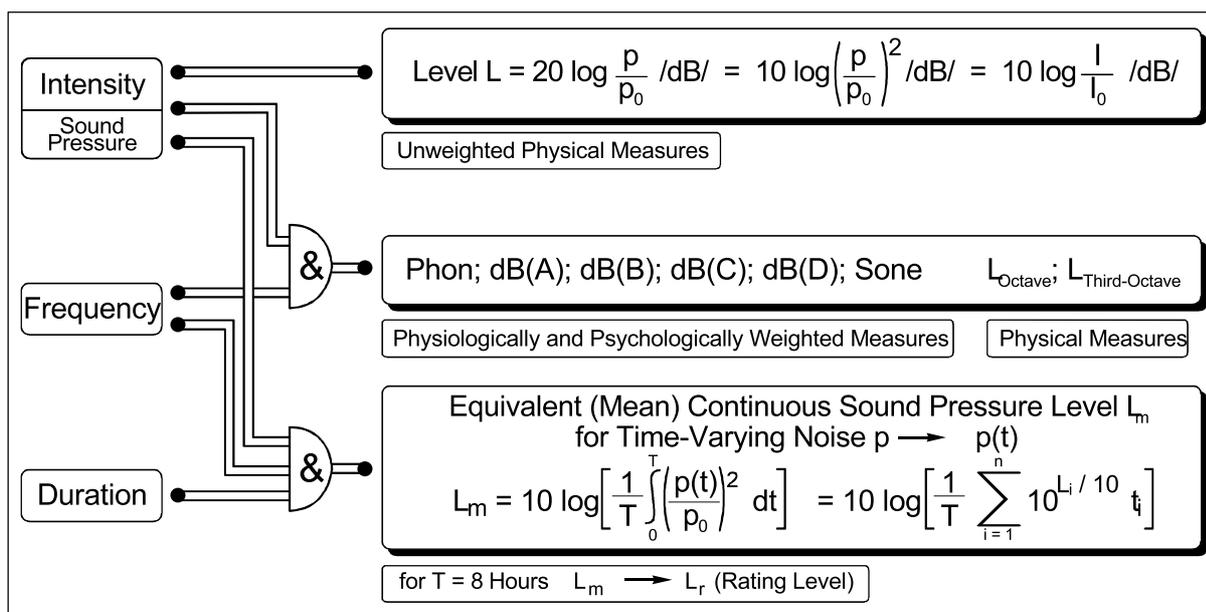
Figure 1. Gradual assessment of the physical dimensions intensity, frequency and exposure time of noise for the development of an integral characteristic value.
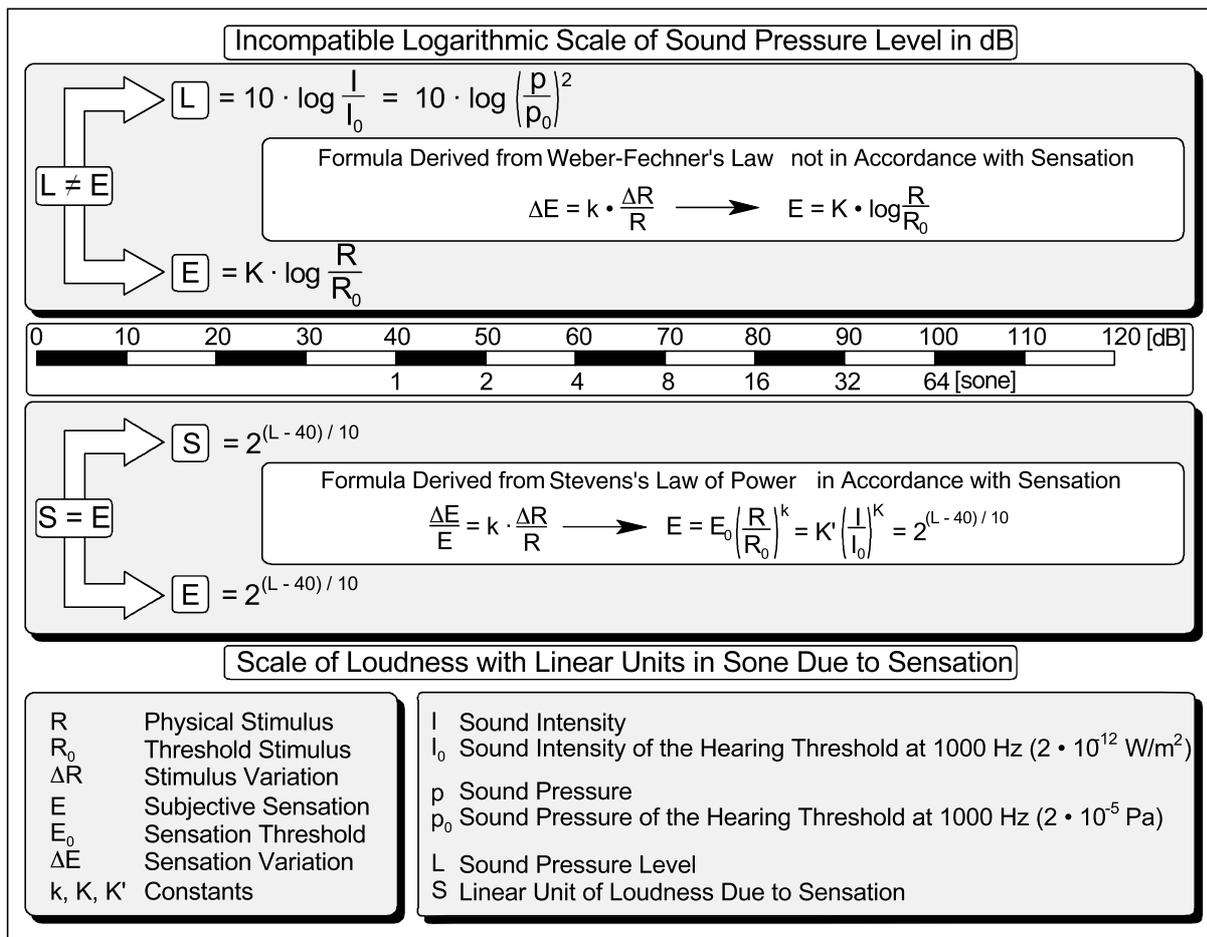
Figure 2. Incompatible logarithmic scale and scale of loudness according to Stevens (1957).

would seem to be not that tremendously much more than the money that "have-nots" have in their pockets.

With the intention of specifying sound immission with regard to intensity and frequency in one single value — as shown in the middle row of Figure 1 — frequency-dependent filters A, B, C or D should take into account the physiological characteristics of hearing.
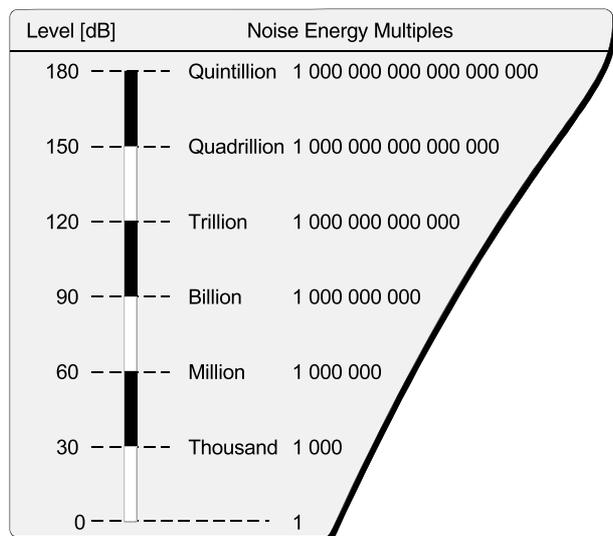
Figure 3. Noise energy multiples.

The filters A–D, however, as a reciprocal approximation of the phon curves in different volume ranges (Figure 4), are based
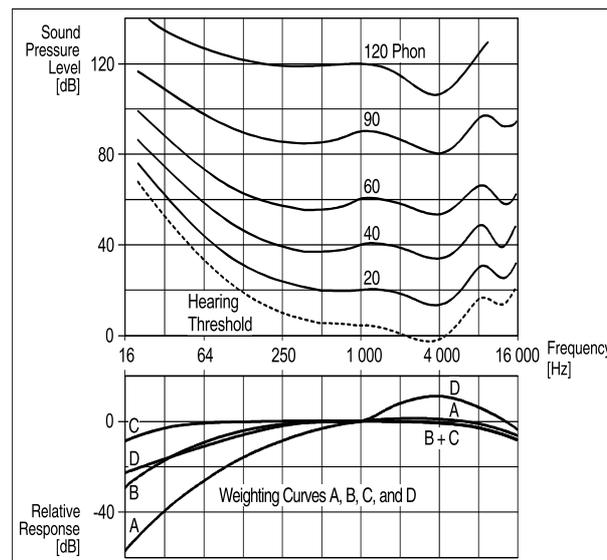
Figure 4. Curves of equal subjective sound level intensity (Phon) in the audible frequency range and dynamic area (upper) and frequency–response characteristics of the weighting networks A–D (lower).

517

on the subjective comparison of sequentially presented tones and, therefore, cannot lead to an adequate assessment of noise, which is normally a mixture of inharmonious sounds. Furthermore, in most cases today, only the A-weighting network is used for all volume ranges, although doing so conflicts with scientific knowledge. This discrepancy sometimes leads to the fact that, to the disadvantage of man, sound pressure levels of some noise sources do not represent the real sensations of man. A paradoxical contradiction sometimes results which can be expressed in the statement: lower dB(A) but higher loudness. This applies, for example, to some types of motorcycles, because the A-filter — which is strictly speaking only applicable to a dynamic range of up to 60 dB — reduces the sound pressure levels more than allowed when it is used to evaluate higher noise levels.

For noise control in particular, it is beneficial to obtain knowledge about the frequency ranges where noise is emitted from a source in order to select suitable protection measures. Therefore, frequency analysis or at least octave or third-octave band measurements of noise are more useful than the application of the weighting curves A, B, C, or D.

Sound pressure levels mentioned in ergonomics and in all legal regulations, standards, and prevention instructions (e.g. NN 1990, 1996, 1998, ISO DIS 1999) do not refer to a momentary sound event; they normally refer to the rating level L_r as an average value for the noise exposure associated with an 8-h working day.

3. RATING LEVEL AS AN INTEGRAL VALUE OF NOISE AND ITS DIVERGENT EFFECTS ON MAN

The energy equivalent calculation of the mean value — as shown in the lower row of Figure 1 — is, of course, applicable to a great many working situations. However, situations also exist where a purely formal calculation yields peculiar results that lead to a serious misinterpretation. This will be visualized and explained by the Figures 5 and 6.

When applying energy equivalence, 85 dB for 8 h are equivalent to 88 dB for 4 h, 91 dB/2 h or 94 dB/1 h.

This mutual settlement of noise level and exposure time (Figure 5) is correct as far as sound dose and sound energy are concerned. However, with regard to physiological and psychological aspects of work, inevitably some discrepancies result.

Ninety-four dB/1 h — as in the previous case — is energetically equivalent to 85 dB/8 h, i.e. it correspond to an $L_r =$ 85 dB. If only the energy, i.e. the sound dose, is considered, what is shown in the upper part of Figure 6 also holds true. In this case, 94 dB for 1 h and an additional 75 dB for the remaining 7 h also result in an $L_r =$ 85 dB. Physically seen, this is correct, but it is comparable with the situation of filling up quiet periods with noise, and from a psychological point of view it is likely that nobody would prefer a situation as described in the upper part of Figure 6. Provided that the noise distribution shown here would stem from two machines, strange effects would also result with respect to technical approaches of noise control. If an
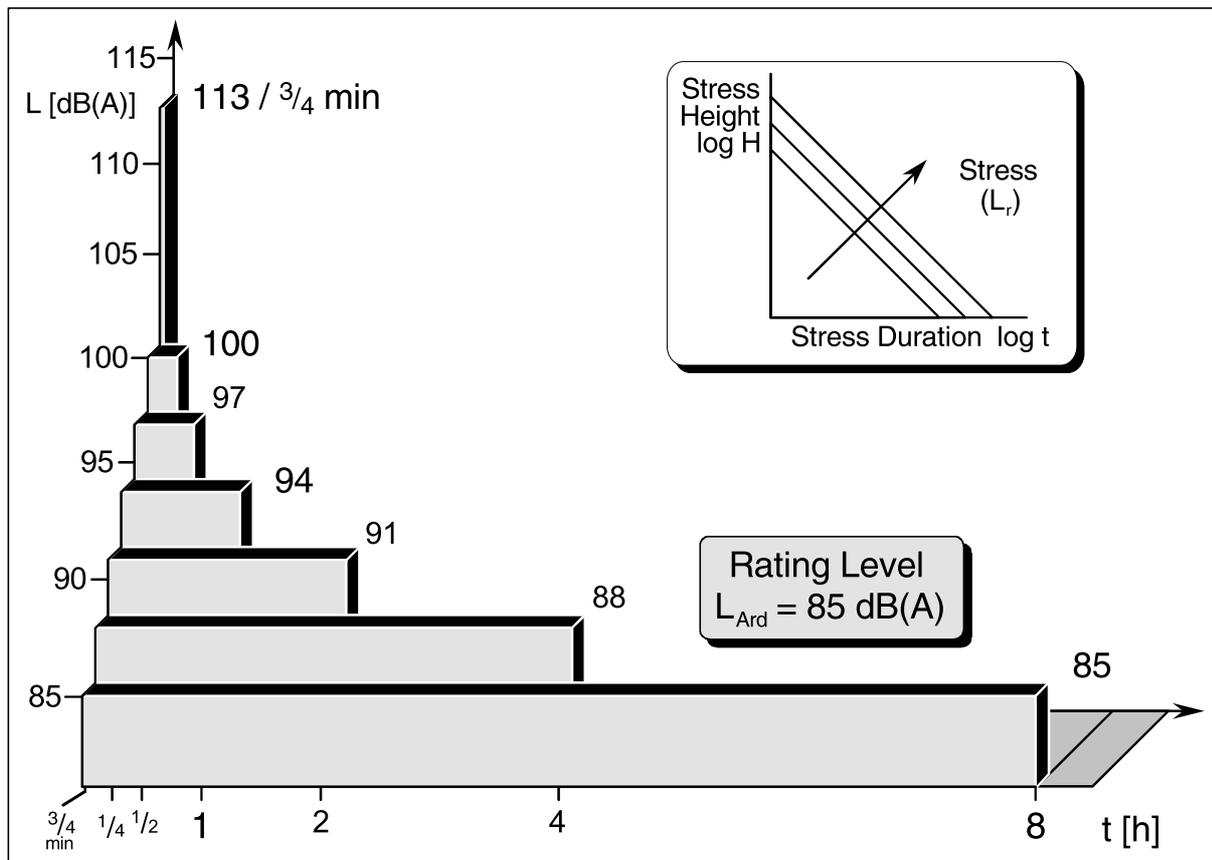
Figure 5. Sound pressure levels of different durations leading to an equal rating level (in this case 85 dB(A)) when applying the "3-dB exchange rate."

Rating Level L_r for T = 8 h

$$L_r = 10 \log \left[\frac{1}{T} \sum_{i=1}^{n} 10^{L_i/10} \cdot t_i \right]$$
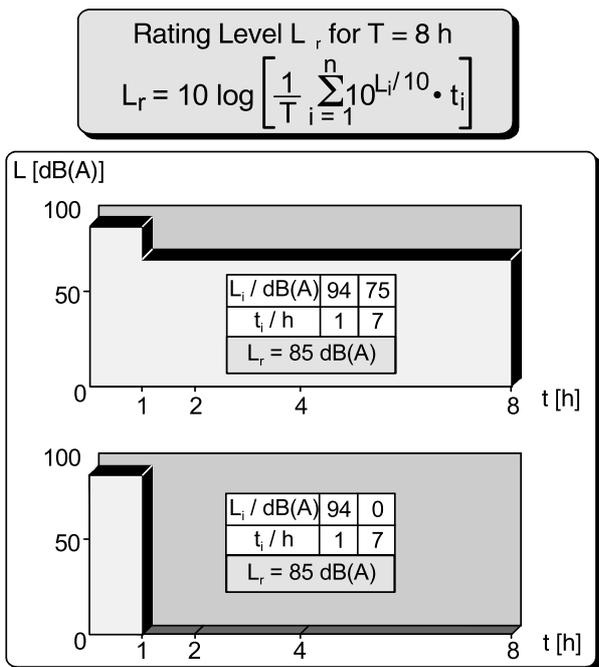
Figure 6. Elucidation of discrepancies in rating noise via the 3-dB exchange rate.

engineer in this case would decide to completely insulate the machine which emits the lower level, the rating level would not be influenced at all. The application of the measure "rating level" consequently allows these curious ratings, as long as the lower

value of noise remains a certain amount below the peak levels. For an equal exposure time, a difference of only 10 dB between the two levels is already enough to neglect the lower level, which absolutely agrees with legal regulations, standards, and national or international guidelines.

When continuing to halve the exposure time and when applying the "3-dB exchange rate" — from a purely arithmetical point of view — even 15 min of dance music in a club at 100 dB would correspond to an 8-h working day at 85 dB (Figure 5). This exposure is still tolerated in the production sector according to almost all international standards (NN 1997). Nevertheless, physiologically seen, high sound levels for a short period, e.g. 100 dB over a 15 min or consequently 113 dB for 45 s have to be assessed much more advantageously than continuous noise. This can be demonstrated, for example, by temporary threshold shifts (TTS) resulting from different noise levels with corresponding exposure times in an energy equivalent arrangement (Irle *et al.* 2000).

4. EQUAL ENERGY PRINCIPLE AND THE CONVENTIONAL RATING OF CONTINUOUS AND IMPULSE NOISE

But may the mutual compensation of stress height and exposure time to the advantage of man be applied without limit? Can 120, 130, 140, 150 or even 160 dB at an adequately reduced exposure (of 10, 1, 0.1, 0.01 or 0.001 s) be assessed to be identical to or even more advantageous than, for example, the above-mentioned 113 dB/45 s? From a physiological point of view the answer must be 'no,' even though TTS may level off completely. Nevertheless, in the past, the energy equivalent compensation of a halving of
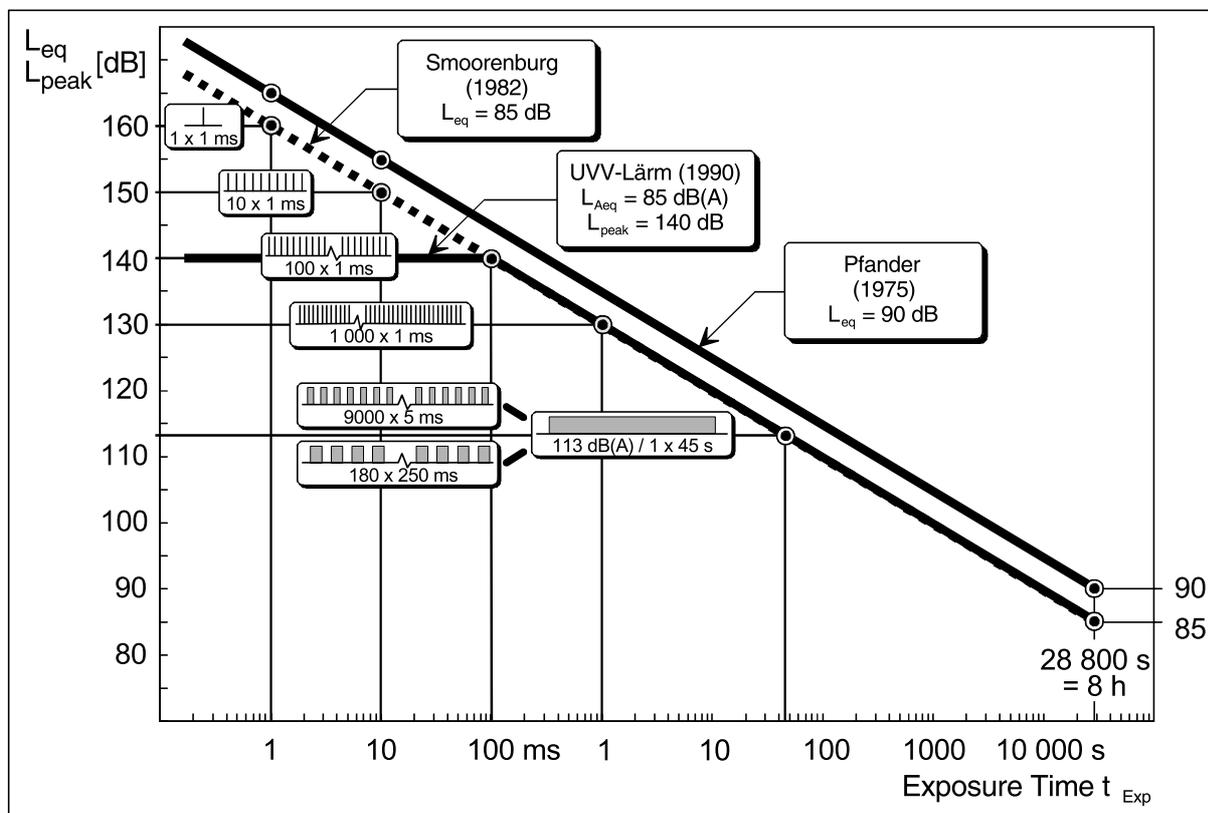
Figure 7. Comparison of various noise rating procedures.

the duration with a level increase by 3 dB and vice versa (or the factor 10 in duration versus level) has become the basis for cut-off level diagrams to avoid hearing impairment which are applied in civil as well as in military sectors (NN 1987).

In the case of impulse noise, exposure times even reach down into the range of ms. When establishing a logarithmic scale for the exposure time in addition to the already existing one for the noise level in dB, the lower straight line in Figure 7 illustrates the energy equivalence for the rating level of 85 dB, e.g.:

- 1 _ 1-ms impulse of 160 dB;
- 10 _ 1-ms impulses of 150 dB;
- 100 _ 1-ms impulses of 140 dB;
- 1000 _ 1-ms impulses of 130 dB;
- 9000 _ 5-ms impulses of 113 dB; and
- 85 dB for 8 h (28 800 s) respectively.

Although a rating level of only 85 dB(A) which has long been tolerated in civil sectors is now provided as limit in the military sector, too (contrary to the former tolerable rating level of 90 dB), and although the unevaluated noise level in industry may not exceed 140 dB according to revised noise regulations (e.g. NN 1990, 1997) due to the lower line in Figure 7, the varying time structures are not considered.

Figure 8 shows cut-off level diagrams that were used to avoid hearing impairment in case of bang noise. The lower limiting line was utilized (in the military sector) for assessing noise constellations, e.g. 165 dB for 1 ms or 90 dB for 8 h, which would still be tolerated without ear protectors. Assuming a 20-

or even a 40-dB noise insulation of an applied protective equipment, the "border line" can be moved upwards by 20 or 40 dB with the consequence that any acoustic stress is allowed, provided that the intersection points of peak sound pressure levels and the exposure times remain below the limit line in the diagram being considered.

The following example will help to illustrate the conventional procedure of rating impulse noise. For details see Strasser (1990) and Strasser and Hesse (1993). When shooting live ammunition from a rifle, a peak level of ~160 dB during an exposure time of ~1 ms can be measured. Similar exposures can be found for policemen's target shooting or when using signal-cartridges or bolt setting guns in the building trade. According to the previous cut-off level diagram, without an ear protector, it can be calculated that two shots per day are the utmost to be allowed; three events would already be beyond the limit. In the case of a total exposure time of 2 ms from two impulse events, the measured peak level indeed leads to a peak-duration constellation (160 dB/2 _ 1 ms), which is just below the limit line. The use of practice ammunition with a peak level of ~140 dB, which is 20 dB less than when using live ammunition, should — from a purely mathematical point of view — allow ~100 times more noise events. That would correspond to ~200 shots per day. Since the exposure time is ~1.5 ms, "only" 186 tolerable impulse events per day have been calculated. Similarly, when using a subcalibre barrel with a peak level of 30 dB less than when using live ammunition, 1000 times more events (2000 events) would be allowable; the number of
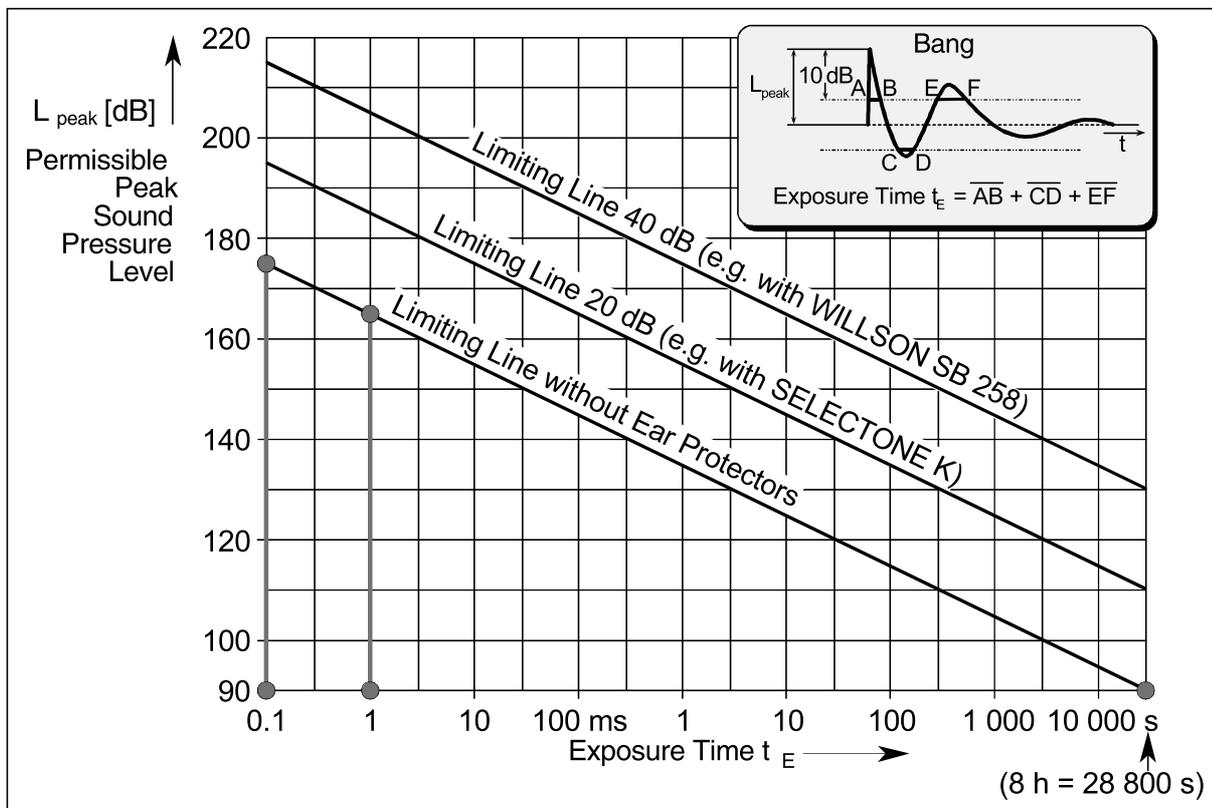
Figure 8. Cut-off level diagram to avoid hearing impairment in case of bang noise. Valid for exposures per day with a following recovery of at least 8 h (Strasser 1990).

shots was finally reduced to 1250 because of the increased exposure time of 1.58 ms in this case.

5. EQUAL WORK OR ENERGY, A PRINCIPLE BEYOND ERGONOMICS LIMITS

But can these calculations, claiming a high level of exactness in the procedures of traditional occupational health and safety, really guarantee a safe degree of hearing protection? The answer is no, because they are based solely on the principle of equal energy, and this is a principle beyond ergonomics limits. Traditional cut-off level diagrams as well as the determination of the rating level can only represent an aid for the evaluation of the sound energy acting on man. But when stress is quantified only with regard to physical aspects, man and his physiological characteristics are principally not included in the approach of the assessment. The calculation of the total stress by a multiplication of intensity and duration indeed is a common procedure for the assessment of other kinds of the physical environment, too. However, if man is involved in work, which is, of course, unalterable in ergonomics, plausible limiting conditions may never be neglected in the domain of stress. This may be visualized in a simple, well-known example.

A performed output (physical work) can be calculated by the height and duration of work, e.g. the performance in Watts on a bicycle ergometer. However, this is only reasonable within physiological limits of the endurance level. A male worker may be able to perform about 50 Watts for extended periods of time. For instance, the product from 50 watts and 60 min working time (e.g. $B_3 = H_3 \_ T_3$ as seen in the front part of Figure 9) is identical to the work resulting from 100 W and a working time of 30 min or also 200 W performed for 15 min (analogous to $B_1 = H_1 \_ T_1$). Yet, can the strategy of mutual compensation be continued arbitrarily ad libitum? The answer must be no. At least in the limiting ranges, human nature does not play along. Nobody will comply with a demand of 500 W for 6 min or 1000 W for 3 min. The principle of equal work, i.e. the mutual compensation of intensity and time, cannot meet physiological laws.

This can be shown via the general responses of the heart rate to the above-described workload constellations, which are identical with respect to the demanded output (upper part of Figure 9). The well-known course of a work-related increase of the heart rate during and after a physical workload representing the physiological costs which have to be paid for physical work by the organism in the domain of strain clearly shows that the "steady-state" during work is lost when exceeding the endurance level. Furthermore, the time after work needed for recovery processes increases over-proportionally. This well-known fact in ergonomics makes it necessary to allocate strain-based measures of recovery time. But, finally, overload cannot always be compensated by rest pauses and restitution periods. For physiological reasons, occupational situations must not occur which, e.g. result in critical heart rates of about 180 beats/min, even if this occurs for only a short time. In this case, the mutual settlement of high workload and short duration must reach an upper limit. But it is exactly this procedure — a procedure which is based solely on physics rather than ergonomics — which is practiced when applying the equal energy
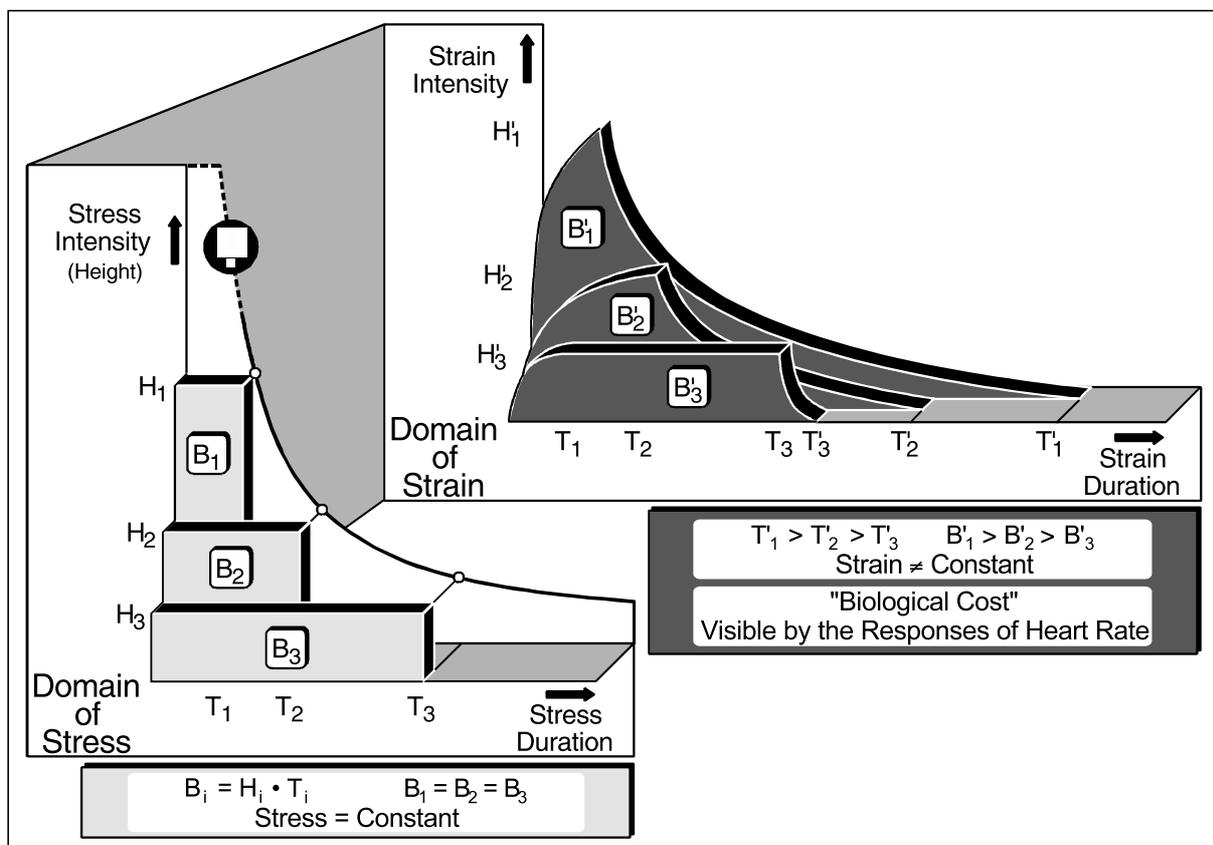
Figure 9. Equal dynamic work, i.e. product of stress intensity and duration (lower frontal part) resulting in varying "biological cost," i.e. work-related increases of heart rate during and after stress (upper part behind).

principle, i.e. the 3-dB exchange rate, to assess short duration continuous and impulse noise.

6. RISKS IN OCCUPATIONAL SAFETY AND HEALTH BY THE APPLICATION OF THE EQUAL ENERGY HYPOTHESIS

Drawing inferences about all the reported theoretical considerations on the one hand (which are presented by Strasser 1995 in much more detail), and regarding the results of several experimental investigations into the physiological costs of noise on the other hand (Irle *et al.* 2000), it should become evident that real risks exist in occupational safety and health rules which are based on the equal energy hypothesis or the principle of equal work. Critical workload areas exist not only in the case of cardiovascular system loading or in the case of hearing; they also exist for most organs and senses. The following examples illustrate that the mutual balancing of the intensity and duration of stimuli is always a questionable procedure:

- Should not local burning be expected when skin temperature is increased, even if it is increased for only a very short period of time? The local increase in temperature resulting from touching the hot-plate of a cooking stove cannot be converted into a cozy warmth over the whole organism at that very moment.
- We should remember that a glaring light beam is always irritating, whereas a balanced optical environment in terms of luminance and illumination enhances visual perception.
- Is not the prick of a needle, e.g. into a finger, always one and only one singular event of a mechanical irritation that causes pain? This cannot be converted into the caressing of this very point over a longer period.

The above-mentioned rhetorical questions also raise the inevitable issue whether our sensory organ "ear" may represent an exception in the case of noise. Can the ear actually be expected to tolerate 160 dB for 1 ms or 100 noise events of 140 dB/1 ms in the same way as it tolerates continuous noise of 85 dB for 8 h, which is what the principle of equal energy suggests? When considering the density of energy acting on the ear, impulse noise simply cannot be compared with continuous noise. Equalizing 160 dB/1 ms or 1000 noise impulses with a level of 130 dB and a duration of 1 ms, each, with, e.g. 85 dB for 8 h according to the 3-dB exchange rate is in accordance with energy equivalence. However, this does not mean that the term 'dose' as a datum level is in fact acceptable in this case. During continuous noise, the sound energy acting on the ear is distributed over the 28 800 s of an 8-h working day. However, in the case of impulse noise, the total sound energy is forced on the human sensory organ in extremely high doses within a fraction of a second. Just as a drug which in a small dosage over a longer period can induce healing and one single "overdose" of that drug can be lethal, the impacts of impulse and continuous noise simply are not identical. An impulse noise event is comparable, for example, with a hurricane in the inner ear which can hardly pass the carpet of hair cells without any adverse effect, just as a short-time gust blowing over a cornfield or being caught in a sail cannot have the same effect as a light wind over a longer period of time.

7. CONCLUSIONS

In the context of the energy equivalence principle in rating the

physical environment (e.g. Martin 1976), one must not forget a mechanical analogue where deformations of materials are the intended aim of an energy concentration. Fast, energetic manufacturing operations, such as, e.g. beating, bumping, or punching, are the essential presuppositions for deformations of materials. Therefore, it is only a stringent consequence that short but intensive events of environmental stress must involve a greater potentiality of health hazards for man. Therefore, the validity of acceptable equivalences of environmental stress to guarantee health protection must be called into question. There should be no doubts that the effect of a dose which is dispensed within two different time intervals is more striking within the shorter one. Also, unquestionably, in the case of increasing density of energy or concentrations of harmful agents, the exceeding of physiological barriers with simultaneous intensifications of the effects becomes much more probable. This is especially true when the organism does not possess effective potentialities of temporal and/or spatial buffers. Therefore, the well-known endeavor for simplification and standardization which drives attempts to squeeze the rating of complex environmental situations into simple models or integrated measures as is done, e.g. also for ultraviolet radiation, mechanical vibrations, and carbon monoxide, cannot be adopted by ergonomics. Via this procedure, multidimensional connections get lost. In this context, a simple but slightly meditative comparison may be convincing for skeptics: The leveling of short lasting high intensity stress, based on physical rudiments, indeed seems to be as trustworthy as the statement that nobody can drown in a river with a statistical average depth of 50 cm.

REFERENCES

IRLE, H., HESSE, J.M. and STRASSER, H., 2000, Noise: physiological costs of noise exposures assessed via measurement of temporary threshold shift. In W. Karwowski (ed.), *International Encyclopedia of Ergonomics and Human Factors* (London: Taylor & Francis).

ISO DIS 1999. 2, 1985, Acoustics — Determination of occupational noise exposure and estimation of noise-induced hearing impairment.

MARTIN, A.M., 1976, The equal energy concept applied to impulse noise. In D. Henderson, R.P. Hamernik, D.S. Dosanjh and J.H. Mills (eds), *Effects of Noise on Hearing* (New York: Raven), 121–53.

N.N., 1987, Effects of Impulse Noise, NATO Document AC/243 (Panel 8/RSG 6) D/9, Final Report of the Research Study Group on the Effects of Impulse Noise.

N.N., 1990, Accident Prevention Regulation — Noise, UVV Lärm, Unfallverhütungsvorschrift der gesetzlichen Berufsgenossenschaften, VBG 121 (Cologne: Carl Heymanns). [in German]

N.N., 1996, *Preventing Occupational Hearing Loss — A Practical Guide*, revised October 1996. DHHS (NIOSH) Publication no. 96-110 (Cincinnati: US Department of Health and Human Services).

N.N., 1997, Technical assessment of upper limits on noise in the workplace — Final report. Approved by the International Institute of Noise Control Engineering, *Noise/News International*, 203–16.

N.N., 1998, *Criteria for a Recommended Standard — Occupational Noise Exposure*, revised criteria 1998. DHHS (NIOSH) Publication no. 98-126 (Cincinnati: US Department of Health and Human Services).

STEVENS, S.S., 1957, On the psychophysical law, *Psychological Review*, 64, 153–81.

STRASSER, H., 1990, Ergonomic considerations with reference to the notion of dose and the equal energy hypothesis in environmental workload, *Zbl. Arbeitsmed.*, 40, 338–54. [in German]

STRASSER, H. (ed.), 1995, *Arbeitswissenschaftliche Beurteilung von Umgebungsbelastungen — Anspruch und Wirklichkeit des präventiven Arbeitsschutzes* (Landsberg: Ecomed).

STRASSER, H. and HESSE, J.M., 1993, The equal energy hypothesis versus physiological cost of environmental workload, *Archives of Complex Environmental Studies*, 5, 9–25.

Occupational Stress Mechanisms

B.M. Gwóźdz

Institute of Physiology, Silesian Medical University, 41-808 Zabrze, Poland

1. INTRODUCTION

The concept of stress has more than one interpretation (Hincle 1979). In physiology, stress means physical or mental tension when the organism is threatened or challenged. Stress is evoked by a variety of physical factors, extremely potent or damaging (injury), and psychological factors which pose a threat to a person's life situation (fear, anxiety, feeling of guilt) (Margolis *et al.* 1974).

There are many factors in the workstation that may provoke non-specific response of the organism involving hyperactivity of the central and peripheral nervous system with a subsequent increase in endocrine function. In this way, excitation causes changes in the organism's internal environment and disturbances in homeostasis.

Homeostasis is the ability to maintain a stable internal environment (plasma and interstitial fluid). Disrupted homeostasis is responsible for decreased work performance and efficiency, and has a negative effect on a person's well-being and state of health. Prolonged lack of homeostasis results in illness (Chrousos and Gold 1992).

2. STRESSORS

Factors which induce stress are known as stressors. There are different categories of stressors associated with occupational stress. Physical stressors include: cold, heat, noise, radiation, sounds and ultrasounds of high intensity, vibrations, mechanical injury, high and low barometric pressure, chemicals, toxins, air pollution, hypoxia and high-force exercise.

Social stressors related to an individual's interaction with the environment are, among others: poor workplace conditions, interpersonal relation problems, relations of domination and subordination, conflicts, loss of somebody or something significant (a person, job, home, money), retirement, loneliness, crowding, high responsibility, as well as unexpected success at work or substantial financial gain.

Psychological stressors form a significant group of job stressors. They are responsible for excitation of the central nervous system under real or imaginary threat accompanied by strong emotional states such as: anxiety, frustration, feeling of guilt, anger, hate, worry, love, and envy. Emotional excitation results in hyperactivity of the sympathetic nervous system and an increase in heart rate, arterial blood pressure, sweating, peristalsis, and metabolic rate.

Powerful job-related psychological stressors also include: inability to fulfill one's ambitions, lack of promotion, low self-esteem, lack of self-confidence, making mistakes, poorly defined role, ambiguous goals, work overload, time pressure, decision-making or exclusion from the process.

One of the most potent occupational stressors is differentiation, i.e. decision-making, particularly in unfavorable circumstances or when the nervous system is overloaded with signals from the rapidly changing work environment. This usually results in prolonged reaction time, which is of crucial importance in some jobs. In the complicated world of signals, the individual is faced with difficult choices, being at the same time aware of the unpleasant consequences of wrong decisions. This stress state is biologically threatening, especially when long-lasting and causing discomfort and neurosis. Therefore, differentiation requires some time and the individual acting under time pressure is at a higher risk of stress-related health problems.

3. STRESS EFFECTS

In the work place, numerous stressors may act simultaneously causing a more intense response which involves nearly all functional systems, i.e. the sensorimotor, vegetative, and endocrine systems. Excitation of the central nervous system leads to an increase in skeletal muscle tone and blood flow in the muscles, while the cardiovascular system responds with elevated stroke volume and cardiac output. The metabolic rate is enhanced to mobilize energy reserve. In response to a stimulus defined as energy unit of high intensity, the central nervous system becomes activated, which results in stimulation of the hypothalamic–pituitary–adrenal axis (Ganong *et al.* 1987) and, in consequence, elevated levels of stress hormones as well as hormones like adrenocorticotropic hormone (ACTH), glucocorticoids, epinephrine, norepinephrine, dopamine, endorphins and enkephalins. In such a state the organism is on the alert for fighting. When the energy reserve is utilized, e.g. during exercise, it is beneficial for the organism, but in sedentary jobs, the increased activity of many organs, including the heart and blood vessels, means unnecessary load for the organism.

In prolonged states of alertness, the system involved becomes exhausted and more susceptible to illness. In the majority of events, and regardless of the type stressor, the response is always manifested in hyperactivity of the functional organs, resulting in overloading of the organism and increased physiological cost.

However different the stressors, their mode of action is similar. Stress at work is evoked by excessive reaction to a stimulus. Therefore, not every factor related to the work process fulfills the stressor criteria. Assessment of stress threat also depends on the outside observer's negative emotions.

Over-excitation of the nervous system and the resulting motor and vegetative responses usually hinder the individual's adjustment to changes in the work environment. Conditioned reflexes are the most commonly inhibited responses in stress, both through external and internal inhibition.

Seemingly simple job tasks require integration of incoming information in the central nervous system as well as formation of action program. In many stress situations, this is not easy, particularly when the individual has to cope with a lot of information in a short time. In such cases, stress becomes more intense and work performance decreases.

4. PHYSIOLOGICAL EFFECTS

Excitation of the nervous system in stress contributes to an increase in skeletal muscle tone. The level of muscle tone regulation changes, manifesting itself in uncontrolled and usually uncomfortable position of the body and, in consequence, leading to central nervous system overload.

Increased stress-induced attention is accompanied not only by elevated skeletal muscle tone but also by higher heart rate

and intense sweating; this results in reduced work efficiency despite the individual's efforts.

Physiological changes accompanying the control processes of the central nervous system in individuals exposed to occupational stress are used to assess the overloading of organs or the nervous system, e.g. heart rate, arterial blood pressure, and the effect of the sympathetic system on heart rate. In stress, reduced work efficiency, ill-being, and health problems depend on a person's individual characteristics and the degree of reaction to factors associated with work itself and its environment. The same stressor may provoke a completely different response in individuals of the same or opposite sex. In some, it will cause stress, while others may experience pleasurable emotions. Prolonged acute stress may result in various diseases when the regulatory mechanisms become inefficient.

5. ADAPTIVE MECHANISM

In response to the same stimulus, some individuals may develop adaptive mechanisms of regulating homeostasis at a different level. The mechanism of response to stress involves the nervous and endocrine systems as well as the immune system. The latter largely determines the activity of cellular immune mechanisms (Dorian and Garfinkal 1987). During stress, the immune response is inhibited, while on cessation of stress the immune system is stimulated, unless the stress has been long term and intense. The response is possible due to the action of glucocorticosteroids, epinephrine, and norepinephrine which are released in stress and usually inhibit cellular immunity. Stimulation of cellular immunity in the post-stress period is a manifestation of the organism's adaptation and mobilization of immune defense forces under threat. Activation of the nervous, hormonal, and immune systems clearly differs and is related to the individual's age.

6. STRESS RESPONSE

Response to occupational stress largely depends on individual tolerance, i.e. a characteristic reaction to a given situation. Response to stress, irrespective of its health effects in individual cases, is particularly important in group work characterized by differences in individual work performance. The relationship between the efficiency of the best and the worst performers is 2:1, regardless of the tools, equipment, technology used.

Work performance and efficiency are biologically determined. In group work, the overall results depend on the performance of the individual most susceptible to stress. If it is impossible to exclude this individual from the group, the sources of occupational stress should be eliminated through ergonomic, technological, and organizational improvements to compensate for the individual's biological defects. It is essential, therefore, to eliminate the recognized stressors.

REFERENCES

CHROUSOS G.P., and GOLD P.W., 1992, The concepts of stress and stress system disorders: overview of physical and behavioral homeostasis. *J.Am.Med.Assoc.*, 267, 1244–52.

DORIAN B., and GARFINKAL P.E., 1987, Stress immunity and illness. *Psychol.Med.*, 17, 393–407.

GANONG W.F., DALLMAN M.F., and ROBERTS J.L., 1987, The hypothalamic-pituitary-adrenal axis revisited. *Ann.N.Y.Acad.Sci.*, 5, 12.

HINCLE L.E. Jr., 1979, The concept of stress in the biological and social science. *Sci.Med. and Man.* 1, 31–8.

MARGOLIS B.L., KROES W.H., and QUINN R.R., 1974, Job stress: an unlisted occupational hazard. *J.Occup. Med.*, 16, 659–61.

Physical Demands of Work

A. Genaidy

Department of Mechanical Industrial and Nuclear Engineering, University of Cincinnati, OH 45221-0116, USA

W. Karwowski

Department of Industrial Engineering, University of Louisville, KY 40292, USA

C. L. Shoaf

Department of Mechanical Industrial and Nuclear Engineering, University of Cincinnati, OH 45221-0116, USA

1. INTRODUCTION

Work content refers to those activities that constitute the nature of the job composition. The work context, the setting (i.e. physical, organizational, social, and individual growth) in which the work activities take place, and the work content characterize the spectrum of the work demands input to the worker. Work content can be subdivided into physical and mental demands components. The physical demands component describes the set of activities that require dynamic body movement and static exertion to maintain a given posture. The mental demands component describe the set of activities executed by the brain's perceptual and cognitive processes. Assessment of physical work is an essential component of the health and safety programs as well as job redesign and work improvement efforts. However, while overwhelmingly regarded as essential, detailed general purpose tools for the assessment of the entire spectrum of industrial physical demands do not exist.

2. PHYSICAL DEMANDS AND EVALUATION METHODS

Physical demands and evaluation methods can be categorized into three major classes: (1) mathematical models (i.e. statistical, empirical, biomechanical, fuzzy set); (2) tabular data; (3) rating methods (see Karwowski and Marras 1999).

2.1 Statistical Models

Statistical models use experimental data to formulate numerical models. Work factors and their interactions in these models are subject to conformance to certain statistical criteria such as the maximum value of multiple-squared correlation (i.e. R^2). Genaidy *et al.* (1988) provided a review of psychophysically based regression models that can be used in predicting human capacity for object handling work. On the basis of this kind of psychophysically based regression models, situations can be evaluated where the human object handling capacity is exceeded. Genaidy and Asfour (1987) reviewed physiological cost-based regression models for object handling work. These models predict the oxygen consumption and/or heart rate values associated with the handling of objects. The obtained values are then compared to physiological tolerance limits (Asfour *et al.* 1988). If these limits are exceeded, the job or task at hand will be a candidate for redesign. Recently, Marras *et al.* (1993, 1995) developed logistic regression models to predict the risk of lower back injuries as a function of trunk motion (e.g. average twisting velocity, maximum lateral velocity) and task variables (e.g. weight of load, frequency)

from lifting tasks. Trunk motion factors were measured using the Lumbar Motion Monitor (Marras *et al.* 1992).

2.2 Empirical Models

Empirical models postulate a certain form for equations and then fit data to them. They are used to predict worker capacity under various conditions. Examples of empirical models are the work of the US National Institute of Occupational Safety and Health (NIOSH 1981; Waters *et al.* 1993), Hidalgo *et al.* (1997), Shoaf *et al.* (1997), Karwowski and Marras (1997), and Karwowski *et al.* (1998). These equations were patterned after the Drury and Pfeil (1975) model which is multiplicative in nature (i.e. interactive work factors). It assumes that there is a maximum load to be handled under ideal conditions. When one deviates from the ideal conditions, discounting multipliers are applied to reduce the amount of load handled.

2.3 Biomechanical Models

In static and dynamic biomechanical models, the human body is treated as a mechanical system made up of various links. Forces and torques imposed upon various joints are estimated using biomechanical models, then compared with tolerance limits of the workforce (Genaidy *et al.* 1993). The job is considered hazardous if the imposed forces and torques exceed the specified biomechanical tolerance limit of either an individual or an agreed percentage of the working population. As an example, a simple static model was devised by Tichauer (1978) on the basis of the following prerequisite for biomechanical work tolerance: "minimize the moments acting upon the vertebral column". The model was expressed mathematically as: $M_E = \{[8 + (1/2) * L] * W\}$ where: (1) 8 is the approximate distance (in inches) from the joints of the lumbar spine to the front of the abdomen — a constant for each individual according to Tichauer; (2) L is the length (in inches) of one side of the object; (3) W is the weight (in pounds) of the object; (4) M_E is the biomechanical lifting equivalent (lb-in). If the moments around the lower back are 350 lb-in or less, the work is considered "light" and can be performed with ease by untrained individuals, either males or females, regardless of body build. Between 350 and 750 lb-in, the work can be classified as "medium-heavy" and requires good body structure as well as some training. Tasks resulting in moments between 750 and 1200 lb-in are considered "heavy", requiring selective recruiting of employees and attention to rest pauses. The work is considered "very heavy" when 1,200 lb-in is exceeded. Very heavy work cannot be performed on a continuous basis and requires great care in worker recruitment and training. A more thorough discussion of biomechanical models can be found in Chaffin and Andersson (1991).

2.4 Fuzzy Set Models

Karwowski *et al.* (1986) presented a detailed fuzzy set methodology to analyze lifting activities. According to this methodology, each variable was defined numerically and qualitatively. Distribution functions were then established to determine the degree of certainty with which the numerical variable belongs to the qualitative variable. For example, our certainty that a 50 lb load belongs to the class "moderate heavy" load is 0.8. Rules in an ''IF THEN'' format (*e.g. If A is low and B is high THEN C is moderate*) were then established on the basis of

approximate reasoning. The method developed by Karwowski *et al.* (1986) is an appropriate methodology which can be extended to analyze complex physical demands in industry. Such an approach allows the involvement of both users and engineers in the analysis of muscular work. On the one hand, users are usually the most knowledgeable about the work tasks in question. Although they are not formally trained in detailed quantitative methods as engineers, their knowledge will allow the quick analysis of work tasks through the use of linguistic descriptions (*e.g.* load is heavy). On the other hand, engineers, by virtue of their training, can transform these linguistic descriptions into numerical scores which can be used later for evaluation and design/redesign purposes.

2.5 Tabular Data

Several ergonomic databases have been detailed in the form of tables. Snook and Ciriello (1991) published detailed tables on the manual handling capacity of individuals for lifting, lowering, pushing, pulling, and carrying activities. Physiological outcome measures such as energy expenditure or oxygen consumption and heart rate have been summarized for various types of tasks, mostly object handling work (e.g. Asfour *et al.* 1986a, 1986b). Astrand and Rodahl (1986) reported that Passmore and Durnin (1955) published tables of daily energy expenditures for various types of activities. On the basis of their experience and review of prior findings, Astrand and Rodahl classified work into five categories on the basis of physiological responses:

(1) In terms of oxygen consumption (i.e. indirect measure of energy expenditure): light work (up to 0.5 liter/min), moderate work (0.5–1.0 liter/min), heavy work (1.0–1.5 liters/min), very heavy (1.5–2.0 liters/min), extremely heavy work (over 2.0 liters/min); and

(2) In terms of heart rate: light work (up to 90 beats/min), moderate work (90–110 beats/min), heavy work (110–130 beats/min), very heavy work (130—150 beats/min), and extremely heavy work (150–170 beats/min).

Webb Associates (1978a, 1978b) published extensive tables on anthropometric measures. These tables included body size and range of motion data for various percentages of the population. These measures were mostly obtained for military personnel and were primarily used for equipment design.

2.6 Rating Methods

Rating methods have ranged from being relatively simple to very complex. The checklist approach represents the simplest methods of evaluating muscular demands. Traditionally, this approach consists of a number of questions in which the answers are binary, i.e. "yes" or "no" answers. Similar to mathematical models (i.e. statistical, empirical, biomechanical), this approach has focused on analyzing selected individual activities. An example is the work of Lifshitz and Armstrong (1986) for evaluating repetitive upper extremity activities. Their checklist consisted of twenty questions requiring yes or no answers. They found good agreement between the evaluation scores obtained from the checklist and injury statistics. Another example is the PLIBEL checklist developed by the Swedish National Institute of Occupational Health (Kemmlert 1995). The checklist consisted of seventeen yes or no questions and was intended for general purpose use.

More elaborate rating methods attempted to analyze various physical activities. Examples are the "AET" (Rohmert and Landau, 1983) and "Job Profile" (Wagner 1985). These are among the most detailed rating methods found in the published literature. The AET method was designed to analyze any type of industrial activity and has been used extensively in Germany. The H-AET was added as a supplement to AET to enhance its level of details. The Job Profile method, on the other hand, was used to evaluate short-cycle activities at Renault plants in France. In contrast to the design of simple checklists, detailed rating methods included more work factor detail and assigned more than two values to each rated work factor. Some methods had more detail than others.

Many efforts, particularly European research, focused on the assessment of postural loading in industry. The Job Profile system, for example, provided a detailed account of postural loading (Wagner 1985). In general, each posture was ranked in terms of difficulty and its intensity was assessed using the time spent in a particular static posture. The RULA was recently devised as a method to analyze: the postures of the upper extremity joints (i.e. hand, lower arm, and upper arm), neck and trunk; the load on the musculoskeletal system caused by excessive static muscle work; repetitive motions; and the requirement to exert force or maintain an external load (McAtamney and Corlett 1993). The main advantage of the RULA system is its practicality for work posture analysis. A comprehensive review on assessment of postural loading in industry can be found in Genaidy *et al.* (1994).

3. CLASSIFICATION SYSTEM FOR PHYSICAL DEMANDS

To support assessment of physical work demands in industry, all activities must be classified and evaluated. A classification and evaluation systems are based on the physiological and epidemiological concepts.

3.1 Physiological Concepts

3.1.1 Dynamic and static work

Muscular work consists of combinations of dynamic and static contractions and releases. Dynamic work is characterized by a rhythmic alternation of contraction and extension, tension and relaxation (Grandjean 1988). Static work, on the other hand, consists of a prolonged state of muscular contraction lasting longer than four seconds (Rohmert and Landau 1983). Both dynamic and static components of muscular work must be analyzed to account for physical requirements.

3.1.2 Human limits: strength, endurance, and range of motion

In the study of muscular work, the following limits must be observed (Genaidy 1996):

- Endurance is a major requirement for continuous work. Repetition and work duration are important factors in the evaluation of continuous work.
- Strength governs the ability to perform infrequent work. Here, the maximum load handled or force exerted is one of the important factors.
- Range of motion and constrained postures resulting in prolonged static effort are determinants of static postural load. The angular deviation of a body part from neutral positions is an indicator of static postural load. The larger

the angular deviation, the higher the static postural load. The range of motion is the maximum angular deviation.

3.2 Epidemiological Concepts

Many studies have documented that overloading the musculoskeletal system at work can result in injuries. Statistics reported by the US Association of Schools of Public Health (1986) indicated that musculo-skeletal injuries are the leading cause of disability of people in their working years and rank first among health problems in the frequency with which they affect the quality of life. Moreover, the cost of musculo-skeletal injuries based on lost earnings and worker compensation exceed that of any other health disorder and they account for one-third of annual worker compensation claims. In general, musculo-skeletal injuries arise due to three primary activities: object handling, repetitive extremity, and prolonged static postural loading.

3.2.1 Object handling injuries

It is well established that manual handling of objects is the largest single subset of musculo-skeletal disorders — particularly the lower back disorders (CDC 1983). In 1993, back disorders accounted for 27% of all non-fatal occupational injuries and illnesses in the United States (NIOSH 1996). Estimates of the total cost of lower back pain to society in 1990 were between $50 billion and $100 billion per year with a significant share borne by the worker compensation system (NIOSH 1996). Troup and Edwards (1985) reported that, in the United Kingdom, object handling accidents were responsible for between 20% and 30% of all industrial accidents in the time period 1930–1977. According to the 1979 statistics, lifting, carrying, and throwing activities were responsible for almost 67% of all over-exertion injuries. An additional 19% resulted from pushing and pulling. The statistics associated with object handling has triggered extensive research over the past several decades. Lifting activities have been studied in depth because they have been reported as one of the principal causes of lower back disorders (e.g. NIOSH 1981; Chaffin and Andersson 1991; Snook 1978; Liles *et al.* 1984; Marras *et al.* 1993, 1995; Ayoub *et al.* 1997).

3.2.2 Repetitive extremity injuries

Although not an entirely new problem, repetitive extremity work, particularly in seated or standing positions, has emerged as a major cause of cumulative trauma disorders (CTD) or repetitive strain injury (RSI) in the past two decades. The magnitude of the problem has been documented by many researchers (e.g. Armstrong *et al.* 1986; Hagberg 1984; Ayoub and Wittels 1989; Burnette and Ayoub 1989; Wallace and Buckle 1987; Kilbom 1988, 1994; Karwowski and Marras 1999). More than $2.1 billion in worker compensation costs and $90 million in indirect costs are incurred annually for upper extremity cumulative trauma disorders (UECTD) (NIOSH 1996). The most frequently reported UECTD affect the wrist/hand region, with carpal tunnel syndrome (CTS) or entrapment median neuropathy occurring at a rate of 5.2 per 10 000 full-time workers.

3.2.3 Prolonged static postural problems

Prolonged static postures have also been linked to work-related musculo-skeletal disorders and productivity losses in the workplace. Punnett *et al.* (1987) and Fine *et al.* (1987) reported strong correlation between the angular deviation of non-neutral shoulder and back postures, and musculo-skeletal disorders. Ryan (1989) found that for standing postures employed in grocery stores, lower back and extremity problems started when the percentages of standing times were 25% and 45—50% of the shift duration, respectively.

On the basis of experience gained from several field studies and his own observation, Grandjean (1988) reported that static loads are associated with a higher risk of arthritis of the joints due to mechanical stress, inflammation of the tendon sheaths, symptoms of arthrosis, and intervertebral disc troubles. He argued that, according to van Wely (1970), persistent musculo-skeletal problems are commonly observed among operators who work all year around at the same machine where the manual controls are either too high or too low.

4. CLASSIFICATION OF PHYSICAL DEMANDS

4.1 Classification Description

Any physical activity can be described in terms of two general classes of muscular work: object handling and extremity postural. A general description of these classes of muscular work is given below.

4.1.1 Object handling classification

Object handling (OH) is an overall body activity which consists of moving objects with one or both hands and involves the use of upper and lower extremities including some action of the trunk. In this study, OH is defined for handling activities performed during standing (i.e. lifting and lowering, pushing and pulling) and walking (i.e. carrying, pushing, and pulling) positions. A typical example is loading and unloading trucks. Handling activities performed in positions other than standing and walking are analyzed under the extremity postural demand classification. An example of such activities is the moving of small parts while the body is in a seated position.

Traditionally, OH has been investigated using three approaches: (1) biomechanical criteria were used to examine non-repetitive (Chaffin and Park 1973; Chaffin *et al.* 1978; Keyserling *et al.* 1980) and repetitive (Marras *et al.* 1993, 1995); (2) physiological approaches were utilized for repetitive activities (Genaidy and Asfour 1989; Genaidy *et al.* 1990; Asfour *et al.* 1991a, 1991b); (3) psychophysical criteria were employed for both repetitive and non-repetitive activities (Snook 1978; Snook and Ciriello 1991; Ayoub *et al.* 1978, 1983, 1997; Liles *et al.* 1984; Mital 1984). Very recently, Karwowski *et al.* (1998) proposed a cognitive engineering approach to setting limits in manual handling tasks.

Karwowski (1983) and Karwowski and Ayoub (1984) demonstrated that psychophysical criteria integrate the acceptability of both biomechanical and physiological approaches for lifting activities. Recently, Hidalgo *et al.* (1997) and Shoaf *et al.* (1997) developed a comprehensive OHW capacity limits which take into account the acceptability of the aforementioned criteria.

Object handling is classified into lifting, lowering, carrying, pushing, pulling, and other types of activities. During lifting and lowering activities, a work object is moved vertically from a lower or a higher level to a higher or lower level. A carrying activity involves holding a load while the person is walking, climbing

stairs, climbing ladders/ramps, or crawling. A pushing or pulling activity involves pushing or pulling a work object either on flat surfaces and ramps (e.g. pushing hand carts) *or* on a work table (including conveyors). The other activity category includes other types of object handling work such as throwing bags and shoveling. For lifting and lowering, the weight of load, height of lift or lower, horizontal distance in front of the body, twisting angle, frequency and work duration are the essential set of variables to determine the physical demands (Waters *et al.* 1993; Hidalgo *et al.* 1997). For carrying, the weight of load, frequency, work duration, travel distance and dynamic posture are the relevant variables (Shoaf *et al.* 1997). For pushing or pulling, force, frequency, work duration, and travel distance are the necessary variables for analysis (ibid.).

4.1.2 Extremity postural classification

Extremity postural work deals with repetitive work in a fixed or dynamic lower body position (e.g. seated, walking) as well as prolonged static postural loading. Many of the variables affecting extremity-postural work have been documented by many researchers (e.g. Armstrong *et al.* 1986; Karwowski and Marras 1997; Hagberg 1984; Ayoub and Wittels 1989; Burnette and Ayoub 1989; Wallace and Buckle 1987; Genaidy *et al.* 1994).

The following fixed and dynamic lower body positions are defined in extremity postural work: (1) seated; (2) standing; (3) kneeling; (4) squatting; (5) one-legged; (6) crouching; (7) reclining; (8) walking; (9) climbing stairs; (10) climbing ladders/ramps; (11) crawling; (12) other postures. For each of these lower body positions, extremity postural work consists of repetitive extremity/head and static postural work.

Dynamic work, which describes the repetitive extremity/head work, can be classified according to the following categories:

- finger work — hand, lower and upper arms maintained in a fixed position (e.g. typing)
- hand work (including palm and fingers) around the wrist — lower and upper arms maintained in a fixed position (e.g. hand tool usage)
- lower arm work (including hand and fingers) around the elbow — upper arm maintained in a fixed position (e.g. hammering with the lower arm only)
- whole arm work around the shoulder (including the hand) — whole arm movement (e.g. hammering with the whole arm)
- head work around the neck — (e.g. driving a car)
- work around the hip or knee — (e.g. operating a floor pedal with the whole leg)
- foot work around the ankle — leg maintained in a fixed position (e.g. operating a floor pedal with the foot only)
- other repetitive work — (e.g. knee movement during carpet laying)

Static work involves two subclasses:

- continuous force exertion using the fingers, hands (assisted or unassisted by the arms), knees, feet (assisted or unassisted by the legs)
- static postural work of the hand, lower arm, upper arm, head *or* back

The frequency, force, travel distance, and work duration are the most important factors which affect repetitive extremity and head work (e.g. Drury and Wick 1984; Armstrong *et al.* 1986;

Silverstein *et al*, 1986; Tanaka and McGlothlin 1989; Burnette and Ayoub 1989; Snook *et al.* 1995). The above postural classification is based, to a large extent, on a recent review conducted by Genaidy *et al.* (1994) with some modifications.

4.2 Mathematical Models

The physical demand classification allows the hierarchical evaluation of muscular work at different levels where higher level demands (e.g. object handling) can be obtained as a function of lower level demands (e.g. lifting, lowering, carrying, pushing, pulling, other activity).

As pointed out before, work demands must be analyzed for strength and endurance requirements. Work factors for strength-related activities can be assessed in terms of the maximum load handled or force exerted with the corresponding work variables (excluding repetition and duration). Nevertheless, all work factors are included in the analysis of endurance-related activities.

The mathematical formulation of each variable in every layer of the hierarchical structure is presented below:

4.2.1 Overall physical demands

$$PD = f_1 (OHD, EPD) \tag{1}$$

where:
PD — overall physical demands
OHD — object handling demands
EPD — extremity postural demands

4.2.2 Object handling demands

$$OHD = f_2 (LID, LOD, CAD, PUD, PLD, OOHD) \tag{2}$$
$$LID = f_3 (WL, HD, F, WD, HD, TA) \tag{3}$$
$$LOD = f_4 (WL, HD, F, WD, HD, TA) \tag{4}$$
$$CAD = f_5 (WL, TD, F, WD, DP) \tag{5}$$
$$PUD = f_6 (WL, TD, F, WD) \tag{6}$$
$$PPD = f_7 (WL, TD, F, WD) \tag{7}$$
$$OOHD = f_8 (WL, TD, F, WD) \tag{8}$$

where:
LID — lifting demands
LOD — lowering demands
CAD — carrying demands
PUD — pushing demands
PLD — pulling demands
OOHD — other object handling demands
WL — weight of load or force
HD — horizontal distance in front of the body
F — frequency
WD — work duration
H — height of lift and lower
TA — twisting angle
TD — travel distance
DP — dynamic posture.

4.2.3 Extremity postural demands

$$EPD= f_9 (SED, STD, KD, SQD, OLD, CRD, RED,$$
$$WAD, CSD, CLD, CRD, OEPD) \tag{9}$$
$$AEPD = f_{10} (DD, SD) \tag{10}$$
$$DYD = f_{11} (FIMD, HMD, LAMD, WAMD, HEMD,$$
$$LMD, FOMD, OMD) \tag{11}$$

$$AMD = f_{12}(F, WL, WD) \tag{12}$$
$$HEMD = f_{13}(WD) \tag{13}$$
$$SD = f_{14}(CFED, SPD) \tag{14}$$
$$CFED = f_{15}(SFFD, SHAFD, SHUFD, SKFD, SFAFD, SFUFD) \tag{15}$$
$$ACFED = f_{16}(SF, WD) \tag{16}$$
$$SBPPD = f_{17}(SHD, SLAD, SWAD, SHED, SBD) \tag{17}$$
$$ASBPPD = f_{18}(WD) \tag{18}$$

where:

EPD — overall extremity postural demands;
SED — seated demands;
STD — standing demands;
KD — kneeling demands;
SQD — squatting demands;
OLD — one legged demands;
CRD — crouching demands;
RED — reclining demands;
WAD — walking demands;
CSD — climbing stairs demands;
CLD — climbing ladders and ramps demands;
CRD — crawling demands;
OEPD — other extremity postural demands;
AEPD — any extremity postural work demands such as seated (SED);
DD — dynamic demands;
SD — static demands;
FIMD — finger movement demands
HMD — hand movement demands
LAMD — lower arm movement demands

WAMD — whole arm movement demands
HEMD — head movement demands
LMD — leg movement demands
FOMD — foot movement demands
OMD — other movement demands
AMD — any movement demands such as finger demands (FIWD)
F — frequency
WL — force exerted during dynamic activities
WD — work duration
CFED — continuous force exertion demands
SFFD — static finger force demands
SHFAD — static hand force demands assisted by the arms
SHFUD — static hand force demands unassisted by the arms
SKFD — static knee force demands
SFFAD — static foot force demands assisted by the legs
SFFUD — static foot force demands unassisted by the legs
ACFED — any continuous force exerted demands such as static finger force demands (SFFD)
SF — static force exerted by the fingers, hands, knee, or feet
SBPPD — static body part postural demands
SHD — static hand postural demands
SLAD — static lower arm postural demands
SWAD — static whole arm postural demands
SHED — static head postural demands
SBD — static back postural demands
ASBPPD — any static body part postural demands such as static hand postural demands (SHD)

Table 1. Evaluation of infrequent object handling demands (i.e. evaluation of activities involving the maximum load or force exerted on an infrequent basis)

Lifting	weight of load	very light	light	moderate	heavy	very heavy			
	horizontal distance	very close	close	moderate	far	very far			
	height of lift	floor–knuckle		floor–shoulder		floor–reach	knuckle–shoulder	knuckle–reach	shoulder–reach
	twisting angle	very small	small	moderate	large	very large			
	lifting demands	very low	low	moderate	high	very high			
Lowering	weight of load	very light	light	moderate	heavy	very heavy			
	horizontal distance	very close	close	moderate	far	very far			
	height of lift	floor–knuckle		floor–shoulder		floor–reach	knuckle–shoulder	knuckle–reach	shoulder–reach
	twisting angle	very small	small	moderate	large	very large			
	lowering demands	very low	low	moderate	high	very high			
Carrying	weight of load	very light	light	moderate	heavy	very heavy			
	travel distance	very short	short	moderate	long	very long			
	dynamic posture	walking		climbing stairs/ladders	crawling				
	carrying demands	very low	low	moderate	high	very high			
Pushing	force	very light	light	moderate	heavy	very heavy			
	travel distance	very short	short	moderate	long	very long			
	pushing demands	very low	low	moderate	high	very h igh			
Pulling	force	very light	light	moderate	heavy	very heavy			
	travel distance	very short	short	moderate	long	very long			
	pulling demands	very low	low	moderate	high	very high			
Other	force	very low	low	moderate	high	very high			
	travel distance	very light	light	moderate	heavy	very heavy			
	other demands	very low	low	moderate	high	very high			
Object handling demands		very low	low	moderate	high	very high			

Table 2. Evaluation of frequent object handling demands (i.e. evaluation of activities involving the load handled or force exerted on a frequent basis)

Lifting	weight of load	very light	light	moderate	heavy	very heavy			
	horizontal distance	very close	close	moderate	far	very far			
	frequency	very low	low	moderate	high	very high			
	work duration	very short	short	moderate	long	very long			
	height of lift	floor–knuckle		floor–shoulder		floor–reach	knuckle–shoulder	knuckle–reach	shoulder–reach
	twisting angle	very small	small	moderate	large	very large			
	lifting demands	very low	low	moderate	high	very high			
Lowering	weight of load	very light	light	moderate	heavy	very heavy			
	horizontal distance	very close	close	moderate	far	very far			
	frequency	very low	low	moderate	high	very high			
	work duration	very short	short	moderate	long	very long			
	height of lift	floor–knuckle		floor–shoulder		floor–reach	knuckle–shoulder	knuckle–reach	shoulder–reach
	twisting angle	very small	small	moderate	large	very large			
	lowering demands	very low	low	moderate	high	very high			
Carrying	weight of load	very light	light	moderate	heavy	very heavy			
	travel distance	very short	short	moderate	long	very long			
	frequency	very low	low	moderate	high	very high			
	work duration	very short	short	moderate	long	very long			
	dynamic posture	walking		climbing stairs/ladders			crawling		
	carrying demands	very low	low	moderate	high	very high			
Pushing	force	very light	light	moderate	heavy	very heavy			
	frequency	very low	low	moderate	high	very high			
	work duration	very short	short	moderate	long	very long			
	travel distance	very short	short	moderate	long	very long			
	pushing demands	very low	low	moderate	high	very high			
Pulling	force	very light	light	moderate	heavy	very heavy			
	frequency	very low	low	moderate	high	very high			
	work duration	very short	short	moderate	long	very long			
	travel distance	very short	short	moderate	long	very long			
	pulling demands	very low	low	moderate	high	very high			
Other	force	very low	low	moderate	high	very high			
	frequency	very low	low	moderate	high	very high			
	work duration	very short	short	moderate	long	very long			
	travel distance	very light	light	moderate	heavy	very heavy			
	other demands	very low	low	moderate	high	very high			
Object handling demands		very low	low	moderate	high	very high			

5. THE LINGUISTIC CHECKLIST APPROACH

The simplest instrument that can be designed to implement the comprehensive physical demand evaluation is the linguistic checklist approach (Karwowski *et al.* 1999). Tables 1 through 6 illustrate the different forms which can used to evaluate the different activities classified under physical demands. The checklist approach employed here differs from traditional checklists in that variables and activity demands are evaluated using five levels rather than a binary (i.e. absent or present) indicator. An example of the linguistic descriptors used in the tables is: "very light", "light", "moderate", "heavy", and "very heavy". Based on experience with the work presented here, we found that three levels are not adequate to provide a linguistic description of the variables and activities evaluated in the workplace. Additionally, five levels and "does not apply" were used in the Position Analysis Questionnaire, a widely used job analysis instrument (McCormick *et al.* 1969).

Tables 1 and 2 are used to evaluate object handling activities for strength- and endurance-related requirements. In table 1, the assessment is on the evaluation of the variables associated with the maximum load handled or force exerted on the job. Frequency and work duration are considered negligible and are not part of Table 1. On the other hand, they are added to other variables in order to evaluate frequent activities or endurance-related requirements in table 2. Under these circumstances, the evaluation is performed with reference to the most common load handled or force exerted on the job during frequent activities.

The evaluation of extremity postural demands is performed for dynamic activities (Table 3), continuous static force exertion (Table 4), and static awkward postures (Tables 5 and 6). It should be noted that static postural loading is evaluated for the most common static posture (Table 5) and the most awkward posture (Table 6) for a given body part. Wagner (1985) reported the evaluation of the most common and awkward postures in the "Job Profile" work assessment instrument.

Because the physical demand model is designed for the evaluation of most muscular demands in industry, there is a need to assess the degree of importance of different activities in order to determine whether they should be included in the physical demand evaluation. Importance here refers to the "significance", "consequential", "critical", "momentous", and "value-added" aspects of work demands relative to job performance.

Table 3. Evaluation of dynamic extremity postural demands (specify body position_____)

Finger	not applicable	very low	low	moderate	high	very high	
	frequency		very low	low	moderate	high	very high
	force/weight of load		very light	light	moderate	heavy	very heavy
	travel distance		very short	short	moderate	long	very long
	work duration		very short	short	moderate	long	very long
	finger demands		very low	low	moderate	high	very high
Hand	not applicable	very low	low	moderate	high	very high	
	freuency		very low	low	moderate	high	very high
	force/weight of load		very light	light	moderate	heavy	very heavy
	travel distance		very short	short	moderate	long	very long
	work durtation		very short	short	moderate	long	very long
	hand demands		very low	low	moderate	high	very high
Lower arm	not applicable	very low	low	moderate	high	very high	
	frequency		very low	low	moderate	high	very high
	force/weight of load		very light	light	moderate	heavy	very heavy
	travel distance		very short	short	moderate	long	very long
	work duration		very short	short	moderate	long	very long
	lower arm demands		very low	low	moderate	high	very high
Whole arm	not applicable	very low	low	moderate	high	very high	
	frequency		very low	low	moderate	high	very high
	force/weight of load		very light	light	moderate	heavy	very heavy
	travel distance		very short	short	moderate	long	very long
	work duration		very short	short	moderate	long	very long
	whole arm demands		very low	low	moderate	high	very high
Head	not applicable	very low	low	moderate	high	very high	
	frequency		very low	low	moderate	high	very high
	force/weight of load		very light	light	moderate	heavy	very heavy
	travel distance		very short	short	moderate	long	very long
	work duration		very short	short	moderate	long	very long
	head demands		very low	low	moderate	high	very high
Leg	not applicable	very low	low	moderate	high	very high	
	frequency		very low	low	moderate	high	very high
	force/weight of load		very light	light	moderate	heavy	very heavy
	travel distance		very short	short	moderate	long	very long
	work duration		very short	short	moderate	long	very long
	leg demands		very low	low	moderate	high	very high
Foot	not applicable	very low	low	moderate	high	very high	
	frequency		very low	low	moderate	high	very high
	force/weight of load		very light	light	moderate	heavy	very heavy
	travel distance		very short	short	moderate	long	very long
	work duration		very short	short	moderate	long	very long
	foot demands		very low	low	moderate	high	very high
Other	not applicable	very low	low	moderate	high	very high	
	frequency		very low	low	moderate	high	very high
	force/weight of load		very light	light	moderate	heavy	very heavy
	travel distance		very short	short	moderate	long	very long
	work duration		very short	short	moderate	long	very long
	other demands		very low	low	moderate	high	very high
Demands of dynamic work	not applicable	very low	low	moderate	high	very high	

Table 4. Evaluation of continous force (i.e. lasting more than 4 sec) extremity postural demands (specify body position:_____)

Fingers	not applicable	very low	low	moderate	high	very high
	force	very light	light	moderate	heavy	very heavy
	work duration	very short	short	moderate	long	very long
	static finger force demands	very low	low	moderate	high	very high
Hand unassisted by arm	not applicable	very low	low	moderate	high	very high
	force	very light	light	moderate	heavy	very heavy
	work duration	very short	short	moderate	long	very long
	static hand force demands	very low	low	moderate	high	very high
Hand assisted by arm	not applicable	very low	low	moderate	high	very high
	force	very light	light	moderate	heavy	very heavy
	work duration	very short	short	moderate	long	very long
	static hand force demands	very low	low	moderate	high	very high
Knee	not applicable	very low	low	moderate	high	very high
	force	very light	light	moderate	heavy	very heavy
	work duration	very short	short	moderate	long	very long
	static knee force demands	very low	low	moderate	high	very high
Foot unassisted by leg	not applicable	very low	low	moderate	high	very high
	force	very light	light	moderate	heavy	very heavy
	work duration	very short	short	moderate	long	very long
	static foot force demands	very low	low	moderate	high	very high
Foot assisted by leg	not applicable	very low	low	moderate	high	very high
	force	very light	light	moderate	heavy	very heavy
	work duration	very short	short	moderate	long	very long
	static foot force demands	very low	low	moderate	high	very high
Other	not applicable	very low	low	moderate	high	very high
	force	very light	light	moderate	heavy	very heavy
	work duration	very short	short	moderate	long	very long
	static other force demands	very low	low	moderate	high	very high
Static continuous force demands		very low	low	moderate	high	very high

Table 5. Evaluation of the most common static postural demands for extremity postural work(specify body position:_____)

Back	posture	bent forward(BF)	bent backward(BB)	twisted(T)	side bent	combined BF & T	combined BB & T
	level of angular deviation	very small	small	moderate	large	very large	
	work duration	very short	short	moderate	long	very long	
	back postural demands	very low	low	moderate	high	very high	
Head	posture	bent forward(BF)	bent backward(BB)	twisted(T)	side bent	combined BF & T	combined BB & T
	level of angular deviation	very small	small	moderate	large	very large	
	work duration	very short	short	moderate	long	very long	
	head postural demands	very low	low	moderate	high	very high	
Whole arm	posture	forward elevation – below shoulder		forward elevation – on or above shoulder		backward elevation	
	level of angular deviation	very small	small	moderate	large	very large	
	work duration	very short	short	moderate	long	very long	
	whole arm postural demands	very low	low	moderate	high	very high	
Lower arm	posture	bent(<=60 deg.)	medial rotation (MR)	side rotation (SR)	combined B & MR	combined B & SR	
	level of angular deviation	very small	small	moderate	large	very large	
	work duration	very short	short	moderate	long	very long	
	lower arm postural demands	very low	low	moderate	high	very high	
Hand	posture	bent forward (BF)	bent backward (BB)	side bent (SB)	combined BF & SB	combined BB & SB	
	level of angular deviation	very small	small	moderate	large	very large	
	work duration	very short	short	moderate	long	very long	
	back postural demands	very low	low	moderate	high	very high	
Overall static postural demands		very low	low	moderate	high	very high	

Table 6. Evaluation of the most awkward static postural demands for extremity postural work (specify body position:_____)

Back	posture	bent forward(BF)	bent backward(BB)	twisted(T)	side bent	combined BF & combined BB & T
	level of angular deviation	very small	small	moderate	large	very large
	work duration	very short	short	moderate	long	very long
	back postural demands	very low	low	moderate	high	very high
Head	posture	bent forward(BF)	bent backward(BB)	twisted(T)	side bent	combined BF & combined BB & T
	level of angular deviation	very small	small	moderate	large	very large
	work duration	very short	short	moderate	long	very long
	head postural demands	very low	low	moderate	high	very high
Whole arm	posture	forward elevation - below shoulder		forward elevation - on or above shoulder		backward elevation
	level of angular deviation	very small	small	moderate	large	very large
	work duration	very short	short	moderate	long	very long
	whole arm postural demands	very low	low	moderate	high	very high
Lower arm	posture	bent(<= 60 deg.)	medial rotation(MR)	side rotation(SR)	combined B&MR	combined B&SR
	level of angular deviation	very small	small	moderate	large	very large
	work duration	very short	short	moderate	long	very long
	lower arm postural demands	very low	low	moderate	high	very high
Hand	posture	bent forward(BF)	bent backward(BB)	side bent(SB)	combined BF&SB	combined BB&SR
	level of angular deviation	very small	small	moderate	large	very large
	work duration	very short	short	moderate	long	very long
	back postural demands	very low	low	moderate	high	very high
Overall static postural demands		very low	low	moderate	high	very high

REFERENCES

ARMSTRONG, T.J., RADWIN, R.G., HANSEN, D.J., and KENNEDY, K.W., 1986, Repetitive trauma disorders: job evaluation and design. *Human Factors*, 28, 325–36.

ASFOUR, S.S., GENAIDY, A.M. and MITAL, A., 1988, Physiological guidelines for the design of manual lifting and lowering tasks: the state of the art. *American Industrial Hygiene Association Journal*, 49, 150–60.

ASFOUR, S.S., TRITAR, M. and GENAIDY, A.M., 1991a, Endurance time and physiological responses for prolonged arm lifting tasks, *Ergonomics*, 34, 335–42.

ASFOUR, S.S., KHALIL, T.M., GENAIDY, A.M., AKCIN, M., JOMOAH, I.M., KOSHY, J.G. and TRITAR, M., 1991b, *Ergonomic Injury Control in High Frequency Lifting Tasks* (Coral Gables, FL: Industrial Engineering Department, University of Miami).

Association of Schools of Public Health, 1986, *Proposed National Strategies for the Prevention of Leading Work-Related Diseases and Injuries: Part 1*, Technical Report prepared for National Institute for Occupational Safety and Health (Springfield, VA: US Department of Commerce, National Technical Information Service).

ASTRAND, P-O. and RODAHL, K., 1986, *Textbook of Work Physiology: Physiological Bases of Exercise* (2nd ed.) (New York: McGraw-Hill).

AYOUB, M.A. and WITTELS, N.E., 1989, Cumulative trauma disorders. *International Reviews of Ergonomics*, 2, 217–72.

AYOUB, M.M., DEMPSEY, P.G. and KARWOWSKI, W., 1997, Manual materials handling. In G. Salvendy (ed.) *Handbook of Human Factors and Ergonomics* (New York: Wiley), 1085–1123.

AYOUB, M.M., SELAN, J. and LILES, D., 1983, An ergonomics approach for the design of manual materials-handling tasks. *Human Factors*, 25 (5), 507–15.

AYOUB, M.M., BETHEA, N.J., DEIVANAYAGAM, S., ASFOUR, S.S., MITAL, A., BAKKEN, G.M. and LILES, D., 1978, *Determination and Modeling of Lifting Capacity* (Lubbock, TX: Department of Industrial Engineering, Texas Tech University).

BURNETTE, J.T. and AYOUB, M.A., 1989, Cumulative trauma disorders: Part II Assessment of risk. *Pain Management*, July/August, 256—64.

Centers for Disease Control, 1983, Musculoskeletal injuries. *Morbidity and Mortality Weekly Report,* 32 (14), April 15.

CHAFFIN, D. and ANDERSSON, G.B., 1991, *Occupational Biomechanics* (New York: Wiley).

CHAFFIN, D. and PARK, K., 1973, A longitudinal study of low-back pain as associated with occupational weight lifting factors. *American Industrial Hygiene Association Journal*, 34, December, 513–25.

CHAFFIN, D.B., HERRIN, G.D., KEYSERLING, W.M. and FOULKE, J.A., 1978, Pre-employment strength testing: an updated position. *Journal of Occupational Medicine*, 20, 403–8.

DRURY, C. and PFEIL, R.E., 1975, A task-based model of manual lifting performance. *International Journal of Production Research*, 13, 137–48.

DRURY, C.G. and WICK, J., 1984, Ergonomic applications in the shoe industry. *Proceedings of International Conference on Occupational Ergonomics, Toronto*, 489–93.

FINE, L.J., PUNNETT, L. and KEYSERLING, W.M., 1987, An epidemiologic study of postural risk factors for shoulder disorders in industry. In P. Buckle (ed.) *Musculoskeletal Disorders at Work* (London: Taylor & Francis), pp. 108–9.

GENAIDY, A.M., 1996, Physical work capacity. In A. Bhattacharya and J.D. McGlothlin (eds) *Occupational Ergonomics: Theory and Applications* (New York: Marcel Dekker), 219–34.

GENAIDY, A.M. and ASFOUR, S.S., 1987, Review and evaluation of physiological cost prediction models for manual materials handling. *Human Factors*, 29, 465–76.

GENAIDY, A.M. and ASFOUR, S.S., 1989, Effects of frequency and load of lift on endurance time. *Ergonomics*, 32, 51–7.

GENAIDY, A.M., AL-SHEDI and KARWOWSKI, W., 1994, Postural stress analysis in industry. *Applied Ergonomics*, 25, 77–87.

GENAIDY, A.M., ASFOUR, S.S., MITAL, A. and TRITAR, M., 1988, Psychophysical modeling in manual materials handling. *Human Factors*, 30, 319–37.

GENAIDY, A.M., KHALIL, T.M., ASFOUR, S.S. and DUNCAN, R.C., 1990, Human physiological capabilities for prolonged manual lifting tasks. *IIE Transactions*, 22, 270–80.

GENAIDY, A.M., WALY, S.M., KHALIL, T.M. and HIDALGO, J., 1993, Compression tolerance limits of the lumbar spine. *Ergonomics*, 36, 415–34.

GRANDJEAN, E., 1988, *Fitting the Task to the Man* (London: Taylor & Francis).

GUO, L., GENAIDY, A., CHRISTENSEN, D. and HUNTINGTON, K., 1996, Macro-ergonomic risk assessment in nuclear remediation industry. *Applied Ergonomics*, 27, 241–54.

HAGBERG, M., 1984, Occupational musculoskeletal stress and disorders of the neck and shoulder: a review of possible pathophysiology. *International Archives of Occupational and Environmental Health*, 53, 269–78.

HIDALGO, J., GENAIDY, A.M., KARWOWSKI, W., CHRISTENSEN, D., HUSTON, R. and STAMBOUGH, J., 1997, A comprehensive lifting model: beyond the NIOSH lifting equation. *Ergonomics*, 40, 916–27.

KARWOWSKI, W., 1983, A pilot study of the interaction between physiological, biomechanical and psychophysical stresses involved in manual lifting tasks. *Proceedings of Ergonomics Society Conference* (London: Taylor & Francis), pp. 95–100.

KARWOWSKI, W. and AYOUB, M.M., 1984, Fuzzy modeling of stresses in manual lifting tasks. *Ergonomics*, 27, 641–9.

KARWOWSKI, W. and MARRAS, W., 1997, Cumulative trauma disorders. In G. Salvendy (ed.) *Handbook of Human Factors and Ergonomics* (New York: Wiley), 1124–73.

KARWOWSKI, W. and MARRAS, W. (eds), 1999, *Handbook of Occupational Ergonomics* (Boca Raton: CRC Press).

KARWOWSKI, W., MULHOLLAND, N.O., WARD, T.L., JAGANATHAN, V., and KIRCHNER, Jr., 1986, LIFTAN: An experimental expert system for analysis of manual lifting tasks. *Ergonomics*, 29, 1213–34.

KARWOWSKI, W., LEE, W.G., JAMALDIN, B., GADDIE, P. and JANG, R., 1999, Beyond psychophysics: a need for cognitive modeling approach to setting limits in manual lifting tasks. *Ergonomics*, 42(1), 40—60.

KEMMLERT, K., 1995, A method assigned for the identification of ergonomic hazards: PLIBEL. *Applied Ergonomics*, 26, 199–211.

KEYSERLING, W.M., HERRIN, G.D. and CHAFFIN, D., 1980, Isometric strength testing as a means of controlling medical incidents on strenuous jobs. *Journal of Occupational Medicine*, 22(5), 332–6.

KILBOM, A., 1988, Intervention programmes for work-related neck and upper limb disorders: strategies and evaluation. *Ergonomics*, 31, 735–47.

KILBOM, A., 1994, Repetitive work of the upper extremity: Part I. The scientific basis (knowledge base) for the guide. *International Journal of Industrial Ergonomics*, 14, 59–86.

LIFSHITZ, Y.L. and ARMSTRONG, T.J., 1986, A design checklist for control and prediction of cumulative trauma disorders in intensive manual jobs. *Proceedings of the Human Factors Society 30th Annual Meeting*, pp. 837–41.

LILES, D., DEIVANAYAGAM, S., AYOUB, M. and MAHAJAN, P., 1984, A job severity index for the evaluation and control of lifting injury. *Human Factors*, 26(6), 683–93.

MARRAS, W., FATHALLAH, F., MILLER, R., DAVIS, S. and MIRKA, G., 1992, Accuracy of a three-dimensional lumbar motion monitor for recording dynamic trunk motion characteristics. *International Journal of Industrial Ergonomics*, 9, 75–87.

MARRAS, W., LAVENDER, S., LEURGANS, S., RAJULU, S., ALLREAD, W., FATHALLAH, F. and FERGUSON, S., 1993, The role of dynamic three-dimensional trunk motion in occupationally-related low back disorders. *Spine*, 18(5), 617–28.

MARRAS, W., LAVENDER, S., LEURGANS, S., FATHALLAH, F., FERGUSON, S., ALLREAD, W. and RAJULU, S., 1995, Biomechanical risk factors for occupationally related low back disorders. *Ergonomics*, 38(2), 377–410.

McATMNEY, L. and CORLETT, E.N., 1993, RULA: a survey method for the investigation of work-related upper limb disorders. *Applied Ergonomics*, 24, 91–9.

McCORMICK, E., JEANNERT, P. and MECHAM, R., 1969, *The Development and Background of the Position Analysis Questionnaire* (Lafayette, IN: Occupational Research Center).

MITAL, M., 1984, Comprehensive maximum acceptable weight of lift database for regular 8-hr workshifts. *Ergonomics*, 27, 1127–38.

National Institute for Occupational Safety and Health (NIOSH), 1981, *Work Practices Guide for Manual Lifting*, NIOSH Technical Report (Cincinnati, Ohio: NIOSH).

National Institute for Occupational Safety and Health (NIOSH), 1996, *National Occupational Research Agenda*, US Department of Health and Human Services, DHHS Publication No. 97-17.

PASSMORE, R. and DURNIN, J.V.G.A., 1955, Human energy expenditure. *Physiological Reviews*, 35, 801.

PUNNETT, L., FINE, L.J. and KEYSERLING, W.M., 1987, An epidemiologic study of postural risk factors for back disorders in industry. In P. Buckle (ed.) *Musculoskeletal Disorders at Work* (London: Taylor & Francis), p. 74.

ROHMERT, W. and LANDAU, K., 1983, *A New Technique for Job Analysis* (London: Taylor & Francis).

RYAN, G.A., 1989, Musculoskeletal symptoms in supermarket workers. *Ergonomics*, 32, 359–71.

SHOAF, C., GENAIDY, A., KARWOWSKI, W., WATERS, T. and CHRISTENSEN, D., 1997, Comprehensive manual handling limits for lowering, pushing, pulling and carrying. *Ergonomics*, 40, 1183–1200.

SILVERSTEIN, B.A., FINE, L.J. and ARMSTRONG, T.J., 1986, Hand wrist cumulative trauma disorders in industry. *British Journal of Industrial Medicine*, 43, 779–84.

SNOOK, S.H., 1978, The design of manual material handling. *Ergonomics*, 21, 963–85.

SNOOK, S.H. and CIRIELLO, V.M., 1991, The design of manual handling tasks: revised tables of maximum acceptable weights and forces. *Ergonomics*, 34, 1197–1213.

SNOOK, S.H., VAILLANCOURT, D.R., CIRIELLO, V.M. and WEBSTER, B., 1995, Psychophysical studies of repetitive wrist flexion and extension. *Ergonomics*, 38, 1488–1507.

TANAKA, S. and McGLOTHLIN, J.D., 1989, A conceptual model to assess musculoskeletal stress of manual work for establishment of quantitative guidelines to prevent hand and wrist cumulative trauma disorders (CTDs). In A. Mital (ed.) *Advances in Industrial Ergonomics and Safety, I* (London: Taylor & Francis), 419–26.

TICHAUER, E.R., 1978, *The Biomechanical Basis of Ergonomics: Anatomy Applied to the Design of Work Situations* (New York: Wiley).

TROUP, J.D.G. and EDWARDS, F.C., 1985, *Manual Handling and Lifting* (London: HMSO).

VAN WELY, P., 1970, Design and disease. *Applied Ergonomics*, 1, 262–9.

WAGNER, R., 1985, Job analysis at ARBED. *Ergonomics*, 28, 255–73.

WALLACE, W. and BUCKLE, P., 1987, Ergonomic aspects of neck and upper limb disorders. *International Reviews of Ergonomics*, 1, 173–200.

WATERS, T.R., PUTZ-ANDERSON, V., GARG, A. and FINE, L.J., 1993, Revised NIOSH equation for the design and evaluation of manual lifting tasks. *Ergonomics*, 36, 749–76.

Webb Associates, 1978a, *Anthropometric Source Book: Vol. I Anthropometry for Designers* NASA Reference Publication 1024, National Aeronautics and Space Administration (NASA) Scientific and Technical Information Office.

Webb Associates, 1978b, *Anthropometric Source Book: Vol. II A Handbook of Anthropometric Data*, NASA Reference Publication 1024, National Aeronautics and Space Administration (NASA) Scientific and Technical Information Office.

Psychophysical Risk Assessment in Manual Handling

P. G. Dempsey

Liberty Mutual Research Center for Safety & Health, 71 Frankland Road, Hopkinton, MA 01748, USA

1. INTRODUCTION

Psychophysics, biomechanics, and physiology represent the three primary approaches to the ergonomic design of manual handling tasks. All three approaches consider the acute responses of the human body. Physiology and biomechanics typically consider the responses of the cardiovascular and musculoskeletal systems, respectively. Psychophysics, on the other hand, considers the perceptions (or sensations) of humans in response to stimuli of different magnitudes (typically load weight or forces) while performing manual handling tasks.

The primary goals of psychophysical data are for design purposes and to assess manual handling tasks at the workplace. At the design stage, such data can be used for relative comparisons of alternative designs. Workplace assessments most often involve a determination of whether or not a particular task, or set of tasks, should be redesigned through alteration of task variables, workplace geometry, equipment, and work organization. This entry provides an overview of performing such assessments in conjunction with psychophysical data as risk assessments.

2. RISK ASSESSMENT

A clear definition is required to discuss risk assessment in the context of the psychophysical approach to manual handling. Risk assessment is comprised of several components. The National Research Council (1983) has defined these components as (a) hazard identification, (b) dose-response assessment, (c) exposure assessment, and (d) risk characterization.

Hazard identification involves determining whether the manual handling tasks performed at the workplace include risk factors for musculoskeletal disorders (particularly of the lower back), risk factors such as awkward postures and high forces or loads handled. This step of the assessment does not usually employ psychophysical data and can be accomplished through a variety of means, including checklists. In some cases, passive surveillance data may have indicated that musculoskeletal disorder hazards exist. Then the hazard identification is sometimes skipped and the analyst proceeds to a detailed assessment of exposure to manual handling tasks.

The dose-response assessment involves examining the results of previous epidemiological studies. The dose-response relationships provide an assessment of the expected value of the outcome of interest (e.g., lower back disorder incidence) for a given dose. Exposure is often used in place of dose, as most studies of musculoskeletal disorder risk examine exposure-response relationships. As will be discussed in the next section, the major barrier to performing risk assessments with psychophysical data is the lack of dose-response (or exposure-response) relationships.

Exposure assessment involves measuring the frequency, magnitude, and duration of exposure to manual handling. The approach using psychophysical data to assess exposure will be discussed in the next section.

The final step in the risk assessment is the risk characterization. By this stage, the exposure assessment has been completed. These results are then used in conjunction with the exposure-response relationships to characterize the risk potential.

3. PSYCHOPHYSICAL RISK ASSESSMENTS OF MANUAL HANDLING TASKS

Hazard identification is usually fairly straightforward for manual handling tasks. The observer will typically look for factors such as awkward trunk postures (bending, twisting), reaching while handling, the handling of heavy loads, or high force exertions required by pushing and pulling. To some extent, the method of hazard identification is independent of the manual handling approach. The analyst may choose to use a checklist or other available tools. At this stage, it is best to be conservative, and if a task or set of tasks appears to pose marginal risk, the analyst should choose to proceed with a more detailed analysis.

A critical aspect of risk assessments is the exposure-response relationship. Unfortunately, we have limited knowledge of exposure-response relationships when exposure is quantified with psychophysical data. The available literature will be briefly summarized.

The study by Snook *et al.* (1978) provided the first epidemiological investigation of the psychophysical approach. It used data on 191 compensable lower back cases. The jobs in which the injuries occurred were analyzed using psychophysical data (Snook 1978), and a determination was made regarding whether or not the manual handling tasks accommodated more or less than 75% of the population. Using this approach, Snook *et al.* (1978) concluded that workers were three times more likely to file a worker's compensation claim when the job accommodated less than 75% of the population than when the job accommodated more than 75% of the population.

The second relevant study had a very different design. Liles *et al.* (1984) did a field study to determine whether the job severity index (JSI) was predictive of future injury. JSI is a time- and frequency-weighted ratio of worker capacity to job demands. Worker capacity was predicted with the models developed by Ayoub *et al.* (1978), which predict individual psychophysical capacity from worker strength and anthropometry values. A total of 453 subjects was used. The results of the field study indicated that both incidence and severity of recordable lower back disorders rose rapidly at JSI values greater than 1.5. Liles *et al.* concluded that JSI is an effective screening method and job design aid. JSI can be reduced to a desirable level by increasing worker capacity (e.g., selecting a worker with higher capacity) or altering task and job parameters to reduce JSI to an acceptable level.

Herrin *et al.* (1986) investigated the relationships between the percentage of the population accommodated according to Snook's (1978) data (most stressful task within a job and average of all tasks) and medical dispensary visits. There was a fairly strong negative correlation between percentage of the population accommodated by the most stressful task and visits for overexertion injuries and contact injuries. There did not appear to be

enough "back" incidents to make a reliable conclusion, but the relationship was also a negative correlation.

The exposure assessment step requires measuring the manual handling task parameters required to apply the psychophysical data being used. The first step of the exposure assessment requires the analyst to perform a task description of the job, i.e., each individual manual handling task needs to be identified. Once all tasks have been identified, the critical parameters of each task need to be measured. Examples of parameters include task frequency, lifting range, vertical hand height, initial pull forces, and carry distance.

The final step is the risk characterization, which uses the exposure-response relationships and the exposure assessment. The studies discussed earlier do not provide explicit quantitative exposure-response relationships. The Snook *et al.* (1978) study used a qualitative exposure measure in the analyses (exposure was dichotomized as above or below 75% of the population being accommodated). Liles *et al.* (1984) presented quantitative analyses, but the JSI represents stress relative to psychophysical capacity for an individual, not in terms of the percentage of the worker population accommodated. The Herrin *et al.* (1986) study used visits to the medical dispensary as the response measure, which is fairly qualitative since this outcome encompasses a very wide range of outcomes, especially in terms of severity. Thus, the risk characterizations using the exposure assessment relationships from these studies must be performed with the limitations of the study designs in mind, i.e., the exposure assessment results need to be expressed in terms compatible with the results of the study or studies used.

4. CONCLUSIONS

Using the formal definition of risk assessment, it is arguable whether risk assessments can actually be conducted when using the psychophysical approach. This problem is not unique to psychophysics, but is pervasive in the field of manual handling (Dempsey 1998). Perhaps the problem is that the term "risk assessment" is used too often without careful consideration of its meaning.

Given the limitations in performing risk assessments, one possible alternative is to call assessments performed with psychophysical data (and other manual handling criteria) compliance assessments. That is, the assessment determines whether the tasks are in compliance with the criterion of interest.

Aside from compliance assessment, psychophysical criteria also provide valuable comparisons of different work designs. This use is too often overlooked when the manual handling criteria are discussed. Whether at the design stage, or when comparing alternate redesigns for existing tasks, psychophysical criteria can provide useful relative comparisons. These relative comparisons are not risk assessments, but provide guidance to help the designer select the best alternative with respect to the acute responses of the human body to the different designs.

REFERENCES

AYOUB, M.M., BETHEA, N., DEIVANAYAGAM, S., ASFOUR, S., BAKKEN, G., LILES, D., SELAN, J., and SHERIF, M., 1978, Determination and modeling of lifting capacity, Final Report, HEW (NIOSH) Grant No. 5R010H00545–02.

DEMPSEY, P.G., 1998, A critical review of biomechanical, epidemiological, physiological and psychophysical criteria for designing manual materials handling tasks. *Ergonomics*, 41(1), 73–88.

HERRIN, G.D., JARAIEDI, M., and ANDERSON, C.K., 1986, Prediction of overexertion injuries using biomechanical and psychophysical models. *American Industrial Hygiene Association Journal*, 47, 322–30.

NATIONAL RESEARCH COUNCIL, 1983, *Risk assessment in the federal government: managing the process* (Washington DC: National Academy Press).

LILES, D.H., DEIVANAYAGAM, S., AYOUB, M.M., and MAHAJAN, P., 1984, A job severity index for the evaluation and control of lifting injury. *Human Factors*, 26(6), 683–93.

SNOOK, S.H., 1978, The design of manual handling tasks. *Ergonomics*, 21(12), 963–85.

SNOOK, S.H., CAMPANELLI, M.S., and HART, J.W., 1978, A study of three preventive approaches to low back injury. *Journal of Occupational Medicine*, 20(7), 478–81.

Rating Scales for Perceived Physical Effort and Exertion

G. Borg

Department of Psychology, Stockholm University, Sweden.

1. INTRODUCTION

Interest in subjective aspects of physical work has increased during the last decades. One main reason for this is the improvement in rating methods designed to measure perceptual attributes. Another reason is the increased understanding of how important it is to listen to the worker who is doing the job. At the beginning of the last century, it was common to simply select individuals for a demanding task, instead of changing and adapting the task to the individual. This approach caused a lot of problems because individuals differ in terms of gender, age, and culture, and because educational and training opportunities are also variable. In the long run, it also led to risks for the individual and expensive medical care. Scaling subjective somatic symptoms and difficulty-of-work tasks are, therefore, important and necessary complements to physical and physiological measurements.

2. THE PURPOSE OF SCALING PERCEIVED EFFORT AND EXERTION

2.1. ''Psychophysical Scaling''

A main purpose of "psychophysical scaling" (meaning measuring perceptual magnitudes) is to complement objective measurements with subjective estimates. Valid physiological measurements are not always easy to obtain in field situations. This is especially true of certain short-term work carried out using bad postures. An advantage in using a rating scale to estimate physical work load is that a direct and individualized measure is immediately obtained.

2.2. Perceptual Estimates

Perceptual estimates integrate many cues from the peripheral musculoskeletal system and the cardiovascular and respiratory organs into a kind of "Gestalt" — a representation of the exertion as a whole. The human sensory system is an important information system that integrates the cues and give weights to those that are most important. Still it is possible to focus on separate aspects, such as local muscular effort and fatigue, breathlessness, or joint pain.

2.3. Subjective Evaluation

A subjective evaluation is, however, not merely a complement to "objective" measures. It has a significance and value in itself, since all individuals act according to what they perceive, how they evaluate a work task, and according to how motivated they are to learn, perform and cooperate. By adding subjective estimations one more "information channel" is added.

3. REQUIREMENTS OF A SCALING METHOD

3.1. Basic Demands

Basic demands on a rating method are that it must be reliable and valid to be useful for analyses and predictions. It should be simple to be practical in most situations. But more than a good scale is needed. The total methodology is very important, including good administration and instruction, both to the test leader and to the person making the ratings.

3.2. Classification of Scales

S.S. Stevens at Harvard University made a great contribution to the field when he classified scales according to their metric properties. Ordinal scales can only measure responses arranged in rank order. Interval scales have scale steps distributed at equal intervals, but no absolute zero. Ratio scales have the best metric properties, that is they have an absolute zero and scale values are at equal intervals. Only ratio scales allow for most kinds of mathematical calculations.

3.3. Scale Properties

It is important to distinguish between mathematical requirements and abstract scale properties, and the actual scaling and its purpose. Stevens' goal was to construct a method that could generate responses along a ratio scale. This goal was never reached, mainly because of cognitive difficulties and "response biases". However, if most people use numbers in a similar way, that is, if there is a high degree of intersubjectivity in number conceptions and usages, then reliable invariances and valid predictions may be obtained.

4. DIFFERENT KINDS OF SCALING METHODS

4.1. Ordinal Rating Methods

The ordinal scales are often called category scales or just simple rating scales. A typical category scale has numbers from, for example, one to five, with verbal anchors such as "very weak", "weak", "medium", "heavy", and "very heavy". To evaluate the effort and exertion of a work task, ordinal methods are mainly used to give very rough evaluations of intensity levels, not to scale relations between work tasks. The different scale values serve primarily as alternatives, such as in a multiple choice task.

4.2. Acceptability and Preference Scaling

In many work tasks it is important to obtain individualized measurements of preferred or acceptable workloads. There is a range of subjective intensities and workloads within which an experienced worker can adapt his or her level and work comfortably and efficiently for a long time-period. In contrast to most scaling methods in which stimuli are rated, with acceptability/preference scaling individuals are asked to adjust or produce a physical intensity until it is perceived to be just right according to the instruction. These methods are very useful in many situations, especially when it is desirable to obtain a certain level or short range of intensities (see Mital *et al.* 1993).

4.3. Equidistant Partition Scaling

4.3.1 The visual analogue scale (VAS)

A visual analogue scale usually consists of a simple line, often ten centimeters long, anchored at one end with the word "Minimum" and at the other with "Maximum". The subject is asked to make a mark on the line that shows the intensity of the perception. Sometimes the lowest end is named "Nothing at all" or "None at all" and the highest "Worst exertion (or pain)

imaginable". If a VAS is anchored in this way, it may be inappropriate for making interpersonal comparisons, since people can imagine so many different things, and from many different points of view. With appropriate administration VAS may be a useful instrument.

4.3.2 Paired comparisons

The method of paired comparisons may be classified into this group of equidistant scales. If the distribution of perceptual estimates is well known, e.g. a normal distribution, results from a comparison task may be placed on an interval scale based on standard deviations. However, such a scale is not a direct scaling method (as dealt with in this article), but an indirect one belonging to the set of discrimination scaling procedures. It is time consuming, but in some situations very useful (see Gescheider 1997).

4.3.3 The Borg RPE scaling

Very few category scales belong to the class of interval scales. If people are instructed well and can understand the concept of equisection, a scale of interval character may be obtained. The "Borg RPE scale", or just simply the "RPE scale" (from Ratings (R) of Perceived (P) Exertion (E), sometimes also called the "Borg

6   **No exertion at all**

7

  Extremely light

8

9   **Very light**

1 0

1 1   **Light**

1 2

1 3   **Somewhat hard**

1 4

1 5   **Hard   (heavy)**

1 6

1 7   **Very hard**

1 8

1 9   **Extremely hard**

2 0   **Maximal exertion**

Figure 1: The Borg RPE scale, the 15-grade scale for ratings of perceived exertion

scale") is commonly used in ergonomic evaluations of perceived workloads. Constructed by Borg in the 1960s and slightly modified later (see Borg 1998), it may be characterized as an interval scale. The RPE scale differs from the old rating scales and also from the "ratio scales" (see S.S. Stevens 1975) with regard to determinations of levels of intensity and their interrelations. A numerical scale from 6 to 20 was selected, to cover a roughly similar number range as that of heart rate (HR), from 60 to 200 bpm (divided by 10). Verbal expressions were placed on the scale such that a linear growth function was obtained between given numerical ratings and workloads for exercise on a bicycle ergometer. In this way, RPE values were also made to increase linearly with HR and oxygen consumption.

High reliability coefficients have been obtained, especially if a whole range of intensities is used in the evaluation. When expressed as correlation coefficients from data obtained with parallel testing or re-testing, most coefficients are above 0.90. Also, very good validity coefficients are obtained, e.g. correlations between HR and RPE, between 0.70 and 0.90. Good predictive power to performances has also been obtained, similar to that of HR.

4.4. Relative Ratio Scaling

S.S. Stevens developed the "ratio scaling" methods in psychophysics (see Stevens 1975). In analogy with scaling in the natural sciences, an arbitrary unit was used on a ratio scale. Several methods were developed, the most common being "magnitude estimation" (ME). In ME a physical standard is chosen and a certain number is assigned as the "modulus", e.g. ten (10) or hundred (100). Subjects assign numbers to each stimulus in relation to the standard such that the numbers reported represent the relative perceptual magnitudes. In "absolute magnitude estimation" (AME) no standard and modulus are given, and subjects report the magnitudes according to some "absolute" inner feeling. The aim was to improve estimates of "absolute" intensity levels. However, AME does not function much better than ME in this respect.

Stevens' "ratio scaling" methods were not as good as he had hoped. This is because human beings, when estimating intensities with numbers, are using their own number conceptions and ideas instead of abstract concepts. However, the methods functioned — and still do — well enough to make possible reliable descriptions of relative growth functions for different sensory modalities. For perception of effort and exertion, positively accelerating functions have been found. This is true of the perception of weight in lifting tasks, of subjective force of handgrip or pedal resistance, and of subjective exertion in short- and long-duration work on an ergometer. These functions can be described by power functions with exponents between 1.3 and 1.8. The differences depend both upon kind of work and scaling methods.

4.5. Level-anchored Ratio Scaling

4.5.1 Basic principles

The latest generation of direct scaling methods are the category (C)-ratio (R)-scaling methods. That a rating scale, such as ME, has properties of a ratio scale may be a necessary condition, but it is not a sufficient one. The CR scales constructed by Borg incorporate the best properties of category scales and those of ratio scales. They are constructed for measuring most kinds of perceptual attributes.
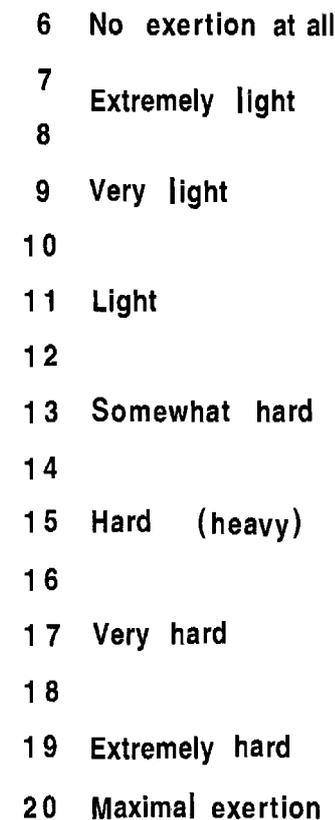
An advantage with "ratio scaling" is that it permits rough descriptions of relative growth functions. However, a main drawback is that valid determination of direct intensity levels is not possible. (That one weight is perceived to be twice as heavy as another does not tell us if the weight is heavy or light). The old category-rating methods, however, were to get estimates of intensity levels.

There are several fundamental principles behind the development of the CR scales. They were constructed according to internal psychophysical criteria, without considering any special relations to physiological correlates. A basic starting point was the results obtained using "ratio scaling", and a new method should give a similar growth function as obtained with these methods. The range principle was another fundamental principle, implying that the whole perceptual range from zero to maximum should be covered. Simple verbal expressions showing a high degree of inter-individual agreement were then chosen. The expressions were positioned on a ratio scale so that a congruent relation was obtained between the meaning of the numbers and the meaning of the expressions. Possibilities were also opened up to give subjects the opportunity to go above the highest scale value and below the lowest one, thus encompassing extreme situations — for example, if a person experiences exertion that is greater than ever previously experienced. A special dot was therefore placed above the scale to encourage people to give higher ratings if necessary.

4.5.2 The Borg CR10 scale

The first CR scale commonly used internationally is the CR10 scale. It has a numerical range from 0 to 10, with the first verbal anchor above 0 set to 0.5. In this way a suitable numerical range

0	Nothing at all	"No P"
0.3		
0.5	Extremely weak	Just noticeable
0.7		
1	Very weak	
1.5		
2	Weak	Light
2.5		
3	Moderate	
4		
5	Strong	Heavy
6		
7	Very strong	
8		
9		
10	Extremely strong	"Max P"
11		
⚡		
●	Absolute maximum	Highest possible

Figure 2: The Borg CR10 scale, a category (C) scale with ratio (R) properties

of 1:20 was obtained (see Borg 1998 and figure 2 below). The dot above 10 means that people may use 11 or 12, or an even greater number. There is usually no need to go above 12 in ratings of physical exertion. If higher ratings are given, they usually depend upon special personality characteristics and rating behavior, or pathological circumstances, such as extreme pain, and should be further investigated.

The reliability and validity of the CR10 scale is as good as that of the RPE scale. However, one special advantage is that the CR10 scale allows determinations of relative growth functions of a more absolute character than the RPE scale.

4.5.3 The cM scale

In some special cases, a more finely graded scale may be needed. A hundred point scale, the cM scale (centi-Max scale, also called CR100), has therefore been constructed by Borg and Borg in 1994; (see also Borg 1998). The responses obtained with this scale closely follow percentage ratings when tested with subjects that have a good ability to perform percentage estimates. To help individuals with difficulties in this respect, and to ensure a high degree of interpersonal agreement, numbers are anchored with verbal expressions. These are placed at about the same positions as on the CR10 scale. Of special interest is that physiological and performance measurements may also be transformed to one and the same CR scale.

5. BASIC FACTORS "BEHIND" RPE RESPONSES AND INDIVIDUAL DIFFERENCES

5.1. Variance Factors

Most of the differences in RPE responses ("RPE" here used independent of kind of scale) can be attributed to physiological factors. In homogeneous groups of healthy individuals as much as 60–70% of the variance can be "explained" by factors such as oxygen uptake and HR, ventilation and respiration rate, EMG measures, perspiration rate, blood lactate, blood pH, monoamines, and catecholamines, among others. A very difficult problem, then, is how to integrate and give weight to all factors for different kinds of work. An advantage with the perceptual system is that it can integrate information related to most of these factors.

5.2. Psychological Factors

A rather large part of the variance in RPE responses may, however, be attributed to psychological factors. In normal cases, one major factor influencing ratings is personality and rating behavior. Emotional and motivational factors may also play an important role.

5.3. Working Capacity Factor

Because of the strong dependence on physiological "causes", one main factor explaining the level of a given rating is the individual's working capacity. A given RPE-rating thus decreases systematically with increasing capacity. However, when the capacity is adjusted there is no real difference in obtained ratings. The same is also true when ratings are plotted against HR. In those cases there is no gender difference. When RPE values are studied as a function of age, they increase with decreasing working capacity. However, since maximal HR also decreases with age, relative HRs have to be calculated for different age groups. The relation is then the same as in age homogeneous groups.

5.4. Individual Differences and External Factors

Differences between individuals are found depending upon kind and duration of work, and the influence of external factors, such as temperature and altitude, and also upon actual somatic factors related to food intake, liquid and medication. Various somatic problems, such as musculoskeletal and cardiovascular diseases, may have a strong influence on RPE values and the profile of different symptoms (e.g. muscle ache, breathlessness and joint pain).

6. DIFFERENT APPLICATIONS

6.1. Physical Workload Evaluation

In order to evaluate the exertion and difficulty of a work task in manual materials handling, RPE-values are often registered. The purpose may be to compare different performance techniques, such as lifting according to the squat or the stoop technique, cleaning the floor by mopping or swabbing, or to evaluate the "absolute" magnitude of effort and exertion. Of special interest are gender differences. It is, of course, of utmost importance that all ergonomic improvement be carried out so as to avoid gender discrimination. Another target group is the elderly or handicapped, who need help with better managing their daily activities and improving well-being.

6.2. Mental Workload Evaluation

Another field of application of the CR10 scale is in human factors studies of mental workload. The difficulty of tasks involving fast and correct perceptual identification of signals and symbols, or psychomotor performance is, thus, analogous to physical work load. Simple evaluations of traffic noise, electric lighting, bad air, etc. may also be made using the CR-scaling technique. Still another field of application concerns determinations of the intensities of moods and emotions, such as the discomfort of traffic noise.

6.3. Risk Assessments

Because ratings give important information about actual strain, they are also used in risk assessments. No regular exertion should be "strong" and "heavy". "When it's hard it's too hard", meaning that RPE values should be less than 15, or CR10 values less than 5. A "moderate" (CR10 "3") level may often be just right (see Colombini *et al.* datum). Because there may be "silent symptoms", a medical examination is always necessary. Subjective ratings can never guarantee that no risks are involved in a certain performance. The sensory system is far from perfect, nor do physiological signs always provide enough information. Work considered to be risky includes low intensity but long-lasting repetitive activities of a "monotonous" kind as, for example, work in the fast check-out line or with steady computer work. For this kind of work, even "light work" may be too hard in the long run. Ratings are often utilized in clinical diagnostics, and in training and rehabilitation.

CONCLUSION

In most situations, ratings of effort and exertion give important information. They are valuable complements to physiological and biomechanical measurements, and are also important in themselves, since people act according to their perceptions of work activities. It is necessary to use a rating scale with good metric properties. This is, however, not a sufficient condition. If a method is to provide valid results, correct instruction and administration are essential.

REFERENCES

AYOUB, M. M. and DEMPSEY, P. G. , 1999, The psychophysical approach to manual materials handling task design. *Ergonomics*, 42(1) 17–31.

BORG, G., 1998, *Borg's Perceived Exertion and Pain Scales* (Champaign, IL: HK).

COLOMBINI, D., OCCHIPINTI, E., DELLEMAN, N., FALLENTIN, N., KILBERN, A. and GRIECO, A., 1998, Exposure assessment of upper limb repetitive movements: A consensus document by the IEA Technical Group for WMSD's. Research Unit Ergonomics of Posture and Movement, Milano, Italy.

GESCHEIDER, G.A., 1997, *Psychophysics: The fundamentals* (3rd ed.) (London: Lawrence Erlbaum).

STEVENS, S.S., 1975, *Psychophysics: Introduction to its perceptual, neural and social prospects* (New York: Wiley).

Signal Detection Theory

J. D. Balakrishnan and J. A. MacDonald

Department of Psychological Sciences, Purdue University, West Lafayette, IN 47907, USA

1. INTRODUCTION

Many of the jobs carried out by human operators in the workplace have a "detection" component to them. Various kinds of watchkeeping tasks (e.g. radar monitoring, airport security, product quality control) are some of the more obvious examples. Other seemingly more complicated tasks, such as medical diagnosis and personnel evaluation, can also be reduced in many instances to a set of all-or-none decisions about the need for an intervention of some kind. In virtually all of these cases, the potential for human error exists, and a critical question for the human factors specialist is whether some or all of these errors can be attributed to operator training, system design, or other aspects of the job that could be retooled in some way.

To answer this question, many researchers prefer to use well-controlled experiments and objective performance indices, if they are feasible. In addition to a rigorous measure of the difficulty of an operator's task and a numerical scale for comparing different approaches to a problem, these indices make it possible for the researcher to examine how factors such as attention and strategy may interact with system design or other factors to determine the overall performance pattern of the operator. Unfortunately, none of the performance indices currently available to researchers are "theory-independent." Essentially, they are byproducts of a relatively simple, skeletal model of the operator's behavior in a goal-directed setting. To be used appropriately, therefore, their assumptions and limitations need to be properly understood. Here we briefly review the several measures and experimental techniques that originated with one of the most successful performance theories in psychology, the so-called Theory of Signal Detection (or, Signal Detection Theory; e.g. Green and Swets, 1966). Elsewhere we consider some alternatives to signal detection theory and some recent empirical results that may fundamentally change the way operator performance is assessed in the future.

2. FORMAL MODELS OF DETECTION

In most real world tasks, the operator can and will make different kinds of errors, with different consequences, and the rate at which some kinds of errors are committed will often be inversely related to the rate of other kinds of errors, presumably because of attention shifts or other changes in the operator's approach to the task. In other words, there will be trade-offs. Changes in the environment or system design presumably can affect the decision-making strategy of the operator, exclusively or in addition to their effects on the quality of the information exchange between humans and machines. Formal models such as signal detection theory represent attempts to provide measures of information processing capacity that are not confounded by the operator's decision-making biases.

The first step in the development of such a model is to analyze the logical structure of the operator's task. With few exceptions, human errors in the workplace are incorrect decisions about the state of a system or real world event, or about the best possible course of action in response to a situation. Signal detection theory encompasses both kinds of errors by dividing the decision-making process into two discrete, non-overlapping stages. In the first stage, the operator collects information from the outside world, and in the second stage the operator applies a decision-making strategy or rule to the information state to arrive at a decision about the nature of the circumstances and/or the best course of action. Of course, in most workplace environments, the situations faced by the operator are constantly changing and so this encoding/decision-making sequence must continuously repeat itself. The absence of any overt action on the part of the operator at any given point in time presumably represents a decision (conscious or otherwise) not to react on the basis of current information about the environment.

The terms "collect information" and "apply a rule" are very general, which is one reason why signal detection theory can be so widely used. Any kind of information processing task will involve information collection of some kind followed by the selection of an appropriate action. Errors can be due to poor information (encoding errors) or faulty decision-making (response selection) strategies. The main objective of the signal detection theory analysis is to separate and quantify these two kinds of errors so that their relative frequencies and their dependence on the properties of the system can be studied.

To do this, the theory borrows some fundamental concepts from the statistical decision-making literature. First, the information that reaches the decision-maker is assumed to be ambiguous or "noisy": the same information state can be produced by more than one of the physical events that need to be discriminated. The decision-maker's problem is, therefore, to determine which event is more likely to have produced this information state and from this the action that is most likely to be correct. In this way, the decision-making process is analogous to a statistical hypothesis test. The operator makes an error when the information output of his/her encoding process is sufficiently misleading, or when the statistical decision rule is misconstrued.

At least in principle, this theoretical framework can be adapted to arbitrarily complex decision-making tasks, involving many different situations and many possible decisions that might need to be taken (and to some extent, Thurstonian Scaling is an example of this; Torgerson 1958). However, most of the efforts to develop the theory have concentrated on relatively simple decision-making tasks in which there are only two possible states and, therefore, two possible classification responses. In this two-choice classification (or discrimination) task, the two possible states of the world can be arbitrarily labeled "noise" and "signal," and there are two possible correct responses (correct signal responses or "hits" and correct "noise" responses, or "correct rejections") and two kinds of errors (incorrect "signal" responses to noise or "false alarms" and incorrect "noise" responses to signals, or "misses"). Signal detection theory assumes that the information state of the operator can be represented by a point on a one-dimensional, bipolar continuum or "decision axis." Values on the one end of this continuum represent instances during the experiment in which the operator is highly certain that the correct response decision should be "noise," and values the other end represent high confidence "signal" responses. Somewhere between
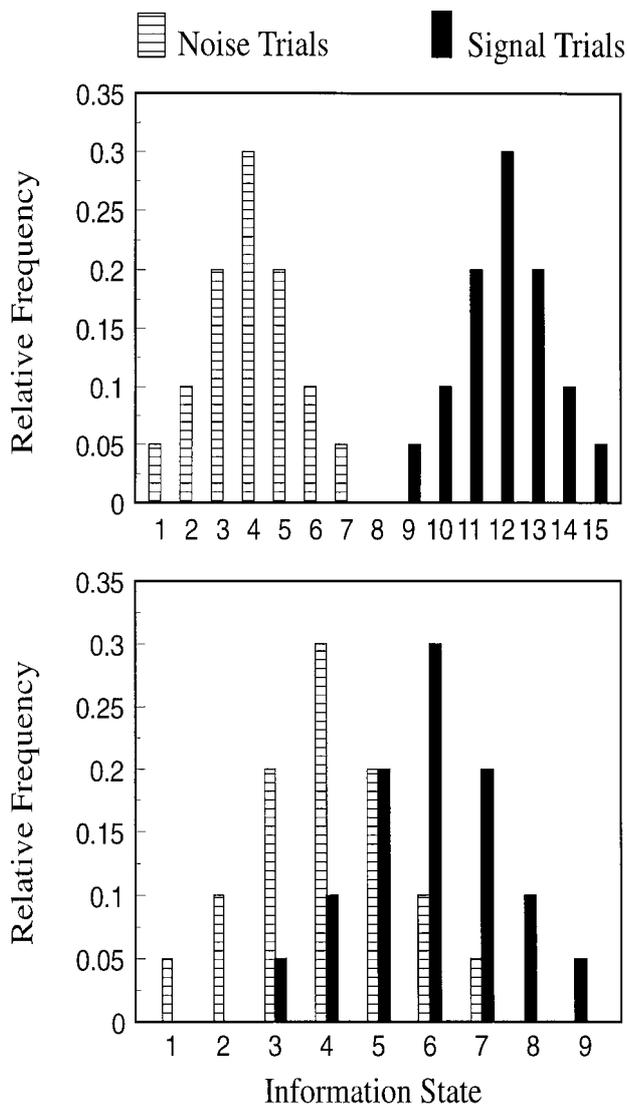
Figure 1. Concept of sensitivity in signal detection theory. (upper) The presentation of a signal puts the operator into one of seven states (9–15 on the abscissa). Noise alone never causes any of these information states, but instead may cause one of a different set (1–7). In this case, the operator can (with an appropriate decision rule) perfectly discriminate between the two events. (lower) Some of the information states (3–7) are caused by both signal and by noise alone (i.e. the distributions overlap), making discrimination errors inevitable for all possible decision rules.

these two extremes is the "complete uncertainty" or "indifference" point.

2.1. Encoding Distributions and Detection Criteria

If the presentation of a signal (or noise alone) does not always have the same effect on the operator, then it follows that the information state of the operator has some univariate probability (relative frequency) distribution. The difference between this distribution of states on noise trials from its distribution on signal trials presumably depends on the physical differences between

the two events and the "sensitivity" of the operator to these differences. In fact, the degree overlap of these two distributions would be the degree to which the noise and signal stimuli are confusable when they are presented equally often. This property is illustrated in Figure 1. In the upper panel, the distributions are completely non-overlapping. In this case, the decision-maker can always correctly identify the stimulus because the information states caused by the noise stimulus are never caused by the signal stimulus, and vice versa. In the lower panel, the distributions overlap. In this case, information states lying close to the middle of the continuum have similar relative frequencies under the two stimulus conditions, and perfectly equivalent relative frequencies for one value (information state "5"). No matter which response is assigned to information states "3" through "7," the operator's decision will sometimes be incorrect.

Using a probability distribution to describe the effect a stimulus on the operator was standard practice long before signal detection theory was developed. The new contribution of this theory was in its emphasis on the sophisticated decision-making processes that should be (and presumably are) applied by human operators to minimize the problems caused by noisy encoding. In addition to the relative frequency of a state on signal trials (or on noise trials), the statistical decision-maker would also consider the relative frequencies, or "base rates," of the signal and noise events during the experiment, the relative costs of the two possible errors (false alarms and misses), and the rewards for the two possible correct decisions (hits and correct rejections). For example, if the signal occurs very infrequently (i.e. its base rate is low), then a state that occurs with moderate frequency on signal trials and moderator frequency on noise trials would actually constitute fairly strong evidence in favor of a "noise" response. Presumably, the operator combines knowledge about the probability distributions with the base rates and payoffs to select an appropriate "criterion" or "decision boundary" to divide the information state continuum into two response regions. As the base rate of the signal event increases, for example, the operator presumably shifts the criterion to the left, increasing the size of the signal response region and hence the relative frequency of the signal response.

3. PERFORMANCE INDICES

Different assumptions about the shapes of the information state distributions partition signal detection theory into several different detection models. The most popular of these assumes that the distributions are normal with equal variance. In this case, the distance between the means of the two distributions in standard deviation units, or d', is a suitable index of the degree to which the distributions overlap ("sensitivity"), and the ratio of the two distributions at the criterion (b) is an index of preference for one of the two responses ("response bias"). Thus, b < 1 would indicate a bias towards the signal response and > 1 would indicate a bias towards the noise response. The decision rule is unbiased (b = 1) when the criterion is set at the point of intersection between the two distributions, which occurs at the midpoint between their means. Both d' and b can be calculated from the hit and false alarm rates of the operator.

If the distributions are normal but have different variances, then an additional parameter is needed to fix the sensitivity scale. Unfortunately, these indices cannot be obtained from a single

pair of hit and false alarm rates, which presumably explains why they are not widely used, despite the fact that the equal variance assumption is generally untenable. To fit the unequal variance normal model to data, estimates of the hit and false alarm rates must be obtained for several different values of the decision criterion.

3.1. Criterion Shifts and the ROC Curve

Running the same experiment under several different base rate or payoff conditions is one way to obtain the extra data needed to allow for unequal variance. Another method is simply to instruct the operator to favor one type of response over the other to some degree. Each manipulation should cause the operator to shift the detection criterion. The different pairs of hit and false alarm rates can then be graphed together in a single parametric plot (or scatterplot) called the "receiver operating characteristic" (ROC) curve. The points on the graph are the hit rates (on the ordinate) corresponding to each observed false alarm rate (on the abscissa). Two examples are shown in Figure 2. For several reasons the shape of the ROC curve is the most powerful diagnostic from the point of view of signal detection theory. The area under the curve, for example, increases as the sensitivity of
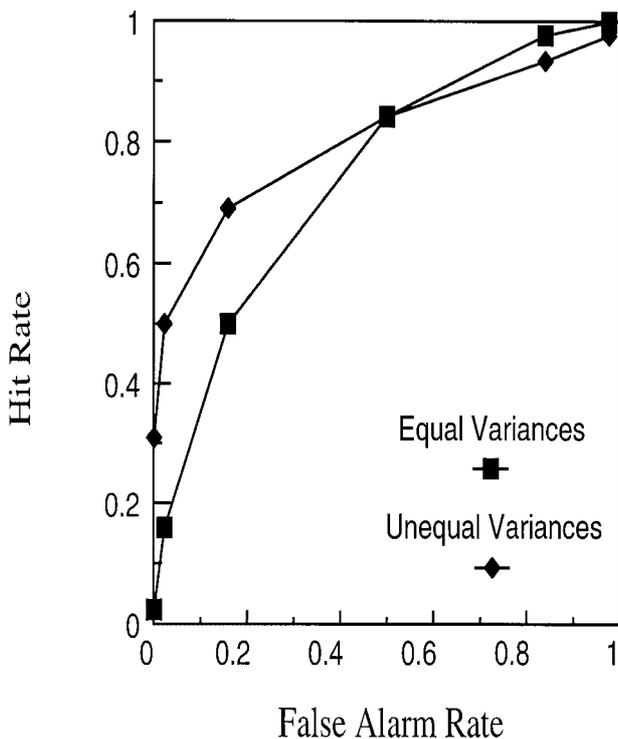
Figure 2. Two possible ROC curves when the distributions are normal with equal variance (squares) or with larger variance in the signal distribution (diamonds). Each of the six different points on the curve corresponds to a different placement of the detection criterion. Area under the complete ROC curve (when the hit and false alarm rate are calculated for every possible location of the detection criterion) can be estimated by computing the area under the these incomplete functions. In this case, however, the operators' placement of the criteria will have some effect on the measure.
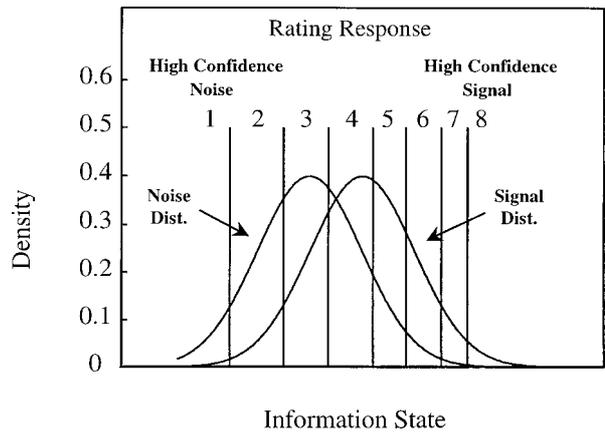
Figure 3. Signal detection theory representation of confidence rating data when there are eight possible rating responses on a single bipolar rating scale. Small values on the scale (e.g. "1") are high confidence "noise" responses and large values (e.g. "8") are high confidence "signal" responses. Each of the seven criteria is a possible value of the detection criterion, making it possible to calculate seven different points on the ROC curve from the operator's ratings data. For example, the proportion of rating responses > 3 on noise trials and the proportion of rating responses > 3 on signal trials are the false alarm and hit rates respectively when the detection criterion is placed at the same location as the upper bound on a "3" response.

the operator increases. Unequal variance in the distributions causes the function to be skewed with respect to the negative diagonal (the diagonal line connecting the upper left corner of the graph to the lower right corner), and the direction of the skew indicates which distribution variance is larger (assuming that the distributions are normal). Alternatively, the researcher can plot the z-transform of the false alarm rate against the z-transform of the corresponding hit rate and observe the shape of this "z-ROC" scatterplot. If the normality assumption is satisfied, this function will be linear with a slope equal to the ratio of the standard deviation of the noise distribution to the standard deviation of the signal distribution.

3.2. Ratings Method

Empirical studies of the shape of a ROC curve can be time consuming and expensive if each point of the curve must be estimated from a different experimental condition. An alternative approach that is perfectly consistent with the signal detection theory assumptions, and at the same time considerably more efficient in terms of costs, is to elicit confidence rating responses from the operator. According to signal detection theory, the operator's information is "graded" and stochastic, varying from strong and reliable on some occasions to weak and uncertain on others. Asking the operators to choose one of two responses is equivalent to asking them to separate their confidence states into two "response bins": "more confident signal" and "more confident noise." Asking them to also rate their confidence in their detection judgment is equivalent to asking them to separate their confidence states into more than two confidence bins. The results are illustrated in Figure 3. The proportion of rating responses larger

than a given value, k, on the rating scale is equal to the proportion of times that the confidence state fell above the upper bound of response bin k. Separating the data into signal and noise trials, these proportions are equal to the hit and false alarm rates that would be observed if the detection criterion were placed at the same location as this upper bound. Thus, a number of points along the ROC curve can be estimated from a single experimental condition. An estimate of the area under the ROC curve is easily obtained from these "incomplete" curves by connecting the points and calculating the areas under each segment.

4. RECENT APPLICATIONS

Watchkeeping (or vigilance) is one of the most natural applications of signal detection theory and probably the most common one. In the laboratory vigilance task, subjects follow a sequence of stimulus presentations and after each one of these they decide whether or not to sound an alarm (detect a signal). A great deal of effort has been extended to understand how biases and sensitivity change over the course of the watch (e.g. See *et al.* 1995, Parasuraman *et al.* 1987 for reviews). Many other variables affecting vigilance performance have also been examined using the signal detection theory approach. For example, Matthews (1996) used d' and b to examine the effects of signal rates on vigilance in high workload (fast presentation rate) conditions. The results suggested that sensitivity declined with time on task, and more precipitously when the signal rate was high (i.e. when the operator must explicitly respond more often). Response biases, on the other hand, apparently remained constant across different signal rates. Results of this kind are not uncommon and seem very plausible. In fact, few researchers take the trouble to qualify their inferences about decision-making strategies in vigilance even though they ultimately depend on the validity of the signal detection theory indices.

Another illustration of the wide range of potential applications of signal detection theory is its routine use as an arbitrator when two relatively informal theories or hypotheses make different predictions about the effects of a factor on the operator's performance. If one theory predicts that there should be no effect of this factor or the opposite kind of effect, then the d' measure is typically recruited to verify that observed differences in overall percent correct or in the hit and false alarm rates are not merely due to effects of the factor on response bias. The direction and size of the operator's response bias are not usually specifically predicted by theories of operator performance (response biases are, by definition, "subject-controlled" factors) and so the d' statistic is usually the measure of most interest. Patterns of bias, however, can sometimes inform the researcher about variables such as cognitive style or motivation. For example, in a recent study on aircraft recognition and recognition training, Goettl (1996) looked at two memory models and their predictions about aircraft recognition under different learning conditions. Recognizing whether an aircraft is or is not a member of a predefined class (e.g. commercial versus noncommercial) is a two-response discrimination task, making it possible to apply a signal detection theory analysis. The signal detection theory measures were used to estimate memory strength (sensitivity) and the results indicated that for male subjects, there was a

difference between two different types of learning schedules on memory strength and not merely on bias, but for the females,the differences between the two schedules could be attributed to bias effects alone.

Another natural application for signal detection theory has been in the various kinds of expert decision-making problems involved in medical and clinical diagnosis. ROC curves are used fairly routinely in these areas to characterize the extent to which some quantifiable property of a medical image informs the physician about the presence or absence of a pathology. Researchers also make use of these measures to compare experienced and inexperienced diagnosticians and to show how new imaging technologies can combine with or replace traditional methods of diagnosis (e.g. Jiang *et al.* 1999, Tsuda *et al.* 1999). Quite often, new approaches to diagnosis do not increase both the "hit" and "correct rejection" rates, making bias effects a very important factor to consider.

In each of these examples, and in many others, some interesting and important discoveries about human performance, including losses of sensitivity with time on task, idiosyncratic response biases, and other well-documented phenomena, would not be possible without the benefit of a formal measurement system. However, these discoveries are only "conditional" statements of fact because of their dependence on the specific assumptions of signal detection theory. Other interpretations of the data are possible, and occasionally other indices are adopted in addition to or in place of the signal detection theory measures. In the next chapter, we look at the evidence in support of signal detection theory and then discuss some of the alternatives and their motivations.

REFERENCES

GOETTL, B.P., 1996, The spacing effect in aircraft recognition. *Human Factors*, 38, 34–49.

GREEN, D.M. and SWETS, J.A., 1966, *Signal Detection Theory and Psychophysics* (New York: Wiley; repr. Huntington: Krieger, 1974).

JIANG, Y., NISHIKAWA, R.M., SCHMIDT, R.A., METZ, C.E., GIGER, M.L. and DOI, K., 1999, Improving breast cancer diagnosis with computer-aided diagnosis. *Academic Radiology*, 6, 22–33.

MATTHEWS, G., 1996, Signal probability effects on high-workload vigilance tasks. *Psychonomic Bulletin and Review*, 3, 339–43.

PARASURAMAN, R., WARM, J.S. and DEMBER, W.N., 1987, Vigilance: taxonomy and utility. In L.S. Mark, J.S. Warm and R.L. Huston (eds), *Ergonomics and Human Factors: Recent Research* (New York: Springer).

SEE, J.E., HOWE, S.R., WARM, J.S. and DEMBER, W.N., 1995, Meta-analysis of the sensitivity decrement in vigilance. *Psychological Bulletin*, 117, 230–49.

TORGERSON, W.S., 1958, *Theory and Methods of Scaling* (New York: Wiley).

TSUDA, K., YU, K.K., COAKLEY, F.V., SRIVASTAV, S.K., SCHEIDLER, J.E. and HRICAK, H., 1999, Detection of extracapsular extension of prostate cancer: role of fat suppression endorectal MRI. *Journal of Computer Assisted Tomography*, 23, 74–8.

Signal Detection Theory – Alternatives

J.D. Balakrishnan and J. A. MacDonald

Department of Psychological Sciences, Purdue University, West Lafayette, IN 47907, USA

1. INTRODUCTION

In *Signal Detection Theory*, we discussed the rational and the basic concepts behind the "formal model" approach to human performance assessment and then briefly sketched out the assumptions and methods associated with the field's most popular formal measurement system, signal detection theory. Over the past several decades, the experimental methods and statistics associated with the signal detection theory framework have become so deeply ensconced in the measurement literature that relatively few theorists would see any serious reason to question them. Competing views have also been developed, however, and should not be ignored merely because they are less familiar. After all, the validity of signal detection theory has never been established, and from time to time some cracks in its foundation have been discovered. In fact, here we review some very recent empirical results that purportedly show that the signal detection theory measures have actually grossly misled human factors researchers about the nature and limits of human performance. First we review the experimental data that make signal detection theory seem so compelling and then consider the strength of this evidence and some alternative interpretations.

2. EVIDENCE IN SUPPORT OF SIGNAL DETECTION THEORY

For most human performance researchers, the important question about signal detection theory is not whether it is a valid theory or not, but whether it is valid enough for their purposes. Typically, the performance statistics are recruited to identify changes in sensitivity and response bias, not their absolute values. The signal detection theory measures are trusted in this role because the way they change under different experimental conditions is usually predictable and consistent with the theory. Most importantly, perhaps, the theory correctly predicts the effects of changing the base rates or payoffs in a yes–no detection task: increasing the frequency of the signal trials almost invariably increases both the hit and the false alarm rates. Stronger manipulations lead to stronger effects of the same kind. Similarly, the sensitivity indices (e.g. d' and area under the ROC curve) seem relatively unaffected by base rates and payoffs, while the bias measure b is not: it increases and decreases appropriately when the base rates or payoffs are manipulated, indicating that the operator's decision-making strategy is rational. Other kinds of evidence seem to justify the 'technical" assumptions of the theory while rejecting the underlying assumptions of many other statistics (Swets 1986b). Empirical *z*-ROC curves, for example, usually are almost perfectly linear, consistent with the distributional assumptions (normality) of signal detection theory.

Using these statistics for more fine-grained analyses of operator performance also leads to a plausible and cohesive account of human behavior. For example, studies have consistently shown

that the operator shifts the detection criterion when the base rates are changed, but the size of the shift is presumably insufficient to cause the decision process to be "optimal" (i.e. to maximize the percentage of correct decisions). In other words, the decision process is "conservative" (e.g. Creelman and Donaldson 1968, Macmillan and Creelman 1990). In laboratory studies of watchkeeping (i.e. the vigilance paradigm), the measures have been used to "establish" that changes in the detection rate are due to changes in the operator's willingness to make a detection response under some conditions (e.g. relatively slow paced detection tasks) and losses of sensitivity under others (e.g. relatively fast paced detection tasks with a memory load; Parasuraman 1979). These accounts are interesting and plausible, and certainly do not raise any special concerns about the validity of the model that gives rise to them, even though it could easily have been otherwise. If the model was untenable, one might expect a more inconsistent or a more confusing pattern of results.

3. ALTERNATIVE INTERPRETATIONS OF THE DATA

Although the classical findings undoubtedly do tell us something important about human performance, they can also be interpreted in other ways, some of which are at least as plausible as the signal detection theory explanation. Response time models, for example, can reproduce all of these properties of the data, even though they represent the decision-making processes in a profoundly different manner (e.g. Townsend and Ashby 1983, Luce 1986). Instead of adjusting a detection criterion, these models assume that the operator accumulates information until enough evidence has been collected to justify one of the two possible responses. Because the decision is reached only after the accumulated information crosses a boundary, the number of samples (the encoding time) will depend on the quality or strength of the evidence as it is collected. If the operator receives weak information early on, for example, s/he will wait longer before responding so that additional information can be obtained. In signal detection theory, the decision-making process only plays a role after the encoding process is completed. Thus, this "static" model assumes that the amount of information collected is entirely independent of the information quality.

In some circumstances, such as clinical or medical diagnoses, the "fixed sample" assumption of signal detection theory is reasonable, or perhaps even necessarily true if the amount of physical evidence available to the decision-maker is not under the decision-maker's control. In many other situations, however, the decision to stop collecting new information and formulate a response is a crucial aspect of the operator's decision-making process, which may itself be biased in some way. In the real world detection problems associated with watchkeeping, for example, the operator must respond quickly as well as accurately to changes in the status of the system. Attempting to decide more quickly whether a signal or a non-signal event has taken place will generally increase both the false alarm and the miss rate (the so-called "speed–accuracy trade-off"), causing d' to decrease. Such biases might show up in mean response times of the operator or in the estimates of the parameters of a dynamic model, but would be invisible to signal detection theory. The difference between tasks that lend themselves to a signal detection theory analysis and those that do not is not always recognized: many researchers

Double Exponential

Gamma
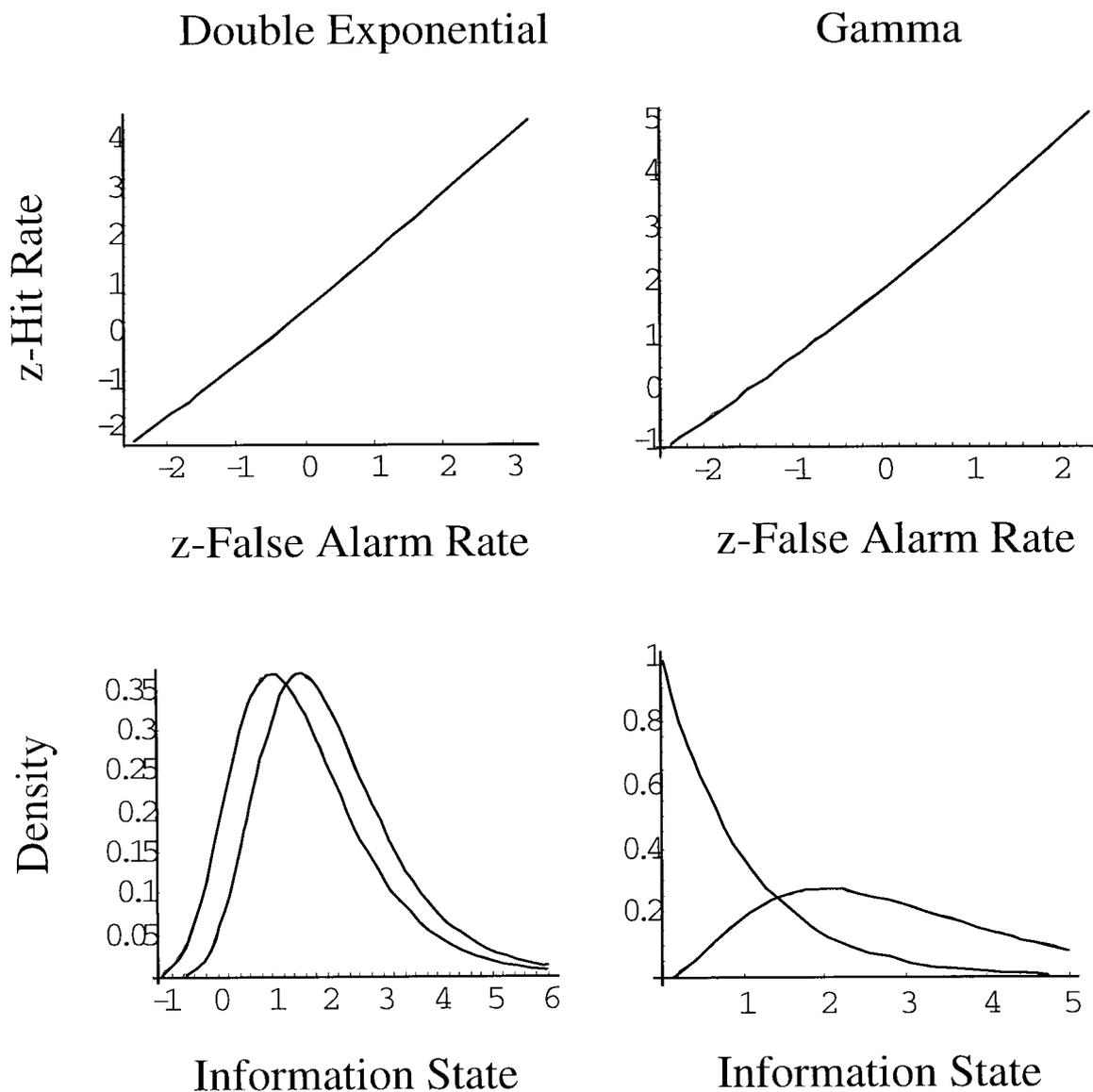
Figure 1. z-ROC curves obtained when pairs of hit and false alarms rate from non-normal distributions are converted to z scores from the standard normal distribution. The two "double exponential" distributions in the lower left panel have equal variance but are positively skewed (e.g. Luce 1986: 508). The "gamma" distributions are unequal variance and positively skewed (e.g. Luce 1986: 507). Many other distributions also predict quasi-linear z-ROC curves.

use d' or similar measures even when the operators are under time pressure.

The evidence purportedly supporting the normality assumptions of the signal detection theory model — i.e. the shape of the empirical ROC curve — is also very weak. Many other distribution models predict a virtually linear z-ROC curve. In fact, the small deviations from linearity in empirical z-ROC curves are actually larger than the deviations predicted by these alternative distributions. Examples of the z-ROC curve predictions of two other distributions, vastly different from the normal (and not "transformable to the normal"), are shown in Figure 2.

4. OTHER APPROACHES

Apart from the response time models, very few of the alternatives to signal detection theory were specifically developed to repair any specific or known deficiencies of d'. Instead, they were proposed as

alternatives that might be more reliable if complete list of these alternative indices would be quite long (>20). Fourteen sensitivity indices are reviewed in Balakrishnan (1998a). Five different bias measures were recently compared by See *et al.* (1997). Not all of these measures were derived from an explicit description of the decision-making process, making it more difficult to evaluate them. However, the supposedly non-parametric indices make some testable predictions about the shape of the ROC curve and other predictions related to the effects of base rates and payoffs (e.g. Swets 1986a, Macmillan and Creelman 1996). In his review of 10 different sensitivity indices, Swets (1986a) concluded that a variable criterion measure that assumes normal distributions with different variance (area under the normal ROC curve) was the most viable index. Similarly, See *et al.* (1997) recommended a variant of the criterion value from signal detection theory over several supposedly non-parametric bias indices.

Rating ROC
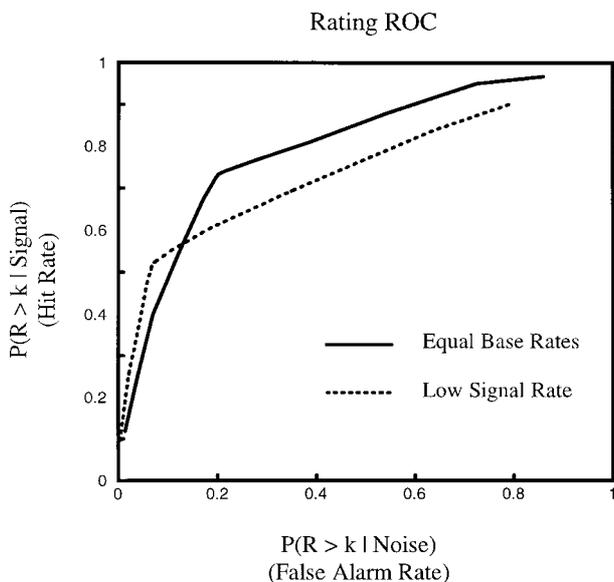
Figure 2. Empirical ROC curves from a vigilance experiment with two different signal rates. When the signal rate is low, the curve is sharply skewed with respect to the negative diagonal (the line connecting the upper left to the lower right corner), consistent with larger variance in the signal distribution. According to signal detection theory, the two curves should be identical. For Figures 2–5 are revised versions reprinted with permission from *Human Factors*, 40(4). Copyright 1998 Human Factors and Ergonomics Society. All rights reserved.

Evidence against the signal detection theory approach is somewhat hard to come by. However, a significant part of the success of this theory may be due to the lack of strong empirical tests of its basic assumptions. Recently, some new methods developed by Balakrishnan (1998a, b, 1999) have made it possible to directly test, in an assumption-free manner, the fundamental principles of signal detection theory and other two stage detection models, including the assumption that (1) response biases exist, (2) the decision process tends to be conservative and (3) the encoding and decision-making processes are independent. To avoid the assumptions required by *d'* and *b*, these tests make use of some extra information about encoding and decision-making contained in confidence ratings data. In many respects, these new methods are a relatively straightforward extension of the ratings paradigm developed earlier by signal detection theorists. However, they differ from the traditional signal detection theory methods in three ways:

- A cutoff between the "noise" and "signal" responses must be explicitly defined on a bipolar confidence rating scale. For example, if the rating scale has 10 values numbered from 1 to 10, rating responses five and six would be labeled "lowest confidence noise" and "lowest confidence signal" responses respectively. The two extremes of the scale (responses 1 and 10) would represent "highest confidence noise" and "highest confidence signal" responses respectively. Alternatively, the researcher may elicit first the yes–no detection response and then a "confidence level" response. For each confidence level in this two-response method, there is a corresponding confidence rating in the one-response method (e.g. five levels of confidence in the two-response method would be

equivalent to 10 rating responses on the bipolar scale in the one-response method). For purposes of data analyses, the confidence levels should be transformed into confidence ratings on a single bipolar scale (e.g. confidence level "2" in a "noise" response would become confidence rating "4" on a 10-point bipolar scale with a cutoff between responses five and six).

- To test the assumptions of signal detection theory, the signal rate or some other factor presumably tied exclusively to the decision process should be set in such a way as to induce a response bias.
- The operators should be instructed to limit their use of the extremes of the rating scale (i.e. the two lowest confidence responses on a bipolar scale) to situations in which they are extremely uncertain. If the operator uses these responses frequently, the new measures will not misrepresent the decision-making process, but they will be uninformative.

5. DISTRIBUTION-FREE MEASURES OF BIAS

The utility of signal detection theory as a data interpretation tool rests on two fundamental precepts. First, the location of the decision criterion should be strongly dependent on the base rates or payoffs, shifting away from the "unbiased" location (i.e. the point at which the two distributions intersect). Second, the encoding distributions should be independent of the base rates. Both assumptions appear to be incorrect. Figure 3 shows the ROC curves from two different base rate conditions of a vigilance experiment (Balakrishnan 1998b) in which the operators were asked to discriminate between two patterns varying in overall size. The larger of the two patterns was the "signal" and occurred on half of the trials in one condition (Equal Base Rates) and on
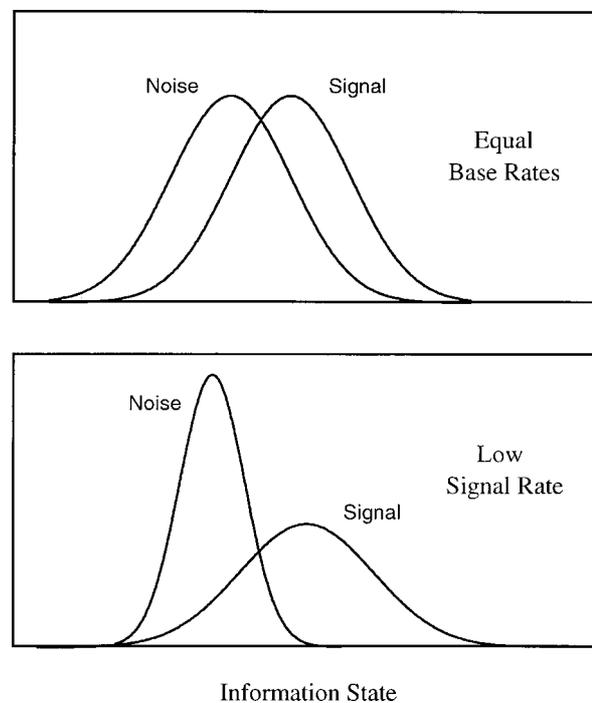
Figure 3. One possible distribution model that could account for the ROC curve data in Figure 2. Decreasing the signal rate increases the variance of the signal distribution while decreasing the variance of the noise distribution.
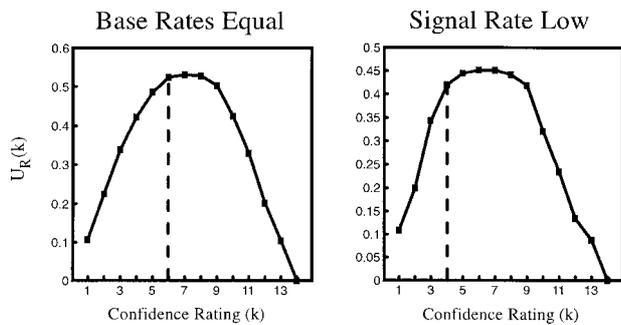
Figure 4. Empirical estimates of the bias test function, $U_R(k)$, in two base rate conditions of a vigilance experiment. The maximum of the function occurs at the value 7 in both conditions, indicating that the detection criterion is set at the intersection of the two encoding distributions, regardless of the base rates. The dashed vertical lines indicate the point at which the function would have reached its peak if the distributional assumptions of signal detection theory were satisfied.

only 10% of the trials in another (Low Signal Rate). Confidence rating data were elicited using a 14-point bipolar rating scale with a cutoff between responses seven and eight. If the encoding distributions do not depend on the base rates, the empirical ROC curves obtained under different signal rate conditions should fall along a single curve. Instead, the two curves clearly map out two very different functions. Their shapes are consistent with the assumption that the variances of the distributions depend on the base rates of the stimuli (Figure 4). When the signal is infrequent, the variance of the signal distribution is relatively large compared with the noise distribution. When the two events are equally frequent, the variances of the distributions are equal.

Clearly, the distributions in this experiment were not independent of the base rates. The next question is therefore whether base rates affect both the distributions and the decision criterion, or just the distributions. Confidence rating data can also be used to determine whether there is any shift of the criterion, without involving any assumptions about the shapes of the encoding distributions. The test is based on the difference between the cumulative relative frequency histograms of the operator's confidence rating responses,

$$U_R(k) = F_N(k) - F_S(k),$$

where $F_N(k)$ and $F_S(k)$ are the proportions of rating responses less than or equal to k on noise and on signal trials respectively, and the argument k is the rating value on a bipolar scale with a cutoff at k^*. If this function is decreasing for any k associated with a noise response ($k \pounds k^*$), or increasing for any k associated with a signal response ($k > k^*$), then the decision rule is biased (i.e. the criterion is not set at the point where the two encoding distributions intersect). The total proportion of these "biased rating responses" in the data, or W_p, is an estimate of the proportion of times the participant makes a biased response. For example, if the rating scale has 10 values with a cutoff between responses five and six, and $U_R(k)$ reaches it maximum value at rating response three, then the total proportion of four and five responses during the experiment is the (estimated) proportion

of trials on which the operator's decision differed from the unbiased decision rule. According to signal detection theory, this value should increase as the signal and noise base rates diverge.

If the peak of $U_R(k)$ occurs at the cutoff (and hence $W_p = 0$), the decision rule could still be biased to some degree. However, in this case, the proportions of the two lowest confidence responses (e.g. responses "4" and "5" on a 10-point bipolar scale) provide upper bounds on the probability of a biased response (e.g. the proportion of "4" responses is an upper bound on the bias towards the noise response, and the proportion of "5" responses is an upper bound on the bias towards the signal response). Instructions to the operator to be conservative in the use of extreme values on the rating scale are intended to keep this upper bound to a minimum. If it is large when $W_p = 0$, then the test will be relatively uninformative (but not misleading).

Illustrative results from the vigilance study described above are shown in Figure 5. In these examples and in many other experiments in our laboratory, the peak of the $U_R(k)$ function occurs at the cutoff value (response "7"), causing $W_p = 0$ when the base rates were equal and when they were unequal. The proportion of lowest confidence "noise" responses was < 0.01 in both conditions. Thus, even when the signal occurred on only 10% of the trials, the subjects' decision rules were not biased towards the noise response.

The absence of bias in the decision rule implies that the subject's decision-making strategy is suboptimal (i.e. does not maximize the percentage correct decisions). Suboptimality was already implied by the supposed "conservatism" of the decision rule that was uncovered by classical signal detection theory studies. However, it is also possible to test for suboptimality without making any assumptions about the structure of the discrimination process or the encoding distributions. To do this, the researcher calculates the proportion of correct responses associated with each rating response given by the operator. If the decision rule is optimal (maximizes % correct), then all of these correct response proportions will be greater than one half, regardless of the base rates. If any are less than one-half, then the decision rule is suboptimal. For example, considering only the trials of the experiment on which the operator chose rating response "2" on the 10-point bipolar rating scale (i.e. the "noise
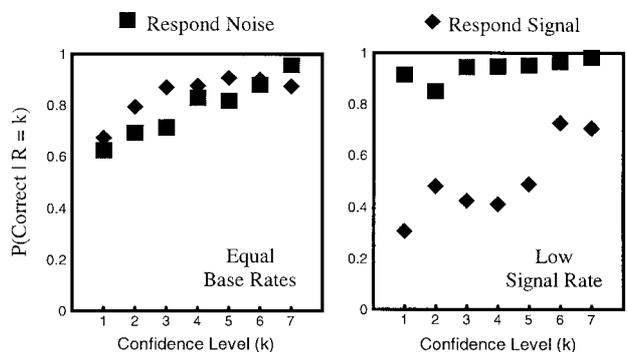
Figure 5. Proportion correct conditioned on the discrimination response and confidence level for two base rate conditions of a vigilance experiment. Any value < 0.5 indicates that overall proportion correct could be improved by changing the decision rule (i.e. the decision rule is suboptimal).

responses" given at confidence level 4), more than half of these responses should have occurred on "noise" trials. If this is not true, then the operator could improve his/her performance by simply switching from a "noise" to a "signal" response whenever s/he would normally make this "2" response.

Examples of this optimality test are shown in Figure 6. Notice that for several of the lower confidence "signal" responses, the subjects are more often incorrect than correct. Reversing these "suboptimal" responses to "noise" responses "corrects" for the suboptimality of the decision rule, providing a measure of the performance level that the subjects could have achieved if their decision rules were optimal. This *post-hoc* correction process has some important potential applications in the workplace: in principle, a trainer can make use of this information to give the trainees focused feedback that will allow them to optimize their decision-making strategies.

6. IMPLICATIONS FOR PERFORMANCE ASSESSMENT

Unless and until some other interpretation of these new empirical tests can be found that is consistent with the basic tenets of signal detection theory, the legitimacy of the *d'* and b analysis of discrimination performance is seriously open to question. Apparently, correcting for the effects of bias on the performance of a human operator is not merely a matter of adjusting for the value of a decision criterion, but instead involves explaining how and why biases affect the two distributions that describe the operator's information states. Area under the ROC curve, *d'* and other indices associated with signal detection theory may still provide some useful information about the operator's behavior, but it is not clear how this information should be interpreted. Similarly, most of the other indices developed to complement or replace the signal detection theory measures also rely heavily on the assumption (implicitly or explicitly stated) that the two encoding distributions are not affected by response bias.

Correcting for any suboptimality in the decision rule using the methods described above will make it possible to compare different operators or systems without the results being confounded by biases in the operator's decision rule. However, this method does not control for biases in the decision processes that influence other aspects of behavior, including how much information is collected before a decision is reached. If the encoding time is operator controlled, then biases can easily mimic the effects of sensitivity changes in signal detection theory. In such a case, the relationship between the operator's accuracy and response time may be the best source of information about the contribution of biases to overall performance. Fitting a response time model to the data is one way to quantify these effects, but of course there is no guarantee that the assumptions of a dynamic model will be any more accurate than those of signal detection theory. If accuracy (hit and correct rejection rate) is increasing and response time is decreasing under the two conditions being compared, than it would be difficult to find any formal model, static or dynamic, that would not attribute this effect to a change in sensitivity. However, if accuracy increases but response times also increase, then the result could be attributed to biases in the data collection process, even though *d'* and other performance indices are likely to suggest a change in sensitivity. Until more facts are known about the effects of bias on operator performance, strong inferences about relative sensitivity levels or the effects of a factor on sensitivity should probably be limited to situations in which response accuracy and response speed are either independent or positively covarying.

REFERENCES

BALAKRISHNAN, J.D., 1998a, Some more sensitive measures of sensitivity and response bias. *Psychological Methods*, 3, 68–90.

BALAKRISHNAN, J.D., 1998b, Measures and interpretations of vigilance performance: Evidence against the detection criterion. *Human Factors*, 40, 601–23.

BALAKRISHNAN, J.D., 1999, Decision processes in discrimination: fundamental misrepresentations of signal detection theory. *Journal of Experimental Psychology: Human Perception and Performance* (in press).

CREELMAN, C.D. and DONALDSON, W., 1968, ROC curves for discrimination of linear extent. *Journal of Experimental Psychology*, 77, 514–16.

LUCE, R.D., 1986, *Response Times: Their Role in Inferring Elementary Mental Organization* (New York: Oxford University Press), 3, 164–70.

MACMILLAN, N.A. and CREELMAN, C.D., 1990, Response bias: characteristics of detection theory, threshold theory, and "non-parametric" indexes. *Psychological Bulletin*, 107, 401–13.

MACMILLAN, N.A. and CREELMAN, C.D., 1996, Triangles in ROC space: history and theory of "non-parametric" measures of sensitivity and response bias. *Psychonomic Bulletin and Review*,

PARASURAMAN, R., 1979, Memory load and event rate control sensitivity decrements in sustained attention. *Science*, 205, 924–7.

SEE, J.E., WARM, J.S., DEMBER, W.N. and HOWE, S.R., 1997, Vigilance and signal detection theory: an empirical evaluation of five measures of response bias. *Human Factors*, 39, 14–29.

SWETS, J.A., 1986a, Indices of discrimination or diagnostic accuracy: their ROCs and implied models. *Psychological Bulletin*, 99, 110–17.

SWETS, J.A., 1986b, Form of empirical ROCs in discrimination and diagnostic tasks: implications for theory and measurement of performance. *Psychological Bulletin*, 99, 181–98.

TOWNSEND, J.T. and ASHBY, F.G., 1983, *Stochastic Modeling of Elementary Psychological processes* (New York: Cambridge University Press)

Situation Awareness

M. R. Endsley

SA Technologies, Inc., 4731 East Forest Peak, Marietta, GA 30066 USA

1. INTRODUCTION

Situation awareness (SA) is important for effective decision-making and performance in many domains, including aviation, nuclear power and chemical processing, automobiles, air traffic control, medical and health systems, tele-operations, trains, space operations, maintenance and advanced manufacturing systems. In these complex and dynamic environments, human decision-making is highly dependent on SA — a constantly evolving picture of the state of the environment. Situation awareness can be described broadly as people's state of knowledge or mental model of the situation around them.

Many definitions of SA have been developed, some very closely tied to the aircraft domain and some more general (Dominguez 1994 for a review). A general, widely applicable definition describes SA as "the perception of the elements in the environment within a volume of time and space, the comprehension of their meaning and the projection of their status in the near future" (Endsley 1988). SA involves perceiving critical factors in the environment (Level 1 SA), understanding what those factors mean, particularly when integrated together in relation to the decision maker's goals (Level 2), and at the highest level, an understanding of what will happen with the system in the near future (Level 3). These higher levels of SA allow operators to function in a timely and effective manner.

An operator's understanding and classification of the situation he or she is in forms the basis for all subsequent decision-making and performance. Even the best-trained operators will perform poorly if their SA is incorrect. One study of aircraft accidents found that as much as 88% of all accidents attributed to human error had an underlying problem with SA (Endsley 1995c).

1.1. Level 1 SA — Perception

The first step in achieving SA is to perceive the status, attributes and dynamics of relevant elements in the environment. This includes relevant system parameters, characteristics and actions of other individuals, and features of the external environment. In many systems, the basic information needed may not be present or may be obscured. In other cases, the information is clearly present; however, it is not attended to due to omission from the scan pattern, attentional narrowing, directing attention to other tasks or information or outside distractions. Failures in SA at this level have also been found to occur due to memory errors (where information is initially detected and then forgotten), and initial misperception of information (often due to distractions of other tasks or falsely succumbing to expectations) (Jones and Endsley 1996). Thus system designs which leave human operators susceptible to such errors may encourage low levels of SA.

Just knowing the value of system parameters is not sufficient, however. Data perceived also must be integrated and compared with operational goals to provide an understanding of what it really means, forming the second level of SA.

1.2. Level 2 SA — Comprehension of the Current Situation

Comprehension of the situation is based on a synthesis of disjointed Level 1 elements. Level 2 SA goes beyond simply being aware of the elements which are present, to include an understanding of the significance of those elements in light of one's goals. The decision-maker puts together level 1 data to form a holistic picture of the environment, including a comprehension of the significance of objects and events. Level 2 SA is developed over time based on the observed dynamics of the system (how variables are changing in relation to each other). It is furthermore highly goal-oriented. The values perceived are only meaningful as a function of the operator's goals.

Failures in comprehension of perceived information (Level 2 SA) have been found to be due to poor mental models for combining and interpreting information, use of the wrong mental model, and over-reliance on default values (normal expectations) in the mental model (Jones and Endsley 1996).

1.3. Level 3 SA — Projection of Future Status

It is the ability to project the future actions of the elements in the environment, at least in the very near term, that forms the third and highest level of situation awareness. This is achieved through knowledge of the status and dynamics of the elements and a comprehension of the situation (both levels 1 and 2 SA). This ability to predict what will be happening with the system, at least in the near term, allows them to behave proactively instead of reactively in dealing with the system. A high level of expertise may be needed to develop Level 3 SA. Failures in projection have been found to occur due to poor mental models and over-projection of current trends (Jones and Endsley 1996).

2. MODELS OF SA

Several researchers have put forth theoretical models depicting the role of numerous cognitive processes in the development of SA (Endsley 1988, 1994, 1995b, Fracker 1988, Taylor 1990, Taylor and Selcon 1994, Adams *et al.* 1995, Smith and Hancock 1995). There are many commonalties in these efforts that point to key mechanisms that are important for SA.

Endsley (1988, 1994, 1995d) proposed a framework model based on information processing theory. SA is regarded as a stage separate from decision-making and performance. SA is the main precursor to decision-making, however, many other factors can come into play in turning good SA into successful performance, including individual strategies, decision-making, training, and features of the environment (such as workload and system capabilities).

Several factors relevant to the individual and to the system/environment effect the accuracy and completeness of SA that operators derive from their environment. First, humans are limited by working memory and attention. The way in which operator attention is employed in complex environments that are often full of multiple competing cues is essential in determining which aspects of the situation will be processed to form SA. Once taken in, information must be integrated with other information, compared with goal states and projected into the future — all heavily demanding on working memory. Thus, limited attention and working memory can negatively effect the completeness and accuracy of SA.

Long-term memory stores in the form of mental models or schemtaa are described as playing a major role in overcoming these limitations. With experience operators develop internal models of the systems they operate, the individuals they work with (or against), and the environments they operate in. These models serve to help direct limited attention in efficient ways, provide a means of integrating information without loading working memory, and provide a mechanism for generating projection of future system states. Associated with these models may be schemata of prototypical system states. Critical cues in the environment may be matched to such schemata to indicate prototypical situations that provide instant situation classification and comprehension. Scripts of the proper actions to take may be attached to these situation prototypes, simplifying decision-making as well. Schemata of prototypical situations are incorporated in this process and in many instances may also be associated with scripts to produce single-step retrieval of actions from memory, thus providing for very rapid decision-making such as has been noted by Klein (1989). The use of mental models and schemata in achieving SA is considered to be dependent on the ability of the individual to pattern match between critical cues in the environment and elements in the mental model.

Goals are also shown in the model as being highly important for SA. SA is seen as resulting from dynamic alternating between data driven (bottom-up) and goal driven (top-down) processing. In goal driven processing, attention is directed across the environment based on current operator goals. The operator actively seeks information needed for goal attainment and the goals simultaneously act as a filter in interpreting the information that is perceived. In data driven processing, perceived environmental cues may indicate new goals that need to be active. Dynamic switching between these two processing modes is important for successful achievement of good SA.

In addition, preconceptions or expectations are shown to influence the formation of SA. These expectations may be formed based on mental models, prior experiences, instructions or other communications. These expectations influence how attention is deployed and the actual perception of information taken in.

Finally, automaticity is another mechanism developed with experience that is shown to influence SA in this model. With experience, the pattern-recognition/action-selection sequence can become highly routinized and developed to a level of automaticity. This provides good performance with a very low level of attention demand in certain well-understood environments. Thus, automaticity can positively effect SA by reducing demands on limited attention resources. SA can also be negatively effected by automaticity due to a reduction in responsiveness to novel stimuli, however.

Fracker (1988) similarly discusses the importance of both working memory and schemata in long-term memory for SA. He points out that while schemata may be very useful for facilitating situation assessment by providing a reduction in working memory demands, they can also lead to significant problems with biasing in the selection and interpretation of information which may create errors in SA. Jones (1997) showed that these biases can be very difficult to overcome, despite very overt cues indicating the erroneousness of the initial schemata.

Adams *et al.* (1995) stress the importance of the inter-relationship between one's state of knowledge, or SA, and the processes used to achieve that knowledge. Framed in terms of Neisser's (1976) model of perception and cognition, they make the point that one's current knowledge effects the process of acquiring and interpreting new knowledge in an ongoing cycle. This agrees with Sarter and Woods (1991) statement that SA is "the accessibility of a comprehensive and coherent situation representation which is continuously being updated in accordance with the results of recurrent situation assessments."

Smith and Hancock (1995) further support this proposition by stating that "SA is up-to-the-minute comprehension of task relevant information that enables appropriate decision-making under stress. As cognition-in-action, SA fashions behavior in anticipation of the task-specific consequences of alternative actions." In defining SA as "adaptive, externally directed consciousness," they take the view that SA is purposeful behavior that is directed toward achieving a goal in a specific task environment. They point-out that SA is therefore dependent on a normative definition of task performance and goals that are appropriate in the specific environment.

Several features of the external environment are also extremely important to SA (Endsley 1995d). First the capabilities of the system's sensors and other aids for detecting and presenting critical situational information to the operator is fundamental. The design of the operator interface for managing and presenting this information in a meaningful, easy to process and timely format is also evident. Other system features, such as automation and complexity, have also been found to significantly effect SA (Endsley and Kiris 1995, Endsley and Kaber 1999). Stress and workload are other external factors that relate to SA.

The relationship between SA and workload has been theorized to be important. Taylor (1990) includes a consideration of supply and demand of resources as central to situation awareness. Adams *et al.* (1995) also discuss the task management problem, involving prioritizing, updating task status, and servicing tasks in a queue, as central to SA. Endsley (1993), however, shows that for a large range of the spectrum, SA and workload can vary independently, diverging on the basis of numerous factors. Only when workload demands exceed maximum human capacity is SA necessarily at risk. SA problems may also occur under low workload (due to vigilance problems) or when workload is in some moderate region.

3. SITUATION AWARENESS MEASUREMENT

As SA has become a major design and training goal in many systems, the direct measurement of SA during system evaluation has become important. Several methods have been established for the measurement of situation awareness. (For a complete review see Endsley 1996 and Endsley and Garland 1999.) The impact of a particular design concept on SA can be measured directly through either objective or subjective means, or it can be inferred through less direct performance measures.

3.1. Objective Measurement

The most commonly used means of objectively evaluating a design concept's impact on SA involves directly questioning operators as to their perceptions of critical aspects of the system they are operating. The Situation Awareness Global Assessment Technique (SAGAT) (Endsley 1988, 1995b) is a technique in which an operator-in-the-loop real-time simulation of the system of interest

is frozen at randomly selected times, the system displays blanked and the simulation suspended while operators quickly answer questions about their current perceptions of the situation. A number of these "snap-shots" are collected across a range of operational conditions and events. Operator perceptions are then compared with the real situation based on simulation computer databases to provide an objective measure of SA with that particular design concept.

SAGAT includes queries about all operator SA requirements, including Level 1 (perception of data), Level 2 (comprehension of meaning) and Level 3 (projection of the near future) components. This includes a consideration of system functioning and status as well as relevant features of the external environment. SAGAT provides an objective, unbiased assessment of operator SA that overcomes memory problems incurred when collecting data after the fact, yet minimizes biasing of operator SA due to secondary task loading (from asking questions during system operation for instance). Empirical, predictive and content validity has been demonstrated for this technique (Endsley 1990a, b, 1995a). A certain degree of measurement reliability has been demonstrated in a study that found high reliability of SAGAT scores for four individuals who participated in two sets of simulation trials (Endsley and Bolstad 1994). SAGAT has been used successfully in a wide variety of domains, including cockpits, air traffic control, medicine, driving and power plant operations.

3.2. Subjective Measurement

Subjective measures of SA are easier and less expensive to administer than objective measures, but may lack the same degree of accuracy and diagnosticity. The most commonly used method is to have operators provide ratings of their SA with system concepts along a designated scale.

Taylor (1990) developed the Situational Awareness Rating Technique (SART) which has operators rate system designs on the amount of demand on attentional resources, supply of attentional resources and understanding of the situation provided. As such, it considers operators' perceived workload (supply and demand on attentional resources) in addition to their perceived understanding of the situation. While SART has been shown to be correlated with performance measures (Selcon and Taylor 1990), it is unclear whether this is due to the workload or the understanding components. SART and SAGAT have been found to have low correlation and are most like measuring very different aspects of the construct (Endsley *et al.* 1998).

3.3. Performance Measurement

In general, performance measures provide the advantage of being objective and are usually non-intrusive. Simulation computers can be programmed to record specified performance data automatically, making the required data relatively easy to collect. As many other factors can act to influence performance measures, however, such as individual strategy differences, variations in decision-making, and inability to carry out desired actions, performance measures can be limited for inferring SA by themselves. In addition, finding sensitive and diagnostic performance measures may be difficult in many systems.

Another limitation of this approach stems from the interactive nature of SA subcomponents. A new system or training technique-enhancing SA on one factor may simultaneously reduce SA on

another, unmeasured, factor (Wickens 1995). One way of dealing with these issues has been the careful development of scenarios that incorporate a "testable response" (Pritchett *et al.* 1995). This approach assumes very specific measurable performance outcomes that can be predetermined to be correct, given a particular level of SA. Researchers must be prepared for anomalous outcomes, however, because individuals may not react as expected to events and interpretation of their actions can sometimes be ambiguous. Performance measures can be used to augment the information provided by other SA measures, subject to these limitations.

4. SUMMARY

Following a structured approach from analysis to design to testing, SA can be incorporated as a significant and attainable design goal. Models of SA can be used to guide the design of systems and training programs to support operator SA. While not all human error is preventable, a significant portion of those problems that are currently labeled as human error can be addressed proactively and successfully by tackling the major design factors that lead to SA problems.

REFERENCES

ADAMS, M.J., TENNEY, Y.J. and PEW, R.W., 1995, Situation awareness and the cognitive management of complex systems. *Human Factors*, 37, 85–104.

DOMINGUEZ, C., 1994, Can SA be defined? In M. Vidulich, C. Dominguez, E. Vogel and G. McMillan (eds), *Situation Awareness: Papers and Annotated Bibliography*. AL/CF-TR-1994-0085 (Wright-Patterson AFB, OH: Armstrong Laboratory), 5–15.

ENDSLEY, M.R., 1988, Design and evaluation for situation awareness enhancement. In *Proceedings of the Human Factors Society 32nd Annual Meeting* (Santa Monica: Human Factors Society), 97–101.

ENDSLEY, M.R., 1990a, A methodology for the objective measurement of situation awareness. In *Situational Awareness in Aerospace Operations.* AGARD-CP-478 (Neuilly sur Seine: NATO–AGARD), 1/1–9

ENDSLEY, M.R., 1990b, Predictive utility of an objective measure of situation awareness. In *Proceedings of the Human Factors Society 32nd Annual Meeting* (Santa Monica: Human Factors Society), 41–5.

ENDSLEY, M.R., 1993, Situation awareness and workload: flip sides of the same coin. In R.S. Jensen and D. Neumeister (eds), *Proceedings of the Seventh International Symposium on Aviation Psychology* (Columbus: Department of Aviation, Ohio State University), 906–11.

ENDSLEY, M.R., 1994, Situation awareness in dynamic human decision-making: theory. In R.D. Gilson, D.J. Garland and J.M. Koonce (eds), *Situational Awareness in Complex Systems* (Daytona Beach: Embry–Riddle Aeronautical University Press), 27–58.

ENDSLEY, M.R., 1995a, Direct measurement of situation awareness in simulations of dynamic systems: validity and use of SAGAT. In D.J. Garland and M.R. Endsley (eds), *Experimental Analysis and Measurement of Situation Awareness* (Daytona Beach: Embry–Riddle University), 107–13.

ENDSLEY, M.R., 1995b, Measurement of situation awareness in dynamic systems, *Human Factors*, 37, 65–84.

ENDSLEY, M.R., 1995c, A taxonomy of situation awareness errors. In R. Fuller, N. Johnston and N. McDonald (eds), *Human Factors in Aviation Operations* (Aldershot: Avebury Aviation, Ashgate), 287–92.

ENDSLEY, M.R., 1995d, Toward a theory of situation awareness in dynamic systems, *Human Factors*, 37, 32–64.

ENDSLEY, M.R., 1996, Situation awareness measurement in test and evaluation. In T.G. O'Brien and S.G. Charlton (eds), *Handbook of Human Factors Testing and Evaluation* (Mahwah: Lawrence Erlbaum), 159–80.

ENDSLEY, M.R. and Bolstad, C.A., 1994, Individual differences in pilot situation awareness, *International Journal of Aviation Psychology*, 4, 241–64.

ENDSLEY, M.R. and Garland, D.J. (eds), 1999, *Situation Awareness Analysis and Measurement* (Mahwah: Lawrence Erlbaum).

ENDSLEY, M.R. and Kaber, D.B., 1999, Level of automation effects on performance, situation awareness and workload in a dynamic control task, *Ergonomics*, 42, 462–92.

ENDSLEY, M.R. and Kiris, E.O., 1995,. The out-of-the-loop performance problem and level of control in automation, *Human Factors*, 37, 381–94.

ENDSLEY, M.R., Selcon, S.J., Hardiman, T.D. and Croft, D.G., 1998, A comparative evaluation of SAGAT and SART for evaluations of situation awareness. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Santa Monica: Human Factors and Ergonomics Society), 82–6.

FRACKER, M.L., 1988, A theory of situation assessment: implications for measuring situation awareness. In *Proceedings of the Human Factors Society 32nd Annual Meeting* (Santa Monica: Human Factors Society), 102–6.

JONES, D.G., 1997, Reducing situation awareness errors in air traffic control. In *Proceedings of the Human Factors and Ergonomics Society 41st Annual Meeting* (Santa Monica: Human Factors and Ergonomics Society), 230–3.

JONES, D.G. and Endsley, M.R., 1996, Sources of situation awareness errors in aviation, *Aviation, Space and Environmental Medicine*, 67, 507–12.

KLEIN, G.A., 1989, Recognition-primed decisions. In W.B. Rouse (ed.), *Advances in Man–Machine Systems Research (5)* (Greenwich, CT: JAI Press), 47–92.

NEISSER, U., 1976, *Cognition and Reality: Principles and Implications of Cognitive Psychology* (San Francisco: W.H. Freeman).

PRITCHETT, A.R., HANSMAN, R.J. and JOHNSON, E.N., 1995, Use of testable responses for performance-based measurement of situation awareness. In D.J. Garland and M.R. Endsley (eds), *Experimental Analysis and Measurement of Situation Awareness* (Daytona Beach: Embry–Riddle University Press), 75–81.

SARTER, N.B. and WOODS, D.D., 1991, Situation awareness: a critical but ill-defined phenomenon, *International Journal of Aviation Psychology*, 1, 45–57.

SELCON, S.J. and Taylor, R.M., 1990, Evaluation of the situational awareness rating technique (SART) as a tool for aircrew systems design. In *Situational Awareness in Aerospace Operations*. AGARD-CP-478 (Neuilly sur Seine: NATO–AGARD), 5/1–8.

SMITH, K. and HANCOCK, P.A., 1995, Situation awareness is adaptive, externally directed consciousness, *Human Factors*, 37, 137–48.

TAYLOR, R.M., 1990, Situational awareness rating technique (SART): the development of a tool for aircrew systems design. In *Situational Awareness in Aerospace Operations*. AGARD-CP-478 (Neuilly sur Seine: NATO–AGARD), 3/1–17.

TAYLOR, R.M. and SELCON, S.J., 1994, Situation in mind: theory, application and measurement of situational awareness. In R.D. Gilson, D.J. Garland and J.M. Koonce (eds), *Situational Awareness in Complex Settings* (Daytona Beach: Embry–Riddle Aeronautical University Press), 69–78.

WICKENS, C.D., 1995, The tradeoff of design for routine and unexpected performance: Implications of situation awareness. In D.J. Garland and M.R. Endsley (eds), *Experimental Analysis and Measurement Of Situation Awareness* (Daytona Beach: Embry–Riddle Aeronautical University Press), 57–64.

Situation Awareness in Teams

E. Salas, E. J. Muñiz†, and C. Prince*

Department of Psychology and Institute for Simulation and Training, University of Central Florida, PO Box 161390, Orlando, FL 32816-1390, †Department of the Navy, Naval Air Warfare Center Training Systems Division, 12350 Research Parkway, Orlando, FL 32826-3275, USA and *409 Balmoral Road, Winter Park, FL 32789, USA

1. SITUATION AWARENESS

The term "situation awareness" came from the aviation operational environment where it has a personal meaning for crew members of all experience levels. In crew members' informal definitions, situation awareness (SA) equates to knowing what is going on with all relevant elements in the flight. For human factors research and theory, definitions of SA have been developed primarily for aviation. In general, definitions emphasize "awareness" over "situation". Perhaps the most quoted definition is that of Endsley (1995a): "the perception of the elements in the environment within a volume of time and space, the comprehension of their meaning, and the projection of their status in the near future."

Endsley (1995a) used an information processing model as a base for her framework, building an explanation for SA that centers on the use of mental models and pattern matching. She characterized SA as a "snapshot" of the crew member's awareness of the situation or situation model. Endsley and other researchers have explained that SA leads to the state of awareness and emphasized the temporal aspect in a dynamic situation (Endsley 1995a, Salas et al. 1995). In a different view, SA has been explained as an accessible situation representation that results from ongoing SAs (Sarter and Woods 1995). That is, a person's state of awareness is based on integrating and updating the knowledge which comes from multiple SAs. This view emphasizes SA as an ongoing cognitive process (Sarter and Woods 1995), as opposed to a state (Endsley 1995a).

Another position was proposed by Adams et al. (1995), who suggested that, because they are interdependent, it is difficult in practice to differentiate between the process and the product (or state). Adams et al. used Neisser's (1976) perceptual cycle to explain SA at the most basic level. In this view, exploratory activities directed by the individual's existing knowledge result in new information. That information is incorporated into the individual's existing knowledge, subsequently helping to direct further exploration. So this position views SA as an iterative process. In summarizing definitions and explanations of SA, Salas et al. (1995: 125) concluded that it "occurs as a consequence of an interaction of an individual's pre-existing relevant knowledge and expectations; the information available from the environment; and cognitive processing skills that include attention allocation, perception, data extraction, comprehension, and projection."

2. TEAM SITUATION AWARENESS

Although explanations of SA have focused on the individual, most aviators fly as part of a crew or team. The impact of the team on each crew member's SA is acknowledged in crew resource management (CRM) programs, where SA is regularly included as a topic. There is evidence that SA is an important aspect of performance in nonaviation teams as well (Salas et al. 1995). Team SA is considered sufficiently relevant to have research projects dedicated to its study and to have published guidelines for its training (Prince 1998).

Despite its recognized importance, there is not a defined scope for team SA. Definitions of team SA have been similar to those for individual SA, with the addition of some team interaction elements as part of team awareness (Salas et al. 1995). There are some who describe team SA as the sum of the awareness of all the crew members; others who consider team SA as only the state of awareness of crew members that are similar; and still others who define team SA as the awareness of each crew member necessary for accomplishing their tasks (Salas et al. 1995).

Each of these positions, as well as information from other relevant literature, was reviewed and considered by Stout et al. (1996) to propose a theoretical framework that explained how cognitive mechanisms facilitate the development of SA in teams performing in dynamic environments (e.g., aviation teams). They suggested that team SA is composed of each team member's SA (based on preexisting knowledge bases and cue and pattern assessment) and the degree of shared understanding (developed via compatible mental models and team interaction behaviors).

The state of team SA at any given time was proposed to be affected by a number of factors, including the situational context (e.g., mission to be performed, aircraft type), environmental elements (e.g., terrain and/or weather conditions), and temporal elements (e.g., time pressure, team being ahead or behind the game plan). This theoretical framework also considers the dynamic quality of team SA by suggesting that it can change instantaneously, based upon how the situation context and environmental and temporal factors affect team processes and the achievement and quality of shared understanding present in the team at any point in time.

3. RESEARCH EFFORTS

As theorists have explained, SA and team SA are primarily cognitive processes, and therefore research is arduous, deficient, and complex. Despite the difficulties, researchers have taken on the challenge of developing conceptual frameworks and conducting empirical investigations that generally have one of three major goals: to increase understanding of the construct, to develop SA instructional strategies, and to enhance SA measurement. Some of these research efforts and their implications are presented below.

3.1. Research to Understand the SA Construct

Several techniques have been used to identify the cognitive activities required for performing highly complex tasks, e.g., knowledge elicitation interviews, multidimensional scaling, concept mapping (Cooke 1994). These techniques, known collectively as cognitive task analysis (CTA), are used to identify task-related knowledge (e.g., concepts, rules, strategies, plans) and cognitive processes (e.g., recognize, anticipate, interpret). Information derived from this approach can have various applications, such as theory evaluation, expert/novice differences, human performance modeling, interface design, and instructional design.

Also CTA would be beneficial for learning about cognitive activities required for effective SA. For example, the ability to recognize relevant cue sets and patterns that facilitate the rapid assessment of a given task situation is critical for effective SA (Stout *et al.* 1999). Based on this notion, Stout and colleagues used knowledge elicitation interviews to uncover cue sets critical for naval aviators to achieve SA (Stout *et al.* 1999). These techniques were successful in identifying trends in the use of common and divergent cue patterns which aviators would attend to while performing a given task. So this data could become the basis for designing and delivering training.

3.2. Research to Develop SA Instructional Strategies

Information derived from a CTA can have implications for designing training strategies to instruct operators so they develop effective cue and pattern assessment that will enhance their SA. One approach that has emerged is cue recognition training, which involves prompting trainees to attend to relevant information in the task environment to increase the probability they will be aware of these cues in the future (Stout *et al.* 1999). Cue sets can be pointed out by using one of several subsets of cueing instructional strategies: passive system prompting, active prompting, behavioral coaching, and instructor-guided practice.

For passive system prompting, cue demonstration can occur via a static system demonstration that prompts relevant information. With active system prompting, cues are pointed out online as the individual or team is practising tasks. Behavioral coaching is a strategy in which a passive demonstration system is used for an instructor to verbalize cues they attend to in accomplishing a task, the relevant processes, and necessary steps taken. Instructor-guided practice is used to demonstrate cues online as the trainee is practising the task. This information is provided with instructor comments on what cues to attend to, the processes to attend to, and the necessary steps to accomplish the tasks. Each of these methods directs the trainee to relevant information in their task environment, and each is believed to enhance SA. Given that these strategies are still at a conceptual stage and have not been empirically evaluated, the choice of one method over another is probably best made by considering available resources.

3.3. Research to Enhance SA Measurement

The strategies, such as training, designed to enhance SA need to be evaluated for their effectiveness. This requires tools to assess individual SA and team SA. Researchers have proposed assessing cognitive processes and behavioral manifestations that indicate presence or absence of SA. Several strategies have been developed to capture a component of SA, but the effectiveness of each one has been criticized. It seems SA is difficult to measure.

SA measures have typically been clustered by several researchers into three categories: subjective measures, implicit measures, and explicit measures (Fracker 1991). Subjective measures are used to obtain an operator's impressions of their own SA or their impressions of another person's SA. An example is the situation awareness rating technique (SART), which requires individuals to rate statements based on the perceived demands of attention to different resources and perceived understanding of specific situations experienced during a given scenario (Taylor 1989). These measures have been critiqued as being ambiguous about whether individuals actually used or integrated information to understand the dynamics of a situation, so it is difficult to determine what these ratings actually capture (Fracker 1991). Despite these criticisms, SART type measures are commonly employed because they are easy to use and to document.

A second cluster of measures are the explicit measures, designed for individuals to self-report information from conscious memory. One of the most popular is the situation awareness global assessment technique (SAGAT); a "freeze" measurement technique, it was developed to assess the awareness of a pilot at various times during a scenario (Endsley 1995b). SAGAT stops a pilot as they fly a scenario, blanks all instruments, and asks a number of questions that relate to awareness (e.g., current heading, altitude, winds at destination). Answers to these questions are used to indicate the pilot's level of awareness. However, this technique has been criticized for intruding into the scenario. Not only is the pilot "removed" from the scenario's actions during questioning, but the questions asked may influence the information to which they are attending. On the other hand, prompting individuals to attend to specific information is believed to have potential as a training strategy for making individuals aware of the extent to which they missed relevant information in the situation.

The third cluster, implicit measures, attempt to use aspects of performance as an index of SA. One example is SALIANT, an event-based methodology whose acronym stands for situation awareness linked instances adapted to novel tasks. This approach uses a number of theoretically derived, predetermined markers; the markers are associated with specific events and they indicate the awareness of an individual or crew (Muñiz *et al.* 1998, Bowers *et al.* 1998). SALIANT was developed to be applicable in a variety of operational settings. The utility and validity of SALIANT have been evaluated with teams performing simulated aviation scenarios. Results of two preliminary validation studies indicated that SALIANT was related to measures of team communications and indices of performance, as suggested by theory (Muñiz *et al.* 1998, Bowers *et al.* 1998). A main criticism of this technique is it assumes a direct link between SA and performance. This problem may be addressed in part by generating indicators of SA based on theory and different from performance indices. A main advantage of the implicit methodology is that it facilitates online assessment of SA without disrupting an individual's behavior.

In general, a major criticism for SA measures is that each of them appears to have the potential to evaluate a component of SA while remaining insufficient to capture the construct as a whole. This suggests that SA should be measured using multiple evaluation instruments to capture its various dimensions.

4. PRINCIPLES OF INDIVIDUAL AND TEAM SA

SA has become important in various operational settings, therefore much research is dedicated to this topic. These efforts have generated a great deal of relevant information, and we can use it to extract principles or fundamentals about the current state of our knowledge.

SA is primarily a cognitive construct

At a high level there are several cognitive elements involved in the formation of SA, such as memory, perception, and action (Neisser 1976, Adams *et al.* 1995). Specifically, knowledge about

a particular situation guides exploratory and perceptual activities used to increase the interpretation about available information. But this is shaped by an individual's preexisting knowledge and expectations and basic familiarity with the current situation. This results in an update in the individual's knowledge, a modification in expectations, and another cycle of information.

SA is not a process but a state

SA is the *state* of knowledge for any given situation at some point in time, whereas the *process* is the cognitive activities that allow the situation to be assessed (Sarter and Woods 1995, Endsley 1995a, Adams *et al.* 1995). Although there are many useful motivations for establishing this distinction, e.g., clarification of the construct, measurement, and system design, remember there is no clear division at the cognitive level. This suggests an interdependence between the processes and the resultant state (Adams *et al.* 1995).

SA in teams involves individual SA and team-shared knowledge

A recent theoretical framework suggests that SA in teams has two main components: individual SA and shared knowledge. Stout and her colleagues have noted how each team member forms SA based on preexisting knowledge along with cue and pattern assessment. Team members also achieve a level of shared knowledge about a given situation and this is developed using compatible mental models and team interaction behaviors.

SA must be measured with multiple methods

Although there are now many measurement methodologies, they share the same problem of capturing only one element of SA. In this respect, they are *all* considered insufficient. There is no single measure that gives a complete index of SA. This means that we need more than one type of measurement to capture the various dimensions of SA so we can obtain a true sense of its level in an individual or team.

SA measures must be construct valid

Many critics have noted that SA measures are not clear about what they capture. To address this problem, it is important to derive SA components as suggested by theory; this ensures the measures actually assess the relevant construct. According to Muñiz *et al.* (1998) and Bowers *et al.* (1998), to demonstrate its effectiveness, each measure should be compared with other SA measures and with measures of other theoretical constructs which modify or influence SA (e.g., communication, performance, mental models).

SA must be trained in context

Cue recognition training has recently emerged as a way of training individuals to attend to relevant cues delineated a priori with CTA (Stout *et al.* 1999). This makes it critical to determine which cues are more relevant and when they are most relevant, given the specific mission at hand. This information is crucial in determining how the cues are important within a given context, hence it affects how to incorporate appropriate learning activities to make them salient during training. With proper training, an individual can derive why the cues are meaningful in context, and can go on to build appropriate knowledge structures for future situations that require attention to these cues.

Overall there is a great deal of information derived from SA research, suggesting that some strides have been made in this area. Progress in SA-related research is helping us to formulate some principles. The future of SA research seems bright. There is still much more to be learned and many more questions to be answered. But our knowledge about this domain is undergoing continual advancement.

REFERENCES

ADAMS, M., TENNEY, Y., and PEW, R., 1995, Situation awareness and the cognitive management of complex systems. *Human Factors*, 37, 85–104

BOWERS, C.A., BARNETT, J., WEAVER, J. and STOUT, R.J., 1998, Empirical validation of the SALIANT methodology. Paper presented at the 1st Annual Symposium and Business Meeting of the Human Factors and Medicine Panel on Collaborative Crew Performance, Edinburgh, Scotland.

COOKE, N.J., 1994, Variety of knowledge elicitation techniques. *International Journal of Human-Computer Studies*, 41, 801–49.

ENDSLEY, M., 1995a, Toward a theory of situation awareness in dynamic systems. *Human Factors*, 37, 32–64.

ENDSLEY, M., 1995b, Measurement of situation awareness in dynamic systems. *Human Factors*, 37, 65–84.

FRACKER, M.L., 1991, Measures of situation awareness: review and future directions. Wright-Patterson Air Force Base, Ohio; Armstrong Laboratories.

MUÑIZ, E.J., STOUT, R.J., BOWERS, C.A., and SALAS, E. 1998, A methodology for measuring team situational awareness: situational awareness linked indicators adapted to novel tasks (SALIANT). Paper presented at the 1st Annual Symposium and Business Meeting of the Human Factors and Medicine Panel on Collaborative Crew Performance, Edinburgh, Scotland.

NEISSER, U., 1976, *Cognition and Reality: Principles and Implications of Cognitive Psychology* (San Francisco: W.H. Freeman).

PRINCE, C., 1998, Guidelines for situational awareness training, NAWCTSD/FAA/UCF Partnership for Aviation Team Training.

SALAS, E., PRINCE, C., BAKER, D. and SHRESTHA, L., 1995, Situation awareness in team performance: implications for measurement and training. *Human Factors*, 37(1), 123–36.

SARTER, N.B. and WOODS, D.D., 1995, How in the world did we ever get into that mode? Mode error and awareness in supervisory control. *Human Factors*, 37, 5–19.

STOUT, R.J., CANNON-BOWERS, J.A., and SALAS, E., 1996, The role of shared mental model in developing team situation awareness: implications for training. *Training Research Journal*, 2, 85–116.

STOUT, R.J., CANNON-BOWERS, J.A., and SALAS, E., 1999, Team situation awareness: cue recognition training, unpublished.

TAYLOR, R.M., 1989, Situational awareness rating technique (SART): the development of a tool for aircrew systems design. *Proceedings of the NATO AGARD Conference on Situational Awareness in Aerospace Operations*, AGARD–CP–478 (Springfield VA: National Technical Information Service).

Skill Learning: Augmented Feedback

D.E. Young†, R.A. Schmidt† and T.D. Lee‡

†Exponent Failure Analysis Associates, 5401 McConnell Avenue, Los Angeles, California 90066, USA

‡Department of Kinesiology, McMaster University, 1280 Main Street West, Hamilton, Ontario, Canada, L8S 4K1

1. INTRODUCTION

The performance of a skill is influenced, and often determined, by the information provided as feedback from an external source, such as a supervisor, or expert. One class of such information is augmented feedback — movement-related information about the task, which supplements a performer's intrinsic feedback. Over the past century or so researchers have examined the various aspects associated with the informational content of augmented feedback, the schedules to present this information, the temporal placement of this feedback, and plausible theoretical explanations that describe how feedback affects skilled performance (for a review, see Schmidt and Lee 1999). In this chapter we discuss the various classes and experimental methods used to study augmented feedback, then present some of the primary laws regarding how augmented feedback influences performance and learning, and end with a description of some explanations that serve as a basis for applying feedback to human performance.

2. CLASSES OF AUGMENTED FEEDBACK

Of the various classes of information that are received and processed as feedback, several distinctions and characteristics are of importance for defining augmented feedback. First, augmented feedback is concerned with the control of movement, such as the direction of the error that a movement has produced. Second, the information is generated as a consequence of the movement; hence, it is not available before the execution of a movement. Several other important distinctions are presented next.

2.1. Extrinsic Feedback

The information provided about the movement by an external source (e.g. instructor, computer, etc.) is defined as extrinsic feedback. In contrast, the sensation or feel that is inherently provided by our sensory receptors in the normal course of a movement is classified as intrinsic feedback. Often, this information, which includes tactile, auditory, and visual feedback, can provide a sufficiently clear representation of the necessary cues relevant for performance (e.g. turning a potentiometer to move the hands on a watch to the appropriate time). However, there are many instances where additional information is needed for performance (e.g. the information provided by a speedometer to indicate the speed of vehicle) or to adequately learn how to perform skill. In this way, extrinsic feedback supplements or augments the information beyond that which is available from the performer's own intrinsic sources.

2.2. Concurrent Feedback

Augmented feedback, although generated as a consequence of movement, can be provided to a performer during a movement. When delivered during the course of a movement, the information is referred to as concurrent feedback. The information feedback about the glide path that occurs when landing a plane is one such example. Generally, performers can utilize concurrent feedback to minimize errors in continuous tasks that exceed several hundred milliseconds. Although concurrent feedback has obvious benefits for performance, it can be detrimental to the relatively long-term beneficial effects associated with learning (Schmidt and Wulf 1997). One potential problem with the provision of augmented concurrent feedback is that it can be used as a substitute for sensory information — when feedback is not available, as in a later test condition, performance suffers (see Section 3).

2.3. Terminal Feedback

When augmented feedback is delayed until the completion of a movement, it is referred to as terminal or post-movement feedback. Often, feedback is delayed to provide the complete set of information about the entire response or its outcome. The vast majority of research has manipulated some aspect of terminal feedback that is not coincident with the movement. This body of research provides the basis for the principles and laws that govern when, how, and what augmented feedback should be provided.

2.4. Informational Content

There are various classes of augmented feedback that provide different sources of information to a performer. Two of the more important sources of information for both performance and retention of skilled performance are given as outcome information (knowledge of results) and movement-pattern information (knowledge of performance).

2.4.1. Knowledge of results

Information about extent to which the performer achieved the outcome in terms of the environmental goal is typically referred to as knowledge of results (KR). This type of feedback provides information about the success of the movement, but not about the movement pattern or characteristics that led to the outcome. KR can augment intrinsic feedback, but also can be redundant with the information the performer normally receives during the course of a movement (e.g. receiving auditory information from a buzzer about entering a building and visually detecting the same information).

2.4.2. Knowledge of performance

In contrast to KR, which provides information about goal success, feedback concerned with the movement pattern itself or some aspect the human system that leads to the movement is referred to as knowledge of performance (KP). In the broadest sense, this feedback includes information about movement pattern kinematics, the forces that produced the movement (or kinetics), and other aspects of the physiological system that are not easily perceived by a performer.

Videotaped feedback of a performer's action is now a common method used to display KP of limb and/or body kinematics. Although there often is a delay associated with the playback of this augmented feedback, performers can observe their overall movement pattern to gain an enhanced perspective of the spatio-temporal aspects of the action and coordination pattern. One

critical aspect in the provision of videotaped feedback involves directing a performer's attention to specific aspects of the movement that require modification or correction. In this way, a performer is less likely to be distracted by irrelevant features of the movement or environmental stimuli, and can focus on those characteristics of the movement pattern that are critical in achieving the desired behavioral outcomes. Failure to direct a performer's attention to specific aspects of the videotape can diminish the beneficial effects of this presentation method and make KP ineffective for learning.

3. AUGMENTED FEEDBACK RESEARCH METHODOLOGY

Two experimental approaches have played an important role in understanding how feedback influences performance. Each has ramifications for training procedures as well as retention and transfer performance.

3.1. Knowledge of Results Paradigm

Much of what is known about how augmented feedback functions to alter human performance and skill acquisition has been derived from one basic experimental paradigm. Typically, this methodology requires subjects to control a very simple, single-degree-of-freedom task (e.g. pointing), where the outcome is masked from the performer's intrinsic feedback sources. In this way, augmented feedback — typically given as information about the movement's success or KR — becomes a necessity for performance and learning. In this movement environment, albeit a very artificial and contrived situation, manipulations of the precision, schedule and temporal placement of feedback can be made to determine how feedback influences performance. Hopefully, the findings from these experiments generalize to other, more complex movement settings, where some type of augmented feedback is provided to a performer. This body of research serves as the basis for our formulation of laws of augmented feedback.

3.2. Transfer Designs

Another important empirical consideration in the study of augmented feedback includes the utilization of a transfer design. This method for studying the influence of some feedback variable (e.g. precision of feedback) consists of an acquisition phase and a retention or transfer phase that are separated by a rest interval that varies from a few minutes to a week or more. The purpose of using this experimental method is to distinguish the temporary effects of feedback, which are associated with performance in acquisition, from the more permanent effects of feedback that would be displayed in performance during the retention test. Other advantages of this method are presented in the section on practice (see Skill Learning: Conditions of Training, Chapter xxx).

4. LAWS OF PRECISION

The accuracy, in terms of the correctness or specificity, of feedback has been of interest to scientists because of its implications for information processing theory. However, an understanding of the laws of precision will also provide a reliable method for administering feedback in training situations.

4.1. Correct and Erroneous Feedback

Early research on feedback established that correct feedback reduced errors in performance, and was responsible for gains in retention over nonsense or no feedback. Subsequent research further established the powerful influence of feedback with manipulations of incorrect or erroneous feedback. In these studies, incorrect feedback was found to impair performance and learning in tasks where subjects could evaluate their performance via intrinsic feedback mechanisms. Hence, subjects used the inaccurate feedback, instead of their own more accurate internally generated sensory feedback, as a basis for acquiring a new skill. These findings are of particular relevance for the provision of KP, as the incorrect information about the pattern of actions could result in negative or reduced learning.

4.2. Specificity of Feedback

The consideration of how specific the feedback should be provided has typically been separated into two classes: qualitative and quantitative. Qualitative feedback is general information about goal success can be provided in forms that merely indicate the correctness ("right" or "wrong"), direction ("short" or "long), or along some other dimension ("high force" or "low force"). In contrast, to qualitative feedback, specific information or quantitative feedback about the accuracy of the response can be provided with exact deviations from target-goals (e.g. "2 mm too short of the target"). This specific feedback, when compared with qualitative feedback, is generally more effective for both performance and learning. However, studies have shown that extremely precise information about errors is ignored and, therefore, unnecessary. Generally, the evidence suggests that information about the direction and magnitude of the error is most beneficial.

Another variant of feedback precision, bandwidth feedback, provides specific information about the accuracy of a movement only when performance exceeds a specified boundary or tolerance for error. In this method of providing feedback, learning is significantly improved in retention with 10% bandwidth (i.e. providing feedback when the errors are less than 90% correct) when compared with a 0% bandwidth (i.e. providing feedback after every trial). As discussed below, this method for providing feedback may be effective because of the precision of the information provided, but also due to the reduced frequency that the information is given to the performer.

5. TEMPORAL LOCUS PRINCIPLES

The provision of augmented feedback between trials of training, i.e., during the inter-trial interval, has created two additional intervals of importance for performance and learning. Unfortunately, the research is somewhat unclear as to how some of the associated temporal-locus variables influence movement control. Nevertheless, this section provides an overview of the general understanding in this area.

5.1. Feedback–Delay Interval

The first is the feedback-delay interval, representing the time that the feedback is delayed after the movement. It has long been believed that increasing the feedback-delay interval would degrade learning for two reasons. First, a long delay would be similar to the degraded learning effects found in the delaying a

reinforcement or reward in animal learning studies. Second, a lengthened feedback-delay interval would weaken the possibility that the performer could associate the movement commands with the outcome or intrinsic feedback consequences thought to be critical for learning (Adams 1971). These notions, however, are no longer considered correct as empirical research has indicated that increased feedback-delay intervals have not affected learning of human skills. In fact, the opposite effect has been shown — no delay or instantaneous presentation of feedback after the movement degrades learning, and, to some extent, performance as well.

Further research that has involved the completion of interpolated activities has provided support for the notion that subjects process intrinsic information about their movement during this interval. Research has shown that interfering with these processes by "filling" the interval with a non-related, attention-demanding task degrades learning. In other research, which has required subjects to verbally estimate their movement errors during this interval, skill learning is facilitated. Together, these findings provide strong evidence that the intrinsic error-detection processes associated with movement control are important in the learning process.

5.2. Post-Feedback-Delay Interval

In contrast to the feedback-delay interval, the temporal interval following the provision of feedback has not been as thoroughly investigated. Some data support the presumption that two important activities occur here. First, a comparison of intrinsic and extrinsic feedback is completed, which is then followed by a development of a plan for the subsequent movement. Unfortunately, support for these propositions is sparse and potentially related to variations in the inter-trial interval. Nevertheless, there is no evidence to suggest that important processes do not occur here.

6. SCHEDULING FEEDBACK

There are a variety of ways in which one could present feedback over the course of an training session. In this section, we present several of the more well-known and studied methods.

6.1. Frequency of Feedback

Studies have evaluated the "amount" or number of trials for which feedback is presented through absolute and relative frequency manipulations. Absolute frequency is the number of feedback presentations received over the course of an acquisition phase, while relative frequency refers to the proportion of trials on which feedback is provided. Research consistently reports that more feedback in the acquisition phase facilitates performance over lower absolute and relative frequencies. However, contrary to these findings, the effects in retention are very different. The vast majority of research indicates that relative frequencies, some as low as 10%, are more beneficial for retention performance than every-trial or 100% feedback conditions.

6.2. Fading Schedules

The effects of relative frequency are more pronounced when provided in a "fading" schedule, which provides feedback relatively often during the initial stages of practice and systematically withdraws feedback as practice progresses. When

compared with a 100% feedback condition, the fading of feedback generates similar effects to those discussed above: the lower relative frequency reduces acquisition performance, relative to an every-trial feedback condition, but facilitates retention performance.

6.3. Trials Delay

Another method for providing feedback to a performer involves the delay of this information for more than a trial. That is, the feedback for a given trial would not be verbalized or provided to a performer until after a number of intervening trials have been completed. On the surface, this trials-delay procedure would appear to be extremely disruptive to performance as a learner would have difficulty in linking or associating a particular feedback result with a movement responsible for producing it, since several other trials could potentially intervene. Early literature in this area confirms these negative effects on acquisition performance. However, more recent research that has examined the effect of a trials-delay procedure in both acquisition and retention report counter-intuitive results. Instead of lower performance relative to an every-trial schedule, a trials-delay of feedback given in a summary format — where feedback about each trial in a sequence was provided after all trials in that set had been completed — resulted in greater retention performance (e.g. Schmidt et al. 1989). These general "reversal" effects are shown in Figure 1. Similar effects have been found for an average format, which involves providing an average value for a set of trials at the completion of the set.

One aspect of subsequent work on summary feedback has attempted to determine an optimal summary length. This effort was based on the assumption that an extremely long summary length would be of little use to a performer, yet every-trial feedback would be detrimental to learning. Although no particular summary length appears to optimal for learning, there is strong evidence to indicate that summary length is related to task complexity, where longer summary lengths are appropriate for simple tasks and shorter lengths for more complex tasks.

A. No visual enhancement cue   B. A solid reference line

C. A translucent reference cylinder   D. Four solid reference line
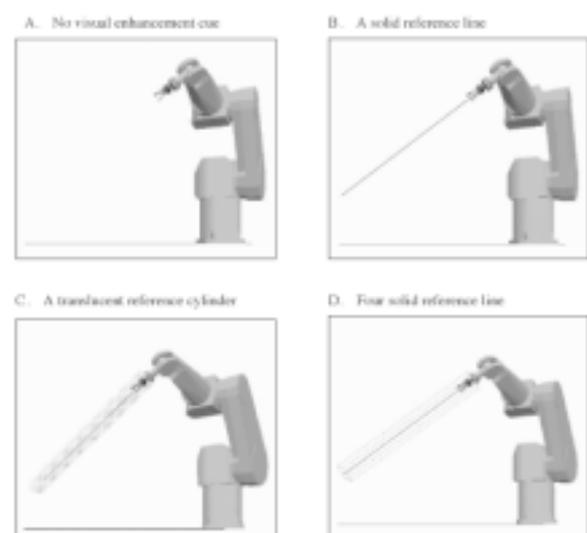
Figure 1. Performance score in acquisition and retention tests for two different schedules of augmented feedback (redrawn from Schmidt et al. 1989).

7. THEORETICAL EXPLANATIONS

Over the course of the last century or so, there have been numerous explanations for how augmented feedback is used in the control and acquisition of skills (see Adams 1987, for a historical review). In this section, we present the theoretical explanations that have substantial empirical support and represent the current thinking about how feedback works.

7.1. Augmented Feedback Provides Information about Response Errors

Whereas augmented feedback has long been believed to function primarily as a reward, augmented feedback also can be used to determine the nature of their errors. This information is critical in planning what do change and how to make subsequent movements more accurate. Feedback may also serve to develop accurate error detection and correction mechanisms, where performers evaluate intrinsic feedback and compare it with extrinsic feedback.

7.2. Augmented Feedback can Motivate Performers

Augmented feedback has a strong energizing role for performers. It can make a task seem more interesting, enjoyable, and often results in individuals setting higher performance goals. As a result, performers are more inclined to practice longer, more often, and with more effort. These motivation effects appear to be performance phenomena, which can be expected to subside after the practice or training period has completed. However, given that learning is significantly influenced by practice and the conditions associated with it, the motivational qualities of feedback can be thought to indirectly influence learning as well.

7.3. Augmented Feedback Works like Guidance

Whereas older ideas suggested that augmented feedback helps to develop associations between movements and their outcomes, more recent evidence suggest that feedback has guidance-type properties (Salmoni *et al.* 1984). First, feedback directs performers about how movements are erroneous and helps to reduce these errors. Second, too-frequent or immediate feedback may prevent the development of important information processing activities. Third, too much information may force performers into developing corrective behaviors that are beneficial for immediate performance, but detrimental for long-term performance. These "short-term maladaptive corrections" occur when performers constantly adjust or change their movements after feedback is received. Finally, as with physical guidance, performers may develop a reliance on augmented feedback, process it as part of the task, and have difficulty with performance on the task later when feedback is not available.

REFERENCES

ADAMS, J.A., 1971, A closed-loop theory of motor learning. *Journal of Motor Behavior*, 3, 111–50.

ADAMS, J.A., 1987, Historical review and appraisal of research on the learning, retention, and transfer of human motor skills. *Psychological Bulletin*, 101, 41–74.

SALMONI, A.W., SCHMIDT, R.A. and WALTER, C.B., 1984, Knowledge of results and motor learning: a review and critical reappraisal. *Psychological Bulletin*, 95, 355–86.

SCHMIDT, R.A. and LEE, T.D., 1999, *Motor Control and Learning: A Behavioral Emphasis* (3rd ed.). (Champaign: Human Kinetics).

SCHMIDT, R.A. and WULF, G. 1997, Continuous concurrent feedback degrades skill learning: Implications for training and simulation. *Human Factors*, 39, 509–25.

SCHMIDT, R.A., YOUNG, D.E., SWINNEN, S. and SHAPIRO, D.C., 1989, Summary knowledge of results for skill learning: support for the guidance hypothesis. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 15, 352–59.

Skill Learning: Conditions of Training

T. D. Lee†, R. A. Schmidt‡ and D. E. Young‡

†Department of Kinesiology, McMaster University, 1280 Main Street West, Hamilton, Ontario, Canada, L8S 4K1

‡Exponent Failure Analysis Associates, 5401 McConnell Avenue, Los Angeles, California 90066, USA

1. INTRODUCTION

The acquisition of new skills is critical for operators who are required to adapt activities to workplaces that are new or which have been updated with different equipment, tasks, or procedures. For situations in which it is not cost-effective or safe for the operator to acquire these skills on the job, special training conditions may be organized so that the operator may acquire these behaviors before entering (or reentering) the workplace. These training conditions are designed to provide the operator with an opportunity to practice the skills that will be necessary for safe and effective performance in the workplace. The purpose of this chapter is to describe how training conditions can be organized to have a significant impact on the effectiveness of performance when the employee enters the workplace.

Applied research, involving specific workplace training conditions in specific job situations, has rarely been conducted or reported in the literature (for an exception, see, for example, Baddeley and Longman 1978). However, a significant amount of research has been conducted regarding the acquisition of skills that are novel to the learner and require practice in order to be learned. We discuss training *conditions* under two general sections: (1) involving hands-on practice (termed *on-task* training conditions), and (2) involving practice away from the actual task (*off-task* training conditions). Before we begin this discussion, however, we describe the important issue regarding how learning is assessed, and distinguish between situations in which training conditions primarily impact temporary changes in performance from the conditions that influence the more permanent changes associated with *learning*.

2. THE LEARNING/PERFORMANCE DISTINCTION

Some training conditions are effective in providing an immediate, short-term benefit to performance. In learning a new procedure, for example, guidance conditions can be arranged that are specifically designed to minimize the amount of error that can occur while practicing a task. However, even though performance may be enhanced while the guidance training condition is in effect, this temporary boost often fails to satisfy the overall training goal, which is to influence *learning* of the skills. By definition, learning involves the relatively permanent change in the capability to perform a skill (Schmidt and Lee 1999). Assessments of learning during the time when a guidance training condition is being implemented can be misleading because of the temporary elevation in performance that is provided by these conditions. A better way to assess learning is to evaluate *retention* or *transfer* of the skills following a period in which the skills have not been specifically practiced, and where the training conditions are no longer providing a temporary benefit to performance.

Performance in retention and transfer tests is used, retroactively, to assess the effectiveness of the different conditions that had used during the training interval.

In this paper, we refer to *effectiveness* in terms of whether a type of training condition results in retention or transfer performance that is qualitatively and/or quantitatively superior to one or more comparison training condition(s). But, training conditions are also concerned with *efficiency*, defined in terms of the amount of training time that is required to produce equivalent amounts of learning under different conditions. In our review of training conditions we will be concerned with both effectiveness and efficiency when discussing the relative merits of specific conditions of training.

3. ON-TASK TRAINING

Physical interaction with the materials, equipment, tools, etc. to acquire new skills is the most critical condition of learning. Below we describe conditions in which the organization of physical interactions with a task during training session(s) has been assessed.

3.1. Distribution of Work and Rest Periods

Practice *distribution* refers to the amount of time spent in training (periods of "work") relative to the amount of time spent not in training (e.g. resting). Experiments of this type often include conditions of *massed* and *spaced* practice, which represent two points on a practice distribution continuum. No rest at all during practice on the task would represent an extreme form of massed practice, and an infinite time between on-task practice would represent the hypothetical extreme of spaced practice. Many types of tasks have been used in this research and the vast majority of them can be classified as *continuous* tasks, in which the work activity continues uninterrupted for an extended period of time (e.g. 20 or 30 s duration). The results of this research are very clear: spaced practice conditions produce very large performance benefits during the training period in comparison with massed practice conditions (reviewed in Lee and Genovese 1988). A similar finding occurs in retention — spaced practice results in better retention performance (and, hence, better *learning*) than massed practice (although the size of the learning effect is usually not as large or dramatic as the performance effect).

Thus, for practice distribution *effectiveness*, the general conclusion is that it is better to provide significant rest periods between periods of work during practice than to provide little or no rest at all. Note however, that longer rest periods will reduce the *efficiency* of the training protocol in terms of the overall time spent in the training regime, due to the increased rest required for effectively spaced practice. Although this could pose a dilemma for deciding on the relative merits of spaced versus massed training, we discuss below a training condition in which practice on other tasks can be substituted in the "rest" interval to improve both learning effectiveness *and* efficiency.

3.2. Task Variability

In many job-related situations there is wide variation in the specific design of the workplace, the tools to be used, and the actions required of the operator. For example, parking lot attendants must be able to adapt to a variety of driver cockpit dimensions, using pedals, knobs, levers and other manipulanda

that vary widely in size, shape, action, etc., and they must perform these actions with varying amounts of movement amplitudes, forces, and accuracy tolerances. How can an effective training regime prepare the operator to adapt to these changing demands in the workplace?

Research on this issue has focused specifically on *transfer* as the main criterion for learning. Typically, studies have contrasted two types of training conditions. In one condition (*low* variability) the operator practices only one version of the task to be learned. In another training condition (*high* variability) the operator practices the same number of total practice trials as in the low variability condition, but disperses these trials across several or more variations of the task to be learned. The assessment for learning effectiveness is a transfer test in which individuals in both experimental conditions are required to perform a novel version of the task. Using our original analogy, such a transfer test might involve an employee parking a completely unfamiliar car for the first time.

As one might expect, performing *n* trials on one version of a task results in better overall performance than practicing the same number of trials spread out over several task variations (see Shapiro and Schmidt, 1982, for a review of this literature). In contrast, transfer to a novel task variation almost always favors the operator who has practiced multiple variations of the task. These findings are consistent with the view that skills are learned as a *schema* — an abstract form of memory representation for the motor skill that allows the operator to generalize beyond the subsets of skills that have been specifically practiced (Schmidt 1975). Development of the schema will be more effective if the practice regime promotes a breadth of task variation experience. It should also be noted that training efficiency will also be facilitated by variability in practice. Not only will training time on new task variations be reduced by variable practice, but evidence also suggests that retention of the specific tasks that have been practiced is also facilitated by variable practice (for details, see Schmidt and Lee 1999).

3.3. Contextual Interference

Above we presented evidence that training several versions of a task facilitates retention and transfer. One additional issue about this research pertains specifically to *how* the training is conducted. If *n* training trials are conducted on each of *x* tasks or skills, then a critical question remains as to how to *structure* the training order of the tasks. A traditional view of skill training advocates concentrated, drill-type practice sequences (termed *blocked* training) in which all of the practice attempts on any one task are practiced in isolation; training trials on another task are introduced into the practice sequence only when considerable mastery of performance has been demonstrated for the current task. More recently, however, researchers have examined the effectiveness of a practice sequence in which practice trials are conducted concurrently on all of the task versions to be trained. In this practice sequence (termed *random* training) there is no concern that a certain level of performance must be achieved before practice on a new task is introduced. In fact, the random ordering of practice trials suppresses the overall level of performance that can be achieved during the training session. This suppression in performance, however, is only an artifact of the training regime. When learning is assessed later, in retention

and transfer, there is overwhelming evidence to suggest that random practice has resulted in far superior learning effectiveness than blocked practice. Although differing views exist for the reasons why random practice is so effective, there is considerable evidence to suggest that it relates to the *cognitive effort* that operators must exert when faced with the difficulties of practicing in such a highly performance-debilitating task order (reviewed in Schmidt and Lee 1999).

The blocked versus random practice issue, termed the *contextual interference effect* (Shea and Morgan 1979), represents a fundamental paradox regarding the performance of operators during training as compared with their later performance in retention and transfer. By its very nature, blocked practice results in superior performance during the training session(s) than random practice, and might therefore be considered the better training order in terms of both effectiveness and efficiency. The paradox is that blocked practice is actually very *ineffective* for learning because much of the improvement in performance vanishes quickly, and *inefficient* because retraining on the same skills will be necessary. This basic finding has been replicated for a large number of motor skills, as well as various perceptual, verbal, and problem-solving skills (for a review, see Schmidt and Lee 1999).

Random practice conditions also represent a potential solution to the problem of inefficiency when practice distributions are spaced, as discussed above. Recall that, in spaced practice, the interval of time between practice attempts is spent resting or otherwise not engaged in practice. Compared with massed practice, this more effective practice schedule is not very efficient because it lengthens the overall total training time. The contextual interference evidence, however, would suggest that both efficiency and effectiveness would be facilitated if, rather than resting, the operator spent this time engaged in practice on a different task or a variation of the task.

3.4. Part- and Whole-task Practice

When introducing a new skill to be learned, a frequent question is whether the operator should practice the task in its entirety (termed *whole-task* training) or break down the task and practice it in its component parts (termed *part-task* training). The issue is primarily one involving transfer, as eventually the operator will be required to practice the whole task, and whether transfer from part-task training will be more effective and/or efficient than practice of the entire task.

The research evidence on this question varies depending on the type of task that is practiced. In some tasks, the operator's goal is to perform a brief action that involves the coordination of various body segments. Shifting while operating a vehicle is a good example as the actions of all four limbs must be timed precisely in order to achieve a smooth shift. In these tasks (termed *discrete* tasks), the whole action of each of the limbs must be learned, but successful completion cannot occur unless they can be performed together, with the correct timing of their sequence, including their mutual (dynamic) interactions. For discrete (and continuous) tasks, the research suggests that training the whole task is more effective and efficient than part-task training; and part-task training has even been shown to produce negative transfer to the whole.

In a different type of task (termed *serial* tasks), any one discrete

component of the task can be considered a unique entity of the whole task to be performed. In serial tasks, the whole task actually is comprised of a series of individual, discrete component tasks. For example, swiping a card with a magnetic stripe through a card reader represents a component task that initiates an automated human–computer sequence. In this situation, it is not critical that the component task be learned as part of any specific whole task. Indeed, for most serial tasks the individual discrete components represent parts of other serial tasks as well (e.g. swiping the card for initiating a purchase, to identify oneself, to enter a secure area, etc.). Research suggests that part-practice is as effective for whole-task transfer and is more efficient due to the transferability of the individual components of the task to the performance of other serial tasks.

4. OFF–TASK TRAINING

Although the effects are not as large as when physically interacting with a task, considerable learning can be gained from different types of training conducted away from the task. Below we describe how training can continue when the operator does not physically perform the task.

4.1. Observational Learning

Cognition, although overshadowed by the dominance of the active, physical component in many tasks, plays a very important role in learning new skills. In fact, most learning theories acknowledge that the cognitive stage is the initial step in the acquisition of any new skill. The operator must understand the goal of the task and the regulatory environmental information that is important to attend to, and be capable of verbalizing some rudimentary plan about how to structure the action prior to performance. In other words, knowledge that is verbalizable and which can be imaged serves a critical role in the acquisition, retention, and transfer of any skill.

Acquisition of the cognitive components of skill can be facilitated through observation. Some of this research has been directed at learning sport skills. An example is the use of videotapes of skilled athletes to model correct or desirable performance. However, research suggests that simply watching a skilled performer has limited benefits. For example, an equivalent benefit to learning can be achieved from observing the performance of a cohort who is at about the same stage of learning as the observer, to ascertain what the cohort has done and how attempts are made to improve performance. In a sense, the observer is drawn into the same problem-solving process as the cohort, but without actually performing the task.

There are many other research issues pertaining to the effectiveness of the observational learning situation that are of obvious importance to human factors, but are beyond the scope of this review (McCullagh 1993 is a good source for further information). Perhaps of most importance however, is the potential for observational learning to improve the efficiency in the training of new skills, especially so when the cost of "hands-on" training with expensive equipment is a factor (e.g. learning to fly an airplane).

4.2. Mental Practice

As the term suggests, mental practice involves the "performance" of a skill in the absence of overt, physical activity. Various forms of mental practice have been investigated, although the most common involves the use of *imagery*. Above we discussed the role that cognition plays in the early stages of learning, and why observational learning is effective. The same logic applies for mental practice. The verbalizable components of skills can be rehearsed through subvocalizations, the visualizable components can be imagined (from either an internalized viewpoint or from a "bird's-eye" viewpoint), and the nature of the action can be subjected to a form of mental rehearsal that runs through both the motor and sensory components of the movement.

Mental practice is quite effective in learning a wide number of tasks. In fact, some studies have shown that, when combined with trials involving physical practice, mental practice is as effective as an equivalent number of additional physical practice trials. Many other factors, such as individual differences, schedules of mental practice, and the like, were reviewed by Feltz and Landers (1983; also Schmidt and Lee 1999). Similar to observational learning, the role of mental practice to human factors has important implications for the efficiency of training programs. If some of the learning benefits from expensive physical interactions with a piece of equipment can be generated by mental rehearsing the task elsewhere, then the overall cost of training could be reduced substantially.

5. SUMMARY COMMENTS

In the foregoing discussion we have presented evidence of a number of situations in which learning can occur in the *absence* of observable physical improvements in performance (e.g. in mental practice and observation). As well, situations occur where performance during practice completely misrepresents how learning is proceeding (e.g. variable versus constant training conditions; random versus blocked training conditions). In these and other instances feelings about the effectiveness and efficiency of a training condition may be an *illusion* if based on subjective evaluations or short-term changes in performance (Bjork 1998). Thus, we feel that it is important not only for the instructor or evaluator of training conditions to be aware of the distinction between performance and learning, but also that the learner be made aware of this distinction as well. Illusions of competence in learning can be potentially dangerous if performance improvements vanish quickly or do not transfer well beyond the training regime.

REFERENCES

BADDELEY, A.D. and LONGMAN, D.J.A., 1978, The influence of length and frequency of training session on the rate of learning to type. *Ergonomics*, 21, 627–35.

BJORK, R.A., 1998, Assessing our own competence: heuristics and illusions. In Gopher and Koriat (eds) *Attention and Performance XVII. Cognitive Regulation of Performance: Interaction of Theory and Application* (Cambridge, MA: MIT Press), 435–59.

FELTZ, D.L. and LANDERS, D.M., 1983, The effects of mental practice on motor skill learning and performance: a meta-analysis. *Journal of Sport Psychology*, 5, 25–57.

LEE, T.D. and GENOVESE, E.D., 1988, Distribution of practice in motor skill acquisition: learning and performance effects reconsidered. *Research Quarterly for Exercise and Sport*, 59, 277–87.

MCCULLAGH, P., 1993, Modeling: learning, developmental, and social psychological considerations. In Singer, Murphey and Tennant (eds)

Handbook of Research on Sport Psychology (New York: Macmillan), 106–26.

SCHMIDT, R.A., 1975, A schema theory of discrete motor skill learning. *Psychological Review*, 82, 225–60.

SCHMIDT, R.A. and LEE, T.D., 1999, *Motor Control and Learning: A Behavioral Emphasis* (3rd ed.) (Champaign: Human Kinetics).

SHAPIRO, D.C. and SCHMIDT, R.A., 1982, The schema theory: recent evidence and developmental implications. In Kelso and Clark (eds) *The Development of Movement Control and Co-ordination* (New York: Wiley), 113–50.

SHEA, J.B. and MORGAN, R.L., 1979, Contextual interference effects on the acquisition, retention, and transfer of a motor skill. *Journal of Experimental Psychology: Human Learning and Memory*, 5, 179–87.

Sleeping Systems: Current Status

B. Haex and R. van Haute

Biomechanics and Engineering Design, Celestijnenlaan 200 A, B-3001 Leuven, Belgium

1. INTRODUCTION

In the Western world, people did not think about sleeping on a well-supported structures until recently: they slept on a sagged, worn down mattress placed on a declined spring base. In the late 1950s a firm bed was promoted for the prevention of low back pain, and many people started to put a wood board under the mattress, it being a typical solution that insufficiently corrects the supporting properties and may cause ventilation problems. Today the definition of a correct sleeping system is much more differentiated, as is discussed in Chapter XX. Nowadays a bed generally consists of a mattress, a mattress support, a frame, a head cushion and a blanket. This Chapter discusses how these different parts can be designed and combined to obtain an adequate sleeping system.

Section 2 discusses different mattress varieties (both core and top layers) and how their material properties influence general characteristics; section 3 illustrates different supporting structures, and explains which kinds of supports can be combined with each of the different mattress classes; finally section 4 briefly discusses head cushions. The bed frame should be strong enough, but apart from esthetical aspects it has no further influence on sleeping comfort, and it will, therefore, not be discussed. Nor will the variety of blankets, because they only affects heat and moisture transportation, which is of secondary importance compared with mechanical properties (especially back supporting qualities) when concentrating on the development of sleeping systems for normal healthy people and thus on the prevention of eventual back problems and discomfort.

2. MATTRESSES

A mattress consists of a core with one or more top layers, surrounded by a cover. Depending on the kind of mattress the cover may be stretchable or detachable. Generally four kinds of mattresses are produced: foam (e.g. polyurethane), latex (e.g. nature latex), spring (e.g. bi-conical-spring mattress) and fluidum-based beds (e.g. waterbeds). Wool, silk and cotton are common top layer materials. Futon or straw mattresses will not be discussed.

2.1. Core

2.1.1. Polyurethane foam mattresses

Polyurethane is a synthetic material that obtains its flexibility through foaming; the material can be applied to perform different functions, e.g. isolation, bedding, coating. Foam mattresses consist of a cellular network giving the material a specific density, elasticity and air permeability. Mattress density measures the weight per unit of volume, and defines the fatigue resistance of mattresses. In principle, foams with higher densities should have a higher stiffness; the fact that one can achieve a wide variety of stiffness characteristics for a given density is one of the main advantages of polyurethane. New chemical procedures (e.g. high-resilient foam) allow producing mattresses with a high density without a stiffness that is too high. A foam mattress core should have a thickness of at least 0.12 m and a density of at least 35 kg/m^3 brut to obtain a reliable fatigue resistance.

Normal polyurethane cells have a homogeneous, non-isotropic, open structure. These mattress cores give a reasonable

Soft zone indentation
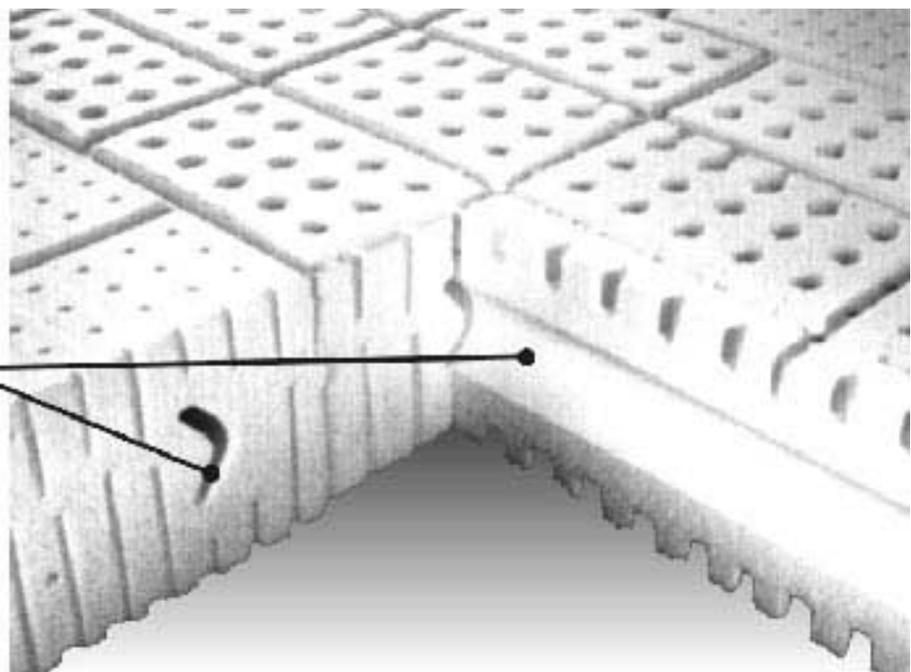
Figure 1. Latex mattress with different stiffness zones.

body support thanks to a small hysteresis and a good elastic behavior. Further, they give good heat isolation, reasonable moisture permeability and they are light and, therefore, easy to manipulate. High-resilient foam cells have a differentiated, isotropic and open structure; material density is much higher. A minimal hysteresis and a perfect elasticity give these mattresses very good supporting qualities; fatigue resistance and moisture transportation are optimal; heat isolation is very good, as is the case for most foams.

2.1.2. Latex foam mattresses

Latex mattresses consist of a block of foamed rubber particles and, therefore, sometimes are called rubber foam mattresses. Rubber particles are of synthetic or natural origin — combinations are possible — and can be foamed firmly or softly. Nature latex mattresses contain ~80% natural latex, completed with synthetic additives, which are necessary to process rubber to latex and to obtain the required elastic properties. Latex is further formed by molding and vulcanized to be dimensionally stable and resistant to temperature fluctuations. As opposed to polyurethane, latex foams show a rather strict relation between density and hardness.

Latex is especially suited for the fabrication of mattresses with different stiffness zones: thanks to adequate mould design a material cut-away obtains a softer elastic behavior when requested. Each zone with well-defined elastic properties can deform independently from other zones: when indenting the hip zone, it only deforms locally without exerting influence on either the shoulder zone or leg zone. Latex, therefore, presents perfect supporting qualities in case the mattress is conceived well; an air chamber with an adjustable volume can even improve the support in the lumbar area. Further, latex is supple, heavy and consequently difficult to manipulate; it has a low air permeability, but offers a very good heat isolation.

2.1.3. Spring mattresses

Spring mattresses exist in all kinds of shapes, dimensions, springs and spring connections. Only the spring core and the top and bottom comfort layers — both synthetic and natural foams — are common. The design of the springs, notably the wire thickness, is mainly responsible for the elastic properties of this kind of mattresses. Their main advantage is that spring stiffness

can be adjusted generally (to suit different population classes) or locally (to combine different comfort zones, e.g. a softer shoulder zone with a firmer pelvic zone). Sometimes springs and foam are combined over the entire volume to combine spring advantages (good ventilation) with foam benefits (heat isolation and good elastic behavior).

Bi-conical-springs — also called "Bonell springs" — have a smaller diameter in the middle (D_2 on figure 2) compared with the extremities (D_1). Spring stiffness is high compared with other types of springs and is constant in the usual range of deformation, resulting in an enlarged resistance against increased loading. Springs are mounted independently next to each other and linked by spiral wires on both sides.

Endless-spring cores consist of one single woven steel wire linking small cylindrical springs together, which guarantees a large flexibility. Pocket springs are mostly cylindrical or tun-shaped (larger diameter in the middle compared with the extremities) and are individually wrapped up into pockets. Pocket springs are mounted into rows — perpendicularly to the cranio-caudal direction — and can deform almost independently from each other, so that they can be used to create different stiffness zones in a mattress. Generally spring mattresses offer a rather low heat isolation and reasonable body support; pocket springs offer a good body support and a slightly better heat isolation (15%) compared with normal springs, but still 50% less than standard foam mattress.

2.1.4. Fluidum-based beds

When sleeping on a too firm surface, body weight will not be distributed homogeneously and the contact area will be reduced, which results in increased pressure (Clark and Cullum 1992) and shear forces (Goossens and Snijders 1995) (parallel to mattress surface) on the skin and the underlying soft tissues, e.g. blood vessels. Sophisticated beds for hospital applications (e.g. alternating pressure or fiber-filled mattresses) generally achieve a good contact pressure distribution to improve the blood and oxygen supply to prevent skin damage and eventual decubitus ulcers. Ordinary waterbeds offer a similar or worse pressure distribution compared with a standard foam mattress.

At the other hand pressure-relieving mattresses do not necessarily support the spine correctly: places where weight is concentrated will sink deep into the mattress resulting in other
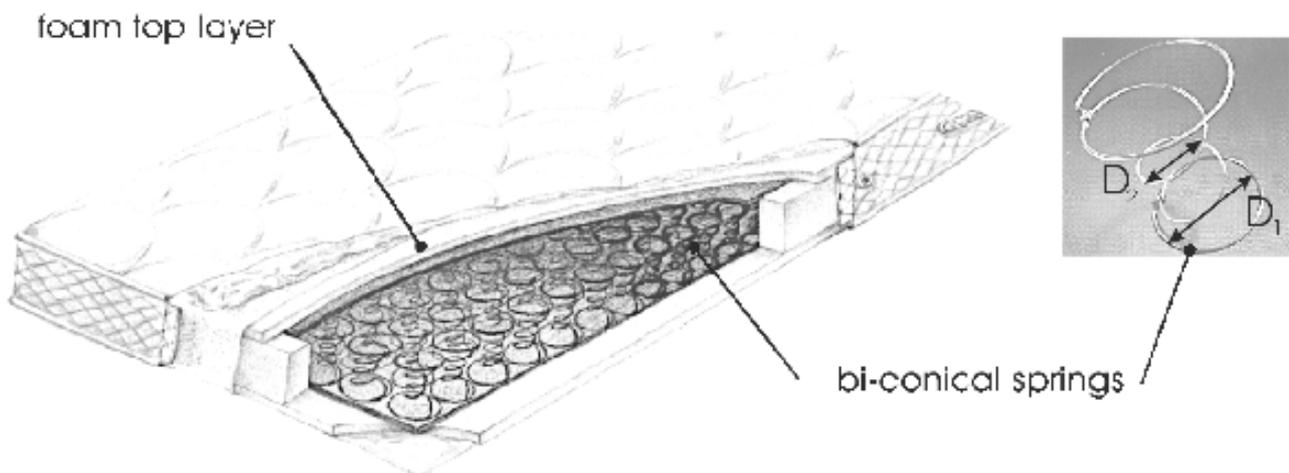
Figure 2. Bi-conical-spring mattress with firmer pelvic zone (left) and single bi-conical spring (right).
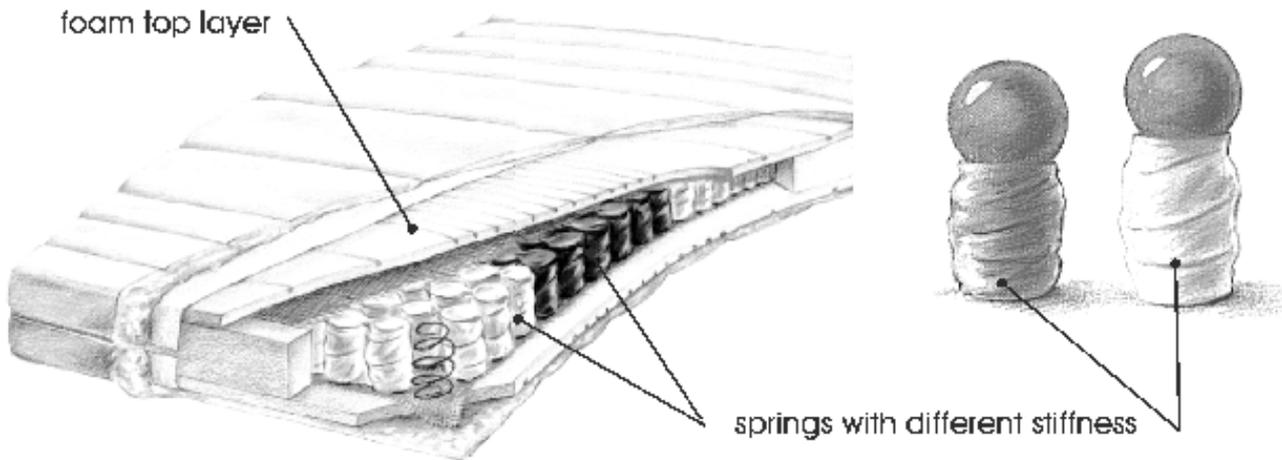
Figure 3. Pocket spring mattress with different stiffness zones (left) and pocket springs with different stiffness (right).

zones to raise. The heavy pelvic zone consequently will cause the mattress to sag, while the lifted shoulder zone will be loaded asymmetrically. Further, the large contact surface will limit mobility — due to the body sagging into the mattress — while water oscillations obstruct stability in case only one fluid chamber is employed. Finally impermeable beds generate a microclimate with much higher temperature and relative humidity (60–70%) compared with a normal mattress (37%); an adequate top layer that is ventilated regularly may prevent humidity transportation problems.

2.2. Top Layer

The humidity regulation of a sleeping system is mainly depending on the top layer of the mattress; the core of the mattress plays a negligible role. Top layers further have isolation and protecting purposes, and are mostly treated against bacteria and mildew formation. Most layers distribute the contact pressure in between the mattress and the human body; some are stretchable to avoid shear forces. Top layers generally can be subdivided into top layers that are stitched to the tick and core covers.

Wool can absorb wet vapor up to 33% of its own weight, without feeling clammy: moisture is first diffused and then slowly evaporated to the environment. Thanks to this absorption quality wool fiber avoids sudden cooling off in case of large temperature changes. Further, it can hold a lot of air for heat isolation purposes, and has a good flexibility to recuperate during daytime.

Silk can absorb moisture up to 40% of its own weight, evaporating it fast to the environment, making it especially suited for people who emit lots of moisture. It cannot hold air within its structure for heat isolation purposes, but has an anti-allergic activity.

Cotton can only absorb moisture up to 8% of its own weight without feeling clammy. It has poor heat isolation properties, which makes it only suited for summer, but it is extremely anti-bacterial and is supported well by allergic people. The use of a top layer — synthetic or natural — that can be stretched off and can be washed > 60°C, is the best prevention of allergy to the house dust mite.

3. MATTRESS SUPPORTS

Generally four kinds of supporting structures can be defined: board bases, spiral bases, box springs and slatted bases. In the Western world slats are becoming more and more common, resulting in an increased research, development and individualized variety for this kind of support. First different supporting structures will be illustrated; further it will be explained how different kinds of supports can be combined with different mattress classes.

3.1. Varieties

3.1.1. Board base

A board base mostly consists of a wooden frame on which perforated plates are mounted at both sides. This kind of mattress support has poor ventilation properties and is very hard. Consequently it offers an inferior support to the human spine, especially when combined with a hard mattress.

3.1.2. Slatted base

A slatted base is composed of a wooden or metal frame on which slats are fixed horizontally and perpendicularly to the cranio-caudal direction. These slats are fabricated in wood, plastic or glassfibre, and are mostly fixed separately onto the frame, comparable to a plank that is cut in pieces. In some cases slats can bend and to cant, to improve the supporting qualities.

Wooden slats are mostly layered and are mounted pre-stressed on the frame. Their number (14–30) and their thickness varies depending on the manufacturer; sometimes diversified slats, e.g. with different thickness, radius, or stiffness, are combined in one single base to optimize general bending properties: heavier slats are used in the pelvic zone where more weight has to be carried; softer slats are applied in the shoulder zone too allow a larger displacement. The same effect is obtained by limiting directly or indirectly the bending of the slats in the pelvic zone, e.g. via a strip or clips linking several slats together.

Recently mechanical properties are optimized by a proper design of the suspension of the slats in the frame: flexible slat supports made of rubber, plastic or steel can now bend and to cant. Canting slats that are mounted in pairs allow a better adjustment to the contours of the human body, especially in case of lateral recumbency; this adjustment can be even improved by adapting the height of the slat suspension. The main part of the flexibility now comes from the suspension, resulting in a mattress
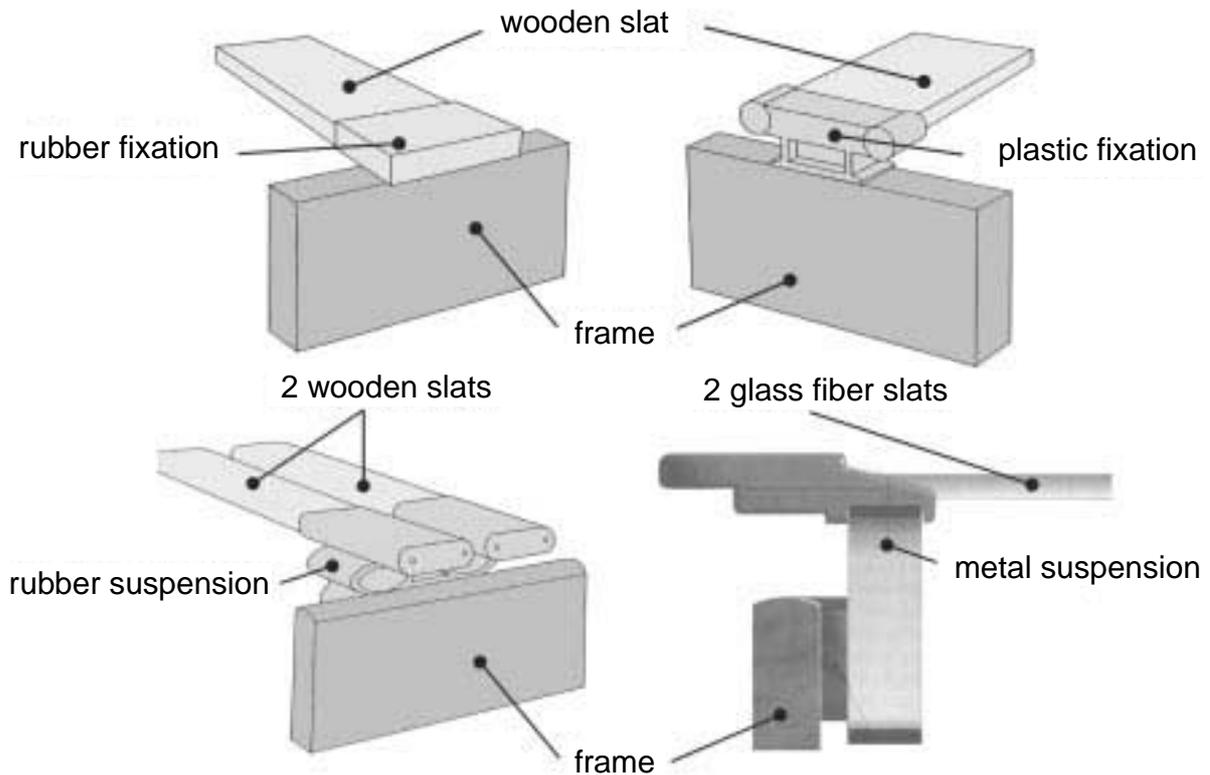
Figure 4. Separate slat fixation (up left) or slat suspension (up right) on the base frame; twin slats (wood) on a rubber suspension (bottom left) and twin slats (glass-fiber) on a steel suspension (bottom right).

support not only bending in the middle, which is especially useful when placing two supporting structures next to each other for a double bed. When glassfiber slats are used, the entire flexibility comes from the suspension (e.g. steel clips or rubber), guaranteeing the same elastic characteristics over the entire width. Steel suspension clips can be fabricated with different material or geometrical properties, which allows to combine several clip classes (e.g. soft pelvic zone) to optimize the supporting qualities over the entire width.

3.1.3. Spiral base

A spiral bed base consists of a metal frame on which ~400 springs are horizontally stretched and weaved into each other. Mostly springs are stretched perpendicularly to the cranio-caudal direction to avoid sagging by spanning only 1 m instead of 2 m. A spiral bed base has perfect ventilation properties and offers reasonable supporting qualities, especially when different body zones are equipped with an adapted spring tension.

3.1.4. Box springs

Box springs generally consist of the same components as a bi-conical-spring mattress (see above), but have different material properties to offer a much firmer support to serve as a mattress supporting structure. Bi-conical springs are linked and mounted vertically, as is the case in a bi-conical-spring mattress, but they are mounted on a stiff wooden or metal base frame. The elastic foam on top and around the springs gives rise to the box springs being softer than other kinds of supports.

3.2. Combination with Mattress Classes

When choosing an optimal combination mattress-supporting structure, several anthropological characteristics are relevant to make a correct combination of different sleeping system qualities. Especially supporting qualities should be measured or modeled adequately before assigning a mattress to a person. Only most significant personal requirements will be discussed here to avoid a large elaboration. Further, not every mattress can be combined with any kind of supporting structure: both evident and unsound combinations will be reviewed.

Body weight is an objective criterion to determine whether a mattress has to be soft or firm: heavier persons need a firmer support, but extreme properties should be avoided. When people have a personal preference, supporting qualities should remain primordial, keeping in mind that it mostly takes only 2 weeks to get used to a new bed. People with perspiration or allergic problems should choose an adequate combination (e.g. spring mattress + slat support). People with pronounced contours (e.g. large shoulder or hip width) are helped with different comfort zones: pocket spring or latex mattresses are perfectly suited here, especially when people sleep in lateral recumbency. In spite of the fact that the mattress is responsible for 60–80% of the support, it can be improved significantly by using an adjusted base, e.g. a canting slat suspension combined with a latex mattress.

A board base only can be combined with a spring mattress to avoid ventilation problems: the combination of a foam or latex mattress with a board base can cause mildew formation. For the same reason a box spring is best combined with a spring mattress. A spiral base has no ventilation or stability problems, and can be combined with any kind of mattress. When combining a slat

base with a spring mattress, the stability of the springs is the prime issue: slats should be wide enough (at least 4.5 cm), they should not cant, and their number should be sufficiently large. Pocket springs can be combined with a slat base if a stretchable layer on the slats prevents the springs from slipping through.

The good supporting qualities of a slat base are best expressed in combination with a supple latex of foam mattress. Latex is less firm in comparison to foam, which makes it more suited to be mixed with an adjustable slat base. Especially twin slat bases with a canting and bending suspension are best combined with latex mattresses to make use of the entire slat flexibility.

4. HEAD CUSHIONS

To support the cervical spine correctly also the head cushion should be designed properly. In case of lateral recumbency the entire spinal column should be a straight line when projected in a frontal plane (Pheasant 1991). This objective can be reached both by correctly positioning and shaping normal deformable cushions (e.g. feather cushions) and by correctly designing less deformable structures (e.g. latex cushions).

Cushions filled with down and feathers are easy to shake up, to sustain the head and the neck in a proper way; they also have a better moisture permeability than polyurethane pillows. Latex cushions offer a good heat isolation and supporting qualities when conceived well: sometimes they are pre-shaped to support the neck, but it can take some time (2–3 weeks) to get used to it. An air chamber with an adjustable volume can even improve the support in the neck area.

5. RECOMMENDATIONS

- Back supporting qualities are of primary importance when assigning a sleeping system to a healthy person.
- People with perspiration or allergic problems should choose adequate materials. By defining material properties correctly (springs, latex, polyurethane) one can obtain correct supporting qualities for different kinds of mattresses.
- By combining different stiffness zones one can optimize supporting properties. Care should be taken in assigning a mattress with comfort zones to an individual: in principle each person needs different zone subdivisions, so an incorrect assignment will do more harm than good.
- Keep in mind that not every mattress can be combined with any supporting structure.

ACKNOWLEDGEMENT

REFERENCES

CLARK, M. and CULLUM, N., 1992, Matching patient need for pressure sore prevention with the supply of pressure redistributing mattresses. *Journal of Advances in Nursing*, 17, 310–16.

GOOSSENS, R.H.M. and SNIJDERS, C.J., 1995, Design criteria for the reduction of shear forces in beds and seats. *Journal of Biomechanics*, 28, 225–30.

PHEASANT, S., 1991, Posture. In *Ergonomics, Work and Health* (London: Macmillan), 98–107.

Sleeping Systems: Design Requirements

B. Haex and R. van Audekercke

Biomechanics and Engineering Design, Celestijnenlaan 200 A, B-3001 Leuven, Belgium

1. INTRODUCTION

One spends about one-third of our life asleep in bed, while our inactive body is given over to the sleeping system (i.e. mattress + supporting structure) we are lying and relying on. The quality of body support during bed rest not only affects the mental quality of sleep, but also has a major influence on physical condition. Low back pain — one of the most compelling problems in the industrialized world — is often caused or worsened by an incorrect body support, which justifies the need to search for an objective and science-based method to determine the right sleeping system for each individual, by guiding the factory design and production of mattresses.

A well-qualified sleeping system has to fulfil many needs depending on personal requirements, and has to be reliable for different sleeping postures. Most authors consider a resting-place being ergonomically justified when the entire musculoskeletal system can recuperate well: muscles have to relax while the unloaded intervertebral disks are rehydrating. Unfortunately this is not the case for many chairs, couches and beds that are available on the market: despite a growing consciousness towards back problems the vertebral column is often supported insufficiently. We discuss below the mechanical properties of a sleeping system that largely determine these supporting qualities.

Although it is influenced by various mattress characteristics, the comfort of sleep is a factor that is often referred to as "subjective." Indeed, specific limitations are easily overlooked when assigning an appropriate sleeping system to a person: next to a house dust mite or latex allergy, other limiting constituents — like respiratory problems or excessive sweating — make it virtually impossible to ignore them while defining sleeping systems. These properties are also discussed below.

The best mattress finally has to be chosen by assigning different weight factors to its qualities, depending on the objectives to be fulfilled, and keeping in mind that most of its properties are not designed for eternity: even a fatigue-resistant mattress looses 10–15% of its elasticity after 10 years.

2. MECHANICAL CHARACTERISTICS

When concentrating on the development of sleeping systems for normal healthy people and, thus, on the prevention of eventual back problems, one will focus rather on mechanical properties (especially back supporting qualities) than on non-mechanical characteristics (e.g. heat transport). Within mechanical objectives back-supporting qualities are primary while all other purposes are secondary, keeping in mind that peak values (e.g. pressure peaks) should be avoided. Only in case of applications involving injured people other issues come first: pressure-relieving qualities become of prime importance in case of hospital applications.

At first sight mattress stiffness seems to be a question of personal preference; in reality mattress properties should be adjusted objectively to personal needs. It is clear that heavier persons need a firmer mattress to avoid the pelvic girdle from sinking too deep into the mattress. Materials have to be developed and combined to optimize general sleeping system characteristics. While the material density mainly affects fatigue resistance, the material elasticity — and the combination of materials with different elasticity — guarantee a correct support of the human body. In fact one can obtain the required characteristics with any material by defining it correctly. Latex mattress elasticity can be adapted by changing mould specifications (e.g. use of indentations); designing spring dimensions enables the modification of pocket spring mattresses; polyurethane mattress elasticity can be adjusted by the use of different kinds of foam (e.g. different densities).

Standardized compression and tensile tests can describe most mechanical properties. Displacement controlled benches measure force at a fixed interval, resulting in a force-displacement characteristic consisting of a loading and a relaxation phase, which can be recalculated to a stress–strain characteristic. Elasticity further can be calculated as the ratio of stress to strain: the more force is needed to reach a certain indentation, the stiffer the mattress will be. In case of a perfectly elastic material, elasticity will be constant; in case of a visco-elastic material (e.g. polyurethane) elasticity is deformation velocity-dependent.

When building a mattress with different elastic properties (e.g. a softer shoulder zone or a harder pelvic zone) it is important that different zones — each with its well-defined elastic behavior — deform independently: when indenting the hip zone, it should only deform locally without exerting too much influence on either the shoulder zone or leg zone. For example this kind of "local elasticity" can be obtained by placing pocket springs with different properties in a matrix, allowing them to deform independently.

Mattress hysteresis can be determined by calculating the area in between the load curve and the relaxation curve. It measures the energy that is dissipated in the mattress and should be minimized to avoid exaggerated energy consumption while moving.

Mattress density measures the weight per unit of volume, and defines the fatigue resistance of mattresses. In principle, foams with higher densities should have a higher stiffness; nevertheless new chemical procedures (e.g. bultex foam) allow the production of mattresses with a high density without a stiffness that is too high.

Of course, the dimensions of a sleeping system and the ability to adjust the mattress support noticeably improve sleep comfort. Bed width should be at least shoulder width plus 0.40 m; length should be at least body length plus 0.20 m; bed height should be at least 0.45 m (for ventilation purposes, see below) while high beds (> 0.55 m) make it more easy to get in or out. When staying in bed for other activities (e.g. reading) a supporting structure can add a considerable surplus when it is large enough and when it guarantees a correct back support in these changed positions.

2.1. Spine Support

It is justified to consider back injuries being a social problem: only a few people (20%) are never confronted with it. Many back problems are induced by posture or movements: an incorrect
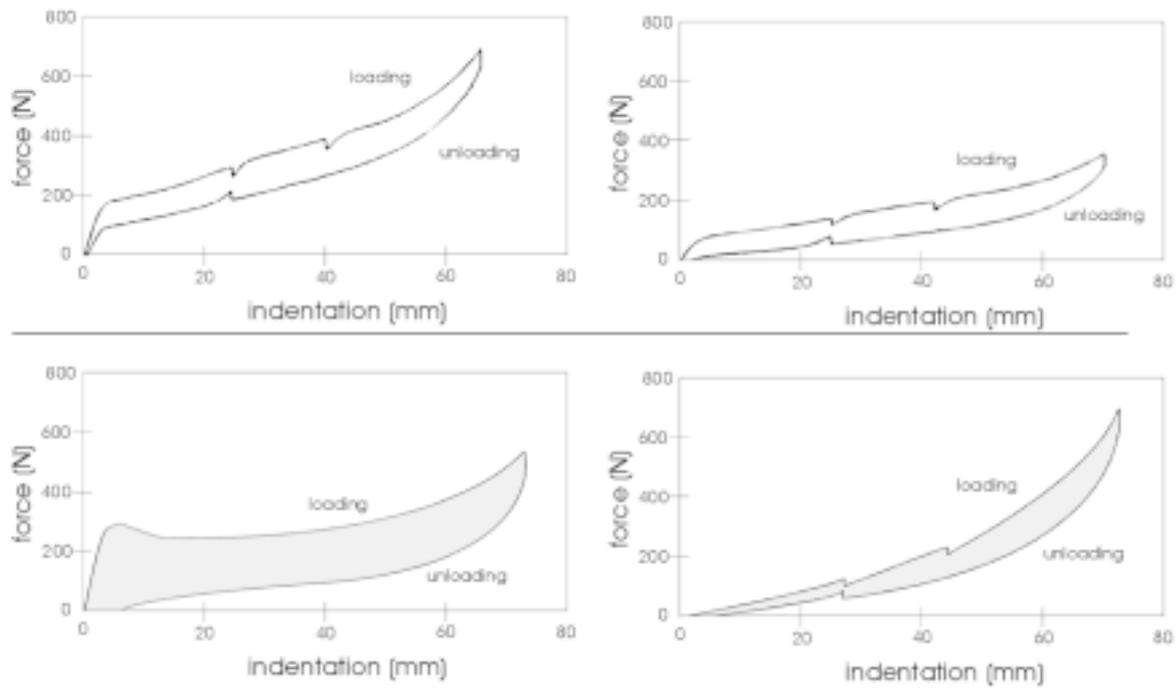
Figure 1. Force-indentation characteristics of a soft (top left) and firm (top right) polyurethane mattress. Unacceptable hysteresis (bottom left) on a visco-elastic foam and agreeable hysteresis (bottom right) on a latex mattress.

support of the spinal column during sitting, working or sleeping often causes or aggravates low back pain. Most authors pose that in a correct sleeping posture natural physiological curves have to be maintained; others emphasize that the lumbar lordosis has to be flattened. When concentrating on the optimal sleeping system for normal healthy people — as it is suggested by most ergonomic

specialists — it has to support the human spine such that it adopts its natural position, which is assumed to be the same as it takes in the upright position (Adams and Hutton 1985). If this thesis is presupposed then an optimal body support for lateral recumbency gives rise to the spinal column being a straight line when projected in a frontal plane. For posterior recumbency an
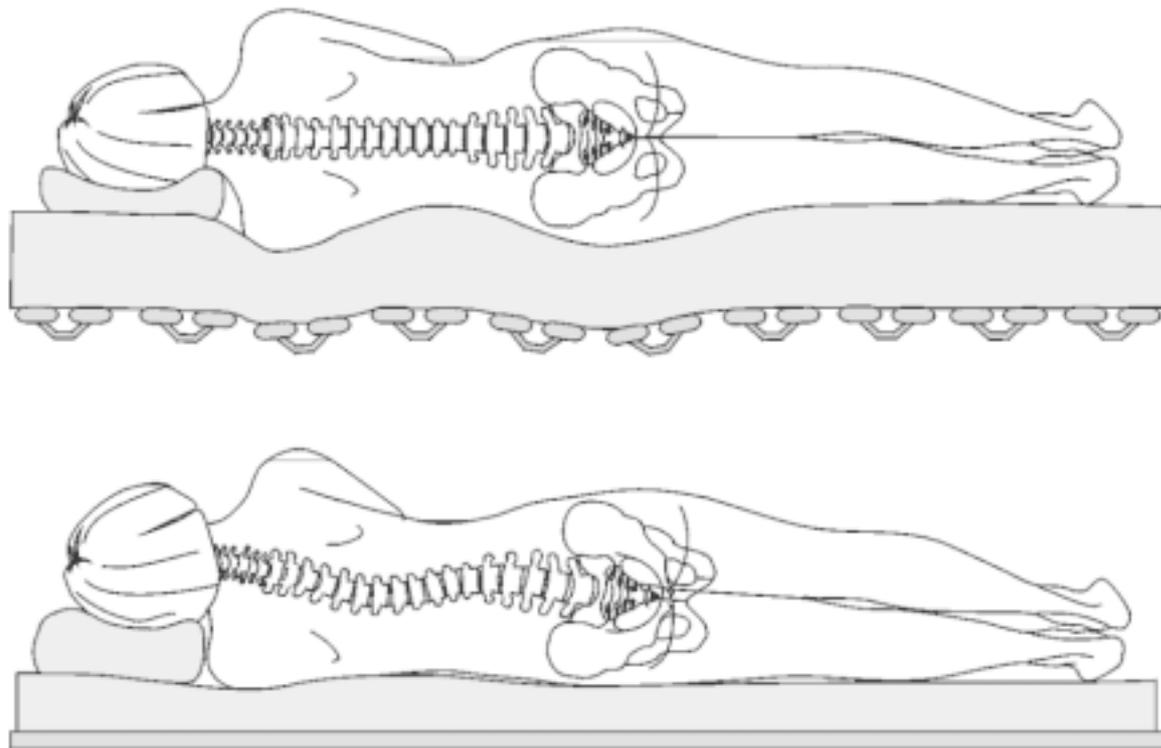
Figure 2. Correct back support (top) and too firm support (bottom) in case of lateral recumbency.

optimal support gives the spinal column the same thoracic kyfosis and lumbar lordosis as in the upright position, yet slightly smoothened by the loss of body weight working in longitudinal direction on the spinal column. Consequently a small prolongation of the spine occurs, as during weightlessness. The incapability of the human body to control actively the spinal column when sleeping justifies the definition of correct sleeping system for each individual, by defining and combining different materials correctly. A perfect support can be realized when different zones of the sleeping system can deform more or less independently.

Intervertebral disk injuries and the way they act on the surrounding soft tissues are responsible for the main part of posture-dependent low back pain; facet joint or ligamentous impairments are of rather secondary importance. Especially the lumbar region is very sensitive to disorders due to the fact that pressure and tension forces mount up under load. During sleep pressure in the intervertebral disks will decrease allowing them to rehydrate, resulting in a prolongation of the vertebral column which is mainly responsible for the increase at night of the body length by 1%. Also muscle relaxation can be obtained on a well-conditioned sleeping system which is the case when mobility and stability are promoted (see below).

When a sleeping system is too soft, places where body weight is concentrated (e.g. the hip zone) will sink deeply into the mattress. Some muscles may be well relaxed in this position, but the spine certainly will not: when lying in posterior recumbency the pelvis will cant backward resulting in a complete and unnatural smoothening of the lumbar lordosis. At the anterior side intervertebral disks will be compressed while soft tissues (e.g. ligaments) will be under tension at the posterior side. When sleeping in lateral position the spine will be loaded asymmetrically as shown on the figure below. Most mattresses that are worn down have the characteristic to sag in the middle; a typical solution is to put a stiff wood board under the mattress, which insufficiently corrects the supporting properties and may cause ventilation problems (see below).

When sleeping too firm the spinal column will be supported incorrectly: in case of lateral recumbency only places with a large body width — the shoulders and the hip zone — will be supported, while the lumbar region will bend down, especially with people who have a more pronounced contour e.g. women. In posterior recumbency the pelvis is first canting forward under influence of tension in the m. illipsoas; after muscle relaxation it will cant backwards as is the case on soft mattresses. The consequent flattening of the lordosis is less pronounced and harmful compared with a mattress that is too soft. Further it is clear that people with major spinal disorders (e.g. people with a spinal injury) need a special treatment: although a hard mattress induces increased pressure peaks and stress at some places by flexing the spine laterally, it can however yield a temporary stress relief at the level of injury.

In order to support the cervical spine correctly also the head cushion should be designed properly. In case of lateral recumbency the entire spinal column should be a straight line when projected in a frontal plane. This objective can be reached both by correctly positioning and shaping normal deformable cushions (e.g. kapok cushions) and by correctly designing less deformable structures (e.g. latex cushions).

As opposed to what some manufacturers of sleeping systems claim, pressure-relieving mattresses do not necessarily support the spine correctly: places where weight is concentrated will sink deep into the mattress resulting in other zones to raise up. The heavy pelvic zone consequently will cause the mattress to fag, while the lifted shoulder zone will be loaded asymmetrically. A correct support of the spine consequently can not be measured with equipment picturing the pressure distribution. Video-raster-stereography (Haex *et al.* 1999) is a possible technique to measure spinal deformations in order to evaluate sleeping systems. It can quantify the spine in any sleeping position in between a prone and a lateral position; for posterior recumbency other equipment was developed (Geelkerken *et al.* 1994).

2.2. Weight Distribution

During sleep a local ischemy will arise in body zones that are in contact with the sleeping system. This ischemy generates metabolic substances that stimulate the sensible nerve extremities, which will cause the person to change his posture before it gets painful (see below). When sleeping on a too hard surface, body weight will not be distributed homogeneously and the contact area will be reduced, which results in increased pressure (Allen *et al.* 1993) and shear forces (Frobin and Hierholzer 1981) (parallel to mattress surface) on the skin and the underlying soft tissues e.g. blood vessels. Blood supply will be attenuated (or even stopped) due to the deformation of these tissues. Normal capillary arteriolar pressure should vary in between 3.3 and 4.6
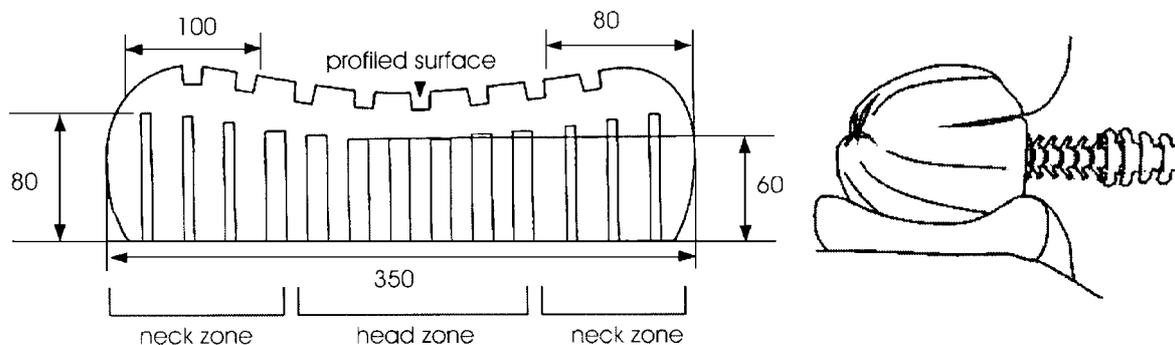
Figure 3. Correct cushion design (left; in mm) and neck support in case of lateral recumbency (right).

kPa (25 and 35 mmHg); pressure in the venules should ~1.6 kPa (12 mmHg) while critical pressure is considered to be 4 kPa (35 mmHg). The combined effect of loading time and intensity may result in the development of decubitus ulcers. Hence, most decubitus ulcers occur with people that stay in bed for a long time, especially when the patient can not be moved because of injuries. Pressure relieving qualities of a sleeping system, therefore, become primordial in case of hospital applications to prevent these ulcers. Also people with symptoms of fybromyalgy, e.g. tender points in bony areas, require an enhanced sleep comfort by optimizing contact pressure.

Many authors analyzed the relation between body posture, the kind of sleeping system, body characteristics and decubitus ulcers. Most of them describe mattress evaluations based on pressure measurements that are easy to perform by means of a mapping system that consists of a blanket containing a matrix of (capacitive) pressure sensors. The general conclusion is that the duration and the level of contact pressure have to be limited to improve the blood and oxygen supply to prevent skin damage and consequent decubitus ulcers. Next to the decreased supplies of nutrients, the reduced removal of waste is a component to reckon with. An increasing number of authors also mention the importance of shear forces to be minimized to prevent stretching and angulation causing thromboses of the blood vessels. Most vulnerable to both phenomena are places where the skin is pinched in between bed and bone, especially because these areas have no fat tissue or muscles to distribute the contact forces. The scapula, the elbow, the sacrum and the heel are, therefore, risk areas in case of posterior recumbency; for lateral recumbency the ankle, the knee, the big trochanter region and the shoulder (acromion) are zones to pay attention to. Women generally have a lower risk to develop decubitus ulcers.

2.3. Mobility and Stability Promotion

Frequent posture changes during sleep require a sufficient mobility: when a mattress is too soft and nearly surrounds the human body, turning around requires lots of energy or even becomes impossible; also the mattress hysteresis (see above) should be minimized to avoid exaggerated energy consumption while moving. At the other hand a stable body support is required to sleep relaxed: when a mattress is too firm, the body is insufficiently surrounded and may roll down uncontrolled. Conclusively a bed guarantees sufficient rest (1) when changing posture is easy and (2) when each adopted posture is stable.

2.3.1. Mobility

Posture changes are necessary to avoid a pressure overloading of soft tissues (see above) and to prevent muscle stiffness; a regular position shift — ~20 times a night — should be sufficient. On some kinds of waterbeds, pressure distributors or too soft mattresses, the pelvic girdle sinks too deep into the mattress. Owing to the fact that a permanent muscle force application is impossible during sleep, the person will roll back in the cavity when trying to change his posture. The consequent extensive periods of immobility are harmful.

Although necessary, posture changes have to be limited: switching back and forth too frequently, e.g. when sleeping on a mattress that is too hard, will increase physical restlessness and the tense to sweat (see below) during sleep. Further a consequent

unsatisfactory muscle relaxation will impede intervertebral disk rehydration, thus, causing back pain indirectly.

In addition, position shifts have to be limited because they influence the consecutive sleeping stadia. Sleep cycles are strongly related to episodes of immobility, which occur mainly during non-REM sleep: immobility mostly starts in stadium 2 and ends in stadium 3 or 4 of the same sleep cycle. Suddenly induced position shifts may cause these stadia — which can be characterized as "deep" sleep — to end prematurely, causing the person not being fit in the morning.

2.3.2. Stability

A stable body position is not guaranteed at all when a sleeping system reacts on body movements by oscillating. In that case the sleeping person will need to apply continuous muscle force in order to stabilize his sleeping posture adequately, resulting in some muscle groups being under permanent stress.

In the event of acute back pain the affected area is protected against sudden body movements — which could provoke pain — by an increased tension of the back muscles. Mattress oscillations might interrupt this stable, back supporting posture, and, therefore, should be prevented.

2.4. Contact Area Optimization

In fact the optimization of the contact area implies several primary objectives which have been — or will be — discussed and, therefore, it can not be considered as an aim on itself. On one hand the contact area should not be too small: when sleeping on a too firm surface, body weight will not be distributed homogeneously, which results in increased pressure and shear forces, giving rise to the symptoms mentioned earlier. It will also be difficult to adopt a stable posture (see above).

At the other hand a too large contact area — which is often pursued in order to prevent blood circulation disorders — causes the person to sink deeply into the mattress and, therefore, limits body mobility (see above). Further a large contact will restrain the skin from breathing and, thus, increases the tense to sweat and the inconvenient feeling to sleep on a clammy surface.

3. NON-MECHANICAL CHARACTERISTICS

3.1. Microclimate Regulation

3.1.1. Humidity

Fluid absorption and fluid transport to the environment (= ventilation) are the main humidity-related mattress characteristics. The human body emits 200–300 ml (up to 1 liter) body moisture each night. One-third is emitted through breathing; the remaining two-thirds is transmitted through the body surface and has to be absorbed by the mattress (25%), the sheets, blankets and head cushion (together 75%). This moisture excretion is not to be confused with perspiration, in which case much more body fluid is lost (e.g. at the time of a fever).

Moisture further has to be transported to the environment to avoid a clammy feeling at the mattress surface, to avert mildew formation at the mattress bottom, and to prevent decubitus ulcers, since a moist skin is rough and, therefore, more sensitive to shear forces (see above). The relative humidity measured between the skin and the blanket or bed should stabilize after 20 min and should not > 65%.

The humidity regulation of a sleeping system is mainly

depending on the top layer of the mattress (80%). The core of the mattress plays a negligible role; even body movements during sleep do not improve its ventilation capability significantly. Both synthetic and natural latex cores have an impenetrable skin, which gives them poor ventilation properties compared with spring mattresses. In case of dry weather room ventilation can be improved by opening a window. Further warm humid air has to be prevented from floating into colder rooms (e.g. bed rooms) where it might condense, thus, obstructing mattress ventilation and creating a seed-bed for house dust mites (see below). Of course, mattress maintenance is a capital issue for the preservation of its ventilation properties; mattresses should, therefore, be easy to manipulate to prevent back problems when turning them around.

3.1.2. Temperature

The main part of the temperature regulation takes place by evaporating water through breathing; the remaining warmth produced by the human body is given out through the skin. Body temperature should stay constant during sleep: when heat isolation is too low, the body will cool off resulting in muscle stiffness and sleep disorders; when heat isolation is too high, transpiration will increase resulting in a too high relative humidity (see above) and consequent sleep disturbances. An optimally isolating sleeping system ensures a bed temperature between 28 and 32°C, allowing the contact temperature to stabilize between 30 and 35°C.

The isolating capabilities of a bed are mainly depending on the core of the mattress and on its top layer(s). A core consisting of natural latex or polyurethane gives a higher isolation than springs (e.g. pocket springs). Further most people prefer warm contact, which largely depends on the capability of the top layer (e.g. wool) to hold air.

Extreme low (–9°C) and extreme high room temperatures (25°C) must be avoided because they affect the duration of REM sleep. It is hypothesized that thermal regulation of the body is suppressed during this phase, resulting in the body temperature varying with the environment. A consequent body temperature variation can act as a stimulus to wake up. Further, the optimal room temperature depends on the sleeping system, the kind of bed textile and the person (e.g. age); generally low temperatures are better supported by most people, but the lower temperatures are, the higher the risk for mildew formation and the better ventilation should be.

3.2. Prevention of House Dust Mite Allergy

House dust mites are minuscule spiders (0.1–0.5 mm) that populate mattresses, carpets, curtains and head cushions, and in large quantities. A house dust mite allergy is not originated by dust nor by the mites themselves; the real cause lies in the allergen "Der PI," which is produced in the intestinal canal of the mites when degrading organic dirt — especially human skin peels — biologically. The mites' desiccated excrements carrying allergen can be assimilated in the breathing air of man and may, thus, penetrate into the cavity of the mouth and the lungs causing allergic reactions (asthma, coughing spells, irritated eyes) which manifest themselves during the night.

The mites prefer a temperature between 20 and 25°C and a relative air humidity between 70 and 75%. A complete extinction of the mites is virtually impossible, but allergic symptoms can
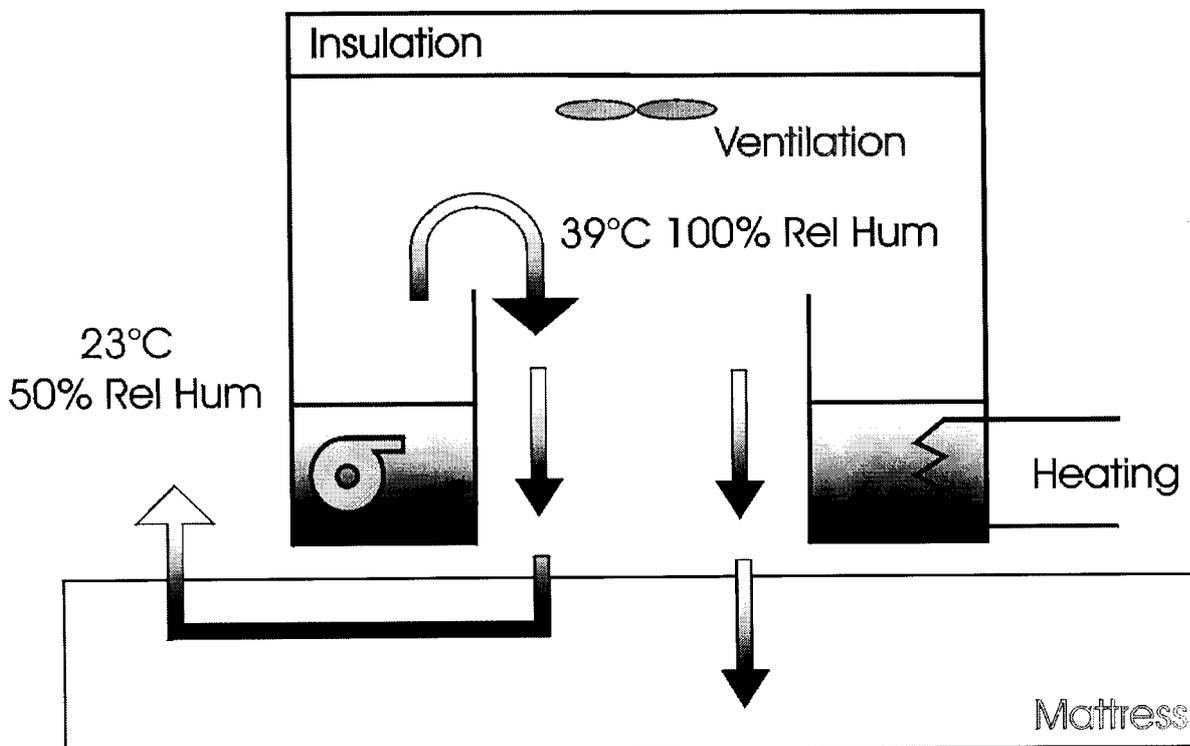
Figure 4. Cross-section view of moisture transportation measurements.

decrease by keeping the air humidity < 55% or by limiting the exposure to the excrements by a correct and thorough sanitation: allergen-proof polyurethane coating around the mattress core and a synthetic top layer that can be stretched off and can be washed > 60°C. Mattresses should, therefore, be easy to maintain.

4. RECOMMENDATIONS

- When concentrating on normal healthy people and, thus, on the prevention of eventual disorders, back supporting qualities are of primary importance.
- The optimal supporting sleeping system has to support the human spine such that it adopts its natural position, which is assumed to be the same as it takes in the upright position, yet slightly smoothened by the loss of weightlessness working in longitudinal direction.
- All other purposes are secondary, keeping in mind that peak values (e.g. pressure peaks) should be avoided.
- Only in case of applications involving injured people other issues come first, e.g. pressure-relieving qualities become primordial in case of hospital applications.
- By defining material properties correctly (springs, latex, polyurethane) one can obtain the same required mechanical characteristics for different kinds of mattresses. By combining different zones one can optimize general sleeping system properties.
- Not every mattress can be combined with any kind of supporting structure.

REFERENCES

ADAMS, M.A. and HUTTON, W.C., 1985, The effect of posture on the lumbar spine. *Journal of Bone Surgery*, 67, 625–9.

ALLEN, V., RYAN, D.W. and MURRAY, A., 1993, Potential for bed sores due to high pressures: influence of body sites, body position, and mattress design. *British Journal of Clinical Practice*, 47, 195–7.

FROBIN, W. and HIERHOLZER, E., 1981, Rasterstereography: a photogrammetric method for measurement of body surfaces. *Photogrammetric Engineering and Remote Sensing*, 47, 1717–24.

GEELKERKEN, R.H., WILLE, J., HAMMING, J.J., HERMANS, J, BRESLAU, P.J. and HOFMAN, A., 1994, Pressure sores and pressure-decreasing mattresses: controlled clinical trial. *Lancet*, 343, 568–71.

HAEX, B., VAN AUDEKERCKE, R., VANDER SLOTEN, J., VAN DER PERRE, G., VAN HAUTE, R. and BAETEMAN, J., 1999, Experimental evaluation of mattress stiffness influence on the spinal curvature during bed rest. *Spine* (submitted).

Standing Work

S. Konz

Department of Industrial and Manufacturing Systems Engineering, Kansas State University, Manhattan, KS 66506 USA

1. INTRODUCTION

Prolonged standing is common in many jobs; some examples are sales clerks, bank tellers, retail checkout operators, barbers/hairdressers and many factory jobs. Generally, they do not stand with "rooted" feet but occasionally move around a limited area. The problems are pain (feet, legs, back) and venous pooling of blood in the legs.

2. BACK/LEG/FOOT

2.1. Anatomy

The mean distance between the inside of the two feet, when standing, is ~100 mm, between foot centerlines is ~200 mm and between outside edges is ~300 mm (Rys and Konz 1994). Yet mean height for males is 1750 mm. Thus, there is a base of only 200–300 mm for a structure of 1750 mm. When standing, the center of gravity passes from the ear opening forward of the spine (even L4–L5) so the body normally has a forward bending moment, counteracted by forces of ligament and back muscle and soleus muscles of the calf.

In supporting the body, the calcaneus (heel) supports 50% of the weight, the 1st and 2nd metatarsals 25% and the 3rd, 4th and 5th metatarsals 25%. Underneath the heel is the heel pad (~18 mm thick). The bottom of the calcaneus is not spherical but has two small "mountains" (upside down!); the pad reduces the pressure on these mountains, and, thus, on the ankle, knee and back.

Sometimes one forgets that everyone does not have a perfect body. The spine is concave backwards in two areas (cervical and lumbar) and concave forwards in one area (thoracic). Lordosis is an increase in lumbar curvature — a swayback posture, with the stomach protruding. (Pregnancy gives a temporary increase in lordosis.) Kyphosis is an increase in thoracic curvature – shoulders roll forward, the person slouches, becomes a hunchback. Scoliosis is a bending of the spine to the side (i.e. from a front view).

Leg length discrepancy (LLD) is the technical name for differences in leg length in the same person. Contreras *et al.* (1993), summarizing studies with $n = 2377$, reported that 40% of people had LLD < 5 mm, 30% had < 9 mm, 20% had < 11 mm and 10% had < 14 mm.

Other imperfections (due to injuries, aging, disease) also can affect posture.

2.2. Physiology

If the legs do not move, the blood from the heart tends to go down to the legs and stay there (venous pooling). Venous pooling causes more work for the heart, as, for a constant supply of blood, when there are lower milliliters of blood/beat, then there must be more beats. In addition, venous pooling causes swelling of the legs (edema) and varicose veins.

Pollack and Wood (1949) give mean ankle venous pressure while standing as 87 mmHg; Nodeland *et al.* (1983) give 80.

Walking ~10 steps drops ankle venous pressure to ~22 mmHg. Nodeland *et al.* (1983) reported standing bench work, with occasional steps around the area, had ankle pressure (48 mmHg) approximately equal to sitting at a desk.

2.3. Dimensions

Weight of leg segments (as % body weight) are: 1.47% for foot, 4.35% for calf, and 10.27% for thigh; a total leg is 16.1% and both legs are 32.2%. For example, the weight of both legs for a 70 kg person would average 70 (0.322) = 22.5 kg.

Mean and SD (cm) of nude US adult female civilians are 162.9 (6.4) for stature height, 74.1 (4.4) for crotch height, 51.5 (2.6) for knee height, 24.4 (1.2) for foot length and 9.0 (0.5) for foot breadth; the corresponding male values are 175.6 (6.7), 83.7 (4.4), 55.9 (2.6), 27.0 (1.2) and 10.1 (0.5). As a rule of thumb, female dimensions are 93% of the corresponding male dimensions. A large proportion of the variation in human stature is in leg length; the torso is relatively constant in height.

3. WORKSTATION DESIGN

3.1. Vision

The viewing angle depends not only on the line of sight of the eyes but also the inclination of the head (usually downward).

Location of a work object is a compromise between visual requirements (seeing object comfortably) and hand requirements (work at elbow height). Reduce the visual requirements by (1) improving lighting, (2) using optical aids (magnification mounted on a worksurface or the head or even using video pictures) or (3) modifying the task (improving size such as use of 12 point text rather than 10-point text, improving contrast such as printing on white paper instead of tinted paper, or increasing viewing time (shooting sitting ducks)).

3.2. Reach

Operator position and object position affect reach distance.

3.2.1. Operator position

While standing, horizontal reach distance is affected by foot position. A barrier extending down to the floor may restrict foot position and, thus, require a bent back while reaching for/manipulating an object. Provide a space of 150 x 150 x 500 mm wide for the feet. (Note that conveyors for seated operators need to have space for the knees and thighs.)

Reduce vertical reach distance by adjusting height of the operator by using platforms; usually it would be a simple wooden structure but it could be motorized to go up and down.

3.2.2. Object position

Minimize horizontal reach distance to reduce forward bending and, thus, stress on the back. For example, when working at a conveyor, keep objects on the closer side of the conveyor; this may require a plow (diverter) on the conveyor.

Minimize vertical reach distance by keeping object position at waist height. Thus, objects on the floor are not acceptable. A simple technique is to stack a loaded pallet onto several empty pallets — thus raising the package heights. Conveyor heights are easy to modify for single operators (adjust conveyor support heights); the situation is more complex if multiple operators use the workstation or there are multiple workstations along the con-

veyor. For gravity conveyors, short sections of belt conveyors may be needed to elevate the load. However, conveyors should be convenient for those working with them, not necessarily easy for engineers to design.

4. POSTURE PROBLEMS WHILE STANDING

As discussed above, not every body is perfect. But even if the body is "perfect," the work activity can cause stress.

Swat and Krzychowicz (1996) assumed that the "free comfort boundaries" are trunk forward < 20°, to the sides < 10°, backward < 5° and twisting < 5°. They calculated force moments from trunk, knee and ankle bending and got five zones. Zone 3 (zero force moments) is the best; it has the trunk bending forward < 20°, the knees straight and the hands below the shoulder. Zones 2 and 4 (2.3 force moments) are next; in zone 2, the hands are above the shoulder; in zone 4, the hands can be as low as knee height (with knees straight). In zone 1 (4.2 force moments), the worker is standing on tiptoe, reaching above the shoulders. In zone 5 (the worst with 7.0 force moments), the hand is below the knees.

Table 1 is another approach to quantifying postural stress while standing; it differentiates neck, trunk and general body as well as combines joint deviations with exposure time.

Table 1. Posture checklist for neck, trunk and legs (Keyserling *et al*. 1992). A zero indicates insignificant risk. An X indicates a potential risk. A star indicates a significant risk.

Job Studied _____	PERCENT TIME POSTURE USED IN JOB		
	Never	<1/3	>1/3
Neck			
1. Mild forward bending (>20°)	0	0	X
2. Severe forward bending (>45°)	0	X	*
3. Backward bending (>20°)	0	X	*
4. Twisting or lateral bending (>20°)	0	X	*
Trunk			
5. Mild forward bending (>20°)	0	X	*
6. Severe forward bending (>45°)	0	*	*
7. Backward bending (>20°)	0	X	*
8. Twisting or lateral bending (>20°)	0	X	*
General body/legs			
9. Standing stationary (no walking or leaning)	0	0	X
10. Standing, using footpedal	0	X	*
11. Knees bent or squatting	0	X	*
12. Kneeling	0	X	*
13. Lying on back or side	0	X	*

Total X = ___ Total * = ___

Comments:

5. SOLUTIONS FOR STANDING

5.1. Changes from Standing

Two basic changes are sitting and walking. The optimum job design will have some sitting, some standing and some walking – not 100% of just one. If a job requires 100% standing (very few do), consider job rotation within the shift.

5.1.1. Sitting

Sitting reduces the load that is exerted on the feet and legs. Thus, there is less energy consumption and less biomechanical loading on the leg/foot joints. However, the person may become a "couch potato," giving inadequate nutrition to the spinal discs (which depend on movement for nutrition) and overadequate nutrition to the flesh, yielding a weight gain.

As pointed out above, the legs are ~32% of body weight. A stool (i.e. no arm or back support) means only 68% of the body weight must be supported by the human structure. It is difficult to quantify the benefits of a chair back support but common experience indicates they are beneficial. Supporting both hands/arms is ~10% of body weight but the percent decreases as the hands are used (e.g. keyboarding, assembly).

Biomechanically, sitting provides a stable support (chair seat) that is closer to the hands than the feet; this reduces sway and tremor. Yet standing permits movement of the legs and feet (not just the torso) so reach distances are improved and ability to twist the torso is improved.

5.1.2. Walking

Some of the problems of standing can be reduced by movement. Occasional walking (say > 10 steps) improves blood flow in the legs, decreasing ankle blood pressure ~60 mmHg. The movement of walking also improves disc nutrition. Walking also improves reach ability over standing.

5.2. External Aids

Two divisions are aids on the body and aids off the body.

5.2.1. Aids on the Body

Body aids are divided into back belts, support hose and shoes.

For such a popular product as back belts, there has been relatively little theoretical analysis or experimental study (Rys and Konz 1995). Back belts come in an amazing variety of designs with an equally amazing lack of theoretical support for the design. Back belts generally are tested on healthy male students doing lifting for short periods (often < 1 h); the results are extrapolated to male and female workers doing a wide variety of tasks for 2000/h year for multiple years. Much more (and better) research needs to be done. However, for simple standing tasks (with no material handling), it is hard to visualize benefits from back belts.

Support hose reduce venous pooling by constricting leg diameter and, thus, not permitting space for the blood to pool.

Shoes also come in an amazing variety of types and styles. Athletic shoes are divided into running shoes (designed for forward movement) and court shoes (designed for quick side-to-side movement); neither would seem applicable to prolonged standing.

Ankle support from boots normally would only be needed in outdoor activities with varied terrain.

A good shoe will have a good innersole that molds itself to the foot, giving support over a relatively large area and, thus, minimizing pressure on the sole. The outer sole of the shoe should be cushioned (either by material such as crepe or by form such as ripple soles). A good coefficient of friction combined with a design allowing liquid to escape from the floor–sole interface should reduce slipping. Shoes worn for standing should be purchased half to one size too large as feet will swell during the

day. Shoe laces permit adjustability; four-lace-hole shoes adjust better than three-lace-hole shoes.

5.2.2. Aids off the body

Two possibilities are standing aids and mats/carpets.

A standing aid allows the person to vary posture and/or support some of the body weight. Can the person lean on a counter or table? Bars have used leaning plus a bar rail for the foot for many years. In an occupational setting, Rys and Konz (1994) recommend, instead of a cylindrical rail, a 100 mm-high platform, either flat or angled 15°. Another alternative is a sit/stand stool; the adjustable height seat should have a forward slope of 10–15°. Nijboer and Dul (1987) reported they were useful in upholstery work even when they could be used for only 15% of the working time.

Cushioning the floor can be done everywhere (carpets) or locally (mats). Carpets/mats also act as a frictional surface (reducing slips).

6. SUMMARIZING

- Avoid surfaces with a low ratio of surface to holes (e.g. grating) as you are standing on "knives."
- Mats/carpets improve comfort over hard-surfaced floors. Comfort may increase in the back as well as the legs/feet.
- Mats/carpets should compress but not too much. Optimum is ~6% under the shoes of a 70 kg adult. Steel and concrete surfaces (i.e. no compressibility) are the worst surfaces for standing.
- Mats should have beveled edges to reduce tripping.

- Mats should be non-slip (both versus the sole and versus the floor).
- Mats that have to be cleaned periodically (e.g. in food service environments) should be smaller for ease of handling.

REFERENCES

CONTRERAS, R., RYS, M. and KONZ, S., 1993, Leg length discrepancy. In Marras, W., Karwowski, W. and Pacholski, L. (eds), *The Ergonomics of Manual Work* (London: Taylor & Francis), pp. 199–202.

KEYSERLING, M., BROUWER, M. and SILVERSTEIN, B., 1992, A checklist for evaluating ergonomic risk factors resulting from awkward postures of the legs, trunk and neck. *International Journal of Industrial Ergonomics*, 9, 283–301.

NIJBOER, I. and DUL, J., 1987, Introduction of standing aids in the furniture industry. In Buckle, P. (ed), *Musculoskeletal Disorders at Work* (London: Taylor & Francis).

NODELAND, H., INGEMANSEN, R., REED, R. and AUKLAND, K., 1983, A telemetric technique for studies of venous pressure in the human leg during different positions and activities. *Clinical Physiology*, 3, 573–6.

POLLACK, A. and WOOD, E., 1949, Venous pressure in the saphenous vein at the ankle in man during exercise and changes in posture. *Journal of Applied Physiology*, 1, 649–62.

RYS, M. and KONZ, S., 1994, *Standing. Ergonomics*, 37, 677–87.

RYS, M. and KONZ, S., 1995, Lifting belts: a review. *International Journal of Occupational Safety and Ergonomics*, 1, 294–303.

SWAT, K. and KRZYCHOWICZ, G., 1996, ERGONON: computer-aided working posture analysis system for workplace designers. *International Journal of Industrial Ergonomics*, 18, 15–26.

Static Load

B. Kapitanik

1. STATIC WORK ARDUOUSNESS

The assessment of work arduousness connected with static efforts becomes more and more significant in the general assessment of work stations. It is due to a widespread introduction of mechanization and automation in modern industry. It leads to an important reduction of dynamic efforts, but also to the appearance of static efforts caused by a still position at work, or by operating self-propelled tools. Such efforts create workloads stemming from both their intensity and duration. A train driver, for instance, has to keep still during a long period of time in order to watch his way ahead and to check the switchboard reading. Similar workloads are experienced by truck drivers, aircraft pilots or operators of huge, multipurpose machines such as excavators, combine harvesters or gantries.

Another kind of workload, which could be called the active static workload, is connected with the use of semi-automatic tools. It is well illustrated by the welder's job, especially in shipbuilding or in erecting steel constructions, presenting on the one hand a very uncomfortable, untypical, still position at work, and on the other hand the necessity to keep the burner fixed. Similar workloads can be found in foundries, where the stamping of the molding mass, the cleaning of the casts, the sand-blasting and the polishing are accomplished with tools, mainly pneumatic, which the worker holds and keeps still in his arms. The same type of active static load is experienced by a woodcutter operating a chainsaw weighing up to some 14 kg, which he has to keep stable. The operators of drifters, graders, grinders, and similar hand tools are often found in many branches of industry.

The terms "workload" and "work arduousness" are often left ambiguous thus leading to misunderstanding and confusion. Therefore, in this article, workload will mean the value of the external strain factor affecting the body, so that it represents the value of the force exerted and the duration of its action as far as the static workloads are concerned, whereas in the case of dynamic workloads, it is the magnitude of the power generated by the body and its duration. This workload can be related to the maximal capacities of a single organism and will then be called relative load; for the static workloads it can be expressed as the percentage of the maximal muscular force which can be exerted by a loaded muscular group (%MVC — maximal voluntary contraction). The workload produces a temporary disturbance in the functional balance of the body, which is compensated by mechanisms of adaptation. Triggering these mechanisms is connected, on the one hand, with the physical capacity (physiological strain) and, on the other hand, with a feeling of discomfort (psychological strain). An objective assessment of the latter is made difficult by its subjective nature. The sum of both physiological and psychological strains will be called work arduousness. Therefore, for the sake of practice, the following simplified assumption can be adopted: the value of the individual physiological response of the body to the workload can constitute an objective measure of work arduousness. Yet, in the present state of knowledge, the physiological assessment of work arduousness based on the appraisal of the value of the individual physiological response of the body to the workload seems to be of paramount importance in ergonomics.

The response of the cardiovascular system is particularly interesting as far as the assessment of work stations is concerned. The parameters characterizing the state of that system are easily accessible for a study with non-invasive techniques, and can be observed unceasingly without interrupting the tasks that the examined person is performing on the workstation.

2. APPROACHES TO ASSESSMENT OF STATIC LOAD

The first scientific recording concerning the physiological response of man's organism to static efforts was made by Beclard in 1861. He studied the changes of temperature in a muscle charged with an intermittent static load, while Tapie published in 1886 a vast monograph on the muscular thermoregulation during static and the dynamic efforts. In 1896, Chauveau presented the results of his research on the energy expenditure of the static work, considering for the first time the practical aspect of his research. Atzler, one of the first researchers to study the possibility of an appraisal of work arduousness connected with the static effort, worked out a method of such an assessment through measuring the energy expenditure for the static effort.

Simonson (1971) proposed a method of assessing the arduousness caused by a static load, basing it on the definition of a so-called "restitution constant", which would testify to the rate of evacuating the oxygen debt after the static work. Thus, he used Linhard's phenomenon described in 1913 consisting of the increase of oxygen consumption after the static work has stopped in relation to the duration of that work. However, the studies by Monod (1956), Rohmert (1961), Laville (1961), Lind (1966) and Wald and Harrisson (1975) showed that this phenomenon is independent of the load value.

Borsky and Hubac (1967) elaborated a nomogram whose construction was based on observing that during static efforts the increase of the heart rate is accompanied by a much lesser oxygen consumption than during the dynamic efforts. The same phenomenon was used by Klotzbucher (1975) to elaborate an index figure which was the quotient of the increase of the heart rate and the net energy consumption. My own research (1980) on work stations in which both general and local static loads were encountered, has proved that Klotzbucher's index figure only allows assessment of work arduousness connected with a general static load, but is unpractical in case of a local load.

The observation made by many authors that, during a static effort, the arterial blood pressure, both systolic and diastolic, grows in proportion to the value of the relative load has led to no practical application as yet. The difficulties of making measurements and a significant individual variability result in a reluctance to use the arterial blood pressure as a parameter in the studies on workstations.

To sum up, it can be said that the present state of research on the cardiovascular response to the static efforts has not laid the foundations for the elaboration of a method of assessing the arduousness caused by the static loads that could suit the needs of ergonomics. The proposed methods allow a haphazard appraisal of the arduousness caused by the static loads and only

within an incomplete range of those loads. Moreover, the first series of methods requires the measurement of oxygen consumption, which causes a certain inconvenience for the subject and cannot be performed continuously. There is a need, therefore, for a method of assessing the arduousness connected with the subcritical static efforts (i.e. involving less than 20%MVC) lasting for a long period of time.

3. STRESS EVALUATION

Since static work generates no visible movement, the static stress (constraint) cannot be measured with physical units of work. Therefore the intensity of work is assessed by the level of the exerted force and the time of the exercise in the case of a continuous task, or the number of repetitions in the case of an intermittent task.

3.1. Application of Force Measurements

During professional work the postural constraints generate neither high levels of force nor high energy expenditure. Conversely, the handling of manual tools (screwdriver, pliers, etc.) or portable tools (drills, saws, etc.) often generates a relatively important percentage of the maximal force.

The most relevant assessment of static stress is done by measuring the forces involved and comparing them with the maximal forces. The measurements are made with a strain gauge placed in a direction of force exercise. Figure 1 presents the measurement of the force of pressing on the pedal of a hospital trolley. The strain gauge (SG) is pressed against the pedal and held in the work position. The same kind of setting allows easy measurement of the forces involved in pushing the trolley, holding a parcel or pulling an object. The measured forces are compared with the maximal forces by using two different methods.

The first method consists in measuring the operators' maximal forces in the same position as at work with the same strain gauge. The advantage of such a procedure is that it gives the real maximal forces of those directly using the machine. The main drawback is that those forces are measured on a small sample and could be altered by measuring conditions which are impossible to standardize. Therefore, a second parallel procedure must always be applied in order to check the normative data. In France, the norm NF X35-105 indicates the limits of forces acceptable for static work. Similar recommendations have been set by NIOSH in the USA and are currently being converted into a European norm EN1005-3 and an international norm ISO11228-2.
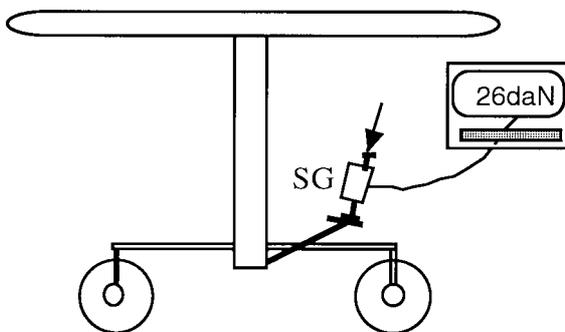
Figure 1. Measurement of the force of pressing on a pedal of a hospital trolley

By applying these procedures we were able to assess the hospital trolley with variable height (see figure 1) and observe that the forces necessary to push it are acceptable for 95% of the female population, whereas the forces necessary to press the pedal are beyond acceptable limits even for the male population.

3.2. Repetivity Calculation

In industrial conditions static work for driving a machine or operating a manual tool is often intermittent. It is crucial in the assessment of workload to have a reliable way of assessing the ratio of static exercise time in relation to rest time. For this purpose it is vital to observe the activity directly with precise time control. In cases of repetitive work, it is advisable to use video recordings which allow precise study later on. Currently, there are several video-processing software packages which facilitate analysis and make results readily available. Some software packages aim mainly at postures (OWAS), others are more general (KRONOS). It is likely that this technology will be considerably improved in the near future.

4. STRAIN ASSESSMENT

4.1. EMG Application

Surface electromyography (EMG) is a totally non-invasive technique which permits the quantification of muscular activity due to portable devices becoming more and more sophisticated and miniaturized. This technology is beginning to be used in ergonomics especially in the case of predominantly static muscular constraints.

With electrodes attached to the skin, the surface EMG permits the recording of bioelectric activity of the part of the muscle close to the electrodes and concomitant with their mechanic activity. The spectral analysis and the integration of the EMG signal allow the quantitative analysis of this activity and comparisons to be made during different exercises of the muscles; integrated EMG (iEMG) represents the intensity of the isometric contraction of muscles. Median frequency (MF), which is among the most common parameters of spectral analysis, is directly proportional to the speed of intramuscular conduction and the mean power frequency (MPF); the two allow the dynamics of muscular recruitment to be seen.

The application of surface EMG during static effort allows the extent of the muscular mass involved in the effort, the intensity of the muscular effort, and the apparition of fatigue to be assessed.

4.1.1. Identification of active muscles

This is important during complex activities, such as using a keyboard. The identification of the muscles involved in this task allows optimization of the keyboard so that the local work of different muscles is minimized and not taken over the threshold of 15% of their maximal force. Similarly, a correct identification of the muscles involved in lifting loads provides a basis for the training of packers and thus contributes to the prevention of musculoskeletal disorders.

4.1.2. Intensity of contraction

During static maintenance, there is a linear relation between the exerted force and the amplitude of electric activity. This relationship becomes slightly exponential when the maintained force reaches values higher than 50%MVC. The EMG increases

more rapidly beyond those values, according to the mobilization of rapid fibers, whose activation threshold is higher. However, for ergonomic applications it is possible to adopt a simplified linear relation because the static efforts exerted during professional tasks are situated in a section lower than 50%, and even for those whose intensity is higher the error is relatively low.

The measurements of the iEMG can only be interpreted in relation to a reference corresponding to the exercise of the maximal force because the recorded electrical intensities depend on the impedance of the electrodes and of the tissue, and on the placing of electrodes. Indeed, it was the evolution of EMG that indicated the intensity of the contraction, not the absolute values recorded.

With this technique we were able to observe that the static effort exerted by a professional violinist in keeping the left hand still during a concert reaches 23% of the maximal activity recorded during the exercise of the maximal force in the same position of the arm (see figure 2).

4.1.3. Effect of fatigue

For a given level of maintained force, the amplitude of the EMG increases with time when the isometric contraction is prolonged. When muscular exhaustion in close, potential of great amplitude and big activity outbursts appear and electrical signals fall to low frequencies because there is a progressive increase in the number of active motor units occurring at the end of the time limit. This muscular fatigue is characterized by an increase of the iEMG and a decrease of MF and MPF.

Muscular fatigue also provokes the apparent emphasis of physiological trembling at 2.5–30 Hz (12 Hz on average), easily detectable in EMG recordings.

4.2. Cardio–circulatory Adaptation

During static exercise the cardiac output remains unchanged initially, and slightly increases later as the isometric contraction of the muscles provokes an important increase of peripheral resistance, which beyond 75%MVC may result in an interruption of the local flow in the contracted muscle. The stroke volume decreases and the heart rate increases because of a reflex of muscular chemoreceptors. This increase is much more important in relation to the increase of oxygen consumption than during a dynamic exercise at a corresponding level of relative load.

By comparing the dynamics of increase for the heart rate and for oxygen consumption, one can assess the influence of the static load. This comparison requires knowledge of the subject's physical capacity. For an average subject (with a VO2 max. at 3 1/min.) the consumption of 110 2/min. during a dynamic exercise generates a heart rate of 120 bpm. During static effort at the same heart rate of 120 bpm the oxygen consumption is approximately 0.6 10 2/min. Yet attempts to establish a practical index of static load have been unsuccessful because of important individual variability and the fact that, although the cardiac reflex does not depend on the muscular mass involved, the oxygen consumption does.

The limit that is proposed for cardiac acceleration during static effort is the absolute value of the cardiac cost of 25 bpm above the rest value. In terms of relative cardiac cost, it is dangerous to go beyond 20% of the maximal cost.

REFERENCES

BOUISSET, S. and MATON, B., 1995, *Muscles, posture et mouvement* (Paris: Hermann), vol. I, p. 735.

BROUHA, L., 1960, *Physiology in Industry* (Oxford: Pergamon), vol. I, p. 145.

MONOD, H. and KAPITANIAK, B., 1999, *Abregé d'ergonomie* (Paris: Masson), vol. I, p. 260.

PENINOU, G., MONOD, H., and KAPITANIAK, B., 1994, *Prévention et ergonomie* (Paris: Masson) vol. I, p. 119.

SCHERRER, J., 1992, *Précis de physiologie du travail* (Paris: Masson), vol. I, p. 585.
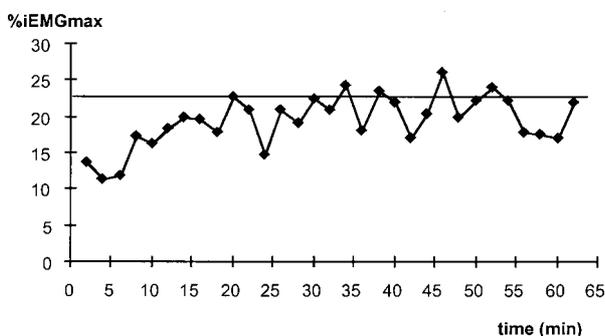
Figure 2. Recording of iEMG expressed in %iEMGmax from left biceps brachi of a violonist playing 1 h continuously. During the first 20 minutes iEMG increases and thereafter rests in a steady state.

Static Work Capacity

B. Kapitaniak

Ergonome, Maître de conférences à l'université Pierre et Marie Curie, faculté de médecine Pitié-Salpêtrière. Paris

1. STATIC WORK DEFINITION

The term "work" is used in physiology in the context of static work in a slightly different way from physics because the shifting occurs only within the muscle with no visible movement outside. The increase of the muscular tension is obtained thanks to the work of the contractile muscular fibers which, in the precise case of an isometric contraction in which the muscle cannot change its length, work against the whole of muscular elastic structures. The work is accomplished by the contractile part of the muscle by stretching the elastic part. In order to characterize this work on must therefore consider the force developed by the muscle during its isometric contraction, the duration of this contraction and the conditions that guarantee the energy supply in the muscle. The product of the maintained force by the time of maintaining is therefore a measurement of static work widely applied in ergonomics.

2. MUSCULAR STRENGTH MEASUREMENT

Since the developed force depends on the length of the muscle, the measurements of the maximal force are made only during isometric contraction when the length is fixed arbitrarily. The maximal voluntary contraction (MVC) of a muscle group involved in a static contraction is a reference to assess the relative forces in terms of %MVC.

MVC depends on physiological, psychological and metrological factors. Above all, it is determined by the quality of the muscles involved in the contraction, characterized by the number of myofibrils per unit of section surface, by the orientation of muscular fibers in the muscle and by the type of fibers dominating in the muscle (at equal force, a rapid muscle exerts a force higher than a slow muscle). It also depends on the quality of the motor command and the nervous transmission, as well as the motor coordination especially in the case of complex contractions.

Beside the quality of the muscles, the MVC depends a lot on the subject's motivation; at equal muscular mass an athletic subject exerts an MVC higher that an untrained subject.

Finally, MVC is subject to an influence of the contention system that allows it to exert a force that can more or less limit the supplementary contractions. This factor may be mastered when measuring the MVC in laboratory conditions, but it is more difficult to realize in the field. When measuring the forces exerted on a workstation one must take all the precautions to define measuring conditions as stable as possible.

The level of the exerted force is expressed either in absolute values (Newtons) or in relative values (%MVC), which allows a fair assessment of the constraint in the case of static work.

During spontaneous activity the static contractions are frequent: the postural bearing, holding or carrying a load, or pressing a pedal. In most of these activities the forces involved may easily be measured with dynamometers and compared with the maximal forces exerted in the same conditions.

3. TIME LIMIT OF STATIC MAINTAINING

The time of the static exercise is expressed either in absolute values (seconds) or in relative values (the percentage of the maximal time of upholding a force called the "time limit"). A quasi-exponential relation was established between the relative force and time limit by Monod (1956) and confirmed by Rohmert (1958).

The time limit of maintaining a given level of force characterizes the capacity of static work. The hyperbolic relation between the force (%MVC) and the time limit of maintaining is presented on figure 1. The origin of the curve situated on the axis of forces corresponds to 100%MVC, the time limit for 50%MVC is ~1 min, and the asymptote of the curve is between 15 and 20%MVC. The critical force corresponding to a theoretically infinite time of maintaining is therefore situated at ~15%MVC. Any contraction realized at a level higher than the critical force leads to local exhaustion, whereas at a lower level it could theoretically be carried on without any time limit. However the prolonged exercise of forces lower but close to the critical force leads to important discomfort, yet without any objective sign of muscular fatigue. On the practical level it would be better to adopt a critical force of 10%MVC for the recommendations concerning the efforts exerted on machinery commands (norm prEN 1005-3).

The time of static exercise determines the intensity of physiological reactions. For forces higher than the critical force reactions increasing proportionally to the time of exercise may be observed. The reactions become intense after 50% of the maximal time limit of maintaining.

4. CHARACTERISTICS OF STATIC WORK CONDITIONS
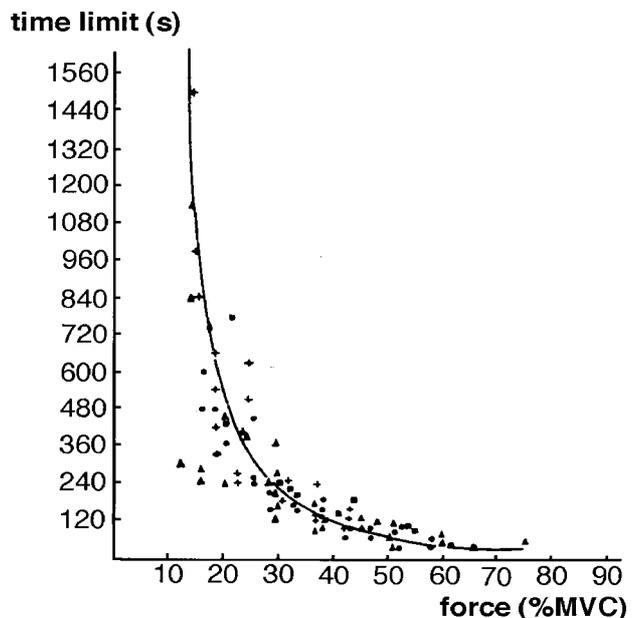
Figure 1. Relation between force (%MVC) and time limit of maintaining (Monod 1956).

4.1. Bioenergetic Consequences of Static Work

The static work of an intensity higher than the critical force is often accomplished in the muscles with an important participation of anaerobic metabolism. This participation depends on the level of relative load and of exercise time which determine the local blood flow in the active muscles. The local flow increases up to 20%MVC, whereas above the increase in no longer appropriate to the energetic needs. It is low at 50 and 70%MVC the flow almost stops.

The low load exercises, < 10%MVC may be realized in aerobic. Those > 20%MVC have an important anaerobic participation and < 70%MVC the exercises are accomplished solely in anaerobic. The work realized during static exercise is there very costly because it carries an important oxygen debt in relation to the mechanical work realized. For intense static exercises low oxygen consumption is observed during the exercise, whereas it significantly increases once the exercise has stopped (Lindhard's phenomenon).

4.2. Cardiac Consequences of Static Work

The mechanical pressure of muscular fibers which is exerted on the vascular compartment and which is clearly reduced as soon as the force is > 25%MVC, provokes an increase of peripheral resistance with heavy hemodynamic consequences. Unlike dynamic work, the vasodilatation generated by the muscular work is impeded by this mechanical pressure which results in an important increase of the systolic and diastolic blood pressure as well as of the heart rate.

The current concept of mechanism concerning cardiovascular response to sustained isometric contraction assumes that the heart rate increases initially by central command via vagal withdrawal and may be modified later by activation of sympathetic nerve activity. On the other hand, the rise in blood pressure at the beginning of isometric exercise is due to increased peripheral resistance and cardiac output, whereas the further increase in blood pressure is due to activation of sympathetic drive originating from muscle chemoreflex. In fact the significant increase in HR associated with a moderate increase in cardiac output and the decrease of stroke volume depending on the decrease of the venous return are commonly observed during different modes of isometric contractions.

Cardiovascular response to sustained static exercise is dominated by the relative workload because the mechanism is essentially reflex. However, it may also be influenced by the mass of muscle involved that can be responsible for a moderate increase of oxygen consumption in the cases of moderate static load.

4.3. Vegetative Modifications on Static Load

The vegetative modifications appear quite early, even for forces scarcely higher than the value of the critical force. They arise as an increase of the plasmatic noradrenaline rate. They seem independent from the muscular mass involved but related to the relative force used. The effect of this "pressure reflex" of muscular origin is to delay the hypoxy due to contraction. Thus, the values of the heart rate noted during static work and resulting from this reflex do not have the same significance as those observed during dynamic work, for purely energetic reasons.

5. INTERMITTENT STATIC EXERCISE
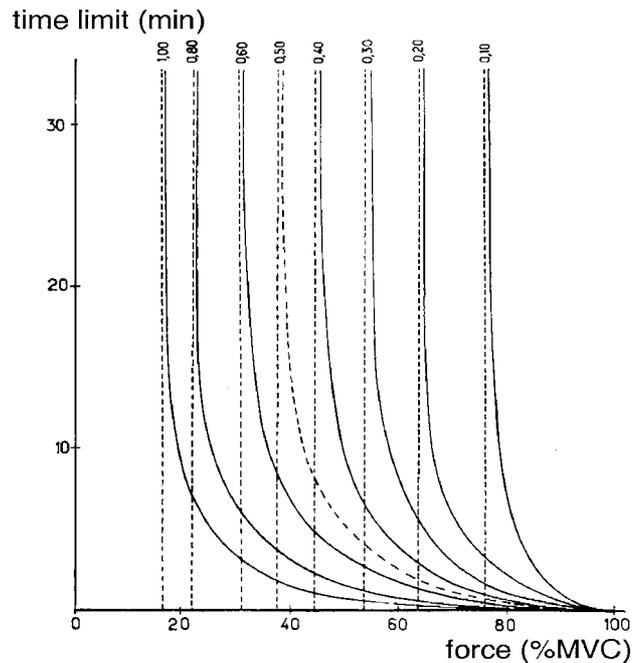
Figure 2. Relation between force (%MVC) and time limit of maintaining in the cases of intermittent static work (Pottier 1969).

The introduction of pauses during which the blood flow is free (intermittent static work) enables the muscle to work longer with higher levels of force. There is a shift of the value of critical force all the more important as the relations between rest period and contraction period are bigger (figure 2).

Intermittent static work is close to dynamic work when the pauses are important because of the liberation of local circulation. On the other hand, repetitive dynamic work may generate the same effects as static work when the frequency of repetition of movements becomes so high that the muscle has no longer the time to restore the local circulation during the pause.

6. LOCAL AND GENERAL STATIC LOAD

In physical professional activities two types of exercises are the most frequent : the use of segmentary force at high levels (holding loads at arms' length, effort on a lever or on a motorized tool) or tasks involving the whole body (lifting and holding heavy loads, pushing a trolley with the whole body). These two aspects correspond to a physiological distinction based on the muscular mass used during the task and which allows to separate local work involving less than a third of the total muscular mass from general work involving more than one-half.

If the intensity of the reflex of muscular chemoreceptors depends mainly on the relative intensity of the isometric contraction, the adaptive reactions also take into account the muscular mass involved. The oxygen consumption increases more during general than during local static work, at equal relative intensity. Similarly, the increase of the breathing output is more important during general work.

7. MIXED WORK

In industrial practice the static load is very often accompanied by dynamic load. Manual handling of loads includes a static load of holding an object and a dynamic load of walking. The result
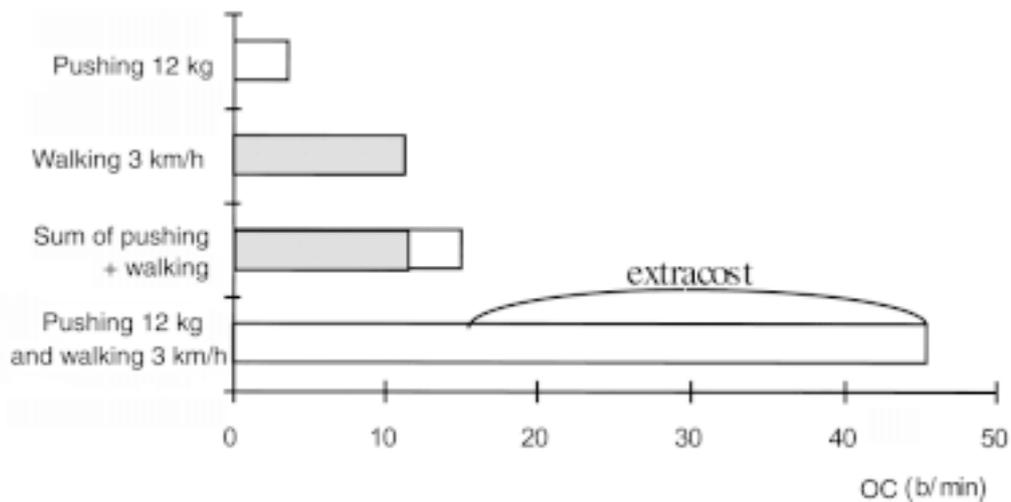
Figure 3. Multiplying effect of mixed load for cardiac cost (CC). The sum of CC from pushing only and walking only is widely under the CC observed for both tasks exerted together (Sanchez 1977).

of this simultaneous double load on the cardiovascular plane is a multiplying effect. Indeed, the reaction observed during mixed load is stronger than the sum of the reactions due to the two separate loads. These multiplying effects were noticed during pushing trolleys experience (figure 3).

8. DETERMINATION OF STATIC WORK CAPACITY

Unlike dynamic work there are no commonly accepted indices of capacity for static work. This is due to the important variability of the tested indices, which does not allow one to reach relevant practical conclusions. The maximal time limit may be considered as the only practically applied index.

There is no correlation between the muscular force and the maximal time limit. Comparing the sex differences, one may observe that women have a maximal time limit higher than men, whereas their maximal voluntary force is ~66% of mens'.

Several studies have been devoted to anisotonic isometric contractions, consisting in maintaining a decreasing level of force.

It seems that the reduction of the maintained force depends on the nature of muscular fibers involved, as well as on the maximal time limit. The studies carried out on trained sportsmen do not systematically bear out this hypothesis. No single proposition of a global index has ever been retained due to the lack of necessary coherence.

REFERENCES

BROUHA, L., 1960, *Physiology in Industry* (Oxford: Pergamon).

FLANDROIS, R. and MONOD, H., 1990, *Physiologie du sport* (Paris: Masson).

LEHMANN, G., 1955, *Physiologie pratique du travail* (Paris: Les éditions d'organisation).

SCHERRER, J., 1992, *Précis de physiologie du travail* (Paris: Masson).

MONOD, H. and KAPITANIAK, B., 1999, *Abregé d'ergonomie* (Paris: Masson).

Stimulus–Response Compatibility

R. W. Proctor

Department of Psychological Sciences, Purdue University, West Lafayette, IN 47907-1364, USA

1. INTRODUCTION

Stimulus–response (S-R) compatibility effects are differences in speed and accuracy of responding as a function of the relation or mapping between stimulus and response sets. Such effects are neither perceptual nor motoric in nature. Rather, they are due to response-selection processes that intervene between stimulus identification and response execution. When the mapping of display elements to controls is highly compatible, an operator can respond faster and more accurately than is possible when the mapping is incompatible.

Compatibility effects have been investigated extensively since classic studies conducted by Paul Fitts *et al.* in the 1950s. Many issues that are fundamental to human cognition, including the nature of spatial coding, the role of attention in coding, and the influence of irrelevant information on performance, have been addressed in compatibility research. The following sections summarize current knowledge regarding the determinants of compatibility effects and discuss the implications for display-control design.

2. PAUL FITTS'S CLASSIC STUDIES

Fitts and Seeger (1953) first demonstrated S-R compatibility effects. Subjects performed an eight-choice reaction task in which three arrangements of stimulus locations were paired in all combinations with three arrangements of response locations. Responses were made by moving a stylus from a centered home location to the response location signified by the stimulus, using the most compatible mappings of the individual stimuli to responses. Although the number of S-R alternatives was identical in all cases, responses were faster and more accurate when the stimulus and response arrangements corresponded than when they did not. A second experiment showed that the compatibility effects persisted across many sessions of practice, a finding that has been replicated in a variety of other task situations. The fact that compatibility effects are resistant to practice implies that the negative consequences of a design that is low in S-R compatibility cannot be overcome simply by providing operators training and experience.

Fitts and Deininger (1954) used one of the spatially corresponding combinations of S-R sets examined by Fitts and Seeger (1953), circular stimulus and response arrays, but varied the mapping of individual stimulus locations to response locations. Responses were much faster and more accurate for a spatially compatible mapping in which each stimulus location was mapped to the corresponding response location than for a mapping in which the relation between stimulus and response locations was random. Although responses with a mirrored S-R mapping (respond at the mirror opposite location) were slower than those with the spatially compatible mapping, they were still

much faster than those with the random mapping. This study illustrates two major points about S-R mappings: responses are fastest when spatial compatibility of display and control elements is maintained, and responding benefits from systematic relationships when the mapping is not spatially compatible.

Through the remainder of the 1950s and 1960s much of the research on S-R compatibility had the goal of discovering the situations under which compatibility effects occur and was relatively atheoretical (Alluisi and Warm). In the 1970s and 1980s, theoretical issues regarding S-R compatibility became of increasing concern in a variety of disparate research areas, including attention and performance, cognitive neuropsychology, motor control, and human engineering, among others. These research efforts culminated in two influential publications in 1990, Proctor and Reeve's edited book that integrated the different strands of research, and Kornblum *et al.*'s dimensional overlap model and taxonomy. In the 1990s there has been a concerted effort by many researchers to examine fundamental theoretical issues pertaining to S-R compatibility. Much of this research is described in the 1997 book edited by Hommel and Prinz.

3. SPATIAL CODING

3.1. Coding of S-R Locations

Most accounts of S-R compatibility rely on the concept of coding in response selection. In the case of spatial compatibility effects, the emphasis is on spatial codes that mediate between stimuli and responses. Much of the basic research on compatibility effects has used two-choice tasks. The standard finding is that responses are faster and more accurate with a spatially direct mapping in which a left stimulus is mapped to a left keypress and a right stimulus to a right keypress than with the opposite mapping. Although most such studies have used horizontally arrayed locations, similar compatibility effects occur if the stimulus and response locations are arranged vertically. Evidence for spatial codes is apparent in the widely replicated finding that the spatially direct mapping is faster than the opposite mapping, even when the hands are crossed so that the left key is pressed by the right hand and the right key by the left hand. Crossing the hands slows response time overall, but this effect is additive with that of spatial compatibility. This finding implies that the response is first determined on the basis of spatial location, and then the appropriate effector is specified to execute the movement at the desired location.

Stimulus location also affects performance when it is irrelevant to the task. Thus, a designer cannot ignore spatial compatibility even when the relevant stimulus information is non-spatial. For example, if a left response is to be made to a red stimulus and a right response to a green stimulus, responses are faster when the red stimulus occurs to the left and the green stimulus to the right. This phenomenon, first observed by J. R. Simon in the late 1960s, is now called the Simon effect (Lu and Proctor 1995, for reviews Simon 1990, Umiltà and Nicoletti 1990). The Simon effect indicates that a spatial code corresponding to the stimulus location is formed automatically and produces response activation when not intended. Because the issue of why spatial codes are formed when location is irrelevant is of theoretical interest, the Simon effect has been the focus of much of the recent research on compatibility effects (Hommel and Prinz 1997).

3.2. Relative Location

Whether stimulus location is relevant or irrelevant to a task, the spatial codes are based on relative location. That is, even if either the display elements or the response keys are both located to one side of the performer in the same hemispace, the mapping of left stimulus to left response and right stimulus to right response is superior to the reversed mapping. Compatibility effects also occur when the spatial information is conveyed by words (e.g. LEFT), symbols (e.g. a left-pointing arrow), or direction of motion (e.g. leftward movement). They occur as well as for vocal responses with spatial content (e.g. the utterances LEFT and RIGHT), movements of a joystick to the left or right, etc. However, the relative advantage for a compatible S-R mapping (e.g. left-to-left and right-to-right) is larger when verbal stimuli are paired with verbal, vocal responses and when non-verbal stimuli are paired with manual responses (Proctor and Wang 1997).

The coding of relative location depends on task goals and the performer's intentions, and not simply the locus or direction of the effector. In one study, subjects performed a two-choice task using sticks to operate the keys, with the sticks crossed such that a press of the left key required movement of the right hand on the right side, etc. Responses were fastest for the mapping in which the key location corresponded to the stimulus location, even though this meant there was no correspondence between the location of effector movement and stimulus location. Another study used a Simon task for which the pitch of a tone (high or low) was relevant and tone location (left or right) irrelevant. The unique aspect of this study was that the keypress response lit a light on the opposite side. When subjects were told to trigger the left light to the low pitch tone and the right light to the high pitch tone, or vice versa, responses were faster when stimulus location and light location corresponded than when they did not, even though the location of the keypress did not correspond with that of the tone. The important point from these studies for system designers is that very different spatial compatibility effects can be obtained as a function of the task goals of the performers.

3.3. Multiple Reference Frames

The results from several studies imply that spatial codes are formed with respect to multiple frames of reference when more than one is present. This point is illustrated by studies that used displays with eight locations, four in the left hemispace and four in the right hemispace. The relevant stimulus attribute was non-spatial, for example, a red stimulus signified a left keypress and a green stimulus a right keypress. On each trial a fixation cross, presented in the left or right hemispace, restricted the possible stimulus locations to the four in that hemispace. This was followed by presentation of two boxes to the left or right of the fixation point (i.e. in the left or right hemifield), one of which contained the stimulus to which a keypress response was to be made. The relative position of the stimulus in the two boxes defined the third frame of reference. Spatial correspondence effects were evident for all three reference frames, hemifield, relative position and hemispace. In other words, responses were fastest when the spatial codes with respect to all frames of reference corresponded to the response that was to be made and slowest when they all conflicted. Conflict between multiple spatial codes can have serious consequences, with such conflict apparently being the primary cause of the crash of a British Midland Airways Boeing

737-400 aircraft in January 1989 in which the pilot shut down the wrong engine following fan-blade failure.

The findings indicative of coding with respect to multiple frames of reference are in accord with the view that spatial codes are generated automatically with respect to referent objects or frames (Hommel 1997). An alternative proposal is that the formation of spatial stimulus codes is linked to shifts of attention involved in the control of saccadic eye movements (Stoffer and Umiltà 1997). Although there is considerable evidence to suggest that attention shifts may play a role in spatial coding, the exact nature of that role has yet to be determined.

It is possible to have stimuli and responses vary along both vertical and horizontal dimensions. That is, with four locations arranged in a square, each location can be characterized as a combination of up or down and left or right. If a two-choice task is used in which stimuli on one diagonal are mapped to responses on the other diagonal, responses are fast if the mapping is compatible on both dimensions and slow if it is incompatible on both. Of interest is what happens when there is compatibility on one dimension and incompatibility on the other. Several studies have suggested a form of what is called right–left prevalence. Performance is often better when there is compatibility on the horizontal dimension than when there is compatibility on the vertical dimension. Current evidence suggests that this prevalence is a function of a strategy that is adopted when either instructions emphasize the horizontal dimension or practice is sufficient to lead the performer to discover the benefit of coding the relations in terms of left and right.

4. SYMBOLIC S–R COMPATIBILITY EFFECTS

Compatibility effects are not restricted to spatial dimensions. Symbolic compatibility effects can be obtained when, for example, stimuli are colors and response keys are visibly labeled with the same colors. Responses will be faster when the assigned response color corresponds to the stimulus color than when it differs. Similarly, verbal responses to alphanumeric stimuli will be faster and more accurate when they are the typical names of the letter, number or word than when they are reassigned. Also, the mapping of an ordered stimulus set, say, the digits 1–4 to four linear response locations is more compatible if the ordering is increasing digit magnitude going from left-to-right than if some other mapping is used. According to Kornblum *et al.* (1990), dimensional overlap between the stimulus and response dimensions, that is, physical or conceptual similarity, is necessary to produce mapping effects.

Even stimulus and response sets that seemingly have no dimensional overlap can produce S-R compatibility effects if the structure of the stimulus set is mapped onto the structure of the response set. According to the salient features coding principle proposed by Proctor and Reeve (1990), when the members of the stimulus set and the members of the response set are distinguished in terms of salient features, performance is best for mappings in which the salient features of the two sets correspond. For example, if four stimuli vary along two dimensions, size (small or large) and letter identity, and letter identity is salient, then performance is better with a mapping in which letter identity corresponds to the salient left–right distinction for a row of four response keys than when it does not. The salient features coding principle has fared well in predicting the relative compatibility

of alternative mappings for numerous stimulus and response sets that vary along two dimensions. In sum, when there is similarity between the dimensions along which the stimulus and response sets vary or structure to the two sets, mappings that maintain the similarity or structure will typically lead to better performance than ones that do not.

5. COMPATIBILITY EFFECTS FOR ORTHOGONAL ORIENTATIONS

One of the more interesting findings in recent years is the occurrence of spatial compatibility effects for orthogonal spatial dimensions. When stimuli vary along a vertical dimension (up and down locations) and responses along a horizontal dimension (left and right responses), the mapping up-right/down-left typically yields better performance than the mapping up-left/down-right. The advantage for the up-right/down-left mapping is obtained for a variety of stimulus types (physical locations, words, arrows) and response sets (aimed movements, keypresses, vocal words). Compatibility effects for orthogonal dimensions are important because there are many situations in which controls and displays are orthogonal, either explicitly or implicitly. For example, the volume control on a car radio may require pressing a right button marked with a right pointing arrow to increase loudness and a left button marked with a left pointing arrow to decrease loudness.

One possible explanation for the orthogonal compatibility effects, based on the notion of salient features coding, is that "up" and "right" are the salient polar referents for the two dimensions and that less translation is required when the salient referents correspond than when they do not. Some authors have suggested that this salient features coding account may apply only to situations in which a person codes the stimuli verbally. The extent to which asymmetric coding is limited to verbal codes or is a more general characteristic of coding is a matter of debate at this time.

A finding with considerable design implications is that when making unimanual left–right switch or finger movements, the preferred mapping varies as a function of where the switch is placed relative to the body midline. If the switch is in the right hemispace, a pronounced up-right/down-left advantage is evident. However, if the switch is in the left hemispace, the advantage shifts in favor of the up-left/down-right mapping. A modified version of the salient features hypothesis says that the relative salience of left versus right is influenced by the relative location of the response switch, such that when the switch is to the person's left, left becomes the polar referent. Evidence has also been presented that the effect can be attributed at least in part to use of the hand as a frame of reference with respect to which the horizontally arrayed response locations can be coded vertically, thus rendering the coding of stimulus and response locations parallel.

Use of frames of reference to align orthogonally oriented stimulus and response arrays has been demonstrated for visual reference frames as well. Specifically, when two vertically oriented stimulus locations coincide with the locations of the eyes of a face rotated 90° to the left or right, responses are faster and more accurate if the location corresponding to the right eye, as viewed by the observer, is mapped to the right response and that

corresponding to the left eye to the left response. The implication is that frames of reference may provide a basis for coding orthogonal dimensions in a parallel manner, allowing a mapping that maintains spatial correspondence to be used.

There has not been much research with orthogonal stimulus and response arrays composed of more than two alternatives, but the limited results that are available suggest that in this case one of the arrays, most likely the response array, is mentally rotated to be aligned with the other array. A recent suggestion is that the effects of response eccentricity on the mapping preference for up and down stimuli mapped to left–right responses in two-choice tasks may be a consequence of changes in preferred direction of rotation. The extent to which mental rotation is involved in performance with orthogonally oriented arrays remains to be determined.

6. ONE OR TWO RESPONSE–SELECTION ROUTES?

Beginning with Fitts, compatibility effects have been attributed in part or whole to a response–selection route that usually is characterized as involving "translation" of the stimulus information into a response based on the instructions given the performer. Translation is considered to be fastest when an identity rule is applicable, for example, respond with the spatially corresponding response, and slowest when a search or look-up process must be used, as when there is no systematic relation between stimuli and their assigned responses. Translation also benefits from other rule-based relations, such as respond with the mirror opposite location.

During the 1990s, the most widely accepted view has been that there is a second response–selection route, characterized as "direct" or "automatic," that contributes to compatibility effects. The general idea is that a stimulus will automatically activate its natural, corresponding response via this route. Kornblum *et al.*'s (1990) dimensional overlap model is a dual-route model of this type. A major reason for proposing an automatic activation route in addition to the translation route is that, as described earlier, stimulus location affects performance even when it is irrelevant to the task. The basic idea is that the irrelevant location information affects performance because the stimulus activates the corresponding response via the automatic route. Measurements of the lateralized readiness potential, an event-related brain potential that shows differential activation of the motor areas in the two cortical hemispheres, in fact show an initial tendency to prepare the spatially corresponding response even when the relevant information specifies the non-corresponding response. As would be expected on the basis of automaticity, the response activation produced by irrelevant stimulus location tends to occur quickly and then, because it is irrelevant, to decay. Thus, the effects of irrelevant location on performance often will be larger when the relevant information can be processed quickly than when it takes longer to do so.

This automatic activation of the corresponding response is frequently characterized as due to links that are hardwired genetically or learned through a lifetime of experience with spatially compatible relations (Umiltà and Zorzi 1997). However, which response is activated can be altered with relatively little experience. Lu and I recently gave subjects given ~1000 trials of

practice with a spatially incompatible mapping of stimulus locations to response locations in a two-choice task. When stimulus location was subsequently made irrelevant, the Simon effect was reversed, that is, responses were faster when stimulus location and response location did not correspond than when they did. This reversal of the typical Simon effect persisted > 600 trials, so it was not extremely short-lived. The point is that the tendency to make the spatially corresponding response can be modified by task instructions that were in effect previously. It is also possible to reverse the effects by using a mapping rule for the relevant stimulus dimension (e.g. color) that allows an "opposite" transformation to be applied (e.g. press the red key if the stimulus is green and vice versa). In this case, the re-coding rule apparently gets misapplied to the location dimension, producing an advantage when stimulus and response locations do not correspond.

There has been less agreement about whether and how this automatic activation route contributes to compatibility effects for relevant stimulus information. Kornblum *et al.* (1990), for example, assume that it is a factor contributing to such compatibility effects whenever the relevant stimulus dimension overlaps with the response dimension. Another view is that the automatic route only comes into play when all of the possible stimuli are mapped compatibly to responses. The finding on which this view is based is that when compatible and incompatible mappings are mixed, the compatible responses are slowed substantially relative to a pure compatible block, but the incompatible responses are not affected much. It has been suggested that the automatic route is inhibited and the slower translation route used when only some of the S-R relations are known to be compatible. The important point for designers is that the benefit of compatible mappings may be lost if other information in the environment is mapped incompatibly.

7. IMPLICATIONS FOR HUMAN FACTORS

1. S-R compatibility effects are pervasive and persistent. They occur whenever there is physical, conceptual, or structural similarity between stimulus and response sets. A good interface must have a high degree of compatibility and allow fast and accurate response selection.
2. With spatially arrayed displays and controls, correspondence in terms of relative spatial locations should be maintained whenever possible. If multiple reference frames exist, compatibility with respect to all frames must be considered.
3. Spatial compatibility effects can occur even when the stimulus location does not provide relevant information.
4. Compatibility effects often occur when displays and controls are oriented orthogonally. These orthogonal compatibility effects

may vary as a function of alternative frames of reference that are available and the location of the controls relative to the operator.
5. Much of the advantage of compatible S-R mappings may be lost when they are mixed with incompatible S-R mappings.

REFERENCES

ALLUISI, E. A. and WARM, J. S. (1990), Things that go together: A review of stimulus–response compatibility and related effects. In R. W. Proctor and T. G. Reeve (eds), *Stimulus–Response Compatibility* (Amsterdam: North-Holland), 3–30.

FITTS, P.M. and DEININGER, R.L., 1954, S-R compatibility: correspondence among paired elements within stimulus and response codes. *Journal of Experimental Psychology*, 48, 483–91.

FITTS, P.M. and SEEGER, C.M., 1953, S-R compatibility: spatial characteristics of stimulus and response codes. *Journal of Experimental Psychology*, 46, 199–210.

HOMMEL, B. (1997), Toward an action–concept model of stimulus–response compatibility. In B. Hommel and W. Prinz (eds), *Theoretical Issues in Stimulus–Response Compatibility* (Amsterdam: North-Holland), 181–208.

HOMMEL, B. and PRINZ, W. (eds), 1997, *Theoretical Issues in Stimulus–Response Compatibility* (Amsterdam: North-Holland).

KORNBLUM, S., HASBROUCQ, T. and OSMAN, A., 1990, Dimensional overlap: cognitive basis for stimulus–response compatibility — a model and taxonomy. *Psychological Review*, 97, 253–70.

LU, C.-H. and PROCTOR, R.W., 1995, The influence of irrelevant location information on performance: a review of the Simon and spatial Stroop effects. *Psychonomic Bulletin and Review*, 2, 174–207.

PROCTOR, R.W. and REEVE, T.G. (eds), 1990, *Stimulus–Response Compatibility: An Integrated Perspective* (Amsterdam: North-Holland).

PROCTOR, R. W. and WANG, H. (1997), Differentiating types of set-level compatibility. In B. Hommel and W. Prinz (eds), *Theoretical Issues in Stimulus–Response Compatibility* (Amsterdam: North-Holland), 11–37.

SIMON, J. R. (1990), The effects of an irrelevant directional cue on human information processing. In R. W. Proctor and T. G. Reeve (eds), *Stimulus–Response Compatibility* (Amsterdam: North-Holland), 32–86.

STOFFER, T. H. and UMILTÀ, C. (1997), Spatial stimulus coding and the focus of attention in S–R compatibility and the Simon effect. In B. Hommel and W. Prinz (eds), *Theoretical Issues in Stimulus–Response Compatibility* (Amsterdam: North-Holland), 181–208.

UMILTÀ, C., and NICOLETTI, R. (1990), Spatial stimulus–response compatibility. In R. W. Proctor and T. G. Reeve (eds), *Stimulus–Response Compatibility* (Amsterdam: North-Holland), 89–116.

UMILTÁ, C. and ZORZI, M. (1997), Commentary on Barber and O'Leary: Learning and attention in S–R compatibility. In B. Hommel and W. Prinz (eds), *Theoretical Issues in Stimulus–Response Compatibility* (Amsterdam: North-Holland), 173–78.

The Substance of Cognitive Modeling

E. Hollnagel

Linköpings Universitet, S-581, 83 Linköping, Sweden

1. INTRODUCTION

One of the hard problems in the study of human cognition is how it should be modeled. While it is taken for granted that models are necessary in the study of cognition, just as in every other field of science, there are two issues of the model problem that need to be addressed. One concerns *what* should be modeled; this is the issue of the substance of modeling. The other concerns *how* the models should be expressed, i.e. the language of modeling, or the technical issue. This article only discusses the former.

The interest in the modeling of cognition goes beyond the common use of models as a tool of science, and can probably best be understood as a backlash to the positivistic view of classical behaviorism. The modern study of cognition is intrinsically linked to the invention of the digital computer and the emergence of scientific disciplines such as cybernetics and information theory, and in large parts derives its legitimacy from those. Within this tradition it has long been the accepted norm that the modeling of human performance requires the modeling of the inner, mental processes and knowledge of the user — in the shape of the ubiquitous user model. The merits of this view notwithstanding, during the 1990s it has been questioned whether cognition should be described primarily as a mental process, and in particular whether it can be considered as a context-free mental process. The aim of this article is to investigate this view, as well as its more recent alternative, and propose a basis for resolving the substance issue.

2. COGNITION WITHOUT CONTEXT

The notion that cognition could be modeled as specific processes or functions without raising the specter of mentalism goes back to the very beginning of the time when computers started to become commonly known. In a seminal paper, Edwin Boring (1946) described a five-step program for how the functions of the human mind could be accounted for in a seemingly objective manner. First, the functional capacities of the human should be analyzed, e.g. by listing the essential functions. Second, the functional capacities should be translated into properties of the organism by providing a description of the input-output characteristics (essentially a black-box approach). Third, these functions should be reformulated as properties of a hypothetical artifact, which in modern terms means expressing them as operational descriptions for a computer system, e.g. in the form of flow charts or programs. The fourth step was to design and construct actual artifacts, which is equivalent to programming the functions in detail and running a simulation. The fifth and final step was to explain the workings of the artifact by known physical principles. This would serve finally to weed out any mentalistic terms and capacities.

Boring believed that the ultimate explanation should be given in terms of psychophysiology, so that "an image is nothing other than a neural event, and object constancy is obviously just something that happens in the brain". Although written more

than 50 years ago, the approach is in all essentials identical to the principles of information-processing psychology, as it became popular in the mid-1970s. The essence of this view was that cognition should be studied and understood as an inner — mental — process rather than as action. More particularly, cognition was explained in terms of more fundamental processes, hence in all essence treated as an epiphenomenon of human information processing. This idea of context free cognition was promoted by people such as Herbert A. Simon, who argued very convincingly for the notion of a set of elementary information processes in the mind. One consequence of this assumption was that the complexity of human behavior was due to the complexity of the environment, rather than the complexity of human cognition (Simon 1972). This made it legitimate to attempt to model human cognition independently of the context, which effectively was reduced to a set of inputs. In the same manner, actions were reduced to a set of outputs, and the inputs and outputs together represented the direct interaction with the context.

As the main interest of human information processing was to model cognition per se, the chosen approach corresponded well to the purpose. Cognitive ergonomics and cognitive systems engineering, on the other hand, are rather more interested in developing better ways of analyzing and predicting human performance. This purpose cannot be achieved by the information processing types of models, and therefore requires an alternative approach.

3. THE LANGUAGE OF COGNITION

The language of cognition, as the term is used here, does not refer to descriptions of how cognition is assumed to take place in the mind or the brain, i.e. the neurophysiological or computational processes to which cognition allegedly can be reduced. It refers rather to the terminology that is used to describe cognition and its role in human behavior. The language of cognition is important because the terms and concepts that are used determine both which phenomena come into focus and what an acceptable explanation is. This is illustrated by the classical work in the study of cognition (e.g. Newell and Simon 1972), which used astute introspection in well-controlled conditions to try to understand what went on in people's heads, predicated on the notion of information processes. While many of the features of cognition that have been found in this way are undeniably correct on the phenomenological level, their description often implies assumptions about the nature of cognition that are ambiguous, incorrect, or unverifiable. This can be illustrated by a small example.

In the daily use of the language of cognition these assumptions are easily forgotten, and the reality of the underlying mechanisms or concepts, such as human information processing, is taken for granted and hence rarely questioned. Consider, for instance, the fact that it is sometimes difficult to recognize people that one knows. In one situation a face may be recognized as familiar, yet it may be impossible to recall the context. In another, a person may be fully recognized, yet the name cannot be recalled. On the other hand, it rarely happens that someone is recognized as familiar and that the name can be recalled, but that nothing else is about the person comes to mind.

The language of cognition determines how this phenomenon is explained. In one case it can be described as a failure to retrieve

the name of the person from long-term memory, and the explanation is that the brain stores information about people's names separately from all other information about them. This explanation implies that there are a number of separate memories, and that the recognition of a person goes through a number of steps, starting by seeing the face and ending by retrieving the name. In another case the phenomenon can be described as the difficulty in remembering people's names. The explanation is, in this case, that names are hard to remember because they are generally meaningless. This explanation does not require an elaborate theory about human information processing, nor a mental model, but simply states a fact (although it also requires an independent definition of what "meaningless" is).

4. COGNITION IN CONTEXT

Since the late 1980s the scientific disciplines that study cognition — predominantly cognitive science, cognitive psychology, and cognitive engineering — have increasingly emphasized the relation between context and cognition. This has been expressed in a number of books, such as Hutchins (1995) and Klein *et al.* (1993). The essence of this "new look", which has been referred to by terms such as "situated cognition", "natural cognition", and "cognition in the wild", is:

(1) that cognition is not confined to a single individual, but is rather distributed across multiple natural and artificial cognitive systems;

(2) that cognitive activity is not confined to a short moment as a response to an event, but is rather as a part of a stream of activity;

(3) that sets of active cognitive systems are embedded in a social environment or context which constrains their activities and provides resources;

(4) that the level of activity is not constant but has transitions and evolutions; and

(5) that almost all activity is aided by something or someone beyond the unit of the individual cognitive agent, i.e. by a tool.

Many people have seen this development as a significant step forward, although the enthusiasm has not been the same on both sides of the Atlantic. Yet while it is praiseworthy that the study of cognition at long last acknowledges that cognition and context are inseparable, it should not be forgotten that "situated cognition" is far from being something new. As long ago as 1976, Ulrich Neisser (1976: 8) wrote that:

(w)e may have been lavishing too much effort on hypothetical models of the mind and not enough on analyzing the environment that the mind has been shaped to meet.

And a few years later Donald Broadbent (1980: 117) echoed the same view when writing that:

one should (not) start with a model of man and then investigate those areas in which the model predicts particular results. I believe one should start from practical problems, which at any one time will point us towards some part of human life.

After many years of gradually moving down a cul-de-sac of human information processing, it is easy to see the shortcomings of this approach to the study of cognition. It is less easy to see what the problems are in the alternative, since its power to solve — or rather dissolve — many of the difficult problems is deceptive. Although it was a mistake to assume that cognition

could be studied without considering the context, it is equally a mistake to assume that there is a difference between cognition in context, i.e. in natural situations whatever they may be, and context-free cognition. The methods of investigation may be widely different in the two cases, but this does not warrant the assumption that the object of study — cognition — is different as well.

The hypothetico-deductive approach preferred by academic psychology emphasizes the importance of controlled conditions, where independent parameters can be varied to observe the effect of predefined dependent variables. This classical approach is often contrasted to the so-called naturalistic studies, which put the emphasis on less controlled but (supposedly) more realistic studies in the field or in near-natural working environments. It is assumed that the naturalistic type of study is inherently more valid, and that the controlled experiments run the risk of introducing artifacts and of studying the artifacts rather than the "real" phenomena.

Many of the claims for naturalistic studies are, however, inflated in a misguided, but understandable, attempt to juxtapose one paradigm (the "new") against the other (the "old") and to support the conclusion that the "new" is better than the "old". Quite apart from the fact that the "new" paradigm is not new at all (e.g. Brunswik 1956), the juxtaposition disregards the essential reality that all human performance is constrained, regardless of whether it takes place under controlled or naturalistic conditions. Given any set of conditions, whatever they may be, some activities are more likely than others – and indeed some may be impossible under given circumstances. In the study of firefighters during actual fires, the goals (e.g. to put out the fire as quickly as possible) and the constraints (resources, working conditions, command and control paths, roles and responsibilities, experience, etc.) will, to a large extent, determine what the firefighters are likely to do, and how they will respond to challenges and events. In the study of firefighters using, for instance, a forest fire simulation game, there will be other goals and constraints, hence a different performance. The difference between the controlled and the naturalistic situations is not the existence or the reality of the constraints, but rather the degree to which the constraints are predefined and controllable. In both cases the performance will be representative for the situation, but the two situations may not be representative of each other.

The hypothetico-deductive approach requires the conditions of a controlled experiment in order to succeed and the *ceteris paribus* principle reigns supreme, although it is well known that this is a strong assumption that rarely is fulfilled. The degree of control is often less than assumed, which may lead to problems in data analysis and interpretation. The important point is, however, to realize that all human performance is constrained by the conditions under which it takes place, and that this principle holds for "natural" performance as well as controlled experiments. For the naturalistic situation it is therefore important to find the constraints by prior analysis. If that is done, then we have achieved a degree of understanding of the situation that is similar to our understanding of the controlled experiment, although we may not be able to establish the same degree of absolute control of specific conditions (e.g. how an event begins and develops). Conversely, there is nothing that prevents a "naturalistic" approach to controlled studies, as long as the term

is understood to mean only that the constraints are revealed by analyzing the situation rather than by specifying it in minute detail. Possibly the only thing that cannot be achieved in a controlled setting are the long-term effects and developments that are found in real life.

5. MENTAL MODELS AND THE LAW OF REQUISITE VARIETY

One answer to the substance issue is provided by the Law of Requisite Variety, which was formulated in cybernetics in the 1940s, and 1950s (Ashby 1956). This law is concerned with the problem of regulation or control and expresses the principle that the variety of a controller should match the variety of the system to be controlled. The latter usually is described in terms of a process plus a source of disturbance. The Law of Requisite Variety states that the variety of the outcomes (of a system) can only be decreased by increasing the variety in the controller of that system. Effective control is therefore not possible if the controller has less variety than the system.

It is consistent with the interests of cognitive ergonomics and cognitive systems engineering, that the study of cognition should focus on problems which are representative of human performance, i.e. which constitute the core of the observed variety. The implication is that the outcome of the regularity of the environment is a set of representative ways of functioning, and that these should be investigated rather than performances that are derived only from theoretical predictions or from impoverished experimental conditions. With regard to the modeling of cognition the substance should thus be the variety of human performance as it can be ascertained from experience and empirical studies — but, emphatically, not from theoretical studies. The requirement to the model is therefore that model variety is sufficient to match the observed variety of human performance. The model must, in essence, be able to predict the actual performance of the user to a given set of events under given conditions.

The difference between the observed variety and the theoretically possible variety is essential. The theoretically possible variety is an artifact, which mirrors those assumptions about human behavior and human cognition that are inherent in the theory. The theoretically possible variety may therefore include types of performance that will not occur in practice — because the theory may be inadequate or because of the influence of the working conditions. (It follows that if the working conditions are very restricted, then only a very simple model is needed. This principle has been demonstrated by innumerable experimental studies.) If research is rooted in a very detailed theory or model we may at best achieve no more than reinforcing our belief in the theory. The model may fail to match the observed variety and may very likely also be more complex than strictly needed, i.e. the model is an artifact rather than a veridical model in the sense of being based on observable phenomena.

The requirements to modeling of human cognition should be derived from the observed variety, since there is clearly no reason to have more variety in the model than needed to account for the observed variety. The decision about how much is needed can therefore be based on the simple principle that if the predicted performance sufficiently well matches the actual performance, then the model has sufficient variety. This furthermore eliminates the affliction of model validation. The catch is, of course, in the meaning of "sufficiently" which, in any given context and for any given purpose, must be replaced by a well-defined expression. Yet even if this problem may sometimes be difficult to solve, the answer to the substance issue should definitely be found by asking what cognition *does*, rather than what cognition *is*.

REFERENCES

ASHBY, W.R., 1956, *An introduction to cybernetics* (London: Methuen).

BORING, E.G., 1946, Mind and mechanism. *The American Journal of Psychology*, 59(2), 173–92.

BROADBENT, D.E., 1980, The minimization of models. In A.J. Chapman and D.M. Jones (eds), *Models of man* (Leicester: British Psychological Society).

BRUNSWIK, E., 1956, *Perception and the representative design of psychological experiments* (Berkeley: University of California Press).

HUTCHINS, E., 1995, *Cognition in the wild* (Cambridge, MA: MIT Press).

KLEIN, G.A., ORAMASU, J., CALDERWOOD, R., and ZSAMBOK, C.E., 1993, *Decision making in action: Models and methods* (Norwood, NJ: Ablex).

NEISSER, U., 1976, *Cognition and reality* (San Francisco: W. H. Freeman).

NEWELL, A. and SIMON, H.A., 1972, *Human problem solving* (Englewood Cliffs, NJ.: Prentice-Hall).

SIMON, H.A., 1972, *The sciences of the artificia.* (Cambridge, MA: MIT Press).

Tolerance to Shiftwork

I. Iskra–Golec

Institute of Management, 13 Pilsudski Street, Jagiellonian University, 31–110 Krakow, Poland

1. INTRODUCTION

Humans seem to be the only diurnal species requested to be active at night and to sleep during the day at certain circumstances, i.e. when they work shifts including night work. Humans evolved in the rhythmic environment resulting from planetary movements (e.g. light/dark cycles). This lead to the development of internal mechanism (a body clock) producing biological rhythms of body processes and behavior (e.g. circadian rhythms) to prepare and enable humans to function in the rhythmically changing environment. The social environment exhibits rhythms structured according to daytime working hours and night time rest. How shift working people react to the displacement of natural sleep–wake pattern in the day oriented pattern of activity and rest of the society and what are the factors which determine their reaction?

2. MODELS OF SHIFTWORK–SHIFTWORKER RELATIONSHIP

There have been developed two groups of models explaining the relationship between shiftwork and working individual; simple, linear ones founded in stimulus based stress theory and more complex models stemming from transactional framework of stress (Colquhoun *et al.* 1996; Härmä 1993).

The first models focus on environmental factors considering the worker as a passive recipient of stress. They regard shiftwork as disturbing environmental factor (exposure to phase shifting of sleep–wake cycle relative to its natural phase) which causes strain in the worker manifesting itself in complaints and health problems. Between stress and strain mediate intervening variables including working conditions, housing, family situation, and personality moderating the strain.

The models of second group take into account apart of biological clock, sleep and social factors also more environmental variables and shiftworker's cognition and behavior emphasizing multi-directional links between them. The most advanced transactional model regards shiftworker as active agent engaged in two-way relationship with the environment. Environmental demands and the workers' capabilities to deal with or manage are cognitively appraised with stress appearing when the demands burden or surpass the capabilities of the worker. Then coping strategies are adopted in response to the judgements that can alter the environment directly and/or can change appraisal process to alter the stressfulness of the situation.

3. CONCEPT OF TOLERANCE TO SHIFTWORK

In the literature an understanding of tolerance to shiftwork is founded on stimulus-based theories of stress. The notion of shiftwork tolerance is often used without any explicit explanation. Judging from the measures used tolerance to shiftwork is assessed according to self-reported complaints concerning ill-health and sleep disturbances (Taylor *et al.* 1997). The only definition of tolerance to shiftwork explains the concept by its reverse, i.e. clinical intolerance to shiftwork, which refers to existence and intensity of a set of following medical complaints: sleep disturbances, persisting fatigue, alteration of mood, digestive troubles and regular use of sleeping pills (Rainberg and Smolensky 1994: 246). Such a conceptualization has been called "biologically restricted perspective of tolerance" and criticized for neglecting psychosocial consequences of shiftwork, and their relations to and interactions with biological shiftwork effects (Nachreiner 1998).

Intolerance to shiftwork has been suggested to be a process following a certain temporal pattern. It is assumed to develop in interactions between health and other factors (biological clock, working and family situations) through four phases from "adaptation" through "sensitization" and "accumulation" to "manifestation" when the diseases and disorders appear.

4. FACTORS INFLUENCING TOLERANCE TO SHIFTWORK

4.1. Shift Systems Parameters and Tolerance to Shiftwork

The shift systems in operation seem to comfort the technological and economical needs of organizations rather than well-being of the workers employed. There is a body of evidence showing a potential of different ways of work scheduling for facilitating tolerance to shiftwork. It should be noted, however, that there is no ideal solution in designing shift systems (apart from excluding night shift) since systems facilitating tolerance in one sphere at the same time neglect the other.

4.1.1. Shift rotation

From chronobiological perspective the best, i.e. the least disruptive for circadian rhythms, is shift system that maximizes or minimizes circadian adjustment. Permanent night shift systems have been assumed to maximize circadian adjustment but when the worker can keep night orientation of circadian rhythms (sleep during the day and be active at night) also during the days off. This, however, would isolate him from the rest of the day-oriented society and this way would be disruptive for his family and social life.

4.1.1.2. *Speed of rotation*

Rapidly rotating shift systems (one, two or three shifts of the same kind in a row) allow circadian rhythms to remain day oriented. This way a continuous state of partial circadian adjustment to and from night work characteristic of slower rotating systems (resulting from insufficient time for the body clock to adjust) is avoided. However, low night time alertness and poor day time sleep are the disadvantages of circadian rhythms day orientation. The poor day time sleep adds to the average sleep loss per day which is the largest in rapidly rotating systems but a short period of night shifts prevents from accumulation of sleep deficit which tends to increase with a number of consecutive night shifts. Rapid shift rotation is also less disruptive for social life since a few free evenings in such systems allow for frequent and regular contacts with family and friends.

4.1.1.3. *Direction of rotation*

There is some support from the research that forward rotating shift systems (order of shifts: morning, evening, night) are

associated with fewer sleep problems, lower fatigue and less social disruption than backward rotating ones (order of shifts: night, evening, morning). Some chronobiologists explain that by a natural tendency of the body clock to run slow (in the absence of time cues its period is > 24 h) what makes people adjust better to delayed hours of sleep induced by forward rotation than to advanced what induces backward rotation. The others argue that backward rotation provides good opportunity to prepare for and to recover from the night shift and that not the work schedule limitations but social demands tend to determine shiftworker's sleep. It has been recently suggested that not the direction of rotation *per se* but "quick returns" (short breaks, e.g. 8 h between the end of one shift and start of the next) common for many backward rotating systems might be responsible for problems reported by shiftworkers.

4.1.2. Duration of work periods

There are not consistent research findings on the effect of extended shift duration (9–12 h) but the opinion that longer shifts are associated with accumulation of fatigue and reduced alertness especially on night shift in difficult working conditions predominates. On the other hand longer shifts are usually followed by longer time spans off providing an opportunity for recovery, social contacts and leisure pursuits so they are popular among shiftworkers. The number of consecutive working shifts in succession seems to be important due to accumulation of fatigue especially among the older workers so 5–7 working days before days off has been suggested as optimal.

4.1.3. Timing of shifts

Early start of morning shift seems not to facilitate shiftwork tolerance since it is associated with reduced sleep (early waking up time) and increased fatigue. In contrast, an early start of morning shift means an early end of night shift allowing earlier onset of day time sleep and its longer duration. The best solution regarding timing of shifts seems to be flexible working time arrangements allowing individual choice of starting and finishing shifts in "time autonomous groups" that experimental implementing has been found to reduce fatigue and complaints about social and family life among shiftworkers.

4.1.4. Distribution of leisure time within the shift system

It has been evidenced that not only the amount but also the positioning of free time within the week is important for shift work tolerance. Short intervals between two consecutive shifts (e.g. 8 h) reduce sleep substantially and increases the fatigue of shiftworkers. Days off during weekends has been suggested to facilitate social life of shiftworkers due to traditional concentration of social events on weekends and 2 days in succession have been advised for recovery from the effects of socializing (e.g. reduced sleep).

4.2. Tolerance to Shiftwork and the Individual

People differ in their ability to cope with shiftwork. It is assumed that relatively stable individual differences between shift workers (e.g. in age, personality) may modify shift work consequences they experience. At present there is not enough knowledge to predict what individual characteristics would make people better

tolerating shift work. Instead there is some evidence on associative relationships between individual difference measures and the amount of problems individuals have in dealing with shift work.

4.2.1. Age

Aging is associated with decreasing tolerance to shiftwork with the critical age being ~40–50 years. Sleep complaints predominate all shift work related problems that tend to increase with age. There also has been found slower adjustment of circadian rhythms, sleep–wake pattern and performance to shift work among middle-aged (> 40 years) shift workers when compared with younger ones.

Decrease in shift work tolerance may result from the aging effect on the body clock that induces changes in sleep–wake pattern (e.g. frequent waking episodes, early morning awakenings) and in the other circadian rhythms (e.g. shortened periods, decreased amplitude) what may hinder adjustment to shift work. Usually a variety of health complains appears with age and adds to the difficulty of coping with shiftwork.

4.2.2. Gender

The opinion that any differences between genders in tolerance to shiftwork result from differences in domestic workload rather than from inherent differences in biology seems to be supported by research findings. There have not been found differences between shiftworking men and women in health complaints but in domestic duties and leisure time activities. Excessive household responsibilities not only shorten women's leisure time but first of all their sleep duration. Female night workers with two children have been found to sleep ~9 h per week shorter than unmarried ones. The amount of symptoms of intolerance among shiftworking women is larger than among men up to 50 years of age. After that age, i.e. when children left home and the amount of duties decreased, the amount of symptoms reported by women decreases and is equal to those reported by men.

4.2.3. Circadian factors

It is assumed that ability to cope with shift work may depend on the circadian adjustment to disruption caused by work schedules. Circadian adjustment to shift work has been found to be associated with changes in circadian rhythms parameters including partial phase shift and decrease of rhythms amplitude which vary within the population.

4.2.3.1. Circadian phase

People differ in phasing of their body and behavioral rhythms. Morning people (morning types or larks) have earlier peaking of body temperature and other body circadian rhythms, wake up early in the morning, feel more alert in the morning and fall asleep earlier that does the population as a whole. The opposite is true for evening people (evening types or owls) who get up late, fall asleep late at night and feel more alert later during the day. Definite morning and definite evening types represent a small percentage (5% each) of the population most of which shows such differences to much less extent. There have been found associations between morningness and indices of shift work intolerance but the correlations achieved were low and inconsistent and with no validity to predict shift work tolerance.

Morningness/eveningness may exert its influence on factors associated with shift work tolerance i.e. on the amount of sleep achieved on different shifts. Natural tendency of morning types to wake up and fall asleep early makes them able to get enough sleep before morning shift but at the same time shortens their day sleep after night shift. The reverse is true for evening types. Neither of these types is advantaged in rotating shift systems.

4.2.3.2. Circadian amplitude

People differ with regard to amplitude between maximum and minimum of their daily body temperatures. There is some evidence suggesting that people with wider body temperature range are better able to tolerate shift work but there are also contradictory research findings. Tested in predictive context the circadian amplitude has not shown validity to predict tolerance. Low circadian amplitude has been found in quite healthy subjects without any clinical complaints so it is not recommended a reliable index of tolerance to shift work.

4.2.4. Personality characteristics

There have been found some association between personality characteristics and individual differences in circadian rhythm phasing so attempts have been made to relate the former to the shiftwork tolerance.

4.2.4.1. Introversion/extroversion and neuroticism

The studies focused on personality dimensions of introversion/extroversion and emotional stability/instability (neuroticism). Introverts (subjects not prone to communicate or socialize) when compared with extroverts tend to have an earlier phase position with regard to the circadian rhythms of body temperature, sleepiness and waking and performance. It is probably why circadian adjustment to night work is slightly faster in extroverts than in introverts. Neurotic extroverts exhibit a greater phase adjustment of body temperature in shiftwork situations than neurotic introverts. There has been found a relationship between extroversion and neuroticism with indices of shiftwork tolerance but the correlations were mainly low and inconsistent. Nether their power to predict tolerance have been demonstrated. Moreover, it has been found in longitudinal studies that neuroticism tends to increase with the shiftwork experience. It is why this is no longer considered a determinant of shiftwork tolerance but rather a symptom of intolerance.

4.2.4.2. Flexibility/rigidity of sleeping habits and ability to overcome drowsiness

A concept of circadian type has been developed to find factors determining adjustment to shiftwork. Two main dimensions were established flexibility/rigidity of sleeping habits referring to ease in changing sleep times and ability/inability to overcome drowsiness concerning the ease in throwing off feeling of drowsiness. There are contradictory results on whether these factors are associated with better circadian adjustment but it has been found their relationship with better long term tolerance to shiftwork. When tested for predictive power in controlled longitudinal studies they both exhibited low but reliable and consistent positive correlations with indices of subjective health.

4.2.5. Behavioral factors

The amount of problems in dealing with shiftwork seems to be associated to greater extent with worker's behavior and cognition (e.g. life style, coping strategies) than with relatively stable individual traits. More research both cross-sectional and longitudinal is needed before any firm conclusions can be drawn.

4.2.5.1. Commitment to shiftwork

Shiftworkers willingness and/or ability to structure their life around working hours has been reflected in better adjustment of both sleep and circadian rhythms (e.g. body temperature, alertness) to shiftwork. It is suggested that it can even overweight the eventual unfavorable for shiftwork tolerance effect of personality and circadian characteristics.

Regularity of behavior in some areas (e.g. eating habits, social behavior) has also been suggested to increase adaptation to shiftwork.

4.2.5.2. Physical conditioning

Physical fitness has been found to facilitate ability to cope with shiftwork. Regular exercising has been associated with increased general fitness, longer self-reported sleep, higher alertness and lower general fatigue during night work but not with better circadian adjustment to shiftwork. Such an influence of physical conditioning may stem from the sleep inducing and improving effect of exercise and/or increase in general strength.

4.2.5.3. Coping strategies

Ineffective strategies i.e. disengagement coping strategies (attempting to avoid or ignore sources of stress) and passive style of coping have been found to be associated to worsening of well-being.

4.3. Tolerance to Shiftwork and Family, Social and Situational Circumstances

4.3.1. Family

Spouses and children tend to both facilitate and hinder tolerance to shiftwork. Existence of spouse or partner seems to facilitate shiftwork tolerance by satisfaction from family interactions. Family acceptance of shiftworking is important for worker's well-being. Wives brought up in families with shiftworking parents tend to accept this work better than others. However, shiftworking seems to hinder fulfilling roles of a partner and a parent since higher divorce rate and lower children school achievements have been found in shiftworkers when compared with dayworkers. Children can make coping with shiftwork more difficult especially for women but also for men. The day sleep after night shift of shiftworking women has been shown to be of the shortest duration among women with small children and of the longest among women without children. The noise made by shiftworkers' own children is one of the most common reasons of day sleep interruptions.

4.3.2. Community

Shiftworkers living in areas where there is a tradition of working shifts seem to tolerate it better than those living in areas where shiftworking is an exception. Working hours of local community members may have an influence of shiftworkers social life since they tend to interact socially with one another rather than with others.

4.3.3. Living conditions and other situational factors

Shiftworkers living in unfavorable living conditions (e.g. unisolated sleeping accommodation, near roads with heavy traffic) complain more about lowering of well-being than those in more favorable ones. In developing countries the health effect of night work tend to be aggravated by bad working conditions, tiring climate conditions, poor and cramped housing and insufficient food.

5. DIRECTIONS OF INTERVENTIONS FACILITATING TOLERANCE TO SHIFTWORK

A number of factors with multidimensional links between them influencing tolerance to shiftwork implies a necessity of complex and multifaceted interventions aimed at its promotion. Research results founded on stimulus based approach to shiftwork stress show one direction of intervention. Emphasizing a potential in proper environment shaping they provide premises for designing shift systems facilitating tolerance to shiftwork. They also recommend regular medical checks and medical criteria for shiftworkers selection. Focus on relatively stable individual traits of this approach proved not to be fruitful at least at the present state of knowledge. Lack of predictive relationship between almost all relatively stable individual traits with indices of tolerance to shiftwork makes impossible any effective strategy of selecting suitable for shiftwork individuals. Transactional approach emphasizing shiftworkers' behavior and cognition suggest second direction of interventions providing premises for facilitating tolerance via enhancing worker's coping strategies by education programs and counseling.

REFERENCES

COLQUHOUN, W.P., COSTA, G., FOLKARD, S. and KNAUTH, P., 1996, *Shiftwork. Problems and Solutions* (Frankfurt am Main: Peter Lang).

HÄRMÄ, M., 1993, Individual differences in tolerance to shiftwork: a review. *Ergonomics*, 36(1–3), 101–9.

NACHREINER, F., 1998, Individual and social determinants of shiftwork tolerance. *Scandinavian Journal of Work and Environmental Health*, 24 (suppl. 3), 35–42.

REINBERG, A.E. and SMOLENSKY, M.H., 1994, Night and shift work and transmeridian and space flights. In Touitou, Y. and Haus, E. (eds) *Biologic Rhythms in Clinical and Laboratory Medicine* (Berlin: Springer), pp. 243–53.

TAYLOR, E., BRINER, R.B. and FOLKARD, S., 1997, Models of shiftwork and health: an examination of the influence of stress on shiftwork theory. *Human Factors*, 1, 67–82.

Visual Display Terminals: Age and Psychophysiology

M. Kumashiro

Department of Ergonomics, University of Occupational and Environmental Health, 1-1 Iseigaoka, Yahatanishi-ku, Kita-Kyushu 807–8555, Japan

1. DEFINITION OF AN AGING WORKER

Aging refers to the manner in which the functions, morphology and other features of living organisms change over time. The period of the process extends from birth through growth to decline and death. The World Health Organization's Study Group on Aging and Work Capacity (WHO 1993) defines the range of chronological age in an "aging worker" as 45 years and older. Similarly, Japan's Ministry of Labor defines persons 45 years or older as middle- to older-aged workers, and persons 55 years and older as older-aged workers, based on a special law to promote the employment of middle- to older-aged persons.

2. PRINCIPLE OF AGING AT THE WORKPLACE

In investigating the aging phenomenon among middle- to older-aged people active in the workplace, remember that the aging worker has two distinct aspects that work in a negative way, and a synergistic capacity that works in a positive way. When evaluating work ability as a person ages, note that synergistic capacity has a large influence in the workplace, whereas the negative aspects may be minimized depending on the type of job and circumstances involved. General remarks about the aging worker fall into three broad categories:

1. Physiological functions enter decline earlier than mental faculties. A general decline in physiological functions begins in the 30s age bracket, accelerating rapidly in the 40s.
2. The decline in muscular strength begins in the legs then proceeds to the upper body via the hips, arms and hands.
3. Artificially acquired capabilities and functions do not wane even after the decline of vital functions with aging. In addition, the more one uses these acquired capabilities — skills, knowledge, judgment, etc. — that have been cultivated through job experience, the more one can continue to use them into old age.

3. EFFECTS OF VDT–BASED WORK ON THE OPERATOR

The effect of visual display terminals (VDTs) on the health of operators has been under study from various angles since the mid-1970s. The findings of this research have brought to light the following six problems related to VDT-based work:

1. Impact on visual function: irritated eyes, blurred vision, burning eyes, and eye strain.
2. Musculoskeletal impact: painful or stiff neck and/or shoulders, and back pain.
3. Mental fatigue, stress and psychosomatic symptoms.
4. Nonionizing and ionizing radiation emissions.
5. Skin disorders linked to the combined effect of airborne chemical contaminants indoors and static electricity generated by low humidity and VDTs.
6. Health risks from the electromagnetic radiation emitted by many VDTs or large-screen VDT equipment in the workplace.

There are a number of methods that can relatively easily resolve the problems related to the effects of VDTs on the human body (items 1, 2, 4, 5) by improving the work environment, redesigning a workstation, and providing an appropriate schedule of work and rest breaks.

However, the effect on the psychoneurotic system — the problem of computerized work and mental stress—remains difficult to resolve.

Furthermore, studies into the effect of electromagnetic radiation on health have yet to establish clearly defined standards of exposure, leaving serious questions that remain to be answered.

4. VISUAL FUNCTION MOST IMPACTED BY AGING

The key feature of computer-based work is the act of viewing the display screen. Cathode ray tubes (CRTs) and flat panel displays (FPDs) are now in widespread use, and the viewing of luminous light is common to both. This is the most significant cause underlying the increasing frequency of subjective symptoms of the eye.

Close observation of an image requires the accommodating ability of the eye to focus light on the retina. However, this is difficult if the characters (letters) or symbols are either luminous themselves or displayed on a luminous background: this is the most significant cause of eye fatigue. Furthermore, it is believed that the eye blinks less frequently during close-range viewing, thereby drying the cornea and causing eye pain. An attempt was therefore made to identify the characteristic features of visual function unique to aged persons.

Earlier work indicates that aged individuals develop an impaired pupillary reflex and opacity of the optic media (Carter 1982, Pitts 1982, Reading 1968), thereby experiencing degradation of light sense, image focusing on the retina, color perception, and glare tolerance, which further impairs the capacity for far vision (Weale 1982), contrast sensitivity (Owsly *et al.* 1983), visual search (Plude and Hoyer 1981), and graphic recognition (Snellen and Sterings, 1924).

Table 1 shows a general overview of aging and changes in visual function from the standpoint of operational efficiency based

Table 1. Changes in visual function closely linked to the aging process

Phenomenon 1	Decline of visual acuity: although both static and kinetic visual acuity are in decline, the deterioration in kinetic visual is especially prominent; night vision also declines
Phenomenon 2	Decline in peripheral vision functions (kinetic field of vision): aging has little effect of the static field of vision
Phenomenon 3	Decline in focusing functions: lengthened focal distances, longer time for focusing
Phenomenon 4	Decline in ocular motion capacity: decline in the speed of ocular tracking motion, difficulty in moving focus

Table 2. Countering the degenerative effects of aging on visual function.

Countermeasure 1	The vision of older persons under bright lighting conditions is good compared with that of younger persons, and the rate of improvement of vision is also superior — adjustment of lighting conditions recommended
Countermeasure 2	Under static vision conditions, improvement is possible with corrective lenses
Countermeasure 3	Limit the field of vision of each eye to within 120° laterally
Countermeasure 4	Adjust the working vision distance

on this knowledge and Table 2 shows a basic set of counter-measures (Kumashiro 1997).

5. AGED PERSONS AND VDT–BASED TASKS USING A LARGE SCREEN

A large screen is useful in presenting a large amount of information at once, and it also has the advantage of facilitating viewing by aging workers. Yet large screens do have drawbacks regarding their effect on the health of users.

Some recent findings related to workloads were obtained from an experiment in which older and younger subjects performed a VDT-based simulation of an instrument monitoring operation and a fuel-control operation representing work within a human–computer system. This task required the dual capacity to input and to sense an external input stimulation largely through visual processes, as well as the ability to make an instantaneous evaluation as a psychophysiological reaction.

5.1. Speed and Accuracy

The response time was affected markedly by aging. Compared to the young group, the aging group required a significantly prolonged response time for an instrument resetting operation, and the incidence of overlooking malfunctioning instruments increased significantly. However, compared with a work setting using 17-inch video displays, functional differences between age groups were less evident when using 110-inch displays, which suggests that increasing the display screen size for VDT-based operations might reduce the burden of work on aging persons and contribute to improved work safety.

5.2. Physiological Effects

In order to predict the effect of aging on cardiac autonomic function, heart rate variation (HRV) was used as a guideline, and this showed a distinct difference between the aging group and the young group. In addition, compared to the young group, the aging group exhibited an apparent reduction in heart rate reactivity to the loaded task. More specifically, for either HRV parameter, the rest-hour value was significantly small in the aging group compared to the young group, and no marked change was observed between the rest-hour value and the value measured during the task. Such heart rate reactive reduction with aging suggests impaired homeostatic control in response to the load of a mental task. In that regard, an aging person may be more susceptible than a young person to the risk of cardiovascular disease in response to the load of a mental task.

Next, an apparent difference was found between the two age groups in the ability of accommodation. Furthermore, no change was observed in the ability to accommodate in response to shifts in optical status in the aging group. Deterioration in the ability of an aging person's eyes to accommodate may not only lead to work safety concerns — stemming from difficulty in alternately viewing far and near objects and more frequent incidences of visual oversight or outright visual error — but also to an increase in the work burden. Therefore, the ergonomic guidelines for VDT-based tasks, which are highly dependent on the sense of vision, should be reviewed in light of the changes in visual function that naturally occur with aging.

REFERENCES

CARTER A., 1982, Visual acuity as measured by dynamic and static tests. *Journal of Applied Psychology*, 50, 460–66.

KUMASHIRO, M., 1997, Disabilities associated with aging in the workplace and their solutions. In Kumar (ed.) *Perspectives in Rehabilitation Ergonomics* (London: Taylor & Francis), pp. 34–68.

OWSLY C., SEKULER R., and SIEMSEM D., 1983, Contrast sensitivity throughout adulthood, *Vision Research*, 23, 689–99.

PITTS D.G., 1982, Visual acuity as a function of age. *Journal of the American Optometry Association*, 53, 117–214.

PLUDE D.J. and HOYER W.J., 1981, Adult age differences in visual search as a function of stimulus mapping and processing load. *Journal of Gerontology*, 36, 598–604.

READING V., 1968, Disability glare and processing load. *Vision Research,* 8, 207–14.

SNELLEN H. and STERINGS S., 1924, Percentage evaluation of macular vision. *Archives of Ophthalmology*, 54, 443–61.

WEALE R., 1982, Senile ocular changes, cell death, and vision. In Sekuler *et al.* (eds) *Aging and Human Visual Function* (New York: A.R Liss), pp. 161–71.

WHO, 1993, Aging and work capacity. Report of a WHO study Group: WHO Technical Report Series 835 (Geneva: World Health Organization).

Visual Measurement: Modern Methods

T. Takeda

Faculty of Engineering, University of Tokyo, 7-3-1 Hongo, Bunkyo, Tokyo, 113-8656, and Core Research for Evolutional Science and Technology (CREST), Japan Science and Technology Corporation (JST), Japan

1. INTRODUCTION

The importance of visual information has been steadily increasing with the progress of the modern computer information revolution. Hence, simultaneous measurement of three major ocular functions — accommodation, eye movement and pupil diameter — in a real working environments, is being demanded in many fields such as ergonomics, ophthalmology, physiology and psychology. The evaluation of visual fatigue induced by prolonged VDT (visual display terminal) work is one such example. Although researchers have tried to achieve this goal, few have fully succeeded.

US researchers O'Neill and Stark tried to accomplish this task by measuring the deformation of lens curvature under accommodation, the diameter of the pupil of the same eye and the convergence of the other eye. But the measured eye was required to remain fixed on a viewing axis. US researcher Crane *et al.* developed eyetrackers using Purkinje images. Though the eye could rotate several degrees when measuring accommodation, it was not stated how the rotation influenced the accommodation measurement, nor how much rotation was permitted. Hence, the instruments were primarily suited for measurement of eye movement. Japanese researcher K. Kasai developed an optometer that used a spherical mirror to measure accommodation while the eye was moving horizontally, but vertical eye movement was not permitted.

In an initial attempt, Takeda *et al.* developed a refractometer that could measure dynamic accommodation easily by modifying a commercially available auto-refractometer. However, it required subjects to fix their eye direction to coincide with the optical axis. Next, a three-dimensional optometer (TDO) was developed. This new optometer eliminates the mechanical chopper used in the older models (Campbell and Robson 1959) and, hence, is much more reliable. The TDO is based on the principle of Kasai's device, but it also allows the eye to move freely for 40° horizontally and 30° vertically. This feature is achieved by using a light relay system that consists of two spherical mirrors and two servo-controlled mirrors that permit the infrared light used for measurement to enter perpendicular to the cornea. The area of the eye's pupil is computed by using a TV monitor image for alignment. The TDO is remarkably easy to operate: the time for alignment is several minutes for most new subjects, normal lighting is allowed and no drug is required for subjects with a pupil diameter of > 2.9 mm. It has proven its usefulness in measuring the accommodation induced by apparent depth sensation while looking at pictures.

The TDO still has one major restriction in that it requires subjects to hold their head still on a chin rest. It was pointed out that this posture restriction might have some influence on visual fatigue during TDO measurement. Takeda *et al.* have since developed an apparatus, called a three-dimensional optometerIII (TDOIII), which allows free head movement as well as free eye movement. The TDOIII is designed to combine the light relay system of the TDO and the optical component of an auto-refractometer into a single compact system. The optical system is mounted on the subject's head, with its weight counter-balanced so that subjects feel no weight and can move their head smoothly. Head movement is measured by a magnetic measurement device.

It is indispensable to have a visual stimulator that can change each visual function independently and combine them arbitrarily. As there was no such stimulator, Takeda *et al.* (1995) developed a three-dimensional visual stimulator (TVS) that can move targets within a 15° x 16° (height x width) rectangular visual field while changing their distance, direction and size with the aid of a relay lens system.

Recently, several new non-invasive measurement methods for human brain functions have been developed, such as PET (positron emission tomography), fMRI (functional magnetic resonance imaging) and MEG (magneto-encephalography). The MEG system is the most promising for investigating dynamic features of the processing of visual information in the brain because it has high temporal resolution. This chapter explains how a 64-channel whole-cortex MEG system is used to deal with visual response. The combination of an optometer with the MEG system allows a more in-depth study of accommodation by analyzing both ocular responses and brain activity. To carry out such research, a special relay lens system was developed to deliver visual stimuli while measuring the accommodation responses with a dynamic refractometer from outside of a magnetically shielded room. With these apparatuses, MEG and accommodation responses have been recorded simultaneously for the first time. This new approach should be highly effective for acquiring new insights into cerebral visual functions.

2. OPTICAL METHODS

2.1. TDOIII (Three-dimensional OptometerIII)

As the principles of TDO and TDOIII are similar, the principle of the TDOIII will be explained here. Subjects can move their head and eyes freely while head movement and the response of the three major ocular functions are being measured simultaneously. The measurement of accommodation is performed by a Campbell type auto-refractometer (Campbell 1959). Two infrared measurement beams are emitted by IR light-emitting diodes (LED). As their central frequency is 880 nm, the subjects sense no measurement light. The reflected images of the beams on the retina make conjugate images on photo detectors. An inner lens is moved to match the images of the two beams. The amplitude of accommodation is calculated from the distance of the lens displacement. Although the subjects are required to gaze into the apparatus and fix their eye position with an ordinary auto-refractometer, in the TDO they are allowed to move their eyes freely to observe objects by the aid of a relay lens system.

In the TDO, an X–Y tracker in a percept Scope (the first function of C3160, Hamamatsu Photonics) measures eye movement, and galvano-mirrors are driven to maintain the

Table 1.   Characteristics of the TDOIII.

Range: Accommodation −12.7D?+26.6D
Eye movement Horizontal ±20? Vertical −25?? +5? Pupil diameter 0 ? 100 %
Speed: Accommodation 4.7 Hz
Eye movement 6.4 Hz Pupil diameter 6.4 Hz
Resolution: Accommodation ±0.02D
Eye movement Horizontal ±0.18? Vertical ±0.25? Pupil diameter ±0.11 mm
Accuracy: Accommodation ±0.25 D
Eye movement ±0.5? Pupil diameter ±0.3 mm

Table 2.   Characteristics of the MEG system.

Dewar
Weight ?200 kg
Volume of liquid He ?70 l
Holding time ?4 days
Helmet Shape Over 95% of Japanese can be measured.
Sensor Coils
Type 1st Gradient
Diameter 2 cm
Baseline 5 cm
Number 64 channels
Distribution Equally distributed over whole cortex.
Software 1-3 order noise reduction
Filters up to 20th order with no phase shift
DC offset reduction

monitored eye in the center of a CRT by a servo controller in the TDOIII. Thus, the infrared measurement light is guaranteed to be injected into the eye with the aid of the servo control mechanism, irrespective of eye movement. Pupil diameter is measured by the second function of the percept scope (area analyzer). It analyzes the image of the eye on a CRT, and counts pixels brighter than a threshold level. Then the pupil diameter is calculated using the assumption that the pupil is a circle. A microcomputer controls a measurement timing, and takes in the data of the three ocular functions and head movement.

Alignment is easily performed with a three-axis movable chin rest that is driven by small motors. The alignment can be done within several minutes with the aid of a CCD camera located on the optical axis and a CCD camera located just below the eye. It can be performed within 10 s for an immediate repetition on the same subject. The final characteristics of the TDOIII are summarized in Table 1.

2.2. TVS (Three-dimensional Visual Stimulator)
The TVS can change the directions, distances, sizes, luminance and varieties of two sets of targets for both eyes independently (Takeda *et al*. 1995). The optical system of the TVS consists of (1) liquid crystal projectors (LCP) that generate flexible visual targets for both eyes, (2) Badal optometers that change target distances while maintaining constant visual angles and (3) an optical relay lens system whose principle is the same as the one in the TDOIII. The ranges of stimulation are as follows: distances 0–20 D; directions ±16° horizontally and ±15° vertically; sizes 0–2° of visual angle and luminance 10^{-2}–10^2 cd/m². The target images are refreshed at 60 Hz and the speed of smooth change of the target (ramp stimuli) are as follows; distance 5 D/s, direction 30°/s, size 10°/s. Some applications of the TODIII and the TVS are explained elsewhere in this volume.

3. BIOMAGNETIC METHODS
3.1. MEG (Magneto-encephalography)
In the MEG system, the sensor coils have a 5 cm baseline distance and are equally distributed over the head with a mean distance of 4 cm. There are reference coils to measure environmental noise and reduce its influence, which is the most important characteristic of the system (Vrba *et al*. 1982). The whole system

is kept cold constantly with liquid helium. The measured data are pre-amplified and digitized by hardware and stored into a microcomputer. A phantom head (diameter 75 mm) evaluated the noise rejection algorithm. It was evaluated in two conditions varying the number of recordings. The door of the shielded room was closed in condition C and was left open with an active CRT placed at the entrance of the magnetically shielded room in condition O. The results revealed that the noise rejection algorithm was effective in condition O, especially for fewer trials in the average. The data confirm that the noise was reduced to an acceptable level with the door open after averaging 30 recordings or more (Takeda *et al*. 1996). The main characteristics of the whole-head MEG system are shown in Table 2.

3.2. Visual Stimulator
To measure accommodation together with MEG, a special relay lens system was developed that optically transferred the subject's eye image to the outside of the magnetically shielded room. It consisted of four identical spherical lenses with 400 mm focal lengths, moving the eye by 3200 mm to form a real image just in front of the dynamic refractometer where the subject's eye was placed in normal measurements. The accommodation measurement was done with the real image of the eye. The dynamic changes of the eye's focus point were measured by the dynamic refractometer.

4. A MEASUREMENT EXAMPLE
4.1. Subjects and Procedure
Subjects were three right-handed volunteers with adequate accommodative power and no ocular problems other than myopia. The accommodative power was measured with the dynamic refractometer by moving the target from a position located farther than the far point (the farthest point which the subject could focus) to another position nearer than the near point. The amplitude of the stepwise stimuli were set to be ~60% of each subject's accommodative power to prevent excessive visual fatigue. Subjects were instructed to try to watch the target as clearly as possible. The target was changed stepwise with a random time interval of 5 ± 0.5 s.

The measurements were done on the right eye, occluding the left eye with an eye patch. The experiments consisted of > 64 recordings; the collected data were averaged after discarding those contaminated by eye blinks. The sampling rate was set to 250 Hz. The recording was started 1.5 s before the onset of the stimulus and lasted 2.5 s. The evoked fields, the accommodation response and the target position were simultaneously recorded by a personal computer. The trigger signal for the MEG measurements was made at a predetermined threshold of the target position. The MEG signals were band-pass filtered between 0.5 and 40 Hz and the DC offset was removed using a 1.5-s pre-recording.

4.2. Results

The accommodation target was moved between –5 and –8 D, because the subject used was –5 D myopic. Figure 1 shows the accommodation and superimposed MEG responses for the same stimulus together with the target movement (Takeda *et al.* 1996). The time lag of the accommodative response was 288±50 ms. It was calculated by averaging the time lags of the 64 records. An arrow in Figure 1 indicates the beginning of accommodative responses. Though the amplitudes do not match with the stimulus, this is natural and is called the accommodation lag. The mean time lag and the shape of the accommodation responses are in good agreement with the literature.

Superimposed MEG responses indicated well-synchronized activation of the brain ~100 and 200 ms. Both peaks showed a phase reversal phenomenon. On the other hand, the brain signal was rather quiet ~300 ms when the average accommodative response began to rise. Several additional control experiments strongly infer that the second synchronized activity in the MEG comes from brain activity which controls accommodation.

The iso-magnetic field map was a very simple pattern which showed outflow of magnetic field from right hemisphere and inflow in the left hemisphere. From a theoretical and biological consideration, a two-dipole model was adopted and the dipole fit was performed by the Nelder Mead Simplex Method. Consequently, it was found that the control center for accommodation, which had not been identified clearly even in electrophysiological studies with mammals, has been
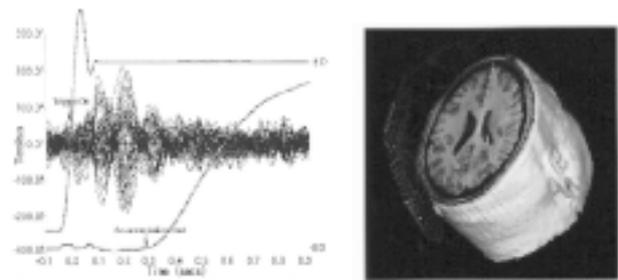
Figure 1. (a) Typical MEG responses, accommodation responses and target change. There were two peculiar synchronized MEG responses ware found before the accommodation response. (b) Locations of the two dipoles near the base of parieto-occipital sulcus, which were identified using the magnetic field at 204 ms after the onset of the stimulus.

identified near the base of the parieto-occipital sulcus, as shown in Figure 1b.

REFERENCES

CAMPBELL, F.W. and ROBSON, J.G., 1959, High-speed infrared optometer. *Journal of the Optical Society of America*, 49, 268–72.

TAKEDA, T. FUKUI, Y. and IIDA, T., 1993, Three-dimensional optometer III. *Applied Optometry*, 32, 4155–68.

TAKEDA, T., FUKUI, T., HASHIMOTO, K. and HIRUMA, N., 1995, Three-dimensional visual stimulator. *Applied Optometry*, 34, 732–8.

TAKEDA, T., MORABITO, M., XIAO, R., HASHIMOTO, K. and ENDO, H., 1996, Cerebral activity related to accommodation: a neuromagnetic study. *EEG Supplement*, 47, 283–91.

VRBA, J. FIFE, A.A. and BURBANK, M.B., 1982, Spatial discrimination in SQUID gradiometers and 3rd order gradiometer performance. *Canadian Journal of Physics*, 60, 1060–73.Figure 1. (a) Typical MEG responses, accommodation responses and target change. There were two peculiar synchronized MEG responses ware found before the accommodation response. (b) Locations of the two dipoles near the base of parieto-occipital sulcus, which were identified using the magnetic field at 204 ms after the onset of the stimulus.

Visual Perception Under Night Driving Conditions

I. Noy

Transport Canada, Road Safety and Motor Vehicle Regulation — Ergonomics Division, 330 Sparks St, Ottawa, ON K1A 0N5, Canada

1. INTRODUCTION

Driving is a complex task, requiring several different processes, vision being the main source of information needed to operate a vehicle. The driver must continuously monitor and respond to the road, traffic, pedestrians, and environment to maintain safe automobile operation. Vision is used to obtain basic information from the driving environment such as where the road is going, one's position on the road, obstacles encountered along the way, and the location and actions of pedestrians and other vehicles. The human eye functions best under high levels of illumination. However, since driving at night is often accompanied by poor illumination, the ability to see is often inadequate.

2. VISUAL PERCEPTION IN NIGHT DRIVING

While a variety of visual factors can affect perception, *luminance contrast* is the principal visual cue at night since colors, surface textures, surface details, and contours are not discernible at low luminance levels. Luminance contrast is the contrast between the brightness of an object and that of its background. It can be positive (bright object on dark background), negative (dark object on bright background) or null. All three types are possible under normal roadway lighting conditions. It is generally preferred to provide positive contrast in busy urban environments as this type of contrast facilitates object recognition.

Luminance is a photometric measure of the amount of light perceived by an observer. It depends on the absolute amount of light energy transmitted, reflected or emitted from the object, the location of the observer in relation to the object, and the relative spectral sensitivity of the eyes. In addition, the sensitivity of the eyes to light energy depends on the level of ambient illumination and their state of adaptation to ambient conditions. (Different light-receptor cells in the retina, rods and cones, are used for day, or photopic vision, and night, or scotopic vision, conditions. A combination of rods and cones is used under dawn or dusk conditions, or mesopic vision.)

The amount of light reflected from an object, such as a pedestrian, depends on the amount of light falling on the object and the reflectance coefficient of the object. The photometric measure of the amount of light falling on an object from all sources is *illuminance*. The amount of light reflected depends on the material and color of the object. Steady-state sources of illumination include luminaries and the moon; erratic sources include headlights from traffic. For a typical pedestrian wearing conventional clothing, a reflectance coefficient of 18% is used.

3. PROCESSING VISUAL INFORMATION

At any given moment a great deal of information is available to the driver. The driver must process the information quickly and make the necessary decisions about which way to proceed. The information detected by drivers is virtually processed automatically, with little conscious awareness or effort. In general the human eyes are attracted to objects of larger size, to bright color, to movement, to objects that differ greatly from their background in terms of color, brightness and size, and to large concentrations of people, signs, lights, cars, etc.

Humans process information in two ways: serially and in parallel. Serial processing refers to dealing with one task at a time. Humans are generally serial processors of information, however, certain overlearned tasks such as eating can be done in parallel with other tasks. Generally, humans are capable of being aware of multiple things at once, yet able to act upon only one thing at a time (Olson 1996). During nighttime conditions, the driver receives and processes information at an even slower rate due to poor visibility and low contrast. Inadequate illumination can also decrease the detection of objects, hence, decrease processing and reaction time. The serial nature of information processing means that an unexpected event must compete with information already being processed to gain the individual's attention. The likelihood of the novel object being processed successfully depends largely on the object conspicuity.

Conspicuity refers to the characteristics of an object or condition that determine the likelihood that the object or condition will capture the observer's attention. Some objects may be present in the visual field but fail to be detected by the observer because it may have less conspicuity than other surrounding features. This is especially true at night, since illumination and contrast of objects is decreased. Nighttime conspicuity of road signs is highly dependent on brightness. When operating a motor vehicle at night, generally only brightness contrast is available. Nighttime conspicuity will increase as sign brightness increases for highly complex backgrounds but not for low complexity backgrounds (Schieber 1998). Headlights and fixed lighting installations solve some of the visibility problems encountered while driving at night; they provide illumination and enhance contrast at night by making the target object appear brighter than its background. In addition, fixed lighting systems may also brighten up the background so that the observer can see the target object in silhouette (Olson 1996).

4. DRIVER SELECTIVE VISUAL DEGRADATION AT NIGHT

Research has shown that most road users, drivers and pedestrians alike, are unaware of the devastating perceptual effects of reduced illumination, low contrast, restricted field of view, and glare. The phenomenon of selective visual degradation at night relates to natural limitations of perception and learning that are not well recognized by road users. For example, at night, drivers can easily see most road signs, road markings, vehicles and dashboard instruments. As a result, they are unaware of visual deficiencies that selectively impair their ability to detect low-contrast objects. Because such hazards are relatively infrequent, drivers have little opportunity to learn of their limitations and, consequently, may not be prepared for a dangerous encounter.

4.1. Factors that Degrade Visual Performance at Night

While many factors can affect visibility while driving at night, fixed lighting installations and vehicle headlights provide much of the illumination. However, the level of illumination provided

by the vehicle's headlights is often insufficient to detect obstacles on the road, especially ones having low reflectance. Fixed lighting installations can also provide adequate illumination especially in conjunction with headlights. However, many streets and highways do not have fixed lighting installations. As a result, the driver is often solely dependent on the vehicle headlamps to provide illumination. There are many factors that degrade visual performance at night, thereby further reducing visibility. Such factors can arise from the environment, the vehicle or the driver.

The atmosphere can greatly affect the driver's ability to detect road obstacles, other vehicles and pedestrians. The most common environmental problems are rain, dust, fog and snow. The result is that light is absorbed and scattered causing less of the light reflected by the object to reach the driver's eyes; and less light from the vehicle's headlamps to reach the target object. In addition, some of the scattered illumination is reflected back into the driver's eyes, causing the atmosphere to appear to light up and target contrast to diminish.

A second type of factor degrading visual performance at night arises from the vehicle and includes the vehicle's windshield and headlamps. The windshield could become badly scratched and pitted with use, reducing visibility and increasing the effects of glare. Headlamps can also be a frequent source of poor nighttime visibility. Out of aim headlamps can increase glare that reduces visibility and can cause discomfort to the driver. In addition, even moderate amounts of dirt on headlamps affect visibility, causing illumination to be scattered and absorbed, often increasing glare to oncoming drivers.

The final factor that can degrade visual performance at night involves problems arising from the vehicle operator. The problems can be temporary, such as fatigue, psychological distress, drugs and alcohol, or permanent, such as aging or night myopia. Nearly every aspects of visual function decreases with age. As a result, driving a night serves as an additional obstacle that can impede safe driving for the elderly.

5. DRIVER EXPECTATION

Driver expectancy relates to the driver's predisposition and readiness to respond to events, situations and information in predictable ways. Driver expectation is a key factor in visibility. When an observer knows what object (i.e. a pedestrian) is up ahead and where it will be encountered, the object can be detected at twice the distance that it will when the observer has no information about the type of object or its location on the road. For example, the key function of road signs is to set driver expectation. When a sign is posted, warning drivers of construction up ahead, the driver can pay close attention to persons working on the road. Hence, adjusting his or her driving and expectations of the road accordingly. Drivers also have preexisting notions of how pedestrians and bicyclist should behave on the road. It is safe to assume that people are generally rational and will not deliberately put their lives in danger. For instance, drivers expect that pedestrians will obey traffic laws, such as walking on sidewalks or crossing the street on a green light, which in turn will largely influence how drivers attend to the road. The importance of preconceived stereotypes and driver expectancy is even more pertinent for drivers under nighttime conditions since visibility is largely reduced.

Much of the emphasis with regard to night driving has been on the physical variables such as headlights, roadways, signs, road marking and delineators. However, psychological variables such as expectancy can have a tremendous impact on visibility. For example, retroreflective tags can provide an improvement in nighttime visibility of pedestrians when drivers expect to see retroreflective tags on pedestrians (Shinar 1985). Retroreflective tags, as safety devices, can be attached to any piece of clothing, thereby increasing the visibility of pedestrians. It is also more likely that pedestrians will wear tags rather than whole clothing pieces to increase their visibility at night since tags can be easily attached and detached from clothing. However, in order for retroreflective tags to serve as safety devices, vehicle operators and pedestrians must be properly informed of the tags purpose and come to associate their presence with that of a pedestrian.

REFERENCES

OLSON, P.L. 1996, *Forensic Aspects of Driver Perception and Response* (Arizona: Lawyers & Judges).

SCHIEBER, F. 1998, Nighttime conspicuity of highway signs as a function of sign brightness, background complexity and age of observer: http://www.usd.edu/~schieber/pdf/conspic.pdf

SHINAR, D. 1985, The effects of expectancy, clothing, reflectance, and detection criterion on nighttime pedestrian visibility. *Human Factors*, 27, 327–33.

Work Ability

M. Välimaa

Tampere University of Technology, Occupational Safety Engineering, BOX 541, Fin-33101 Tampere, Finland

1. INTRODUCTION

Work ability defines an employee's ability to do his work. As a definition work ability is complicated because its character is very individual. It is also not a constant concept but all the time it is changing depending on a close interaction between personal capacities, for example, health, professional skills, motivation and commitment to work, and external factors, like created necessary conditions to do the work and social circumstances.

2. DEFINITION OF WORK ABILITY

Work ability can be considered focusing on the effects of work on employee health or concentrating on an employee's fitness for work. Here questions have been asked about employees' ability to carry out their work in a manner without risk or harm to themselves or others. Such questions have been posed largely within the contexts of medical screening for work, health monitoring of workers in work and rehabilitation (Cox and Cox 1992)

Work ability can also be defined concerning the situation where it has been used. According to Mäkitalo and Palonen (1994) there are three different definitions for work ability, the medical concept, the balanced concept and the integrated concept of work ability.

3. MEDICAL CONCEPT OF WORK ABILITY

Medical concept defines work ability as health, precisely the lack of diseases and injuries. The basis of the medical concept is in the biomedical concept of sickness. Health has been used to refer only to physical health and lack of physical disease, injury or disablement. The degree of work ability and the degree of disability are defined based on that by clinically determined state of psychophysical disorder. According to this concept it is thought that when an individual is healthy (without diseases) he is completely able to work and when sick (without health) partly or completely unable to work. In practice work ability is often defined purely on the basis of clinical examination (injury–disability basis) without information of the demands or possibilities of the employee's environment (Urponen *et al.* 1991).

With regard to work ability in relation to legitimacy for social benefits, for example, disability pensions and sick leave, work ability has traditionally been defined from a very medical point of view (Stone 1984). This is very often true in the assessment of an individual's fitness for work, too.

The assessment of an individual's ability to work is also often done to help the administrative decision-making. Then the measures used and so the contents of work ability are defined by the basis of the administrative and legislative definitions for disability to work. In this connection work ability is strictly a characteristic of individual and the diseases and injuries are considered to be the essential factors affecting work ability (Mäkitalo and Palonen 1994).

The medical concept of work ability does not include the concept of work or the relationship of the individual and his environment. For that reason in the context of the prevention of disability, promotion of work ability and rehabilitation, as well as for assessment of an individual's ability to work, the purely medical concept of work ability is too narrow.

4. BALANCED CONCEPT OF WORK ABILITY

The balanced model of work ability defines work ability as a relationship or balanced state between the recourses of individual workers and the demands of work (Ilmarinen and Tuomi 1992, Järvisalo 1992). This concept is based on the definition of professional disability to work and includes medicine, physiology, psychology and ergonomics. According to the balanced concept of work ability the emphasis is on the capacity of the individual, which is divided into physical, psychic and social capacities (Ilmarinen and Tuomi 1992). Work ability is considered as a pure relationship between adequacy and suitability of these individual capacities and the demands of work. Individual factors affecting work ability go beyond health and are composed of educational ability, skills, experience, motivation to work, work environmental issues such as ergonomics, safety and work hygiene, and organizational factors such as developmental, psychosocial and management issues.

5. INTEGRATED CONCEPT OF WORK ABILITY

Even wider work ability is defined by the integrated concept of work ability. In this concept work ability is a means to describe how an individual copes with work. It forms the concept of work ability for the basis of assessment and adjustment of an individual's competence for work and the promotion of work ability instead of the incomplete fitness for work concept.

Integrated concept of work ability prescribes the contents of work ability by the factors that really are prescribing it in each situation and time. Instead of pure work–need dilemma, work ability is considered as an interaction and entirety formed of individual, community and environment. Besides disabilities also the possibilities and resources of individual, community and environment are taken into consideration. Backgrounds of the integrated concept are in the biopsychosocial human concept as well as in the system theoretic model in which several factors, activities and environmental factors are in a continuous interaction with each other (Mäkitalo and Palonen 1994)

According to the integrated concept, individual resources (like ability to act, professional skills, motivation) or problems related to them are considered necessary to develop and be meaningful in relationship to the structure of an organization, the activities of superiors and fellow workers and the employer's labor and educational policy (Järvikoski *et al.* 1991). Factors of community not only affect work ability but are a part of it (Mäkitalo and Palonen 1994). In this aspect there is also the emphasis on the individual taking an active part in improving his possibilities.

6. DISCUSSION

All factors of work ability act simultaneously and influence each other continuously. Some support it, others weaken it and the current work ability is a state balanced between all these. For that reason the wide aspect of work ability should be considered in order to assess an individual's ability to work properly and to understand disability and the factors behind it.

Individual factors affecting work ability are, for example:

- Health (physical, psychological and social).
- Physical, psychic and social resources.
- Professional skills, education and experience.
- Social and economic situation.
- Motivation and commitment to work.
- Environmental factors include as.
- Organizational culture.
- Management.
- Employer's labor and educational policy.
- Healthiness of the organization, possibilities to develop work community and efficiency, atmosphere and human relationships of work community.
- As well as the values and attitudes of society in relation to a willingness to work, a policy of social benefits.

Work and work environment-related matters are, for example:

- Organizing of work.
- Physical, mental and social strain at work.
- Condition of tools, instruments and environment.
- Safety and health at work and work hygiene.

All these factors are also naturally affected by the economic situation of both the enterprise and the society.

Problem of work ability always includes the question of those who cannot work and those who will not. Ability to work is related to an individual's personal habits and moral choice, how he manages his own life and how it is balanced in relation to all its parts and environmental factors.

The emphasis on work ability issues has traditionally been a little different when concerned with, e.g. society, insurance systems, and the employer and individual worker. Good work ability is a benefit to each part. To achieve this goal the common concept of work ability is significant. Essential is how, based on this common, wide concept, society, insurance systems, employers, as well as individual workers, support the concept of willingness to work and understand that it is possible to create safe and healthy work as a natural precursor of productive work.

REFERENCES

COX, T. and COX, S., 1992, Occupational health: past, present and future. *Work and Stress*, 6, 99–102.

ILMARINEN, J. and TUOMI, K., 1992, Work ability of aging workers. *Scandinavian Journal of Work and Environmental Health*, 18 (suppl. 2), 8–10.

JÄRVISALO, J., 1992, Työterveyteen, terveyteen ja työkykyyn liittyvistä käsitteistä. In *Työkykyä ylläpitävä toiminta ja työterveyshuolto* (Helsinki: Sosiaali-ja terveysministeriö, työterveyshuollon neuvottelukunta), 84–113.

JÄRVIKOSKI, A., TOSSAVAINEN, A. and RISSANEN, H., 1991, Työkyvyn aleneminen ilman sairautta? Keskustelualoite varhaiskuntoutuksen käynnistämisperusteista. *Kuntoutus*, 3, 31–40.

MÄKITALO, J. and PALONEN, J., 1994, Mitä on työkyky: lääketieteellinen, tasapainomallin mukainen ja integroitu käsitystyyppi. *Työ ja ihminen*, 3, 155–62.

STONE, D., 1984, *The Disabled State* (Philadelphia: Temple University Press).

URPONEN, K., HOKKANEN, L., KINNUNEN, P. and YRITYS, E., 1991, *Työkykyarvion ja kuntoutuksen ongelmat* (Rovaniemi: Lapin yliopiston yhteiskuntatieteellisiä julkaisuja C:11).

Work Design: Age-Related Policies

R. Cremer

TNO Work & Employment, PO Box 78, 2130 AS Hoofddorp, Netherlands

1. INTRODUCTION

Over the past decades the workforce has been steadily aging in a large variety of countries. The increase in the average age of workers runs at different speeds in different sectors of work. This demographic trend can be considered from two perspectives: on the one hand that there will be more older workers and on the other hand there will be fewer younger workers. Making use of these facts requires knowledge about the concept of workability in various work sectors and for various age cohorts.

Apart from this knowledge, awareness in policy making is required on several levels of action. In this sense it is important that responsible organizations know how to make best use of both ends of the age spectrum, without under- or overloading their workforce. In this context the word "organizations" means not only employers but also other layers of work-responsible organizations. It is obvious that different actors are involved with a different focus of interest.

This state-of-the-art review uses a number of recent publications on work and age-related policies, mostly published in the context of progression in European work organizations. I have selected literature on policies aimed at maintaining a working life, not on redundancy schemes and the provisions of beneficial pensions. I look at the question of which parties are needed to bring about the changes in work society in order to prolong the working life of aging workers in good health.

2. AGING IN DIFFERENT EPOCHS

Aging carries different significance for different age cohorts. In other words, the process of aging is associated with changes which are typical for age cohorts. In this respect, it involves not just changes in health care and dietary habits but also the influence of historical and cultural determinants. Taking into consideration the cohorts outlined below, it can be imagined how life and work style in various time periods can be affected (epochs are suggestive and bound to the European situation).

- *Born in the epoch 1925–35*: the crisis generation. Nothing is guaranteed, making a daily living is not a self-evident process; it takes a lot of effort to survive and to establish a position in society; people are obedient and conventional; one should be pleased with whatever work is available.
- *Born in the epoch 1936–46*: the (European) war generation, later the profit generation. A privileged generation which believes chances are created and effort is rewarded and for whom many jobs were available; in their youth this generation was still poor but material growth came soon, prosperity grew rapidly, and a lifelong career was often guaranteed.
- *Born in the epoch 1947–57*: the protest generation, also the baby-boom generation. Everything is possible, many positions are available, and it is rather easy to establish a position.

- *Born in the epoch 1958–70*: the lost generation. Most positions in society have been filled in; an economic depression means economic shrinkage and even a halt to growth.

This illustrates that the effects of aging are in a sense also cohort-related. Whatever may be the exact effects of those cohorts on the process of aging, it is clear that (un)used capacities, health and well-being, values and beliefs towards aging are different. Note that cohort effects should be accounted for along with other individual differences.

3. AGE-RELATED POLICIES AND PARTIES INVOLVED

When market mechanisms outsource certain groups of employees, e.g., the disabled but also aging workers to a certain extent, a change in attitude or policy is relevant for various actors. In a review at least four groups of actors can be appointed with their specific focus and responsibility:

- At the governmental level (national or continental), conditions can be ordered by laws and regulations.
- On the level of society at large, changes in attitudes, expectations, and prejudices are usually lively and prevalent (opinions of elderly people, what they can do and how they should behave).
- The people in the older cohorts — the aging workers — can affect their situation by becoming self-aware and improving their positions; at work this may mean that workers' interests are promoted by employee unions or works councils in work organizations.
- On the level of work organizations, the employers' concrete actions can be developed and taken in order to better anticipate for change in workability in employees; employers' interests may also be promoted by trade unions.

Adequate adaptation to change can best be achieved in concerted action by all these actors. Actions can be based on various motives depending on the level of management. For example, at the governmental level, the number of pension claims increases and adds to social security expenses. This generally recognized development becomes too expensive and demands appropriate actions aimed at maintaining work for elderly workers. Similarly, due to poor adaptations in work conditions and provisions, elderly workers are more prone to sickness leave or work incapacity schedules.

Society at large is expected to become more tolerant towards senior citizens. Senior citizens form a major share in residential communities. The modern older person is well aware of their position, knows what to do, and knows how to spend their time. Self-awareness and self-esteem are no longer the exclusive preserves of the young developing person. But most important is the situation in work organizations. This is where age awareness should be better developed and action should be taken. In a World Health Organzation study group on aging and working capacity (WHO 1993) recommendations were given to employers, trade unions, and regulatory agencies. The following issues were addressed:

- Work capability, not age, should be the criterion for hiring and retaining employees.
- Employers should maintain sufficient flexibility in the design of jobs and the work environment to ensure appropriate working conditions for the markedly heteroge-

neous older population. Workers should be given the opportunity to participate in decisions or actions that affect their jobs.

- Work arrangements should be flexible to allow for job sharing, part-time work, and time off for family responsibilities. There should be incentives to encourage part-time work during the years traditionally dedicated to retirement in order to maximize older people's participation in the workforce.
- Every worker should be provided with appropriate education and vocational training as a basic component of work. This education should anticipate technical changes in the workplace, anticipate the redesigning of work as workers age, and allow employees to maintain and increase their skills and achieve job satisfaction.

These guidelines are the result of applied research and they may be useful. Also important are people's experiences in projects dealing with the problems of elderly workers.

4. PROJECTS IN EU MEMBER STATES

A state-of-the-art report published by the European Commission (1998) reviews projects in a large number of member states. In all, 123 examples of initiatives were given, representing activities of 13 EU member states. The study was meant to allow transferability of findings. All projects appeared to have the potential to be applied in other member states. Conclusions and recommendations are summarized below.

The initiatives in EU member states are considered to be at the vanguard of new policies for older workers. Projects serve several purposes, not only to employ older workers but also to overcome negative attitudes towards older workers among employers and the general public. Generally, projects offer considerable additions. Here are some areas mentioned in the EC (1991) report:

- Providing targeted services which generally do not exist elsewhere.
- Providing tailored solutions for both older workers and employers.
- Programs aimed at motivating older people to undertake further education and training.
- Methods of recording and systematizing the knowledge and work experience of older workers.
- Use of motivation and teaching methods which consider the age of participants.
- In-company career planning for older workers.
- Involvement of older people in the delivery of services; older people have a greater understanding of clients' needs.
- Helping people who have retired to reenter employment and to find useful work in the community.
- Providing employers with information about the employment of older workers and examples of best practice.
- Raising awareness about the needs of older workers among policy makers, employers, and the general public.
- Developing long-term relationships with groups of employers.
- Establishing good practice networks in member states.

5. RESEARCH OPPORTUNITIES

Apart from the conditions summarized in Section 4, the EC report concludes that substantial knowledge is missing. It has therefore proposed evaluation studies in which a variety of topics should be examined:

- Tracking of older workers over a period of several months in order to look at the longer-term effects of initiatives.
- Sustainability of jobs found by older workers.
- A detailed look at workplace attitudes following age awareness training — not only among managers but among younger and older workers — searching for concrete evidence of changes in workplace practices.
- Additionality, displacement of younger workers by older workers (particularly in projects providing incentives to recruit older workers) and deadweight.
- Relationships and networking between older workers initiatives and other organizations operating in the labor market, in particular to examine additionality and problems of referral between agencies.
- Take-up of places on initiatives by subgroups of older jobseekers (e.g., long-term unemployed people and disabled people).
- Benefits to employers of recruiting older workers.
- The process of adaptation to a working environment by older workers coming from projects, and the problems encountered by them and by employers.

6. GUIDELINES FOR CHANGE IN THE WORK ORGANIZATION

Recently Walker (1998) compiled a set of four guidelines that are intended to introduce good practice into work organizations in general. The guidelines were based on 22 in-depth case studies in seven European countries. They entail the following issues.

6.1. Backing from Senior Management

Good practice in age management is only possible by support from management.

6.2. A Supportive Human Resources Environment

All of the workplace case studies benefited from a supportive human resources environment. Sometimes sound business reasons meant that older workers were highly valued. Traditional cultures and traditional management styles may create an age-supportive human resources environment.

6.3. Commitment from the Aging Workers Involved

Workplace initiatives with respect to aging workers are mainly the result of top-down policy decisions. However, there is no doubt that the support of the older workers concerned is a vital key to successful age-related policy. In some organizations older workers had to be persuaded to accept and own the initiative.

6.4. Careful and Flexible Implementation

The implementation process is vulnerable and requires a number of steps that should be carefully followed. Here are some of them:

- Careful preparation, including research in recruitment trends and age profiles of employees and labor market projections.

- Open communication both with staff generally and with the target group about the objectives of the initiatives, including appropriate use of seminars, workshops, and newsletters.
- Early involvement of trade unions, works councils, and staff associations.
- Early involvement of workers themselves.
- Education and awareness-raising among line managers.
- Staged implementation, including a pilot phase both to test the initiative and to demonstrate to any doubter that it can be effective; provide regular monitoring and feedback with any required adjustments to the initiative.
- Periodic assessment of impact and feedback.
- Constant communication with all employees in order to avoid stigma and feelings of inferiority among older workers or the development of "them and us" attitudes.
- Attention to other aspects of the working environment, such as arduous tasks and conditions, which may inhibit the good practice initiative from achieving its intended goal.

7. SUGGESTIONS FOR GOOD PRACTICE

Here is an overview given by Walker (1999); it offers guidelines for good practice and it addresses all parties involved in changing the general attitude towards the aging worker.

7.1. National and European policy

7.1.1. European Union

Eliminate age barriers from the European Commission's own recruitment practices. Ensure that the new European Social Fund makes older workers a priority group. Ensure that the needs of older workers are adequately reflected in employment guidelines and equal opportunities policies. Introduce a new European Code of Good Practice covering employment of older workers.

7.1.2. Governments

- *Education*: promote public education to counteract negative images of older workers and the promotion of lifelong learning.
- *Employment policy*: active labor market policies should be designed to enable older workers to remain in employment or return to employment, and they should promote quality employment for this group and other groups.
- *Pensions and social security policies*: eliminate incentives to employers to make older workers redundant.
- *Inclusion of older workers*: encourage employers to establish comprehensive action programs on age and employment by publishing good practice guides and disseminating age awareness literature.

7.1.3. Trade unions (national)

Disseminate examples of good practice to members as part of promoting positive approaches and attitudes to age management.

7.2. Employers

Develop age awareness throughout the organization and ensure that age is not used inappropriately in recruitment and training. Union representatives should take part in age awareness training as a matter of routine; collective agreements should cover training measures that rectify the disadvantages experienced by older workers.

7.3. Aging Workers

Take advantage of training and lifelong learning opportunities and develop the awareness of training and career requirements.

8. CONCLUSION

In considering effectiveness of age-related policies in work environments, parties on various levels can take responsibility for action. However, the responsibility is not always felt or the need for action is not acknowledged. In a sense, actions depend on the preparedness for change and any questions surrounding the action.

In a workforce with an increasing average age, it is often discussed how proactive policies can be designed and implemented in work organizations. Here lies a dilemma: anticipating change usually demands some kind of investment and revenues may only be seen after many years. Therefore it is rare to see proactive age-related policies specifically aimed at individual changes over time. From this perspective, training and education are the tools for improving the quality of future work.

The ability to learn is of major importance. As information processing abilities change over the life span, it is evident that learning abilities also change. Accordingly, deprivation of learning may affect the motivation to study. Learning abilities are related to intellectual capacities, personal style and effectiveness, and preferences in learning strategy. Preparing for modern work at a later age requires learning efficiency which takes into account the factors stated above.

It is clear that age-related policies entail more than creativity with retirement schedules or work flexibility programs. Attention should be devoted to extension of the working career in a way that is rewarding for all actors.

REFERENCES

EUROPEAN COMMISSION, 1998, *A Review of the Findings of Eurowork Age* (Brussels: Industrial Relations and Social Affairs Directorate Unit V/A.2).

EUROPEAN COMMUNITY, 1991, *Proceedings of a European Colloquium*, Paris, 12 June 1991 (Luxembourg: European Community), L-2985.

WALKER, A. 1998, Employment and retirement strategies. In A. Kilbom (ed.) *Ageing of the Workforce* (town: publisher). Keynote presentations and abstracts from a workshop held in Brussels, 23-24 March (Stockholm: Arbetslivinstitutet).

WALKER, A., 1999, M*anaging an Ageing Workforce*: *A Guide to Good Practice* (Dublin: European Foundation).

WHO (World Health Organization), 1993, *Aging and Working Capacity*. WHO Technical Report 835 (Geneva: WHO).

Work Hazards and Risk Assessment in Human Performance

A. Genaidy

Department of Mechanical Industrial and Nuclear Engineering, University of Cincinnati, OH 45221-0072, USA

W. Karwowski

Center for Industrial Ergonomics, Department of Industrial Engineering, University of Louisville, KY 40292, USA

1. INTRODUCTION

Human performance improvement aims to balance work demands with worker abilities and needs. Human performance is often described in terms of health and safety measures with the goal of risk quantification of workplace hazards. Risk, a common assessment parameter in health and safety analysis, quantifies the degree of harm with respect to likelihood (e.g. probability) and severity (e.g. consequences) in response to workplace hazards (Cox and Cox 1993; Manuele 1997). A workplace hazard is defined as an event or situation with the potential for harm (Cox and Cox 1993). Typically, harm has been interpreted as resulting from physical hazards such as those produced by physical task demands (i.e. biomechanical, physiological) and by the environment (e.g. nuclear, radiological, biological, chemical, physical).

The modern concept of work hazard includes the non-physical hazards: specifically, psychosocial, work organization, and mental demands. These hazards address both the interactions among job content, work organization and management, and work environment and organizational conditions, and the workers' competencies and needs (International Labour Office 1986). Work-related mental demands pose challenges which, if excessive, form a source of hazard (e.g. Hancock and Warm 1989).

Occupational stress, the harm resulting from psycho-social and work organization hazards as described in its respective literature, is generally chronic in nature, persisting for a period of time during which the harm may be cumulative or progressive. Evanoff and Rosenstock (1994) reported that estimates of direct and indirect medical costs associated with occupational stress in the United States have ranged from $80 to $150 billion annually. The National Occupational Research Agenda (NORA) also named organization of work as one of its twenty-one research priorities (NIOSH 1996).

For example, the work-related musculoskeletal disorders (WMSDs) have been widely acknowledged as a major part of occupational harm resulting from the acute and cumulative exposure to physical task demands (Centers for Disease Control 1987), are among the most prevalent cause of lost time injuries and illnesses in almost every industry and, importantly, are among the most costly (Bureau of Labor Statistics 1995; National Safety Council 1995). The National Institute of Occupational Safety and Health (NIOSH) has conservatively estimated the cost of occupational musculoskeletal disorders as $13 billion annually (NIOSH 1997a) and has devoted extensive resources in the investigation of risk factors for WMSDs (NIOSH 1981; Waters *et al*. 1993; NIOSH 1996, 1997a, 1997b, 1997c). Recognizing that further research is needed, the National Occupational Research Agenda (NORA), established through a collaboration of NIOSH and over 500 stakeholders in public and private sectors for guiding health and safety efforts, identified low back disorders and upper extremity disorders as two of its twenty-one critical priority areas (NIOSH 1996).

2. INTERACTIVE EFFECTS OF WORK HAZARDS

Due to the highly interactive nature of multiple hazards comprising work systems, it is important that the health and safety assessment system takes into consideration all of the known elements and their potential and actual relationships with each other. While "zero" risk at work is an optimal goal, effective risk can be minimized to levels acceptable to the individual worker and organization members. The hazard and risk assessment methods must clearly describe and assess all fundamental elements, their relationships with each other, the potential for both acute and chronic exposure, and the harm that may result.

The adverse health effects of interactive relationships among physical task demands, mental task demands, and the physical/social/psychological/organizational environment conditions has been suggested through several studies. WMSDs have been reported as exacerbated by the occupational environment (Ulin and Armstrong 1992; Armstrong *et al*. 1993; Bongers *et al.* 1993; Chaffin and Fine 1993; Kuorinka and Forcier 1995; Moon and Sauter 1996; Smith and Carayon 1996; NIOSH 1997a, 1997b). Devereux (1998) confirmed the interactive effects of physical and psycho-social work demands on the prevalence rates of musculoskeletal disorders. However, such interactive effects are themselves contingent upon factors such as age, skill, and fitness, since personal characteristics have also been cited as additional contributors to the development of WMSDs (International Labour Office 1986; Fraser 1989; NIOSH 1997c).

The comprehensive work assessment system should include: (1) characterization of the domains of work demands (i.e. physical/mental demands and physical/social/psychological/organizational environment conditions) with corresponding outcome measures (i.e. effort/perceived risk/actual risk); and (2) assessment of the relationships between work demands and outcome measures in order to optimize human performance with respect to safety and health as manifested through the abatement of risk associated with the development of WMSDs. Work demands include the work content (i.e. the physical and mental job demands), and the work context (i.e. the physical, social, individual growth, and work organization environment).

In the past, safety professionals, ergonomists, and industrial hygienists have historically approached the investigation of the work system for hazard and risk evaluation using a spectrum of investigational methods. These efforts can be clustered into two general groups: (1) job analytic techniques and (2) system safety techniques.

3. JOB ANALYSIS TECHNIQUES

Job analysis techniques have been used to characterize work demands and their effects on workers. Shoaf *et al.* (1998) recently suggested that the term "job analysis" has become outdated as the scope and complexity of "jobs" have increased. Rather, the evaluation of jobs is more accurately referred to as "work system

assessment". Here, we use "work system assessment" throughout to address the traditional "job analysis" concept as well as advanced work-related hazard/risk assessment. As indicated earlier, work system assessment includes the characterization of work demands (i.e. physical/mental demands; physical/social/psychological/organizational environment conditions) and outcome variables (effort/perceived risk/actual risk).

Work system assessment may be applied to any unit of work — that is, to tasks, jobs, or occupations. A "task" describes a distinct part of a job, and a "job" is defined as all the work carried out by a worker or group of workers (British Standards Institution). Therefore, a job may consist of one or more tasks. An "occupation" refers to a job of a general class without regard to organizational lines (McCormick 1979). Shoaf *et al.* (1998) grouped job analysis techniques into one of three classification systems: micro-assessment methods, macro-assessment methods, and comprehensive methods.

Micro-assessment methods are specialized in technique and narrow in scope. They focus typically on one work domain or one of its specific subsets. Examples of micro-assessment techniques are: analysis of lifting demands and their effects on workers (e.g. Waters *et al.* 1993); analysis of mental demands (Hart and Staveland 1988; Reid and Nygren 1988); and psycho-social assessment (Hackman and Oldham 1976; Karasek and Theorell 1990).

Macro-assessment methods are not as detailed as micro-assessment methods and typically involve the evaluation of a particular work demand domain or may span across more than one work domain. A typical macro-assessment method utilizes a checklist approach in the form of a questionnaire. Examples include, but are not limited to, the work of Newman (1977), Keyserling *et al.* (1991), and Guo *et al.* (1996).

Comprehensive methods incorporate characteristics of both micro-assessment and macro-assessment methods. The Position Analysis Questionnaire (PAQ; McCormick *et al.* 1969) and Arbeitswissenschaftliches Erhbungsverfahren zur Tatigkeitsanalyse (AET) (Rohmert and Landau 1983) are examples of comprehensive job analysis methods. They are the among the most thorough systems in the literature, especially because they characterize the entire spectrum of work demands. The PAQ and AET, however, lack many of the detailed findings established within the last two decades and thus do not possess a sufficiently comprehensive framework for hazard and risk assessment at this point in time.

4. SYSTEM SAFETY TECHNIQUES

System safety hazard/risk evaluation techniques have been largely devoted to analysis of the physical environment. System safety techniques, which address occupational hazards, are process-based. They serve as regimented methods of analyzing system design to ensure the intended operation and mitigate possible failures. Thus, system safety hazard evaluation covers a wide spectrum of potential and existing hazards. In contrast, non-process based techniques (e.g. air or water sampling) evaluate only existing physical work-related hazards.

The following is a summary of the most common process-based hazard evaluation techniques (Gressel and Gideon 1991):

- Checklists are among the simplest forms of hazard evaluations. They can identify recognized hazards and ensure compliance with accepted design standards. Checklists can be applied to equipment, procedures or materials.
- Preliminary hazard evaluation analysis lists the hazardous materials, equipment components, and process operating conditions. As each hazard is identified, the possible causes, consequences and corrective measures are listed.
- "What If" analysis can identify both hazards and their consequences and help develop possibilities for potential hazard reduction. The analysis procedure usually starts at the beginning of the process and asks a series of questions concerning process upsets or malfunctions. Additional questions based on the initial analysis may be formulated.
- Safety reviews are conducted to identify plant conditions and procedures that may have deviated from the intended design.
- Failure Modes and Effects Analysis (FMEA) checks each process component individually and describes the function of each component and all of its potential failure modes. The method then determines the causes of these failures as well as the effects.
- Fault Tree Analysis (FTA) determines and displays the cause of a major unwanted event. This method starts with the top or end event and develops a logic tree showing the causes of the problem through the use of "AND" and "OR" gates.
- Event Tree Analysis (ETA) is similar to FTA in several ways. As in FTA, a tree structure is developed to outline the events of a hazard scenario. While FTA develops a vertically oriented tree logic, an ETA tree is constructed horizontally and begins with an initiating event and moves forward rather than beginning with the end event.
- Hazard and Operability Study (HAZOP) is a powerful evaluation technique in terms of identifying complex failure scenarios that involve multiple independent events. By using the plant equipment and instrumentation drawings, the process is broken into small segments or nodes such as the line connecting a pump to storage tank. Deviations of the process from normal operating conditions are evaluated by applying a series of guide words to the node. Recommendations for improvements or for more study are based upon the likelihood and consequences of the deviations.

In addition to hazard identification, most of the above mentioned system safety techniques provide probabilistic risk quantification (e.g. ETA, FTA, HAZOP). They are, however, limited because of their unaccountability for the role of human behavior, particularly human error (Feyer and Williamson 1998).

Human reliability analysis techniques (e.g. influence diagrams, human cognitive reliability models, technique for human error rate prediction) have attempted to improve system risk assessment by quantifying human error probability (Kirwan 1990). Still, the development of risk assessment techniques has failed to realize the enlarged definition of work hazard (International Labour Office 1986) described above in two respects. First, these techniques are designed to address acute hazards. Acute hazard exposure usually results from human error or technical failure and can be characterized as an "off–on" switch (Cox and Cox 1993). Hazards resulting from chronic exposures, to which it is much more difficult to assign a probability value, are largely neglected. Second, the domains of organizational and

psycho-social hazards have been generally disregarded with respect to their contribution to overall system risk. These omissions represent serious deficiencies in the current system safety hazard/risk assessment techniques.

5. A COMPREHENSIVE WORK ASSESSMENT

The main objective of the comprehensive work assessment is hazard identification and risk quantification. In addition, such a system can also be used for:

- Design of rehabilitation and return-to-work programs by health practitioners. Work system analysis provides information to classify, which tasks are essential, and which are non-essential regarding execution of a job. Critical tasks can be used to provide input to physical therapists as recommendations for therapy guidelines. Later, the critical tasks can serve as the performance criteria for determining when the worker possesses sufficient capability to return to work.

- Design of the most appropriate medical examinations by occupational medicine specialists. A work demands model yields task descriptions, which serve as inventories of various occupations. These inventories form a database, which can be used as a functional capacity checklist of essential job tasks, therefore aiding medical specialists. Evaluation of an individual's capabilities for a specific job as demonstrated in a pre-employment health screening is problematic without such data Additionally, an inventory of job tasks serves to identify workers who participate in hazardous tasks, and therefore, alerts medical specialists to monitor such workers for adverse health effects.

- Decision-making with respect to work restrictions and provisions of reasonable accommodations by occupational medicine physicians, human resource specialists, and business managers. In recent years, there has been considerable government attention regarding the needs of the handicapped and disabled in the active workforce. As a result, medical selection procedures for job candidates have undergone considerable scrutiny. A worker's limitations with respect to a given job can only be evaluated relative to a detailed description characterizing the nature of the essential tasks. Work system analysis provides a means for collecting and classifying this information.

- Improvements of the work processes by industrial engineers, business managers, and supervisors and work teams. To understand the complex operation of a work system, all aspects must be defined and their interrelationships described. This knowledge, which is a prerequisite for improvement efforts, can be provided by a work system analysis.

6. CONCLUSIONS

The level of risk at work depends on the intensity, frequency, and duration of the multi-faceted array of factors, which characterize the workplace as they relate to the workers' capacity to respond to work demands. The interactive nature of physical and non-physical (i.e. cognitive, psycho-social, and organizational) work system hazards and the magnitude of the resulting harm to workers, businesses, and the economy, warrant the need for an assessment tool which considers all work system

elements. NIOSH (1997c) recognized that, in general, knowledge of the relationships between risk factors and the level of risk is still incomplete. The magnitude of this problem and the recognition of its impact creates a pressing need for development of the comprehensive work system hazard/risk assessment instrument that integrates all system elements (Shoaf *et al.* 2000).

REFERENCES

ARMSTRONG, T.J., BUCKLE, P., FINE, L.J., HAGBERG, M., JONSSON, B., KILBOM, A., KUORINKA, I.A.A., SILVERSTEIN, B.A., SJOGAARD, G. and VILKARI-JUNTARA, E.R.A., 1993, A conceptual model for work-related neck and upper-limb musculoskeletal disorders, *Scandinavian Journal of Work and Environmental Health*, 19, 73–4.

BONGERS, P.M., DE WINTER, C.R., KOMPIER, M.A.J. and HILDEBRANDT, V.H., 1993, Psychosocial factors at work and musculoskeletal disease, *Scandinavian Journal of Work and Environmental Health*, 19(5), 297–312.

Bureau of Labor Statistics, 1995, *Workplace Injuries and Illnesses by Selected Characteristics, 1993*, Washington, DC: US Department of Labor, Bureau of Labor Statistics, USDL 95-142.

Centers for Disease Control, 1987, *Morbidity and Mortality Weekly Report: Prevention of Leading Work-Related Diseases and Injuries*, US Government Printing Office, 748-122/60526.

CHAFFIN, D. B. and FINE, L. J., (eds), 1993, *A National Strategy for Occupational Musculoskeletal Injuries: Implementation Issues and Research Needs,* 1991 Conference Summary, Department of Health and Human Services (NIOSH), Publication No. 93-101.

COX, S. and TAIT, R., 1998, *Safety, Reliability and Risk Management: An Integrated Approach* (Oxford: Butterworth-Heinemann).

COX, T. and COX, S., 1993, *Psycho-social and Organizational Hazards at Work*, European Occupational Health Series No. 5, Copenhagen: WHO Regional Office Europe.

DEVEREUX, J., 1998, The combined effects of physical and psychosocial work factors. In M.A. Hanson (ed.) *Contemporary Ergonomics* 1998, 25–9.

EVANOFF, B.A. and ROSENSTOCK, L., 1994, Psychophysiologic stressors and work organization. In L. Rosenstock and M.R. Cullen (eds) *Textbook of Clinical Occupational and Environmental Medicine* (Philadelphia: Saunders).

FEYER, A., and WILLIAMSON, A., 1998, *Occupational Injury* (Bristol, PA: Taylor & Francis).

FRASER, T.M., 1989, *The Worker at Work* (London: Taylor & Francis).

GRESSEL, M.G. and GIDEON, J.A., 1991, An overview of process hazard evaluation techniques, *American Industrial Hygiene Association Journal*, 52(4), 158–63.

GUO, L., GENAIDY, A., CHRISTENSEN, D. and HUNTINGTON, K., 1996, Macro-ergonomic risk assessment in the nuclear remediation industry, *Applied Ergonomics*, 27, 241–54.

HACKMAN, J.R. and OLDHAM, G.R. 1980, *Work Redesign* (New York: Addison-Wesley).

HANCOCK, P.A. and WARM, J.S., 1989, A dynamic model of stress and sustained attention, *Human Factors*, 31, 519–37.

HART, S.G. and STAVELAND, L.E., 1988, Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In G.E. Stelmach and P.A. Vroon (eds) *Advances in Psychology*, 52 (Amsterdam: Elsevier), 139–81.

International Labour Office (ILO), 1986, *Psycho-social Factors at Work: Recognition and Control, Occupational Health Series*, no. 56 (Geneva: International Labour Office).

KARASEK, R. and THEORELL, T., 1990, *Healthy Work: Stress, Productivity, and the Reconstruction of Working Life* (New York: HarperCollins).

KEYSERLING, W.M., ARMSTRONG, T.J. and PUNNETT, L., 1991, Ergonomic job analysis: a structured approach for identifying risk factors associated with overexertion injuries and disorders, *Applied Occupational Environmental Hygiene*, 6(5), 353–63.

KIRWAN, B. 1994, *A Guide to Practical Human Reliability Assessment* (London: Taylor & Francis).

KUORINKA, I. and FORCIER, L. (eds), 1995, *Work Related Musculoskeletal Disorders (WMSDs): A Reference Book for Prevention* (London: Taylor & Francis).

McCORMICK, E., 1979, *Job Analysis* (New York: Amacom).

McCORMICK, E., JEANNERT, P. and MECHAM, R., 1969, *The Development and Background of the Position Analysis Questionnaire* (Lafeyette, IN: Occupational Research Center).

MANUELE, F.A., 1997, *On the Practice of Safety* (New York: Van Nostrand Reinhold).

MOON, S.D. and SAUTER, S.L. (eds), 1996, *Beyond Biomechanics: Psycho-Social Aspects of Musculoskeletal Disorders in Office Work* (Bristol, PA: Taylor & Francis).

National Institute for Occupational Safety and Health (NIOSH), 1981, *Work Practices Guide for Manual Lifting*, NIOSH Technical Report (Cincinnati, OH: NIOSH).

National Institute for Occupational Safety and Health (NIOSH), 1996, *National Occupational Research Agenda*, US Department of Health and Human Services, DHHS Publication No. 96-115.

National Institute for Occupational Safety and Health (NIOSH), 1997a, Report of written testimony from Dr Linda Rosenstock, *The Science of Occupational Musculoskeletal Disorders*, submitted to Subcommittee on Workforce Protections Committee on Education and the Workforce, US House of Representatives, May 21.

National Institute for Occupational Safety and Health (NIOSH), 1997b, *Musculoskeletal Disorders and Workplace Factors* (Cincinnati, OH: US Department of Health and Human Services), DHHS Publication No. 97-141.

National Institute for Occupational Safety and Health (NIOSH), 1997c, *Elements of Ergonomics Programs* (Cincinnati, OH: US Department of Health and Human Services), DHHS Publication No. 97-117.

National Safety Council, 1995, *Accident Facts* (Itasca, IL: National Safety Council).

NEWMAN, J.E., 1977, Development of a measure of perceived work environment. *Academy of Management Journal*, 20(4), 520–34.

REID G.B. and NYGREN, T.E., 1988, The subjective workload assessment technique: A scaling procedure for measuring mental workload. In P.A. Hancock and N. Meshkati (eds) *Advances in Psychology*, 52 (Amsterdam: Elsevier), 185–218.

ROHMERT, W. and LANDAU, K., 1983, *A New Technique for Job Analysis* (London: Taylor & Francis).

SHOAF, C., GENAIDY, A. and SHELL, R., 1998, A perspective on work system analysis: classification and evaluation of methods, *Ergonomics*, 41, 881–98.

SHOAF, C. GENAIDY, A., SHELL, R., HAARTZ, J., KARWOWSKI, W., WATERS, T., HANCOCK P.A. and HUSTON, R., 2000, An adaptive control model for assessment of work-related musculoskeletal hazards and risks. *Theoretical Issues in Ergonomics Science*, 1 (1), in press.

SMITH, M.J. and CARAYON, P., 1996, Work organization, stress, and cumulative trauma disorders. In S.D. Moon and S.L. Sauter (eds), *Psychosocial Aspects of Musculoskeletal Disorders in Office Work* (London: Taylor & Francis), 23–42.

ULIN, S. S. and ARMSTRONG, T.J., 1992, A strategy for evaluating occupational risk factors of musculoskeletal disorders. *Journal of Occupational Rehabilitation*, 2, 35–50.

WATERS, T., PUTZ-ANDERSON, V., GARG, A. and FINE, L.J., 1993, Revised NIOSH equation for the design and evaluation of manual lifting tasks. *Ergonomics*, 36, 749–76.

Working with Age: An Ergonomic Approach to Aging on the Job

J. C. Marquié and S. Volkoff*

Laboratoire Travail et Cognition, UMR 5551 CNRS, Maison de la Recherche, Université Toulouse le Mirail, 5 allées Antonio Machado, F-31058 Toulouse Cedex, France and *Centre de Recherches et d'Etudes sur l'Age et les Populations au Travail (CREAPT), 41 rue Gay-Lussac, F-75005 Paris, France

1. CURRENT STATE OF THE PROBLEM

Changes in the age distribution of the workforce in industrialized countries during the past few years are the outcome of two opposing movements. On one side, the aging of the population as a whole and the later age at which young people enter the working world are increasing the percentage of older employees. On the other, retirement policies designed to reduce the number of employees are decreasing the percentage of older workers in many countries, although national differences are still great in this area.

According to the most plausible predictions, the percentage of young workers will continue to drop. And given the uncertain future of the social security system and other social programs, one can assume the current trend toward early retirement will slow down if not reverse itself altogether. In any case, one can predict two things:

- An increase in the percentage of individuals between the ages of 40 and 55 in practically all industrialized nations, although at different times and different rates; the process has already begun in most countries.
- At least stabilization in the percentage of workers age 55 or older for the next two to three decades.

These overall trends are subject to large discrepancies across sectors and firms. The aging trend is particularly clear-cut in manufacturing industry, but the middle-age population bulge is also apparent in the service industries, particularly those where the employment rate soared in the 1970s.

In situations where the working conditions are highly demanding, as in shift work, time pressure jobs, and adaptation to modern technology or skill diversification, this demographic trend may cause serious problems, as demonstrated by the results of national or European surveys on working conditions today. Labor selection mechanisms, which have contributed to protecting older employees against certain types of job constraints, will become more difficult to implement if the demographic weight of individuals in this age range continues to increase.

To solve this problem, some companies are attempting to foresee the impacts of an aging staff. In their previsional management of employment and employee skills, they hope to develop ergonomic policies that will take this new dimension into account while still satisfying ever changing economic demands. They are seeking the necessary knowledge and methods for coping with the sometimes urgent practical problems they are facing.

For the individual, these numerous changes have various consequences, including providing jobs for aging workers, the state of health of employees exposed for many years to strenuous or tedious working conditions, the involvement of individuals over 45 in the technological conversion process, job training, task and workstation design, and more broadly, the overall organization of work in accordance with the characteristics of workers of all ages.

Studies that shed light on the social debate about these issues are still insufficient, but see Laville (1989), Snel and Cremer (1994), Kilbom et al.(1997), and Marquié et al. (1998); research efforts must be expanded. In addition to the disciplines that traditionally contribute to the ergonomic approach, the psychology of aging, occupational demographics, and epidemiology are all being called upon to be part of this development.

2. JOB DEMANDS AND AGE STRUCTURES

The discipline we might call occupational demographics, while still in its early stages, has set out to develop new tools for collecting and analyzing macrodemographic or microdemographic data — macrodemographics at the workforce or production sector scale and microdemographics at the company level, or even within a department or group of jobs — in an attempt to bring out facts and set forth hypotheses that can be approached through other disciplines. This branch of study has shown, in a more detailed way than ever before, that age is often a decisive criterion in the labor selection mechanisms now in effect.

Analyses of the distribution of the population by age, and changes in that distribution over time, have revealed significant differences between sectors or branches of the economy. One can find, for example, "elderly" sectors, where there is little turnover and retention measures that lead to aging of the personnel; "aging" sectors, where quadragenarians predominate; and "young" sectors, where the hiring rate is high and people frequently change jobs before the middle of their career. Many of these younger sectors are "female" and they rely more heavily on short-term contracts than other parts of industry.

Two scenarios have been proposed to schematically describe these differences and patterns. Hypothesis H (for history) is based on the history of the company and its volume of activities and jobs. In hypothesis A (for age), due to the constraints involved, certain types of job maintain a relatively stable age distribution: part of the aging employees are transferred to other positions within the same company or segment of industry via a "selection-reassignment" process (more frequent in sectors with an older age structure) or lose their jobs altogether via a "selection-exclusion" process (more frequent in sectors with a younger age structure).

A more detailed analysis of the situations that trigger this kind of age-based selection process indicates that some job demands play an important role in this process, either individually or to an even greater extent when several constraints occur in conjunction with each other. The many work constraints that contribute to age-based selection of workers include shiftwork (e.g., two-shift or three-shift systems); time pressure variables (e.g., time limits on repetitive or short-cycle tasks, machine-paced work, short deadlines); postural constraints that exert excessive stress on the skeletal, articular, and cardiovascular systems; monotonous tasks and unskilled jobs that do not increase knowledge and may even impair the worker's initial capacities; new technologies that quickly render his/her skills obsolete, or

frequent job transfers to meet the company's needs. However, their increasing impact with age should not cause us to lose sight of the fact that these very same constraints also place heavy demands on younger workers. Situations shown to have adverse effects on older workers are very often indicative of more general problems affecting all employees.

3. WORK, AGING, AND HEALTH

The way in which job constraints and demands are withstood at various ages should also be considered in relation to health, whether implicitly or explicitly, this is often a selection criterion in the workplace. The connection between work and health can rarely be described by a single causal relationship. The same job characteristic may have several effects on health, and a single health condition may be attributable to several occupational causes. Moreover, a health problem can have a feedback effect on the manner in which a job is performed. For example, an organ deficiency or impaired bodily function may lead the operator to "spare themselves" in that area, which may cause additional stress on other organs or functions, with possible new pathological consequences.

The health–job relationship is therefore a complex one, and the work activity itself may be a catalyzer. This relationship is difficult to grasp since health-related symptoms are often sub-pathological, making the task of determining what to study far from straightforward. In addition, the effects in question often lag behind their causes: the epidemiology of occupational risks must look at long-term associations between working conditions and health, taking into account occupational histories fraught with selection mechanisms that invalidate any instantaneous comparative conclusions.

It is easy to understand, then, why taking age into account in this context poses a number of methodological problems, at the same time as it offers some new opportunities. On the one hand, it complicates the analysis by introducing another variable related to health in ways that are neither solid, nor uniform, nor easily modeled, even in probabilistic terms. On the other, it promotes a more unified approach by necessitating a dynamic analysis that accounts for the evolution of the work activity plus the individual's health condition and occupational and pathological background. As such, age is both a "factor" in itself (as there is a real link between chronological age and functional age — the age of the arteries — even though it cannot be precisely determined) and a "transmission variable", since the effects of work histories and past events outside the workplace are combined with the individual's advancing years.

This fact becomes apparent if we look at interindividual variability, at a given age, across the various job categories for the three indicators of biological aging used in epidemiology:

- *Life expectancy*: how long we can expect to live based on statistical predictions and criteria such as job category and sex.
- *Physical performance*: age-related changes in abilities are measured at different functional levels of the organism.
- *Susceptibility to disease*: aging is approached in terms of the growing risk of disease or disablement with advancing years.

The differential aging process, reflected by interindividual variability on these different measures, is detected during routine physical examinations by occupational physicians. Signs of pre-

mature "wear and tear' are commonly observed. But there is little data in today's scientific literature for understanding the links between aging, health, and the characteristics of the work environments to which a given individual has been subjected.

The available findings nevertheless point out that the impact of age and exposure to adverse working conditions can be cumulative (e.g., age and exposure to noise have an additive effect on hearing acuity), interacting (e.g., age, heavy workloads, and rigid job procedures considerably increase the risk of lower back problems), or delayed (e.g., the persistence of sleep disorders in former shift workers). The expansion of research in occupational epidemiology, accompanied by longitudinal follow-up of several age and occupational groups, should help provide insight into these complex links, and thus provide a means for achieving early prevention.

4. AGE–RELATED CHANGES IN WORK ACTIVITY

Although problems linked to age do indeed arise in the areas of work and health, they are usually symptoms of modifications that have taken place in the work activity itself. To identify and understand these changes, one must first be aware of the diverse range of phenomena that age encompasses. Here are some of them:

- *Generation effects*: The cultural, social, and health conditions to which individuals born in the 1940s have been subjected, and the education and occupational training they have received, are not the same as those of persons born 25 years later. Such factors leave a long-lasting and sometimes irreversible mark on an individual's health, attitudes, and behavior.
- *Changes in capacities*: Basic biological, sensory, motor, and cognitive capacities change with the aging process. The alteration is gradual, can begin early or late, and be more or less pronounced, depending on the function under consideration. Although the decline is generally moderate between the ages of 20 and 60, its consequences may be significant when job demands are high.
- *Knowledge acquisition*: As age and work experience increase, so do an individual's knowledge and ability to meet new job requirements. However, overspecialization in some areas can hinder training in the new skills needed to adapt to changes in the work environment.
- *Length of exposure*: Long exposure to various occupational hazards and stressful working conditions can impede the development of an individual's basic physiological and cognitive capabilities, sometimes even more than the aging process itself.
- *Modifications*: Modifications in the subject's expectations, needs, and lifestyle can occur at the personal, social, and economic levels.

Precisely because of the multiplicity of these *time-linked* phenomena, whether it be *the moment in time* or the *amount of time*, chronological age is an arbitrary marker: it can mask the fact that the nature and speed of the changes that take place over time vary considerably across individuals. This makes it impossible to predict solely on the basis of age whether performance and difficulties encountered on the job will increase or decrease. Rather, this rising or falling curve is the result of the combined

effects of the wide range of endogenous and exogenous time-linked factors, unique to each and every individual.

The ergonomic approach nevertheless allows us to improve our understanding of changes in work behavior as age increases, as experience is gained, and as skills are acquired. Men and women on the job are not passive spectators of the good or poor fit between the characteristics of their jobs and their own state of functioning. Consciously or unconsciously, they modify their operating modes (movements, work pace, posture, etc.), reduce their effort level in some subtasks, make more plans to avoid emergency situations, check the outcome of their actions so as to reduce errors that would be costly to correct, and adjust the distribution of tasks in cooperative and collective work situations.

But these strategies can only be implemented if the working conditions and organization foster and promote them. This is why ergonomics stresses the need to improve work environments in order to curb the negative effects of aging. In particular, combinations of two types of environmental conditions are thought to be likely to engender greater difficulty with age: overloading of the individual's sensory, motor, or cognitive systems, and tasks that generate little know-how and require little experience (Marquié *et al.* 1998, Warr 1994). These conditions typically correspond to unskilled, time pressure jobs where the individual has little control over their own work. In the other cases, the predictable correlation between age and job difficulty (in terms of workload and performance) is zero or negative. Zero because the individual has little experience but job demands are moderate, or the worker's basic capacities are overtaxed but efficient compensatory strategies acquired through experience are devised. And negative because there is moderate taxing of basic capacities and high demands for expert knowledge; this is the most favorable situation.

5. AGING, NEW TECHNOLOGIES, AND TRAINING

The massive and rapid development of new technologies, the *knowledge obsolescence* they create, and the ensuing need for the ongoing acquisition or restructuring of knowledge may also pose some specific problems for older workers. Various national and international surveys have revealed that computer use is negatively correlated with age. The same age effect on the use of other types of technology has also been observed in industry (robots, numerically controlled machines, etc.).

Although part of this phenomenon can be accounted for by a *transitory generation effect* (e.g., cultural and educational differences in computer use between the younger generation and individuals over 45), it is also likely that the acceleration of technical change will perpetuate this aspect of the problem into the future. Besides, *resistance to change* — frequently cited as a factor in the alleged lesser ability of older workers to adapt to technological advances — says nothing about the causes of potential difficulties facing the aging worker.

A wide range of factors with seemingly cumulative effects must be explored if we hope to overcome the obstacles currently facing the aging worker. In reality, age seems to be less predictive of unfavorable attitudes toward new technologies than seniority, which in turn is less predictive than a complete lack of experience in using such tools.

A variety of other elements must also be considered in any analysis of this issue. These include the attitudes future trainees develop about the trade-off between the potential career benefits and the cost of training, employees' fears that company restructuring triggered by technological change will have a negative impact on employment, the conditions under which training will take place, the ergonomics of new workstations, the consequences of technical transformations on task content and organization, and last but not least, the practical measures taken by the company to inform, prepare, and involve the concerned employees in the upcoming changes.

This problem can be particularly significant for certain categories of employees. It is true of unskilled workers, for whom employment problems, retraining in a new occupation, eligibility for training programs, and adaptation to current teaching methods are all critical issues. But difficulties can also arise for more highly qualified individuals: overspecialization in a given job may actually be a handicap when extensive technological change takes place.

This brings us to training. Training must be considered in the light of research findings on the attitudes of both employers and employees regarding the training of employees over 45, the participation of older employees in ongoing training programs, their success rate, and the sources, types, and severity of the difficulties encountered. Several factors must be taken into account in creating training programs that are better suited to the needs of older workers plus their cognitive, metacognitive, and emotional specificities, instead of being geared to younger individuals as is often the case. These factors include being unaccustomed to learning and new teaching techniques, age-related alterations in cognitive processes, the role of past experience in the skill to be learned, the effects of mediating variables such as educational background, and the very type of work done over the years, which may turn out to engender more or less confidence in the subject's own learning ability.

Finally, in regard to the period after the working phase of life, *gerontechnology* is a new approach to the relationship between aging users and their technological environment. It aims to determine how older persons can best use the possible aids provided with new techniques and tools (home ergonomics, medical techniques designed for individual use, communication technologies, etc.), so they can maintain an optimal level of autonomy, health, and social life.

6. HOW TO STUDY THE LINKS BETWEEN AGE AND WORK

A number of methods are available for investigating the problems specific to advancing age at work. Each has its own advantages and shortcomings. As is often the case, it is a combination of several methods that is best able to lead to a finer diagnosis and to the appropriate ergonomic action.

We have already mentioned methods like age structure analysis (determining the relationship between the mean age of workers holding a given type of job and the constraints associated with those jobs) and epidemiological research (demonstrating the complex cross-sectional or longitudinal links between age, health, and work, via questionnaires about past and present health and working conditions, or by means of physiological, sensory, and cognitive measures).

Performance analysis is another method. However, in addi-

tion to the fact that no systematic correlation has been found between age and performance, these approaches are subject to a number of biases and limitations. One such bias stems from the fact that they deal with the "survivors", i.e., the workers who have adapted the best to the constraints inherent in the jobs studied, since the less "fit" will have left these positions. They have also been criticized for failing to take into account the diverse components of performance, or for having assessed performance by means of subjective evaluations offered by fellow workers and superiors who may be unconsciously influenced by age stereotypes.

Despite their tendency to simplify reality, and despite the fact they usually focus on age-linked alterations in basic biological and cognitive resources, experimental laboratory studies on aging can be useful in improving our understanding of certain problems encountered at work. They can be designed without specific regard for work-related issues, or they can be aimed at getting an in-depth view of a particular work problem. All of these approaches contribute to untangling the respective effects of the different factors of age-related change, as well as to determining the characteristics of the biological and cognitive development of adults and their capacity to adapt to high levels of demand as they grow older. However, precisely because they usually discount the role of experience, these methods as they stand cannot be used to predict behavior in all work settings.

Job analysis is one method for discerning and enhancing awareness of the important role played by experience in worker-to-job and job-to-worker adaptation and suitability. It consists of studying the actual way in which subjects carry out a task and the strategies they employ to cope with internal and external constraints. As they become increasingly refined over time, these strategies (see the earlier examples) enable the worker to substantially reduce the effects of aging in different ways: they *automate* more components of the activity (automatic processes seem to be resistant to the effects of aging); they use *compensation* mechanisms (declining capacities are compensated by skills that improve with practice); they rely on *accommodation* mechanisms (strategies for avoiding situations where difficulties might arise through modification of subgoals or use of external aids). The ability to identify these strategies is an indispensable key to designing tasks and job organizations that will give workers room to maneuver. The benefits of these means of granting greater control over one's work may be reaped not only by aging employees, but also by all individuals in the company who are in charge of managing human resources and organizing work.

REFERENCES

KILBOM, A., WESTERHOLM, P., HALLSTEEN, L., and FURAKER, B., 1997, Work after 45? *Arbete och Hälsa*, 29, 1–299.

LAVILLE, A. (ed.), 1989, Vieillissement et travail. *Le Travail Humain*, 52(1, 2, 3), 1–286.

MARQUIÉ, J.C., PAUMÈS, D., and VOLKOFF, S., 1998, *Working with Age*. London: Taylor & Francis.

SALTHOUSE, T.A., 1990, Influence of experience on age differences in cognitive functioning. *Human Factors*, 2(5), 551–59.

SNEL, J. and CREMER, R., 1994, *Work and Aging: European Perspective*. London: Taylor & Francis.

WARR, P., 1994, Age and employment. In Triandis, Dunnette, and Hough (eds) *Handbook of Industrial and Organizational Psychology*, Vol. 4 (Palo Alto CA: Consulting Psychologists Press) pp. 485–550.

Part 4

Information Presentation and Communication

Alarm Initiated Activities

N. A. Stanton

Department of Design, Brunel University, UK

1. INTRODUCTION

Little advancement has been made upon the model of alarm handling proposed by Lees (1974) over two decades ago. This model comprised three stages: detection (detecting the fault), diagnosis (identifying the cause of the fault), and correction (dealing with the fault). It appears to be very similar to the process model put forward by Rouse (1983) comprising: detection (the process of deciding that an event has occurred), diagnosis (the process of identifying the cause of an event), and compensation (the process of sustaining system operation). A similar model, comprising: detection (detection of the onset of a plant disturbance), diagnosis (diagnosing the particular disturbance from presented symptoms), and remedial actions (selecting and implementing the appropriate actions to mitigate the disturbance) was proposed by Marshall and Baker (1994) as an idealized three-stage decision model to describe how operators dealt with faults. There appears to be little to distinguish these three models apart from the idiosyncratic labeling of the last stage in all cases. Rouse (1983) offers an expanded version of the process model comprising three levels: recognition and classification (in which the problem is detected and assigned to a category), planning (whereby the problem solving approach is determined), execution and monitoring (the actual process of solving the problem). Arguably this is reducible to the original three-stage model, but rather more emphasis has been place upon the interpretation of the problem. What is not clear from any of the analyses presented with these models is whether they accurately reflect processes undertaken by the operator within the alarm handling task.

2. ALARM INITIATED ACTIVITIES (AIAS)

An alarm handling sequence can be described as consisting of a number of generic activity stages. These generic activities have been assembled into an analysis of alarm handling (Stanton *et al.* 1992; Stanton 1994) as shown in Figure 1. The analysis distinguishes between routine events and critical events involving alarms. Although the two types of events have most activities in common, critical events are distinctive by virtue of an investigative phase. It is proposed that the notion of AIAs is used to describe the collective of the stages in alarm event handling. The term "activities" is used to refer to the ensuing behaviors triggered by the presence of an alarm. It is posited that these activities would not have been triggered without the alarm being present, thus they are alarm initiated activities. The AIAs are linked to other supervisory control activities which can be typified as continuous tasks (such as visual scanning of instruments and fine tuning of plant variables in response to minor variations in plant) and discrete tasks (such as putting plant into service and taking plant out of service). In such tasks, alarm information may be used instead of, or in conjunction with, other information (such as data on plant variables, some internal reference to plant state, comments from other operators, or reports from engineers who are in direct contact with plant). Whilst it may be difficult to
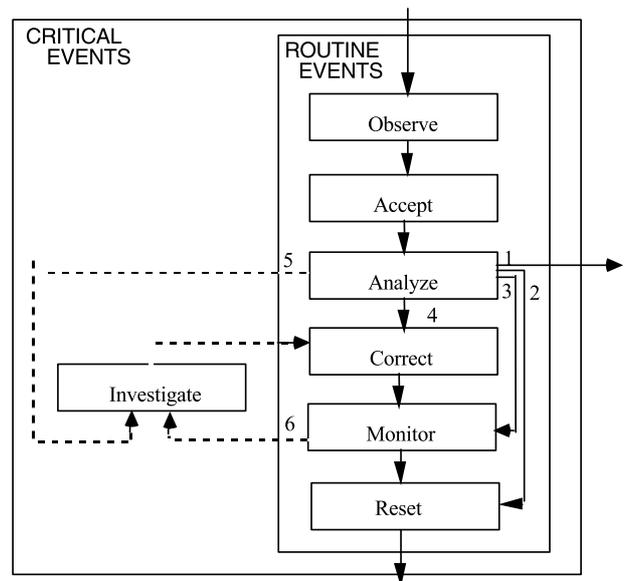
Figure 1.

distinguish between some activities, whether triggered by alarms or otherwise, examination of alarm handling activities can be justified in terms of providing useful information regarding the design of alarm systems. This will, of necessity, involve the consideration of activities where alarm information is of primary importance (such as in a critical event) and activities where the alarm information is of secondary and/or supplementary importance to the task (such as in a routine event).

Operators report that, in alarm handling, when they observe the onset of an alarm, they accept it and make a fairly rapid analysis of whether it should be ignored (route 1), reset (route 2), monitored (route 3), dealt with superficially (route 4), or requires further investigation (route 5). If it cannot be cleared (by superficial intervention), then they may also go into an investigative mode (route 6). In the penultimate mode, the operators will monitor the status of the plant brought about by their corrective actions and ultimately reset the alarm. Routine behavior has a "ready-made" response, whereas critical behavior needs knowledge-based, deductive reasoning. The taxonomy (observe, accept, analyze, investigate, correct, monitor, and reset) is proffered as a working description of alarm handling behavior, rather than intending to represent a fully validated psychological model.

3. A TAXONOMY OF ALARM HANDLING

Activity in the control room may be divided broadly into two types: routine and critical. Incident handling activities take only a small part of the operator's time — approximately 10% (Reinartz and Reinartz 1989; Baber 1991) — and yet they are arguably the most important part of the task. This is particularly true when one considers that the original conception of the operator's task was one of operation-by-exception (Zwaga and Hoonhout 1994). In order to develop a clearer understanding of alarm handling, a taxonomy was developed (Stanton and Baber, in press) on the basis of a literature review, direct observation, and questionnaire data. This taxonomy reveal 24 alarm-related behaviors subsumed under seven categories: observe, accept, analyze, investigate, correct, monitor and reset (see Figure 2).
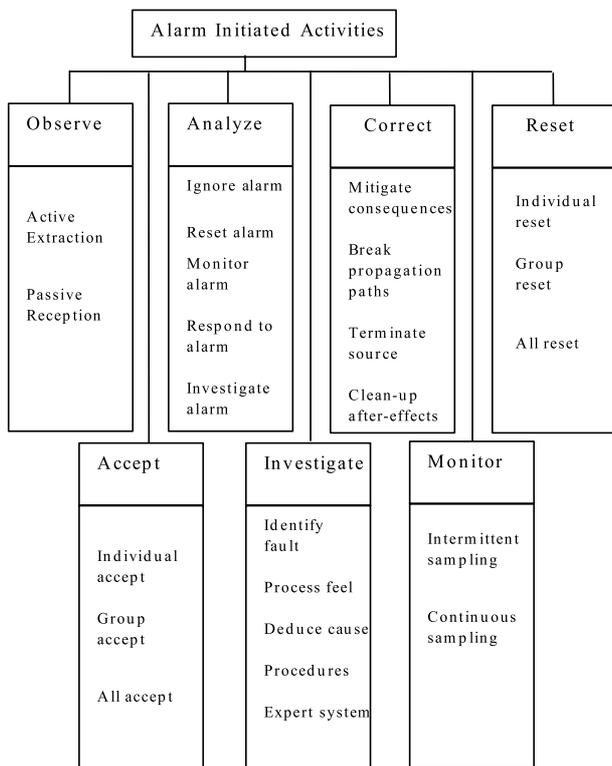
Figure 2.

The development of a taxonomy and model of alarm handling provides a focus on the design formats required to support alarm initiated activities. The human requirements from the alarm system may be different, and in some cases conflicting, in each of the seven categories of the taxonomy. The difficulty arises from the conflicting nature of the stages in the model and the true nature of alarms in control rooms, i.e. they are not single events occurring independently of each other, but they are related, context-dependent, and part of a larger information system. It is not easy to separate alarm initiated activity from general control room tasks, but it is useful to do so. While many of the activities may also be present in more general aspects of supervisory control tasks, their consideration is justified by relating them to the alarm system in particular. This leads to the following observations.

First, designers of alarm systems need to consider all manner of alarm initiated activities. Often there is little consideration of alarm handling beyond the initial 'observe' stage. Second, one could question the need to consider the alarm system as a separate entity from the other displays in the control system. It seems apparent that 'active extraction' of information would be more suitably supported by a permanent visual display. This would suggest three main changes: the integration of alarm information, the move away from the trend for sole reliance upon scrolling text-based alarm displays, and the provision of parallel displays rather than sequential displays.

These changes could support operator heuristics by making the information available to the operator at all times without requiring extensive searching activities. With the continued drive to replace traditional annunciator panels and wall-based mimics with VDUs in control rooms, there is good reason to question the use of text-based alarm displays.

REFERENCES

BABER, C., 1991, *Speech Technology in Control Room Systems* (Ellis Horwood: London).

LEES, F. P., 1974, Research on the process operator. In E. Edwards and F. P. Lees (eds) *The Human Operator in Process Control* (Taylor and Francis: London), 386–425.

MARSHALL, E. and BAKER, S., 1994, Alarms in nuclear power control rooms: current approaches and future design. In N. A. Stanton (ed.) *Human Factors in Alarm Design* (Taylor and Francis: London), 183–91.

REINARTZ, S. J. and REINARTZ, G., 1989, Analysis of team behaviour during simulated nuclear power plant incidents. In E. D. Megaw (ed.) *Contemporary Ergonomics* (Taylor and Francis: London), 188–93.

ROUSE, W. B., 1983, Models of human problem solving. *Automatica* 19: 613–25.

STANTON, N. A., 1994, Alarm initiated activities. In: N. A. Stanton (ed.) *Human Factors in Alarm Design.* (Taylor and Francis: London), 93–117.

STANTON, N. A. and BABER, C., 1995, Alarm initiated activities: an analysis of alarm handling by operators using text-based alarm systems in supervisory control systems. *Ergonomics* 38 (11): 2414–31.

ZWAGA, H. J. Z. and HOONHOUT, H. C. M., 1994, Supervisory control behaviour and the implementation of alarms in process control. In N. A. Stanton (ed.) *Human Factors in Alarm Design.* (Taylor and Francis: London), 119–34.

Auditory Warnings

J. Edworthy and E. J. Hellier

Department of Psychology, University of Plymouth, Plymouth PL4 8AA, UK

1. INTRODUCTION

1.1. Terminology

Auditory warnings can be thought of as a specific type of alarm, although the terms "auditory warning" and "alarm" are not entirely synonymous. "Alarm" may be thought of as the generic term for all forms of sounds which attract attention, including animal calls, burglar and car alarms, equipment alarms and so on. Auditory warnings are sounds which attract attention, but for the specific purpose of providing additional support and information in potentially dangerous situations. They can be more specifically construed as attention-getting sounds that can, when well designed, both attract attention and inform the hearer so that s/he will then take appropriate action. "Auditory warning" can also be construed to include speech warnings, although these are excluded from this article as detailed discussion of intelligibility and comprehensibility are required. Other terms such as "alerts" and "attensons" (attention-getting sound) are used but there is no universally agreed terminology.

Typically, an auditory warning will remain silent unless some critical parameter value is reached (e.g. temperature, pressure or height). If the critical value is exceeded, the warning will sound. Usually an auditory warning will produce only one identifiable sound once activated, although more sophisticated warnings can produce graded levels of urgency, depending upon the criticality of the parameter being signaled. Throughout this article the hazard or danger that triggers an auditory warning will be referred to as a referent (Edworthy and Adams 1996).

1.2. Scope

Auditory warnings vary greatly in terms of the environments in which they are used, the types of sounds which are used and the type of behavior that they are expected to elicit.

1.2.1. Usage

Auditory warnings are used heavily in occupational settings, particularly where people work under conditions of high workload or are continually having to carry out a lot of visual scanning or are constantly moving around. Thus they are used extensively in aircraft, hospitals, control rooms, factories and transportation. Specific types of auditory warnings are also used in more general public facilities and buildings, signifying dangers such as fire.

1.2.2. Warning types

Auditory warnings can be non-verbal or verbal (speech). Essentially there are three categories of non-verbal auditory warnings — abstract, semi-abstract and concrete. The first, abstract, includes traditional types of warning sounds such as bells, horns, klaxons, buzzers and the like. For these types of sound the association between the referent and the warning is learned through stimulus–response association. Hearers know what the warning sound means because of a learned association between the sound and the situation it is signaling, not because of some inherent property of the warning sound itself (although sirens possess an inherent "wailing" quality, the response to which may not need to be learned). The second type, semi-abstract, are usually more technologically sophisticated, and are able convey some feature of the referent through the sound itself (e.g. an increase in pitch might indicate an increase in height or temperature). In this case, the association between the hazard and the warning should be easier to learn. The third class of auditory warning type, concrete sounds, are the sounds of real objects and events. Here, the sounds of everyday objects undergoing specific processes (e.g. the sound of a glass filling up with water) are used to signify processes which may have related features (e.g. the process of copying material from one computer disk to another). This third category of auditory warning may require minimal learning because the recognition and interpretation of these sounds relies on "everyday listening" (Gaver 1993). There is little research evidence to show which types of warning sounds are the most effective either generally or for specific applications.

1.2.3. Warning and monitoring sounds

Typically, auditory warnings exist in a binary on/off state. In many situations (in medicine, for example) they can be preprogrammed to signal when a physical parameter reaches critical state. Some more advanced warning systems are capable of signaling various states of criticality by producing sounds differing in urgency or seriousness, depending on the current state of the referent. Yet other systems will supply auditory information almost continuously, with changes in the auditory signals being used by the operator in order to monitor and track the system. Semi-abstract and concrete warnings (see Section 1.2.2.) are likely to be more effective in these kinds of applications than traditional, abstract auditory warnings. Whether the sounds used in these kinds of application should be considered as auditory warnings is not clear. A more general definition of these sounds as monitoring or tracking sounds is likely to be more parsimonious in the longer term, with the definition of an auditory warning being confined to those sounds that operate only in an on/off state.

2. AUDITORY WARNING DESIGN

The process of tailoring an auditory warning to the appropriate background noise spectrum is a fairly complicated procedure as it depends on a number of factors including the degree of fluctuation in background noise levels, the psycho-acoustic profiles of the people working in that setting, and appropriate modeling of the auditory filter. Guidelines and accompanying software have been written to make the task more straightforward for the ergonomics practitioner (Patterson 1982, Laroche *et al.* 1991). The guidelines deal largely with the acoustic and psycho-acoustic principles involved, but also offer advice on psychological issues although the latter have been embellished in more recent literature.

2.1. Acoustic Issues

2.1.1. Audibility

In order for auditory warnings to be effective, they must be heard. They must also not be too loud as the likely effect will be to startle, cause annoyance, direct attention away from instead of

towards the task in hand, and the risk might be that the warnings will be turned off and then not be available when required. Their audibility will be determined by a number of factors, primarily the background noise in which they are to be heard. Thus it is essential to model auditory filter shape in order to predict masked threshold in a given noise environment. Once this is known, an appropriate band for auditory warnings and auditory warning components can be predicted. It is also possible to predict which warning (and other) sounds will be masked by the ambient noise; it is also possible to predict which sounds will be masked by other warning sounds. Two expert systems are available which allow this to take place. A system developed by Patterson (1982) for aircraft measures noise in small bands across the spectrum, superimposes Patterson's auditory filter model over these measurements to predict masked threshold, and then generates a band 15–25 dB above threshold which is deemed to be the appropriate band for auditory warning components. At 15 dB above masked threshold sounds should be audible; by 25 dB above threshold they will be hard to miss.

A second approach (Laroche *et al.* 1991) aimed at work areas such as factories, is to take one-third octave band measurements of both noise and warnings, then to generate the excitation pattern that would be produced if those sounds were presented to the ear, and then to compare excitation patterns. Here, if one excitation pattern completely covers another then the covered sound will be inaudible if heard simultaneously with the covering sound.

Both systems can be used to show whether auditory warnings in a particular environment will be either inaudible or too loud. They can also be used to predict how auditory warnings should be adjusted so that they are appropriately loud.

2.1.2. Localizability

In practice the localizability of auditory warning signals is an important issue. The ear uses two different mechanisms for localizing sound, depending on the frequency of the sound being heard. At low frequencies phase differences between sound waves entering each ear can be used to localize that sound. At high frequencies the side of the head which receives the sound first casts an "acoustic shadow," which also allows localization. For maximum localizability, auditory warnings should be either of relatively high or relatively low frequency. High frequency sounds will be very irritating and aversive, so ideally only relatively low frequency sounds or at least sounds with a relatively low fundamental frequency, should be used. In practice, the frequency of many auditory warnings is one that allows neither of the two mechanisms to function adequately. This problem can be addressed through redesigning those sounds, particularly adjusting the fundamental frequency to one which will allow localization. The addition of several components will also help in the process.

2.2. Psychological Issues

There is a large range of psychological issues that impinges on auditory warning design. Three central issues in practical terms are the number of warnings that the operator might need to deal with, the relationship between a warning sound and its meaning, and finally its believability. Each of these factors will contribute to determining the efficacy of both single, and sets of, auditory warnings.

2.2.1. Number

Generally, the number of auditory warnings used in any particular occupational or public setting should be kept to a minimum, as problems will ensue if large numbers are used. One problem is that the more warnings there are, the more likely it is that they will mask one another if they sound simultaneously. This may lead to confusion and/or a progressive increase in the loudness of sounds, neither of which is ergonomically desirable. The second of the problems is that the more auditory warnings there are in a set, the harder it will be to identify and learn them. A third problem is that if auditory warnings are used ubiquitously as information sounds (such as they might be in a nuclear power plant control room or an intensive care ward) then this will weaken their role as specific attention-getting sounds.

Ideally only the highest priority situations should be signaled by unique auditory warnings, that is where there is one-to-one correspondence between warnings and specific hazards and events. Situations of lower priority can be signaled by using a single warning denoting the category of risk only (such as second- or third-priority). On hearing the warning, the operator can then further identify the problem through other means, such as visual scanning. By using such a strategy, the number of auditory warnings can be kept to a minimum without compromising safety.

2.2.2. Relationship between auditory warning design and function

At the time of writing there is no clear consensus as to which category of sounds (abstract, semi-abstract or concrete; see Section 1.2.2.) are best suited for use as auditory warnings *per se*. There is little consensus also as to which sorts of sounds are best suited to which sorts of function. Traditional warning sounds have the advantage that they are generally recognized as warnings (although their specific meaning may not be clear) whereas semi-abstract and concrete sounds may be intuitively more easy to identify.

A more clearly defined, but more specific type of mapping between auditory warning and referent is that of *urgency mapping* (Edworthy and Adams 1996). Urgency mapping is a process through which warnings may be mapped on to referents so that the more important, hazardous situations and events are signaled by more urgent-sounding warnings. By such a process, some degree of cognitive compatibility can be achieved between warning and referent. Assessments of importance and hazardousness of risks must be generated by experts in the occupational settings in which the warnings are to be used. Sets of warnings can then be generated from existing databases showing the relationship between perceived urgency and acoustic design (e.g. Edworthy *et al.* 1991, Hellier *et al.* 1993). One advantage of appropriate urgency mapping is that even when the precise meaning of a warning may not be known, the degree of attention that the hearer should pay to the situation being signaled is conveyed through the warning. The same process can be used for visual warnings.

2.2.3. Believability

A warning's believability is a strong determinant of its effectiveness. Any warning, no matter how well designed, will be ignored if it continually sounds when there is no danger or hazard to be found. Thus false alarm rates are an important factor

in auditory warning design and implementation. Research evidence (Bliss 1995) shows that people will match their response rate quite accurately to a warning's reliability. Thus if a warning is 90% reliable (that is, for 90% of the time when a warning sounds, there actually is a problem requiring attention) then it will be responded to nearly 100% of the time. If a warning is only 10% reliable (so 90% of the time it signals when there is no actual problem to be attended to) then the warning itself will only be attended to 10% of the time. Thus it is important to keep false alarm rates as low as possible. Setting an auditory warning so that it sounds only when it should requires close collaboration between designers, engineers, manufacturers and experts in the environments where the warnings will be used.

3. AUDITORY WARNINGS AND OTHER TASKS

Generally auditory warnings are heard and dealt with in conjunction with other tasks. In many occupational settings responding to alarms is only one of many tasks to be performed simultaneously. The balance of providing warning information in an auditory, visual or other mode (e.g. tactile) needs to be carefully considered when systems are implemented. It is thus important to take account of multiple resource models of attention (e.g. Wickens 1992) and dual-tasking approaches to cognition and performance in order to establish the likely relationship between auditory warnings and the environment in which they are to be used.

REFERENCES

BLISS, J.P., GILSON, R.D. and DEATON, J.E., 1995, Human probability matching behavior in response to alarms of varying reliability. *Ergonomics*, 38, 2300–12.

EDWORTHY, J. and ADAMS, A.S., 1996, *Warning Design: A Research Prospective* (London: Taylor & Francis).

EDWORTHY, J., LOXLEY, S.L. and DENNIS, I.D., 1991, Improving auditory warning design: relationship between warning sound parameters and perceived urgency. *Human Factors*, 33, 205–31.

GAVER, W.W., 1993, How in the world do we hear? Explorations in ecological acoustics. *Ecological Psychology*, 5, 283–313.

HELLIER, E.J., EDWORTHY, J. and DENNIS, I.D., 1993, Improving auditory warning design: quantifying and predicting the effects of different warning parameters on perceived urgency. *Human Factors*, 35, 693–706.

LAROCHE, C., TRAN QUOC, H., HETU, R. and MCDUFF, S., 1991, "Detectsound": a computerized model for predicting the detectability of warning signals in noisy workplaces. *Applied Acoustics*, 32, 193–214.

PATTERSON, R.D., 1982, *Guidelines for Auditory Warnings Systems on Civil Aircraft*. No. 82017 (London: Civil Aviation Authority).

WICKENS, C.D., 1992, *Engineering Psychology and Human Performance* (New York: HarperCollins).

Augmented Reality

W. Barfield and E. B. Nash

Virtual Environment Laboratory, Department of Industrial and Systems Engineering, Virginia Tech, Blacksburg, VA 24060, USA

1. INTRODUCTION

An "augmented" or "merged" reality environment represents a scene with synthetic imagery combined with parts or all of a real-world scene (Figure 1). The purpose of this chapter is to review the design and use of augmented reality displays, discuss the advantages of augmented reality systems and to discuss some human factors issues that should be considered when creating augmented reality displays.

2. COMPONENTS OF AUGMENTED REALITY

Generally, to create a computer-generated augmented environment the following equipment is necessary: (1) hardware for creating and displaying visual images, (2) a position and orientation sensing system, (3) hardware for combining computer-generated images with either the real-world scene or video images of the real world and (4) the associated system software.

The two main types of head-mounted displays (HMD) used in augmented reality applications are a opaque or see-through HMD. Both types of displays are commonly used for a variety of tasks and have different advantages and disadvantages associated with them (Holloway 1997). Moreover, an augmented reality environment can also be created using a monocular HMD (one eye), bi-ocular HMD (same scene to two eyes), binocular HMD (two slightly displayed scenes shown to two eyes), or using a desktop computer or large-screen projection system.

2.1. Opaque Head-mounted Displays

As indicated, one type of display used to create an augmented reality environment is an opaque HMD. When worn over one eye, the user must integrate the real-world scene perceived with the uncovered eye with the graphical image(s) projected to the

Figure 1. Graphical image overlaid onto video image. The user is wearing an opaque HMD with a CCD camera and luminance keyer. Courtesy: Virtual Environment Laboratory, Virginia Tech.
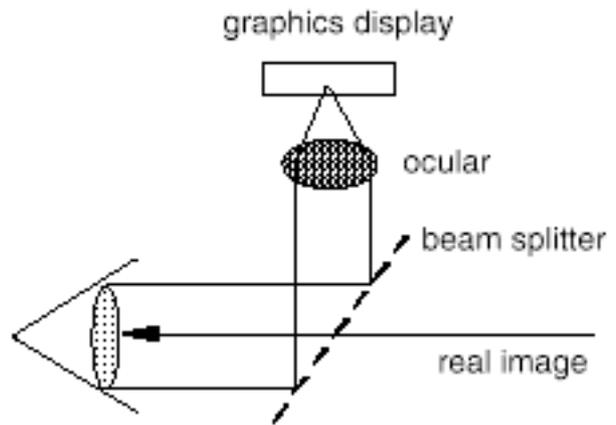
Figure 2. Diagram of a see-through display. The display includes an ocular that magnifies the displayed image that is then projected onto the retina.

other. However, when worn over both eyes, the user perceives the real world through video captured by mounted cameras. A computer then fuses (e.g. using luminance keying) video of the real world with graphical images to create a video-based augmented reality. The location of the cameras depends on the environment one wishes to augment. For instance, if the design objective is to augment the users immediate surroundings, then the ideal position to mount the cameras (typically CCD cameras) is near the user's eyes. In contrast, if augmenting is desired for a more remote environment, cameras may be mounted at the remote scene either on a robot or some ideal location with respect to the robot or other remote manipulator. Although using a single camera produces only a monoscopic (non-stereo) image of the real world, stereoscopic images may be obtained by using two cameras calibrated to provide the appropriate parallax.

2.2. See-through Head-mounted Displays

In contrast to opaque HMD, see-through HMD admit light from the outside environment allowing users directly to view the real world with the naked eye. In addition, half-silvered mirror systems fixed in front of the user's eyes reflect light from computer-generated graphic images (Figure 2). The resulting image is an optically combined view of the real world superimposed with graphical images.

2.3. Screen-based Displays

Provided that video images are used to capture the real-world scene, augmented reality scenes may be viewed using either an opaque HMD or a screen-based system. Screen-based systems can project images to users either using cathode ray tubes or by projection screens. In either case, stereoscopic images may be generated by viewing timed alternations between left and right eye views through a shutter system that occludes the left eye view when the right eye image is displayed and vice versa (Davis and Hodges 1995).

3. IMAGE REGISTRATION IN AUGMENTED REALITY

A typical requirement of augmented reality scenes is that the computer-generated images accurately register with the surroundings in the real world (Janin *et al.* 1993). Registration requires

accurate knowledge of the geometry of the real world, the computer-generated scene and the head movements of the user. One technique commonly used to register separate coordinate systems is to place a position sensing transmitter at a fixed and known location in the real environment while positioning the sensing receiver on a HMD. Then, as the user moves his head, the position and orientation of the user changes with respect to the transmitter. Using this information, updates to the virtual display can be computed based on the location of the virtual display with respect to the optics of the physical display.

4. MULTIMODAL DISPLAYS USED IN AUGMENTED REALITY

Although augmented reality displays typically combine only visual images with the real world, Barfield *et al.* (1995) proposed that auditory and haptic information can also be used to enhance the real world as well. For example, computer-generated sounds could be placed over objects in the real world to enhance their visibility or change their meaning. Furthermore, by overlaying computer-generated haptic feedback with real world objects, one will/might not only experience the sensations of the haptic feedback equipment but also of the real tactile and kinesthetic sensations from the object. This practice will likely lead to novel interface techniques for human interaction with complex systems.

5. ADVANTAGES OF AUGMENTED REALITY DISPLAYS

5.1. Conservation of Computational Resources

A major motivation for the use of augmented reality displays relates to the computational resources necessary to generate and update computer-generated scenes. In computer graphics, the more complex the scene the more computational resources are necessary to render the scene, especially for real-time applications. However, the concept of augmented reality is to enhance the real world with virtual objects. This approach does not require a scene that consists entirely of computer graphics. Instead, the graphics are used as a supplement to the real-world scene. Images created in this way can serve as an enhancement to the knowledge already existing in the real world (Feiner *et al.* 1993). For many conceivable applications the virtual images may require only wire frame graphics or text overlaid on top of the real-world scene (Janin *et al.* 1993). Such an approach may conserve computational resources that can be used for other tasks (e.g. networking).

An additional reason why it may be desirable to combine synthetic imagery with real-world scenes is to maintain the high level of detail and realistic shading that one finds in the real world. As noted, to model the complexity of the real world using entirely computer graphics would not only require tremendous computational resources, but also would be a laborious and time-consuming process to model such complex scenes. Thus, the incorporation of real-world objects in augmented reality will provide the designer with a rich visual scene, which then can be augmented with computer graphics.

5.2. Processing Capabilities of Video

One advantage associated with using video for augmented environments is the ability to further process or to edit the visual information captured by the cameras to enhance or highlight a particular portion of the image (Mann 1997). For example, parts of the video scene can be substituted with synthetic imagery. In addition, because the real-world image can also be easily changed by mounting the cameras at a remote site, this affords wide application for tele-operations tasks, especially when the site does not easily support the habitability of humans.

6. SELECTED HUMAN FACTORS CONSIDERATIONS FOR AUGMENTED REALITY DISPLAYS

6.1. Display Optics of the Camera Versus Display

Because video-based augmented reality utilizes cameras to generate the real-world scene, careful attention must be paid to the relationship between the display optics of the camera(s) and the visual parameters of the real-world scene. Generally, if video of the real world is to appear as if viewed through the user's eyes, the cameras should operate as if they are at the same physical location as the wearer's eyes. This is a difficult design task. For example, to accomplish this, the field of view must be matched between the computer-generated scene and real-world scene. Additionally, if two cameras are used to generate the real-world scene, additional considerations such as the horizontal disparity and convergence angle of the cameras should be considered. These parameters can be manipulated by adjusting the horizontal tilt and disparity of the display. The horizontal tilt represents the convergence angle of the display and the disparity represents the horizontal distance between the center of the two displays.

6.2. Limits in Graphical Display Resolution

When using see-through displays for augmented reality applications, it is desirable to match the resolution of the projected graphics with the spatial resolution of objects in the real world. However, unlike the real world, graphical image resolution is dependent on the pixel resolution afforded by the HMD, and thus is one limiting factor in the quality of visualization of the augmented scene. For example, some commercial HMD provide resolutions of 720 x 480 pixels in a 60° field of view, equivalent to 20/100 visual acuity in the real world. Disparities of this magnitude between synthetic imagery and the real-world scene would be quite obvious to the observer.

6.3. Frame Rate and Update Rate

Piantanida *et al.* (1993) define frame rate as a hardware-controlled variable determining the number of images presented to the eye s^{-1}. To contrast frame rate from update rate, consider a display that has a frame rate of 60 Hz. The update rate is the rate at which new (different) images are generated and presented to the viewer. Thus, if the update rate is only 3 Hz, the system will present 20 images of the same scene to the viewer before 20 images of the next scene are shown (Piantanida *et al.* 1993). In this way, only three different images s^{-1} will be displayed despite the frame rate of 60 Hz. Sufficient update rates will assist designers in accurately registering graphics projected in the real world with changes in the users head movements.

6.4. Sensor Delay

Piantanida *et al.* (1993) noted sensor delay as another type of system delay that affects the registration of the visual scene in

augmented reality. Sensor delay is the time required by the system to determine that the viewer has made a movement requiring an update to the display. In order for image registration to be as accurate as possible, head tracking sensors for augmented reality displays must not only be quick enough to register natural head motion, but also be accurate in tracking to a small fraction of a degree in orientation and a few millimeters in position (Azuma 1993).

6.5. Stereoscopic and Monoscopic Displays

With augmented reality displays, either the graphical images or real-world images can be stereoscopic, meaning that each eye views a slightly offset version of the same image. This quality (i.e. stereopsis) is known to provide enhanced perception of depth and three-dimensionality for a visual scene (Yeh and Silverstein 1992). However, because augmented displays work to combine two otherwise independent visual images, the stereoscopic cues provided by either may not completely support each other, and may even be in conflict (Milgram and Drascic 1997). Monoscopic displays do not suffer these complications because they lack the necessary depth cues. Yet users have been shown to make accurate distance judgements in cases where the head is allowed to move horizontally and the virtual objects are close to the viewer (Ellis *et al.* 1997).

6.6. Importance of Task

It is imperative when considering the capabilities of an augmented reality system that careful attention be paid to the task for which it is designed. For instance, medical applications such as registering CT images with the patient's anatomy will require quite accurate image resolution and registration between the synthetic imagery and the real world environment. In contrast, other applications may only require that the synthetic image or text appear in the observer's field of view, such as a maintenance task.

6.7. Mental Models

As noted, augmented reality displays consist of computer-generated imagery merged with the real-world scene. Thus, the designer must account for both the user's mental model of the synthetic imagery as well as the real image. That is, the scene should be designed in such a way that a single mental model will be formed of the augmented scene. Integrated mental models will likely result from improvements in the display parameters. Specifically, the augmented scene will likely be viewed as one singular environment as the graphics become more realistic, the two images become better matched in terms of brightness and color, and image registration improves.

Whether performance in augmented reality environments will improve as a result of such integration is not yet known. However, some preliminary information is provided by a literature review on head-up displays (HUD) (Stokes *et al.* 1990). This review suggests that it may in fact be difficult to switch attention between projected HUD symbology and the real-world scene. For example, pilots have been shown to turn off the HUD during critical phases of a mission. Thus, a relevant question to ask for augmented reality environments is the extent to which overlaid imagery may interfere with the visualization of the real-world image and performance of a task. Within this issue of interference is display clutter. Without an empirical basis for combing these two types of information (synthetic and real world), synthetic imagery may not enhance information presented in the real world but rather obscure important parts of the real-world scene. How attentional resources are allocated to augmented reality and/or the real-world image, and how display clutter affects performance are important topics for future research.

7. CONCLUSIONS

In this chapter we discussed concepts related to augmented reality. Several display technologies are available for the creation of augmented reality systems, which present advantages over conventional virtual reality systems. However, successful employment of any augmented reality system requires that the capabilities and limitations of the hardware, the task to which the technology is being applied, and the motor, perceptual, and cognitive aspects of the user all be considered. As display and tracking technologies develop along with our knowledge of the human capacity for perceiving and understanding displayed information, augmented reality displays will likely provide rich and innovative enhancements to the knowledge existing in the real world.

REFERENCES

AZUMA, R., 1993, Tracking requirements for augmented reality. *Communications of the ACM*, 36, 50–1.

BARFIELD, W., ROSENBERG, C. and LOTENS, W.A., 1995, Augmented Reality displays. In W. Barfield and T. Furness III (eds) *Virtual Environments and Advanced Interface Design* (New York: Oxford University Press), pp. 542–75.

DAVIS, E.T. and HODGES, L.F., 1995, Human stereopsis, fusion, and stereoscopic virtual environments. In W. Barfield and T. Furness III (eds) *Virtual Environments and Advanced Interface Design* (New York: Oxford University Press), pp. 145–74.

ELLIS, S.R., MENGES, B.M., JACOBY, R.H., ADELSTEIN, B.D. and MCCANDLESS, J.W., 1997, Influence of head motion on the judged distance of monocularly presented virtual objects. In *Proceedings of the Human Factors and Ergonomics Society*, *41st Annual Meeting*, pp. 1234–8, Albuerque, New Mexico. Published by the Human Factors and Ergonomics Society.

FEINER, S., MACINTYRE, B. and SELIGMANN, D., 1993, Knowledge based augmented reality. *Communications of the ACM*, 36, 39–49.

HOLLOWAY, R.L., 1997, Registration error analysis for augmented reality. *Presence: Teleoperators and Virtual Environments*, 6, 413–32.

JANIN, A.L, MIZELL, D.W. and CAUDELL, T.P., 1993, Calibration of head-mounted displays for augmented reality applications. In *Proceedings IEEE Virtual Reality Annual International Symposium*, 18–22 September, Seattle, WA, IEEE Publisher, New Jersey, pp. 246–255.

MANN, S., 1997, Wearable computing: a first step toward personal imaging. *Computer*, 30, 1–10.

MILGRAM, P. and DRASCIC, D., 1997, Perceptual effects in aligning virtual and real objects in augmented reality displays. In *Proceedings of the Human Factors and Ergonomics Society*, *41st Annual Meeting*, pp. 1239–43, Albuerque, New Mexico.

PIANTANIDA., T., BOWMAN, D.K. and GILLE, J., 1993, Human perceptual issues and virtual reality. *Virtual Reality*, 1, 43–52.

STOKES, A., WICKENS, C. and KITE, K., 1990, *Display Technology: Human Factors Concepts* (Warrendale: Society of Automotive Engineers).

YEH, Y.Y. and SIVERSTEIN, L.D., 1992, Spatial judgements with monoscopic and stereoscopic presentation of perspective displays. *Human Factors*, 34, 538–600.

Automatic Speech Recognition

H. M. Khalid† and M. G. Helander‡

†Institute of Design and Ergonomics Application, Universiti Malaysia Sarawak

‡Graduate School of Human–Machine Interaction, Linköping University, Sweden

1. INTRODUCTION

Automatic speech recognition (ASR) is a 25-year-old technology whose time may finally have come. It is a technology for communicating with a machine using spoken words or phrases. Lernout and Hauspie, one of the forerunners in speech recognition technology, have launched a variety of speech products that enables people to control remotely the central heating, lighting, burglar alarms, cookers and video recorders with speech commands. ASR will increasingly be used for home banking, transactions on the Web, and even as security for hole-in-the-wall cash machines. Because every human voice is different, speech is considered secure.

In the past decade, there has been great advances in ASR (e.g. Westall *et al.* 1998). The error rate has reduced by more than a factor of five and the speed in recognition has increased by several orders of magnitude — brought about by a combination of faster recognition search algorithms and more powerful computers. These factors combine to make high-accuracy, speaker-independent, continuous speech recognition for large vocabularies possible in real time. The improvements in speech recognition performance are due to four factors: use of common speech corpus, improved acoustic modeling, improved language modeling, and a faster research experimentation cycle (Makhoul and Schwartz 1994).

1.1. ASR Applications

Speech recognition is now being used in many "real-world" applications (e.g. machine translation services on the Internet), despite the perception that the current technology is still not flexible enough to allow easy and natural communication with machines. Situations in which the use of speech recognition systems may be advantageous include:

- When the user's hands or eyes are busy (e.g. mission pilots, parcel sorters, VLSI circuit designers).
- When only a limited keyboard and/or screen is available (e.g. telephone operator services, banking transactions).
- When the user is disabled (e.g. motor impaired users may control household appliances and wheelchairs).
- When pronunciation is the subject matter of computer use (e.g. foreign language learning, teaching of reading).
- When natural language interaction is preferred (e.g. database access).

To date, efforts are continuing in the advancement of speech recognizers for a variety of telecommunication and computer-
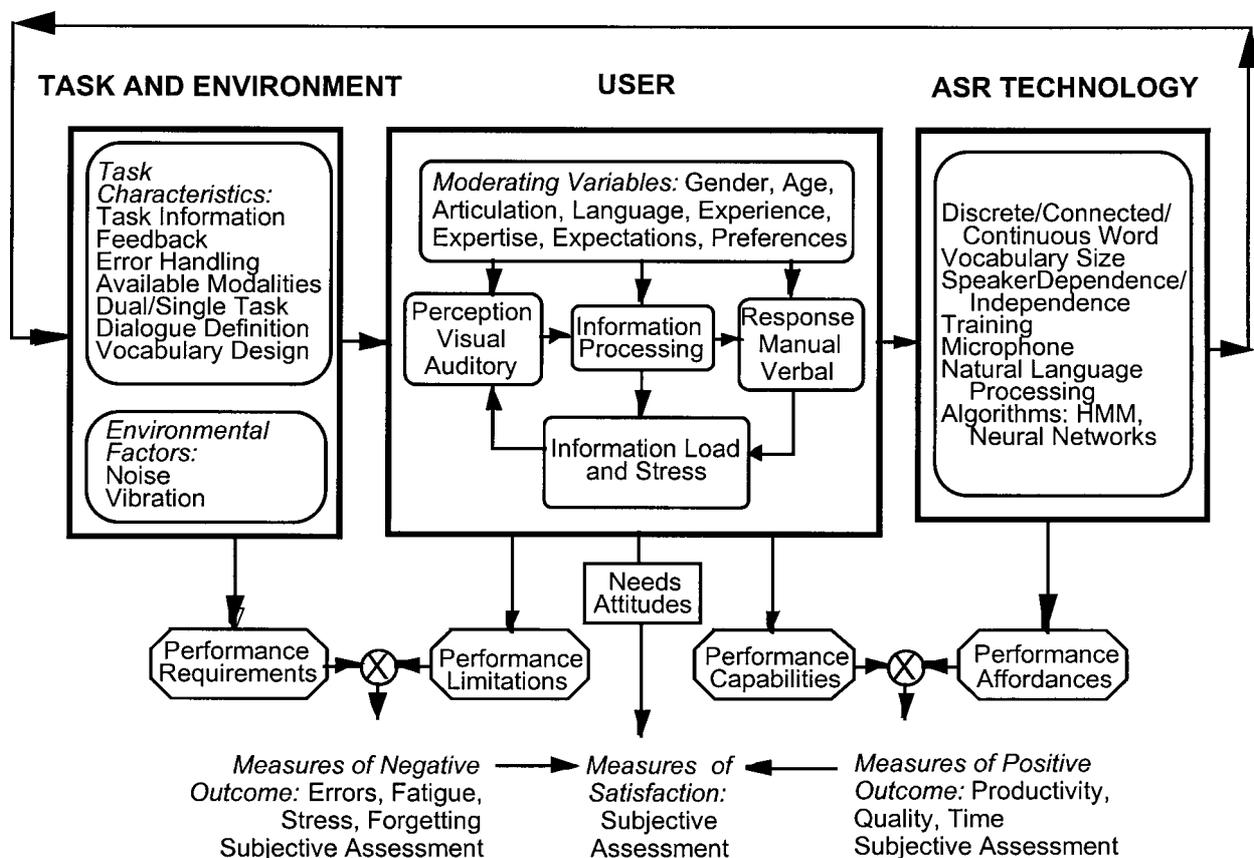
Figure 1. Systems approach to automatic speech recognition.

based tasks. E-mails can now be translated automatically into the language of the recipient. Other facilities include: voice dialing (particularly useful in cars), automatic call-centers, computerized operator services, and messaging that allows users to pick up and send e-mails and faxes on an ordinary telephone. With improved computer processing power, memory and battery life, there will be greater demand to use speech with palmtop computers, dictation machines, organizers and smart telephones, thereby exploiting speech recognition technology in its own right. ASR systems will be more successful in the future if they can solve specific tasks, such as dictation for medical doctors, lawyers and research scientists.

1.2. ASR Types

There are two types of ASR: speech recognizers and speech understanding systems (or natural language understanding). The task of a speech recognizer is simply to identify the words spoken without interpreting their meaning. Recognition would be very useful if speech recognizers could identify words in fluent speech for any talker. Some of the more advanced recognizers on the market today approach this goal. Very few speech-understanding systems are currently being marketed. In addition to associating the acoustic input with a particular vocabulary item, speech understanding systems need syntax (parsers) and semantics (akin to a contextual dictionary) to assign meaning to speech input.

This chapter presents human factors associated with speech recognition only. The discussion is organized around major issues that are highlighted in Figure 1, in particular, the capabilities and limitations of speech technology, the characteristics of speech use, and the design strategies for speech interfaces. We also identify some human factors research that needs to be addressed to further the development of speech recognition systems. At the bottom of Figure 1 are three different kinds of measures of system performance: measures of positive outcome (efficiency), negative outcome (error) and measures of satisfaction (attitudes). These measures derive from comparisons of performance requirements imposed on the user, and performance affordances offered by the system. The measures are central to usability assessment. For example, the measures of satisfaction (i.e. attitudes) rest on subjective assessments of user expectations of ASR performance as compared with the actual measures of negative and positive outcome.

As depicted in Figure 1, the task and environment, the user and the ASR technology form a control system with error feedback. To optimize the performance of this system and to create successful applications, the user interface designer must consider not only the capabilities of the technology, but also the effects of the task requirements, as well as user's expectations, expertise and limitations. The technology will be described first, and then user and task considerations.

2. ASR TECHNOLOGY

A recognition system consists of the hardware and software for storing a library of words and for comparing incoming words to the library. The ASR is first trained, meaning that features of words are extracted and stored in the library. Once a library of words has been stored, new spoken words will be compared with those already stored (Figure 2). A set of algorithms compares a feature pattern representation of the input word to a library of stored reference patterns, which have been obtained by the training process. Algorithms are used to decide which of several possible words gives the closest match.
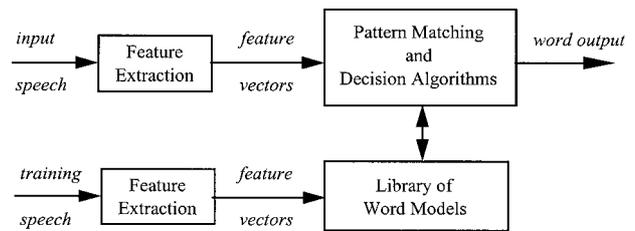
Figure 2. Basic components of a speech recognition and training system.

There are hence four basic components of a speech recognition system. First, to know what a talker has said, the system must have some way of encoding and representing a set of utterances. Many systems represent the spectral-temporal properties of an utterance. Second, during recognition a pattern-matching algorithm compares a representation of a particular input utterance with the representations in the library of words. Many current speech recognizers use the technique of Hidden Markov Modeling (HMM) for pattern matching. An HMM model is based on a statistical representation that helps the recognizer to cope with most of the variability in human speech. The newest algorithms are based on neural networks (Kewley-Port 1995). Third, given the comparison of some input with the library vocabulary, an algorithm decides which utterance in the vocabulary was produced. Finally, to support the functions and operation of the recognition system, there must be a user interface (Nusbaum *et al.* 1995).

2.1. Characteristics of ASR

ASR are classified according to their capabilities. *Speaker-dependent* systems must be "trained" by an individual user who, using a microphone, records samples of each word. *Speaker-independent* systems can recognize speech for many users. This is done by training a large corpus of utterances spoken by many different people with different types of speech characteristics. *Speaker-adaptive* systems are speaker-independent systems that have the additional feature of updating the system during an interaction with a user, based on small samples of user speech. Thereby the match between the models in the system and the user's speech is improved (Kamm and Helander 1997). Most large vocabulary dictation systems use speaker adaptation techniques.

Some speech recognizers consider only isolated-word or *discrete speech*. These require a brief pause after every word or may allow only one word as input. Systems for *connected speech* input accept a sequence of concatenated words. Systems for *continuous speech* accept fluent speech without pauses.

Some systems recognize a large number of words (tens of thousands) or phrases, while simple systems may recognize only a few words, such as the digits 0–9. Some systems permit vocabulary switching so that different vocabularies are available contingent on the outcome of recognizing an utterance. This can be used in a hierarchical menu of choices. After each word the computer will consider a constrained set of new words, which enhances recognition.

2.2. Speech Recognition Accuracy

Vendors often describe their speech recognition hardware as offering very high recognition accuracy, but it is only in the context

of a realistic field test with noise and disturbances, that one can meaningfully compare the performance of recognizers. Recognition error rates below 1 percent are typically obtained for highly constrained vocabulary and controlled speaking conditions, but for large-vocabulary (e.g. 5000 words), connected-speech systems, the error rate may > 5% (Cohen and Oviatt 1994). Clearly, different applications will require different capabilities. Controlling a wheelchair might well be achieved with a simple system using only a few commands, but high reliability is crucial.

The major obstacle to high-recognition accuracy is the large variability in the speech signal characteristics. This has three components: linguistic variability, speaker variability and channel variability. Linguistic variability includes the effects of phonetics, phonology, syntax, semantics and discourse on the speech signal. Speaker variability, that is intra- and interspeaker variability, also includes the effects of *coarticulation*. In this case, neighboring sounds affect the acoustic realization of a particular phoneme, due to continuity and motion constraints on the human articulatory apparatus. Speaker differences such as dialect, speaking rate, and voice characteristics (e.g. pitch, hoarseness) require solutions that enable the recognizer to "learn" to recognize a particular voice. Channel variability includes the effects of background noise and acoustics, and the transmission channel (e.g. microphone, telephone). The recognition process must unravel all these variabilities.

2.2.1. Recognizer errors

Recognizer errors may be of four types (Khalid 1989, Kamm and Helander 1997). *Substitution* (or misrecognition) error occurs when the speech recognizer matches the user's utterance to the wrong vocabulary item (e.g. the user says "line" but the machine identifies the word as "nine"). *Rejection* error occurs when properly spoken words that are part of the active vocabulary are detected but not recognized, *deletion* error when an utterance that was spoken was not detected, and *insertion* (or false acceptance) error when the device fails to reject non-speech sounds (e.g. background noise, cough, sigh) by falsely matching them against a stored template.

A user interface must have a mechanism for error correction. First, the user must be able to see (or hear) that an error has occurred. Second, the user should be allowed to correct the error before it creates a problem in task performance. When the recognition system detects an error, it must present an error message visually on the screen (e.g. a blinking cursor) or aurally by using *synthetic speech* or a beep. For systems using continuous speech, error correction can become cumbersome since there is a question of what word to correct.

2.2.1.1. Vocabulary size

With a large vocabulary more recognition errors are likely. Substitution errors can be reduced by minimizing the number of words in the vocabulary as required by the task, and by ensuring that the chosen words are acoustically dissimilar. Increasing the number of templates per utterance (e.g. three to 10 samples per word) increases the chances of correct recognition. The user can manipulate the acceptance or rejection rate of spoken words by setting the reject threshold to different percentage values.

3. USER ASPECTS

A number of human factors issues arise in the design and use of speech recognition systems. Inconsistencies in verbalization lead to recognition errors. User inconsistencies may be due to several factors:

- *Stress and fatigue*. During a working day, the voice may change because of morning hoarseness and vocal fatigue as well as general fatigue. There may also be stress and psychological fatigue — particularly if there are many error corrections. Fatigue and stress affect the "task" voice. Also, memory failures due to large vocabulary size can induce stress. Using words familiar to the user may possibly alleviate this.
- *Attitude and expectation*. As with much novel technology, users tend to have unrealistically high expectations of the system. It may take a month to learn how to handle ASR, and users must adopt a more appropriate and consistent attitude towards ASR. Users must be aware of how the system works and learn how to compensate for the weaknesses in ASR.
- *Experience*. Experience with ASR and its application has a major influence on recognition accuracy. As an example, users learn to adapt their speech to the computers. With increasing experience, they automatically insert the necessary pause between words that is required for a discrete speech system.

Failure to use ASR systems consistently is also caused by idiosyncratic factors, such as gender, age, articulation, regional accent and speech habits (e.g. hm, huh). In the past, ASR algorithms were developed using the voices of the system developers, typically males with an American accent. They are not representative of the user population, and a generic approach must be taken.

3.1. Prompts and Instructions

Prompts are used to elicit spoken user responses. For example, in a banking transaction via an automated teller machine, the system may prompt the user: "If you have a personal account say 1 and a company account say 2." A careful construction of prompts (e.g. semantic and prosodic cues) can result in a productive interaction. Prompts can also be used after an ambiguous input or error condition, in order to help the user modify the next input to improve recognition. The user must also know what is expected and how the system can provide instructions. For simple tasks, the instructions may consist of a single prompt. For complex interactions, the instructions may be detailed and lengthy.

3.2. Error Recovery Strategies

Error recovery strategies may be initiated by the user or by the system, for example by flagging the errors. Users prefer error recovery techniques that require the least amount of effort. For example, during partial recognition failures, the system could repeat what was recognized and then request, by adjusting prosody (pitch, loudness, duration), that the user only repeat the section of the utterance that it has failed to recognize.

4. TASK AND ENVIRONMENTAL ASPECTS

For applications involving disabled users there may not be a choice available, so the decision to use ASR is well argued. For other applications, ASR must be carefully assessed to ensure that it is more efficient and reliable than manual input. Many early

applications focused on hands-busy, eyes-busy tasks. One typical example is the mission fighter pilot. However, the high recognition error rate of early systems made error correction cumbersome, stressful and time consuming. Since error rates have reduced, there are now new prospects for ASR.

The first assessment could be a task analysis, where necessary user input, system output and feedback from error correction are listed. Given the characteristics of the task, such as frequency of input and feedback, one can allocate control actions to either voice or manual, and feedback to either visual or auditory functions — all for the purpose of optimizing task performance.

Dialogue definition must consider the semantics of the task — is there a persistent mental model that can be utilized to select how voice commands can be sequenced during task performance to improve the "rhythm" of the task. The choice of sentences or words depend on the mental model. In addition, one must also consider that they are easy and "natural" to memorize, and that they produce distinctly different profiles in the ASR equipment, so as to minimize substitution errors.

4.1. Task Knowledge

There are two aspects of task knowledge. The first concerns training the user on how to use the device so as to maintain a consistent "task" voice during performance. This not only provides a knowledge of what constitutes successful recognition but also the procedural knowledge involved in device use. The second relates to the task at hand. By clearly specifying what the task entails and what is to be expected, first-time users may focus on the task itself and not the device.

4.2. Environmental Characteristics

Environmental characteristics relate mainly to background noise. In avionics, for example, high levels of acoustic noise, high vibration, and poor voice-communications can characterize noise. On the factory floor and office, continuous noise from machinery, air conditioning, or intermittent noises from other people, telephones, printers, etc can characterize noise. Because of its variability, intermittent noise is more difficult to compensate for. A hardware solution to background noise is to use unidirectional noise-canceling microphones.

5. DESIGN OF ASR INTERFACES

People commit errors regardless of whether they are interacting with humans or with an automated system. Errors may be reduced through a thoughtful and usable design. In developing a user interface, principles of ergonomics design should be applied, including: focusing early on the users and the tasks they will perform, collecting performance data early in the design process by using simulations and prototypes, improving design at an early stage through design iterations. During the design process, usability principles should be considered, such as: use simple and natural dialogue, speak the user's language, minimize user memory load, provide consistent voice command structure, provide feedback, shortcuts and good error messages, as well as prevent errors (e.g. Baber and Noyes 1993, Karis and Dobroth 1995, Kamm and Helander 1997).

To date, the role of the system image for a speech recognition device has not been explored systematically (Nusbaum *et al.* 1995). As the complexity of recognition systems increases,

efficient use of these devices will depend on interfaces that modulate the user's expectations about how to speak, about recognition performance, and about how to control the recognizer's functions.

6. FUTURE RESEARCH

Research and development has increased during the past decade. This reflects the increased interest, in particular from telephone and computer companies. Below are described a few potential research areas in human factors. For research relating to technological developments, see Atal (1994) and Westall *et al.* (1998).

- *Multimodal interactive systems*. These systems, in which speech is only one component, may provide more flexible and robust interfaces, besides providing more options for error avoidance and error correction by allowing users to switch to another modality when errors occur. Some of the research issues to be addressed include: conceptualizing interactions among communication modalities, defining methods of evaluating the performance of multimodal interfaces, and identifying integration issues that arise when multimodal input is permitted.
- *Improving speech recognition by recognition of user intent*. Research can be undertaken to model task-oriented dialogue to identify the structure of dialogue. Many professional tasks are procedural in nature, and they are therefore fairly easy to model. Given that the system can keep track of user input, it may also predict the next action. Goal-plan graphs have been used to predict the intentions of operators. Such technique could be used to predict what the operator would say in the future. This constrains the dialogue options, and ASR would improve.
- *Spoken language understanding*. Research supported by DARPA (Defense Advanced Research Program Agency) has developed systems that can deal with the idiosyncrasies of natural speech. Challenging problems remain, such as "repair of spontaneous speech," "detection of relevance to spoken topic" and "use of discourse and dialogue theory to create conversational interfaces."
- *Use of mental models for dialogue definition*. Experienced operators have distinct mental models of the task. Given that the mental models of several operators coincide, one might define a dialogue using a natural phraseology and natural concepts. Experienced operators could be trained to build up unitary mental models, whereby one might reduce the training requirements for ASR.

7. CONCLUSIONS

ASR will not succeed unless the system is well designed to consider variations in operator speech as well as task behavior. An appropriate model for ASR should be based on task requirements and should consider environmental constraints. Procedural guidelines for ASR design need to be developed so that implementation of ASR can appropriately consider the human factors and mental models of users.

Given the limitations of the speech recognition algorithms today, further research is needed to improve the robustness of speech recognition, to incorporate higher level linguistic knowledge, and to determine how best to integrate speech into

efficient multimodal interactive systems. The ultimate systems should be capable of robust speaker-adaptive continuous speech recognition. There should be no restrictions on vocabulary, syntax and semantics.

REFERENCES

ATAL, B.S., 1994, Speech technology in 2001: new research directions. In Roe and Wilpon (eds), *Voice Communication between Humans and Machines* (Washington, DC: National Academy Press), 467–81.

BABER, C. and NOYES, J.M. (eds), 1993, *Interactive Speech Technology: Human Factors Issues in the Application of Speech Input/Output to Computers* (London: Taylor & Francis).

COHEN, P.R. and OVIATT, S.L., 1994, The role of voice in human–machine communication. In D.B. Roe and J.G. Wilpon (eds), *Voice Communication between Humans and Machines* (Washington, DC: National Academy Press), 34–75.

KAMM, C. and HELANDER, M., 1997, Design issues for interfaces using voice input. In M. Helander, T. Landauer and P. Prabhu (eds), *Handbook of Human–Computer Interaction*, 2nd edn (Amsterdam: Elsevier), 1043–60.

KARIS, D. and DOBROTH, K.M., 1995, Psychological and human factors issues in the design of speech recognition systems. In A. Syrdal, R. Bennett and S. Greenspan (eds), *Applied Speech Technology* (Boca Raton: CRC Press), 359–88.

KHALID, H.M., 1989, Human factors of integrating speech and manual input devices. The case of computer aided design. PhD thesis, University of London.

KEWLEY-PORT, D., 1995, Speech recognition. In A. Syrdal, R. Bennett and S. Greenspan (eds), *Applied Speech Technology* (Boca Raton: CRC Press), 341–58.

MAKHOUL, J. and SCHWARTZ, R., 1994, State of the art in continuous speech recognition. In D.B. Roe and J.G. Wilpon (eds), *Voice Communication between Humans and Machines* (Washington, DC: National Academy Press), 165–98.

NUSBAUM, H.C., DEGROOT, J. and LEE, L., 1995, Using speech recognition systems: issues in cognitive engineering. In A. Syrdal, R. Bennett, and S. Greenspan (eds), *Applied Speech Technology* (Boca Raton: CRC Press), 127–94.

WESTALL, F.A., JOHNSTON, R.D. and LEWIS, A.V. (eds), 1998, *Speech Technology for Telecommunications* (London: Chapman & Hall).

Chinese Keyboard Input

B. Han and P. Sun

Institute of Psychology, Chinese Academy of Sciences, Beijing 100101, PR China

1. INTRODUCTION

Computer processing information is represented in binary digits. This is suitable for processing basic alphabetic language with a limited number character set, e.g. English. However, it is not so easy to process Chinese information with > 60 000 characters in the whole set. Even if the coding capacity of the binary code system was enough for Chinese characters, the coding combination of up to 60 000 is no doubt an unbearable burden for normal users with limited capacity of attention, memory and cognitive processing. Therefore, a special coding strategy is needed to process Chinese.

The unique square feature in each Chinese character is quite different from the serial, linear word of the alphabetic language. This is possibly the main reason that Chinese word-processing was delayed ~10 years. The first English word-processing system was developed in the 1960s, but the first Chinese coding system developed in mainland China was applied to the computer in 1978. During past two decades, systematic research was conducted with respect to the Chinese language, the computer and the user. Various features in Chinese, such as initial, final, stroke, component and component combination, were used as elements to encode the character/word to the international standard small keyboard according to specifically designed rules.

Recent research has improved the coding technology greatly in the areas of speed, usability and learnability. This chapter will introduce some fundamentals and coding strategy of Chinese keyboard input (CKI).

2. FUNDAMENTALS OF CKI

Chinese can be input through varying size keyboards, by handwriting, by clicking on the computer screen, by optical character reader (OCR) or by recognizing speech. This chapter focuses on coding input using the universal international standard small keyboard with 26 letter keys, 10 number keys and other symbol keys. This is the primary method of most Chinese users in daily word-processing work (Chen and Hu 1994).

The aim of the CKI is to transfer the written text in a manual way or in the user's brain to the binary internal code text that can be processed by the computer. The standard international small keyboard is used as the human–computer interface. Figure 1 shows the program flow of Chinese keyboard inputting.

When compared with English keyboard input, stages 1, 2 and 4 are unique in CKI. Stage 5 is much more complex than the letter-internal code transforming in English inputting. This increases the workload for both user and computer.

Coding systems for CKI can be divided into rational and irrational codes. In the irrational coding system, each character is represented by a certain combination of numbers, e.g. "1412" represents the Chinese character "jiang". This is based on the Wade System invented by Thomas Wade in the nineteenth century for telecommunication using Chinese. There is no coincident code in the irrational coding system for the whole Chinese character set. But the user must refer to a special dictionary to relate the code to a specific character. Obviously, this is a difficult practice in daily Chinese word-processing work.

In the rational coding system, which is the focus here, each character is represented by a type of component combination that can be decomposed from the character be phonological or orthographical element (i.e. initials, finals, strokes or components). Users become familiar with the components through both school education, and daily reading and writing, so the rational coding system is easier to learn, remember and use. However, there are coincident codes for different characters, meaning that the users have to choose a target character between a group of characters sharing the same code. See Table 2 for examples.

The rational coding system was designed based on statistical research on the phonological and orthographical features of Chinese characters/words. The system can use either the phonetic or orthographic feature or use both of them to form the code-element set. This set is based on the operator's idea about how many elements should be used, how each character should be decomposed in the these elements, how many coincident codes are acceptable in the system, how many keys on the keyboard should be used, and how the elements should be mapped on the keys. Table 1 show some example features of the Chinese character/word. A great deal of adjustment is needed before the final arrangement is determined.

No matter how perfect the coding system is, the obvious mental load for the user is great. The user must remember the code-element set, the mapping rule between the code elements and keys in addition to possible exceptions. In a general sense, all the coding systems try to release the user's mental workload through using various coding strategies. As seen below, this has been proved very difficult and the only solution seems to be a type of balance between advantages and disadvantages.
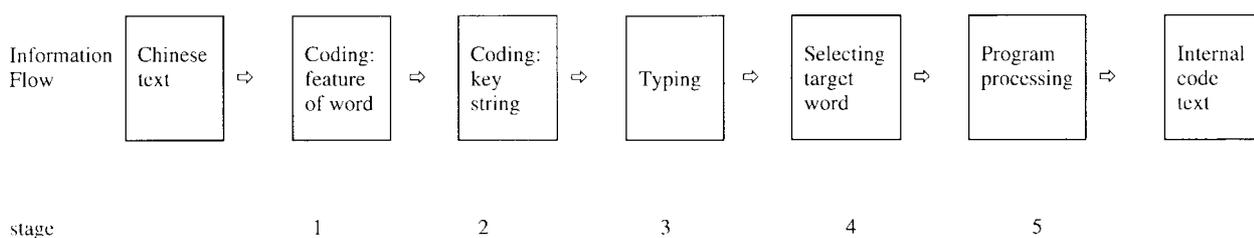
Information Flow	Chinese text	⇨	Coding: feature of word	⇨	Coding: key string	⇨	Typing	⇨	Selecting target word	⇨	Program processing	⇨	Internal code text
stage			1		2		3		4		5		

Figure 1. Flowchart of Chinese keyboard inputting (adapted from Chen and Hu 1994).

Table 1. Features of Chinese character shown by character '将'

General features	Detailed features	Example (Simplified, 将)
Phonological(jiang, qiang, yang)	Initials	j, q, y
	finals	iang, ang
	tone	1, 2, 4
Orthographic (将)	strokes (number and relation)	丶,/,丨,丿...(9, separated, attached, or cross)
	components (radicals)	爿,夕,寸 (爿,寸)
	orthographic structure	left-right
Semantic	word-building ability	general (将军), etc.
	entries	more than 40
	grammatical roles	Adverbial (will), etc.
	context	General as noun can be subject
Token frequency (‰)	character	0.85
	component	2.57 (爿), 7.04(夕), 24.12(寸)
	component combination	1.04 (爿-夕), 0.92(爿-寸), 0.92(夕-寸)

3. CODING STRATEGIES IN CHINESE KEYBOARD INPUT

Some of the key issues in developing the CKI system are: What type of features should be used and how should they be selected? What type of mapping relation between feature elements and keys should be used and how should it be defined? What type of rules should be used to form the key element sequence for inputting Chinese characters/words?

The operator of the CKI system has many tasks thereupon. These tasks include: construct the feature set, compute the coincident code, define the code and encoding rules, define codes for common words/characters, determine code length, determine element set and encoding space, distribute features on the keyboard, and develop the human–computer interface.

There have been three stages in the historical process of technical development concerning Chinese coding and inputting strategies, as revealed by three different encoding units: character, word, sentence. During the 1970s, single characters were the main encoding unit for input, but since the 1980s the word was the main encoding unit for input and in the 1990s the sentence has been the main encoding unit for input. Coding strategy often characterized by its emphasizing on a certain encoding unit.

3.1. Single Character as Encoding Unit

3.1.1. Orthographic versus phonological feature coding

Coding systems developed in the early stages were focused on the feature of single characters. For a character, the orthographic features used as the code-element could be either a stroke, an order of strokes from the writing system, a radical component, a component combination or even a structure type (left–right, up–down, embedded) (Han and Ren 1995). Additionally, the system could also use a phonetic feature like initials or finals according to the Chinese Pinyin system (Scheme of the Chinese phonetic alphabet, proposed by the Chinese government in 1958, in which each character is represented by initial and final) for notating Chinese phonologically.

The idea behind this system consists of mapping the elements to keys according to specially designed rules the user must remember to use the coding system. Different operators used different mapping strategies. Some operators based their systems on orthographic features, such as the Biao–Xing–Ma and Wu–Bi–Zi–Xing coding systems, whereas others used phonetic features such as those in the Zi–Ran–Ma and Zhi–Neng ABC coding systems. But most operators use both orthographic and phonological features to balance the advantages and disadvantages between these two types of elements.

3.1.2. Advantages versus disadvantages

The coding system using phonological elements is easy to learn and remember because it is derived from the Pinyin system which most Chinese students learn early in their schooling. In addition, there is almost no memory load for the decomposing the Pinyin of a character. The user only needs to remember the mapping rules between elements and keys. This memory can be transferred to an automatic kinetic combination after a short period of practice. However, the disadvantage for the phonological elements is the inevitable coincident code since only 21 initials (such as zh, ch, sh, etc.), 35 finals (such as -eng, -ang, -ong, etc.), and four tones can be used as elements to encode thousands of characters.

The coding system using orthographic elements has less problem with coincident codes, fewer characters share the same code. This decreases the possibility of choosing target characters from the coincident code group, thereby speeding the typing process. However, the user must remember the code-element set that may be different between different coding systems. The user must also remember the special mapping rules between elements and keys, especially those elements that share the same key since > 100 elements are mapped onto 26 letter keys.

The phonetic coding system has obvious advantages to the orthographic system with respect to ease of learning and use. Because China is such a large country with many dialects, the people from the northern region cannot communicate with those from the south even though they use the same written Chinese system. These dialect differences have actually caused the phonetic system to loss popularity to the orthographic coding system. So the popularization of Manderian (Putonghua) is necessary for the popularization of CKI system.

3.1.3. Typing situations and coding methods

Daily users of the Chinese coding system type in a variety of circumstances, while transcribing dictation, while typing from a written page, or while mentally contemplating prose to be written. The phonological coding system is suitable for typing from dictation or from mental contemplation since the auditory modality is involved in processing phonological information. On the other hand, the orthographic coding system is best suited to typing from a written page since the visual modality is involved in processing this visual information. The users must determine themselves which coding system they prefer. There is always a balance between ease of learning, remembrance and use. They must determine their personal "pros" and "cons". As the idiom says, "there is no free lunch", so there is no perfect coding system that can simultaneously solve all the problems at the character encoding level.

3.2. Word as Encoding Unit

A Chinese word can be one single character or over two characters, compound words. When the compound word is taken as the encoding unit, i.e. use of the component in each character as a code element by the similar coding strategy as single character inputting, the input efficiency improves compared with a single character input. Table 2 shows some codes for different encoding units.

The phonetic coding system normally uses the initial of each character as an element. The orthographic coding system uses the first stroke or component in each character. However, to decrease the likelihood of coincident coding, the other element, such as the last stroke or last character component, is often used.

When using the word as the encoding unit, the code length is shortened, the input is more efficient and timely, and the coincident code is decreased. The word as encoding unit also corresponds to normal thinking and writing. The coding rule, word as encoding unit, is simplified, therefore, decreasing the mental load during input and lessening the error rate. However, the word as encoding unit cannot completely replace character inputting since most single character words have a high token frequency (over 50%) in the language (Wang *et al.* 1986, Liu *et al.* 1990).

3.3. Sentence as Encoding Unit

For the sentence as an encoding unit, Pinyin inputs the whole sentence and the computer transforms it to the orthographic system automatically based on its knowledge database, selecting of words and grammar-matching technology. To system users, his process actually realizes the no coding input of Chinese characters similar to the alphabetic language. This improvement benefits natural language comprehension and automatic translation. However, there is still a need for character or word input performance in daily word-processing for tasks such as typing tables or figures. This means that three types of encoding strategy must compensate each other.

4. DISCUSSION

4.1. Achievement of CKI Research

The role of research on CKI has evolved over the past two decades by solving some of the main problems in CKI. After the initial developments on technical issues such as program and hardware design, the quantitative research has centered on the Chinese

Table 2. Encoding unit, coding strategy, coding system, code and coincidence code in CKI system

Levels	Endocing unit (example)	Coding strategy (example)	Coding system	Decomposing Code	Characters shared the same coincident code
Character (将)	phonology (jiang)	whole syllabification	jiang	jiang	将,江,姜,浆,僵,疆,缰
		reduced syllabification	j, ang	jh	
	orthography (将)	strokes	丶,/, 丨, 丿, 乛, etc.	123451, etc.	将, 奖, etc.
		components	扌, 夕, 寸	jxc	no
word (将军)	phonology (jiangjun)	whole syllabification	jiangjun	jiangjun	no
		reduced syllabification	jiang, jun	jj	将军, 渐渐, 建交, 监禁, 坚决, etc.
	orthography (将军)	strokes	丶,/, 丨, 丶, 乛, ⼀	123167	no
		components	扌, 冖	jn	no
Sentence (将军指挥打仗)	Phonology (jiangjun zhihui dazhang)	Whole syllabification	jiangjun zhihui dazhang	jiangjun zhihui dazhang	将军指挥打仗, 将军只会打仗, 将军智慧打仗
		reduced syllabification	jiang jun zhi hui da zhang	jjzhdz	将军只好打字, etc. in addition to above three sentences
	orthography (将军指挥打仗)	strokes	丶, 丶, ⼀, ⼀, ⼀, 丿	117774	no
		components	扌, 冖, 扌, 扌, 扌, 亻	sntttr	no

Note. Following description is for the situation when character is the encoding unit, but it can also be used to infer the word or sentence.

1. In phonological coding strategy, whole syllabification code is just the same as the decomposing result; In reduced syllabification, letter 'j' and 'h' correspondent with initial 'j' and final 'iang' respectively.

2. In orthographic coding strategy, stroke '丶', '/', '丨', '丿', '乛', '⼀' correspondent with number '1', '2', '3', '4', '5', '6', '7' respectively; while component '扌, 夕, 寸' correspondent with 'j, x, and c' respectively because of phonological similarity (phonological mapping). Orthographical similarity (orthographic mapping) is shown by mapping '冖' to n. Note that the same code shared by characters '将, 奖' since they have same radical and first stroke.

characters as well as researches on the modern Chinese language. The statistical distribution of orthographic and phonological features elucidated by these researches, has guided operators in their selection of the code element set for their system.

Recent achievements in database technology, programming and CKI itself have eliminated the need for handwriting innumerable cards for sorting and memory in developing CKI system. Presently the computerized listing of internal code brings simplicity, efficiency and precision to the system. The evaluation index of the system is placed on the user's ability to learn and remember the code elements and keys. No longer is the standard unit of measure the simple velocity and coincident code input. It is essential that the user always be foremost in the operators' mind.

Generally speaking, inputting Chinese is not so difficult as before due to the research achievement on simplification, reforming, and technology of hardware and software. Users of early CKI systems have to stand most of the memory and coding workload by themselves. Most of the workload has been transferred to computer right now.

The inputting efficiency of Chinese text by keyboard coding is ~1.5 times higher than that of English text. This stems from the fact that the Chinese character or word has higher entropy value and lower redundancy than alphabetic languages. In addition, there is the ability to use reflexable combinations of characters to form words in different contexts. However, there is still a problem for the general population because of economic issues and dialect differences.

The development of CKI in past two decades is concurrent with the development of statistic research on Chinese, while the issuing of related standards produced reliable basis for the future research (Han and Luo 1999). The developing trends of the CKI encoding system can be summarized as based on a single character, an emphasis on words and an intelligent processing of sentences as future trend. Intelligent processing is regarded as the main stream of the future since it can release most of the users' mental workload for inputting Chinese.

4.2. International Communication

Issues concerning CKI (history, theory, application, achievements, technology) have seldom been presented in international journals. Even some fundamental concepts can be new to foreign colleagues. Reference retrieving in January 1999 revealed this. No match item came out when using "Chinese keyboard input" as keywords through WebSPIRS (includes PsyINFO 1967–98 and MEDLINE EXPRESS [R] 1966–98), Social Sciences Index (1989–98) and WinSPIRS 2.0 (includes ERIC 1966–98, Psyndex plus 1977–98, PSYNDEX plus Tests 1945–97, LLBA Disc 1973–98, Education Index 1983–98 and PsycLIT 1887–98) in several libraries in Stockholm.

There are two papers among the few match items retrieved keywords as "China and ergonomics" or "Chinese and computer" related to CKI directly. Ong and Shahnavaz (1987) introduced some basic statement of CKI with five short paragraphs. Kang and Boon (1987) discussed something about Chinese computing

input methodology and proposed an experimental design. Meanwhile, CKI is also a new field to Chinese ergonomists (Wang 1990, 1993)

It is an ongoing tough process for Chinese researchers to make their language fitable to computer and easy to handle in daily information processing work, although several key technical issues have been solved. Since some issues, such as the human–computer interface, remain challenging and in need of further research, international communication can be helpful to overcome the remain difficulties. The research achievements of Chinese and foreign colleagues in Ergonomics need better communication. Relevant research concerns, especially ergonomic issues, need international communication and consideration. This in a task not only for the Chinese academic field, but also for the international community which can easily exchange information with respect to scientific, economic and cultural issues.

ACKNOWLEDGEMENT

REFERENCES

CHEN, Y.F. and HU, X.H., 1994, *Fundamentals of Technology and Theory for Chinese Character Inputting* (Beijing: Qinghua University Press and Nanning: Guangxi Science and Technology Press). [in Chinese]

FU, Y.H. *et al.*, 1998, *International Standard Chinese Dictionary — Multimedia version* (Beijing: Peking University Press). [in Chinese]

HAN, B.X., 1994, Development of the database of Chinese constituents information-statistical analysis of the frequency of constituents and their combinations. *Acta Psychologica Sinica*, 26, 147–152. [in Chinese with English abstract]

HAN, B.X. and LUO, L., 1999, Statistical research of modern Chinese and related national standards for information exchange. *International Encyclopedia of Ergonomics and Human Factors* (submitted).

KANG, A.B.H. and BOON L.H., 1987, Issues in software ergonomics: user interface design and Chinese computing input methodology. *Ergonomics*, 30, 1645–53.

LI, G.Y., LIU, R. *et al.*, 1988, *Chinese Information Dictionary* (Beijing: Science Press) [in Chinese]

LIU, Y., LIANG, N.Y., WANG, D. *et al.*, 1990, *Frequency Dictionary of Frequently Used Words in Modern Chinese* (Beijing: Aero). [in Chinese]

ONG, C.N. and SHAHNAVAZ, H., 1987, Ergonomics in China: a historical perspective and some recent development. *Ergonomics*, 30, 1631–44.

WANG, H., CHANG, B.R., LI, Y.S., LIN, L.H., LIU, J., SUN, Y.L., WANG, Z.W., YU, Y.X., ZHANG, J.W. and LI, D.P., 1986, *Frequency Dictionary of Modern Chinese* (Beijing: Beijing Language College Press). [in Chinese]

WANG, Z.M., 1990, Recent developments in ergonomics in China. *Ergonomics*, 33, 853–65.

WANG, Z.M., 1993, Editorial: Strategies for ergonomics in developing countries. *Ergonomics*, 36, 597–9.

Convergence of Telephony, Television and Computing in the Home

A. M. Lund

U S WEST Advanced Technologies, 4001 Discovery Drive, Boulder, CO 80303, USA

1. INTRODUCTION

Nearly all US homes have one or more televisions and telephones. Most have cable or satellite access to television programming. About one-half of US homes have personal computers, and almost one-half of those have access to the Internet. While penetration of personal computers into the home has slowed, it is continuing and it is expected to increase even more as easier to use information devices and computers priced closer to entertainment systems become available. A recent survey of Web use (www.gvu.gatech.edu) suggests that the demographics of Web users continue to converge on the demographics of the US population in general. When the number of people who have access to the Internet from outside the home is combined with these figures, it becomes clear that telephony, television and computing can be treated as almost ubiquitous in the experience of most people. Future experience, however, will probably be shaped at least as much by the evolution and convergence of the technologies that lie behind the three media and the impact of that evolution on the design of new applications.

2. MEDIA USE IN THE HOME

The default location in the USA for the telephone is usually the kitchen. The kitchen plays a key role in the organization of the home. The kitchen is in use much of the day for food preparation and often for eating, and therefore functions such as scheduling activities, messaging among members of the family and other communication with family, friends, organizations and businesses become associated with the kitchen. The telephone fits well into this environment, and the most feature-rich telephone in the house is often seen in the kitchen. Recently, cordless telephones that allow conversation while other tasks are being performed have been attractive to consumers. Telephony features that support the effective management of communications such as minimizing unwanted interruptions are also popular. When a new telephone is purchased a useful older telephone typically migrates elsewhere in the house. The fastest growing area of telephony is wireless (e.g. cellular and PCS phones, and pagers), and when users leave home wireless technologies often serve as an extension of the home communications environment.

The main television in the home tends to be located in one of the common living areas (with 75% of television viewing taking place in the living/family room or dining area). The default location of the television is driven by the need for a shared entertainment space, where interaction with the entertainment tends to be passive (Logan *et al*. 1995). Viewing usually happens during the week after 17:00 hours (although weekend viewing is relatively higher during the day). Of viewing, 60% is with other family members and 6% is with friends. Interestingly, while ~40%

of a person's free time is spent watching television almost two-thirds of the time television is being watched people report doing something else as well (Kubey and Csikszentmihalyi 1990). The properties of being a largely passive medium and viewable from a distance are what support the social functions of television viewing as well as the ability to use viewing time in a variety of ways.

The general-purpose, personal computer has been relatively expensive until recently and is most typically associated with a 'home office'. The home office may be something as simple as a desk near the family room or in the kitchen. The computer is generally purchased for telework or homework, but once installed its interactive entertainment role becomes one of its most frequent functions. Word processors and spreadsheets account for most software sales; entertainment applications are the next most popular purchases (and growing quickly); and home education programs are the next most frequent purchases. As the cost of entry-level computers approaches the cost of entertainment systems more families will be able to justify the purchase simply for entertainment and educational reasons.

Recent statistics suggest that Internet usage is cutting into television watching and sleeping, with ~37% of active Web users reporting watching less television than before they had access to the Internet. Nevertheless, Web usage is still more likely to range between 2 and 10 h per week compared with 10–20 h per week for television watching. E-mail and other forms of electronic communication are currently the most popular use of the Internet. A recent longitudinal study of on-line computer users, however, has found some evidence that communication within households and the size of users' social circles may decrease with Web use, with attendant increases in depression and loneliness (Kraut *et al*. 1998, homenet.andrew.cmu.edu). These findings have reinvigorated a long-standing debate concerning the social and psychological impact of computers in the home, reminiscent of a similar debate about television that took place in the 1950s.

3. TECHNOLOGY TRENDS

One way of looking at the trend that is leading to the convergence of telephony, television and computing is that it is based on the digital conversion of information to a common form (packetized data). The digitized information is being made available to an increasing variety of devices and applications through an evolving, ubiquitous network of networks. The technological barriers to creating applications that are designed to satisfy user needs and desires more effectively are rapidly falling. These trends primarily represent improvements in networking, in access to the networks and in the interfaces that people use to interact with the applications.

3.1. Networking

The principal technologies for building the networks that carry information can be divided into wired and wireless. The wired networks are typically fiber-based, and they support two-way communication with essentially unlimited bandwidth. New records for the amount of information that can be carried in a fiber are set on a regular basis. Increasingly, information and conversations will be moved in packets rather than depending on circuit-switched connections. Problems with network congestion and managing networks whose growth is accelerating

and whose capacity is unbounded are being addressed. Satellite is expected to be the most important wireless networking technology for covering broad geographic areas. Currently, the satellites are at high altitude but a variety of low altitude satellite data networks are likely in the future. The future will also see significant growth in wired and wireless networks within the home, with the opportunities for new applications such networks will bring.

3.2. Access

People access the public networks to exchange digital information using technologies such as analogue and ISDN modems, and increasingly they are using digital subscriber loop and cable modems. The bandwidths currently associated with several of these access technologies are shown in Table 1. As a comparison, it takes 1.5 Mb.s^{-1} to deliver a VCR-quality video using MPEG-2 compression today. Video of the type used for desktop video conferencing applications should have ~384 Kb.s^{-1} available for natural video that is free of artifacts and that supports enough detail to allow real-time recognition of expression. Stereo audio at CD quality requires 1.4 Mb.s^{-1}, although near-CD quality audio can be provided with 256 Kb.s^{-1}. Audio at the level used by the telephone system today and AM radio quality audio is carried at 64 Kb.s^{-1} (although with compression good quality speech can be provided at considerably lower rates). As compression technologies continue to improve higher quality will be available at lower speeds.

The wireless access technologies include satellite and the radio spectrum used by devices such as cellular phones and pagers. Satellite networks have the advantage of coverage and the ability to deliver a great deal of information quickly. Their disadvantages include the bandwidth that is available for the typical user to send data, and limits on real-time interaction. In other words, satellite networks are best for supporting broadcast-oriented content (whether video, audio, images or information) with limited interactivity. The radio spectrum used by cellular phones and pagers (and more recently PCS phones and wireless modems) tends to support relatively lower bandwidths compared with alternative technologies, but it lends itself to applications for mobile users particularly well. Efforts continue to find ways to increase the wireless bandwidth between mobile devices and networks (e.g. to support video more effectively). New ways of integrating wireless with wired applications are also being created.

The wired access technologies include the twisted pair copper that connects virtually all homes to telephone networks and the coax used to connect the majority of homes to cable networks.

Table 1. Examples of bandwidths available today

Technology	Bandwidth from the user (Kb s^{-1})	Bandwidth to the user (Kb s^{-1})
Satellite	28	400 Kb s^{-1}–30Mb.s^{-1}
Radio (e.g. CDPD)	9.6	9.6
Analogue modem	56	56
ISDN	128 (2b-channels)	128 (2b-channels)
ADSL — distance-dependent	16–800	256 Kb s^{-1}–8Mb.s^{-1}
Cable modem — load-dependent	20–200	300 Kb.s^{-1} –8 Mb.s^{-1} (10 Mb.s^{-1} single user)

Analogue modems connected to the telephone lines are supporting greater data speeds over the telephone network and can be multiplexed to boost the bandwidth available. However, it is the introduction of digital subscriber loop (DSL) and cable modem technologies that is promising to enable the convergence of telephony, television and computing. DSL and cable modem technologies provide bandwidth that allows channels for information and communication to be continuously available, and it allows synergies between the media to be exploited for new applications. It also allows hardware and software at the users location to be enhanced with the capabilities enabled by servers and processors on networks, making the location of those servers and processors essentially irrelevant.

3.3. Interface Technologies

Moore's Law (Moore 1996) predicts that the number of transistors on a microchip will double every 1.5 years. This prediction has described the evolution of traditional computers well and appears true for the near future. It has been predicted that at this rate computers will surpass the processing power of the human brain by 2047 (Bell and Gray 1997). The evolution of other components (e.g. memory) has developed with similar speed. As information and communication are converted to data, and the components inside the devices that provide the interfaces for interacting with the data become smaller, faster and cheaper, the variety of forms these devices take and the applications they support should explode.

Telephones will essentially become network computers and will be capable of receiving information from the Web and sending messages. Televisions will increasingly be used for accessing information on the Web, checking e-mail and voice mail, and managing calls (e.g. seeing who is calling during a program and sending calls to a voice mail system). Computers will receive video, and will be used to view background information associated with a given point in a video stream. Computers already support many functions that in the past have been associated with high-end telephones.

Behind the functionality that is moving across platforms, however, is the fact that inside each of the devices is, in essence, a computer that is linked to a network. Increasingly, each has a display and the resolution of the displays is improving. Each increasingly has a processor and memory; and each has some kind of an interface for entering data (including commands). Whether a given small hand-held device should be classified as a telephone, an entertainment device or a computer can be ambiguous. The cellular phone may also function as an information retrieval device or as a vehicle for receiving broadcast radio; and a child's necklace may be designed to allow a parent to track a child, send distress signals and receive messages from home. The applications supported by the devices and their interfaces will be shaped more by the users' needs and relatively less by the limitations of the technologies available for building the devices. The applications may be synchronous or asynchronous and increasingly they will be mobile. The trend of moving from a primarily mass-market approach to designing and delivering customized and customizable applications should continue.

It is unlikely that telephony, television and computers will disappear. The functionality that crosses the traditional

boundaries of the media will be successful only if it complements the ways in which people use the media. For the foreseeable future, there will be a large screen, largely passive, entertainment device in the shared living area that will have the label "television". "Largely passive" means that while functions such as interactive entertainment, information retrieval and communications will be supported, the value of the device most of the time will be as a shared and sometimes private entertainment experience.

Even though a variety of new computing devices will be available to support specific tasks in the home, it is reasonable to expect that in near future there will continue to be a higher-end general purpose computing device in the home. It will be used for homework, work-at-home (including communications associated with work) and home management functions, even if the device is used in the evening for interactive entertainment. Producers of content intended to be delivered over the television will continue to try to create a feeling of presence to touch the heart. Producers of content that is largely intended to be delivered to a personal information retrieval device like a workstation will be successful when the content is information rich.

Finally, there will be a device whose primary purpose is to be the communications center for the household. The communications center's role, however, may include accessing directory information pulled from the Internet, checking on the weather, accessing a family calendar and sending and receiving messages. The functionality of all communications devices will increase and it will be possible to communicate and get needed information from wherever one is and whenever it is needed. Nevertheless, the communications center's role will be to help make management of the home and family easier and more efficient.

4. IMPLICATIONS

4.1. Design

The domains of user interface design for telecommunications, computing and entertainment have been quite separate in the past, and each culture has been unique. For example, the software industry has been heavily influenced by methodologies that involve deploying products that may not be "perfect" and then correcting the imperfections based on feedback from the field. The telecommunications industry has had a history rooted in the old Bell System where a product was not released unless it met a high criterion for quality. Both the software industry and the telecommunications industry, however, are used to an iterative design process in which an incomplete design is gradually improved and developed through customer feedback. The entertainment industry has been based more on the culture of the artist in which feedback is not obtained until the design is largely complete. Entertainment applications are designed to provide a specific experience rather than to meet a list of requirements. Increasingly, user interface professionals will need to become familiar with the design issues and techniques in the three domains and will need to draw on their diverse methodologies and strengths.

While many of the design principles that apply to applications emerging from converging technologies apply across domains, the technologies are evolving so quickly that it may not be clear how to apply a specific guideline. Principles of layout, for example, may guide user attention to appropriate information in

each case, but color choices for appropriate contrast may differ between televisions and computer monitors. When designs increasingly involve realistic images and multimedia, guidelines about using colors in an interface may need to be redefined to be less technology-dependent and to be driven more by the properties of human perception and attention.

An additional challenge arises as the technologies converge. A given application can be accessed through a variety of devices, and designers may therefore have to create several interfaces that are both compatible with each other and also are customized for the devices and for the context of use. For example, a grocery shopping application accessed through a computer might be information intensive and allow a comparison of products based on nutritional value and cost per gram; and might support a recipe sharing bulletin board. The grocery shopping application accessed through a screen telephone might provide an easy way to place the weekly order for staples such as milk, bread and cereal. Finally, the shopping application accessed through the television might allow users to view a variety of dining experiences (e.g. a romantic evening for two or a birthday party), and then to order everything required to duplicate the experience.

4.2. Methodology

The convergence of these media requires a more complete understanding of the user and an understanding of the user in new contexts such as the home. The traditional task analysis that precedes design in industrial settings does not capture the complexity of many of the new environments in which applications based on the converging technologies will be used. In addition to methodologies designed to improve usability by reducing errors and improving efficiency, it is important to assess whether a given application supports the needs for which it was designed. It is important to determine whether the application fits the context of use. It may be important to assess aesthetic and emotional aspects of the interface. The issue of whether an interface is compelling is new to a field that has long worked hard at making the interface "disappear" for users. Thus, design for converging technologies more than ever will increasingly require a collaboration of traditional human factors and behavioral science expertise, design expertise, expertise in assessing user needs as a function of context (e.g. ethnography) and expertise in shaping a product to be desirable to users.

4.3. Standards

Technology is changing so rapidly that it is difficult for formal standards and the research that should drive the standards to match the pace of evolution. Many design standards, therefore, will evolve as *de facto* standards driven by successful designs in the marketplace and the marketing power of manufacturers. The touch-tone pad on the telephone and the IBM-style keyboard are two examples of the impact interface standards can have on application development. With appropriate standards, a wider range of people will benefit from converging technologies and they will experience the benefit sooner.

An increasing number of standards bodies are directly or indirectly addressing user interface issues in the convergence of television, telecommunications and computing. The Information Infrastructure Standards Panel (ISSP; www.ansi.org/public/iisp), for example, is attempting to identify all the standards activities

that are needed to support the evolution of a Global Information Infrastructure and to identify organizations to create those standards. Many of the issues that have been submitted to ISSP address technology convergence. The Human Factors and Ergonomics Society's ANSI-200 committee is writing standards for software interfaces. While originally the focus was on text and graphical interfaces, more recently speech interfaces have been included and standardization has been taking place with an eye towards emerging multimedia applications. The International Telecommunication Union ITU-T committee is the body developing telephony standards (www.itu.ch). But even here the working groups increasingly have been identifying and working on issues associated with new forms of communications and multimedia information.

Industry standardization activities include the Advanced Television Enhancement Forum work on the future of television and its relationship to the Internet (www.atvef.com). Standards to integrate telephony and computing (including emerging set-top boxes for televisions) include the Telephony Application Programming Interface (TAPI) defined by Microsoft and Intel, and the new JTAPI (Java-TAPI) standard developed by Sun, IBM, Lucent, Intel, Nortel and Novell. The goal of the JTAPI effort is to create an object-oriented application–programming interface (API) that allows applications to be ported across multiple platforms. While these and the many other standards bodies are often involved in the definition of technical standards, the specific

decisions made can impact the flexibility interface designers have in designing interfaces. As a result, it is important to influence the standards to reflect the needs of users, anticipating the applications that will arise from converging media.

REFERENCES

BELL, G. and GRAY, J. N., 1997, The revolution yet to happen. In R. Denning and C. Metcalfe (Eds) *Beyond Calculation: the next fifty years of computing* (New York: Springer), pp. 5–32.

KRAUT, R., LUNDMARK, V., PATTERSON, M., KIESLER, S., MUKOPADHYAY, T. and SCHERLIS, W., 1998, Internet paradox: a social technology that reduces social involvement and psychological well-being? *American Psychologist*, 53, 1017–31.

KUBEY, R. and CSIKSZENTMIHALYI, M., 1990, *Television and the Quality of Life: how viewing shapes everyday experience* (Hillsdale: Lawrence Erlbaum Associates).

LANDAUER, T. K., 1996, *The Trouble with Computers* (Cambridge, MA: MIT Press).

LOGAN, R. J., AUGAITIS, S., MILLER, R. H. and WEHMEYER, K., 1995, Living room culture — an anthropological study of television usage behaviors. In *Proceedings of the Human Factors and Ergonomics Society 39th Annual Meeting*, 326–30.

MARTIN, J., 1976, *Telecommunications and the Computer* (2nd ed.) (Englewood Cliffs: Prentice-Hall).

MOORE, G., 1996, Nanometers and gigabucks — Moore on Moore's Law. University Video Corporation Distinguished Lecture (www.uvc.com).

Cross-cultural Issues in Human–Computer Interaction

Y.-Y. Choong

GE Information Services, Inc., 100 Edison Park Drive, Gaithersburg, MD 20878, USA

1. INTRODUCTION

As businesses become more international and global oriented, a large percentage of computer software developed in one country is being exported to other countries. For example, in the USA, about one-half of all software products developed by US companies are sold overseas. Human–computer interaction designers are faced with the challenges of insuring that their products are with the same level of usability and quality in foreign countries as in the country where the products are originally designed. Many companies have expanded their marketing and usability efforts to understand the new cross-culture user requirements in the vastly growing global market.

This chapter provides an overview of the design issues involved in cross-culture human–computer interaction, as well as some interface design guidelines for international users.

2. THE WORLD IS DIFFERENT

When conducting cross-culture studies, it is most obvious, though not enough, to examine the differences in the language, orthography, symbols, images and number formats used by people with different cultural backgrounds. Culture is defined as "the total pattern of human behavior and its products embodied in thought, speech, action, and artifacts, and dependent upon man's capacity for learning and transmitting knowledge to succeeding generations through the use of tools, language, and systems of abstract thought". People from around the world may be different in their appearance, perception, cognition, and style of thinking. They may hold different cultural assumptions and values. They may view the world differently and they may have very different customs that make people of a specific culture unique and distinctive from people of other cultures. To consider designing computer interfaces for international users, it is important to study cultural traits of different cultures in a more profound fashion other than simply language or symbol translation.

2.1. National Languages

Language is a key to meaningful communication among people. The communication will be most effective when the first languages or "mother tongues" of the peoples of the world are used. There are between 3000 and 4000 spoken languages, with numbers ranging from many millions of speakers down to a few dozen or even fewer. There are hundreds of different written languages in use around the world.

2.1.1. Scripts

Scripts are a collection of characters and glyphs that represent a written version of a spoken language. In many cases, a single script may serve to write tens or even hundreds of languages; for

Table 1. Popular Scripts in the World

Script	Country
Roman	English and other European languages
Chinese characters	China, Japan, and other Asian countries
Cyrillic	Russia and Eastern Europe
Arabic	Middle East
Kana	Japan
Devanagari	India
Korean	Korea
Bengali	India
Thai	Thailand
Telugu	India

example, the Latin scripts. In other cases, only one language uses a particular script; for example, Hangul, which is used only for the Korean language. The writing systems for some languages may also use more than one script; for example, Japanese makes use of the Han (or Kanji), Hiragana, and Katakana scripts.

In Table 1, it shows examples of some popular scripts and the countries where the scripts are used (Fernandes 1995).

2.1.2. Directionality

Written language can be bi- or unidirectional. Most languages are unidirectional. For example, English; it is written from left to right in a unidirectional fashion. Chinese is another example of unidirectional scripts, but the direction can be left to right, right to left, or top to bottom. A bidirectional script, such as Arabic, can be written from right to left and left to right (in certain situations, such as numbers) in the same context.

2.1.3. Collation sequence

Collation sequence is the sorting order that people prefer using a particular written language. It should be noted that even in languages that use the same script, the collation sequences might vary. For example, in Finnish, A is the first letter but Å comes after Z; in French, Å comes after A. Far-Eastern languages like Han (or Kanji) have ideographic characters that represent concepts. The western alphabetic notion of collation does not apply for scripts like Kanji. There are three different ways of sorting Kanji characters: by radicals, by strokes, and by phonetic order. Often, a combination of the sort orders is used: by strokes within radicals, or by radicals within strokes.

2.2. Cultural Issues

As discussed previously, culture consists in patterned ways of thinking, feeling and reacting, acquired and transmitted mainly by symbols, constituting the distinctive achievements of human groups. The essential core of culture consists of traditional ideas and especially their attached values. Different human groups can be different in their cognitive characteristics, in their surrounding societies, and may carry various cultural biases. When designing for cross-culture HCI, all these cultural-related issues will greatly affect the effectiveness of the design if the issues are not taken into account carefully.

2.2.1. Cognitive styles

The ways of thinking vary from culture to culture. For example, the Americans tend to classify things on the basis of functions, while the Chinese tend to group things based on their interrelationships and thematic relationships. The American way

of thinking tends to be analytic, abstract, and imaginative or beyond the realm of the immediately apprehended. The Chinese way of thinking tends to be synthetic, concrete and remains on the periphery of the visible world. In a recent study, it was reported that the Chinese users performed better when the computer system had concrete representation and the user interface was structured in a thematic fashion, in comparison with using a system with abstract representation and functional interface structure (Choong 1998).

2.2.2. Color coding and affect

There is no difference for the perception or coding of colors among people since there are common physiological bases for color vision. However, while color vision is universal, what people feel about color is definitely more subject to cultural variation. Color associations with various concepts or conditions revealed different population stereotypes due to cultural variation. For example, in the USA, it is very common to associate *red* with *stop* or *danger*, *yellow* with *caution*, and *green* with *go*. In Japan, *red* is usually associated with *anger*, *yellow* with *grace* or *nobility*, and *green* with *future* or *energy*. In France, *red* may mean *aristocracy*, and *green* means *criminality*; yet in Egypt, *green* could mean *fertility* or *strength* and *yellow* mean *happiness* or *prosperity*.

2.2.3. Cultural patterns

Human behaviors can be greatly influenced by their social patterns. One of the major dimensions of cultural variation is individualism and collectivism. Individualism pertains to societies in which the ties between individuals are loose. People are expected to look after themselves and their immediately family. Its opposite, Collectivism, pertains to societies in which people from birth onwards are integrated into strong, cohesive in-groups, which throughout people's lifetime continue to protect them in exchange unquestioning loyalty. Examples of countries that value individualism include USA, Canada, France and Germany. Examples of countries that value collectivism include Japan, Korea, Mexico and Costa Rica.

The cultural patterns will sometimes have impacts on computer system usage and evaluation. As pointed out by Ito and Nakakoji (1996), the collective nature of Japanese culture has impacted their use of groupware, especially group decision-supporting systems. Individual workers' contributions to the company are appreciated as teamwork, not as individual contribution, even if a single individual has come up with a really good idea. Individuals do not want to be recognized among the others. This makes it important to keep anonymity in using groupware since Japanese users would not speak up or object to any opinions especially when their bosses were present.

2.2.4. Cultural biases

There are four major components defining a specific culture: values, rituals, heroes and symbols. These components are built one on top of the other.

Values represent a culture's belief about what's right or wrong, what's good or bad, and what's proper and improper. Issues such as gender, race, formality, hierarchy and family values can have great impacts on cross-culture HCI design if not taken into account carefully. For example, some cultures would find images depicting such common things in the US as women in senior executive roles as offensive.

Rituals are established procedures taken by people that often reflect a culture's value. Rituals include mannerism, work ethic, political processes, problem-solving processes, and religion. User interface designers often develop metaphors from rituals based on their own cultural backgrounds that could be problematic when deigning cross-culturally. For example, some American products use terminology or concepts such as the "penalty box" from hockey and being "at bat" from baseball that could be totally unheard of by people in other countries.

Heroes in a specific culture often represent an idea or are valued for their talent in one of the cultural rituals. Heroes can be real or fictitious. When products with cultural-specific heroes or icons of this type embedded, they are prone to misunderstanding and misinterpretation for cross-culture HCI.

Symbols can represent positive messages of national identity, good luck and many other beliefs. Symbols can also represent negative messages such as death, bad luck, or threat. Even the same symbols may mean different things in different cultures, for example, hand gestures. Some American products utilize the "thumb-up" hand gesture to represent "OK" or "Good". However, the same "thumb-up" gesture is used to insult people in Australia. When designing for international users, be sure to use internationally accepted symbols and avoid using taboos in icons, logos, sound effects or clipart.

3. DESIGN FOR CROSS-CULTURE HUMAN–COMPUTER INTERACTION

3.1. Input and Output Devices

When designing for the users in different cultural contexts, it is the designers' responsibility to make sure that the products provide adequate input and output mechanisms.

The most common text input device is keyboard. Keyboard layouts vary throughout the world. The layout used in the USA is called QWERTY. Germany uses a QWERTZ layout and France uses an AZERTY keyboard. For Asian languages such as Chinese, Japanese and Korean, some input-method-editors (IME) need to be introduced to make text entry possible with the Western keyboards. There are other text-input devices available such as handwriting recognition devices and speech recognition devices. These alternative approaches are especially promising for Asian languages users due to the difficulties they have experienced using the Western keyboards for text input.

An international product must be able to provide the appropriate scripts used in the target cultures in all user output formats such as displaying or printing.

3.2. User Interface

From our discussions on the cultural-related differences from around the world, there are some components of the user interface that need to be considered when designing cross-culturally.

3.2.1. Language issues

As discussed above, people are using difference languages around the world. Often times, there are dialect differences within the same language used in different locales; including spelling, word usage, grammar and pronunciation. Cross-culture HCI design will also need to take into account physical language variations such as directionality, hyphenation, stressing, fonts, sizes, orientation, layouts, spaces, wrapping, and justification. The

various sort orders are also imperative factors for effective cross-culture interface design.

There are some key linguistic differences that need to be considered when translating the user interface form one language to another. Adequate screen space needs to be allocated for possible text expansion due to translation. For example, German words are usually longer than their counterpart English words. Composite messages should be avoided, such as warnings with a word or words dynamically determined. The composite messages will not be translated well since sentence structures could vary dramatically across cultures. The rules for word wrapping may also be very different from one language to another.

Sometimes, it is essential to support multiple languages simultaneously. It should be avoided using national flags to toggle among languages or using words in one language for selecting among languages. Using national flags for language selection may be offensive to some users since there are often times more than one country using the language. For example, English is used in the US, UK, Canada and many other countries. Using one language for language selection is also inappropriate since the users will have to know that language to make the selection. For example, in an English user interface, a Chinese user will have problem in picking out the Chinese from a selection list consisting of language options written in English if he or she does not understand English.

3.2.2. Format conventions

Formats are an artifact and are specific to different locales. Numeric values can be represented in different ways. Separators are not always used. And, even they are used, different locales are using different symbols for separators and with different formats. For the same number, "1,234.56" is used in the US, Canada; whereas "1.234,56" is used in Germany, Holland and Italy; and "1 234,56" is used in France and Sweden. The numeric glyph shapes can be different as well.

Other format conventions that need to be taken into account include currency, calendars, date and time formats, names and addresses, and telephone numbers.

3.2.3. Cultural references

Cultural-specific references should be avoided for they are prone to misunderstanding and misinterpretation cross-culturally. Visual and verbal puns should not be used since they won't be translated well. For example, a very popular icon with a "light bulb" is used to represent "bright idea" in the USA. However, to the rest of the world, it might merely mean a "light bulb" after all. Another example is the use of a US mailbox to represent electronic mails. Interface designers cannot make any assumptions about what people will look at and understand as a mailbox. Mailboxes have very different looks in other parts of the world. A better alternative for representing electronic mails is probably to use an envelope that is more recognized around the world.

3.2.4. Symbols and taboos

As discussed above, the designers for cross-culture HCI should stay alert with any potential political, religious or other symbolism which may affect the meanings perceived in a particular country. Though it should be avoided using human body language, extra care needs to be practiced if any human body language is introduced into the user interface. The designers should try to use internationally acceptable symbols wherever possible, and allow for localization where necessary.

3.3. International Usability Testing

Usability testing with real users is the most fundamental usability method that provides direct information about how people interact with computers and what problems they experience with the system being tested. The importance of usability testing becomes more prominent with cross-culture HCI since there may be many more unforeseen problems in the interface due to national and cultural issues.

There are some key points to consider when planning for an international usability testing. The first step is to pick a target locale. Pick a city in the specific locale that you have a contact, that has many representative users, and that can provide you an adequate test site.

The second step is to determine what equipment you will need for the testing such as audio, video, logging software, etc. It is better if the equipment is available onsite; otherwise, use highly portable equipment with adequate plugs and voltage converters.

The third step is to recruit target users for testing. Most likely, someone other than you will be doing the recruiting task since the target users may speak a different language. Make sure criteria for recruiting the users are developed and well explained to the recruiter.

The preparation is keen to the success of the test. Care must be given prior to leaving for the target locale since international traveling can be expensive and time-consuming. If a foreign language is involved, an interpreter needs to be arranged. Often times, since the interpreters are not trained usability professionals, they can interfere by giving users advice, thinking that they are being helpful. You need to make sure enough time is allocated before the test for you and the interpreter to go through the usability testing process in details.

The last step is to analyze and to report the results. Make sure you go over the data collected right after each session with the interpreter since it will be costly if you need to contact the interpreter after you come back from the test site. In reporting the results, you will need to refer both to usability problems experienced by the participants and to issues you observed yourself while preparing for the test (Nielsen 1996).

4. CONCLUSION

As the computer technology becomes more globally accessible, how to effectively design for the cross-culture HCI becomes greater challenges to the designers. The designers will need to acknowledge the cultural related differences among users for whom the systems are designed. The designers will also need to take into account all issues involved when planning the design for international users. International usability testing is irreplaceable for the success of delivering a usable product for cross-culture use.

REFERENCES

CHOONG, Y.Y. and SALVENDY, G., 1999, Implications for design of computer interfaces for Chinese users in mainland China. *International Journal of Human Computer Interaction*, 11, 29–46.

DEL GALDO, E.M. and NIELSEN, J. (eds), 1996, *International User Interfaces* (New York: Wiley).

FERNANDES, T., 1995, *Global Interface Design: A Guide to Designing International User Interfaces* (Chestnut Hill: AP Professional).

HALL, P.A.V. and HUDSON, R. (eds), 1997, *Software Without Frontiers: A Multi-platform, Multi-cultural, Multi-nation Approach* (New York: Wiley).

HOFT, N.L., 1995, *International Technical Communication: How to Export Information about High Technology* (New York: Wiley).

ITO, M. and NAKAKOJI, K. 1996, Impact of culture on user interface design. In E.M. del Galdo and J. Nielsen (eds) *International User Interfaces* (New York: Wiley), 105–26.

NIELSEN, J. 1996, International usability engineering. In E.M. del Galdo and J. Nielsen (eds) *International User Interfaces* (New York: Wiley), 1–19.

Cultural Aspects of User Interface Acceptance

D.L. DAY

Computer & Information Sciences Department, Towson University, Towson, MD 21252 USA

1. INTRODUCTION

In a 1980 film classic, *The Gods Must Be Crazy*, a bushman in the Kalahari Desert is bombarded by an empty Coke bottle thrown overboard from a private plane. To the bushman, the bottle represents a totally alien technology. However, he and his extended family assess its properties and soon find productive uses for this artifact. Among other things, it is ideal for rolling out tortilla-like foodstuffs — and for making whistling sounds. Neither use was intended by the creators of the thing, but that is unknown and irrelevant in the bush. But in the end the bushmen reject the bottle despite its usefulness, because this technology becomes a source of conflict among members of the family.

The adaptation of artifacts by one culture that were created in another is instinctual. Culture — the sum of all shared values, practices and expectations of a socially defined group — brings often unique perspectives to the assessment and use of technology. In many cases, e.g. a Coke bottle in Botswana or a personal computer in Mongolia, the peculiar perspectives brought to technology by local culture can profoundly affect the degree to which it is accepted — and, therefore, how it is used. They can also lead to uses for artifacts that were never dreamed of by their creators.

This chapter describes an expanding effort on several continents to understand how cultural factors affect users' acceptance of computerized technology — in particular, computer interfaces. This effort is seen both as it draws upon several contributing disciplines, such as anthropology, and within the context of a widespread attempt to globalize (and localize) computer information systems.

2. THE ANALYSIS OF TECHNOLOGY ACCEPTANCE

User acceptance of technology is a complex concept. Unlike the Kalahari bushman, modern users of technical systems — whether they be automated garage door closers, cellular phones or computers — often have no choice in whether to use a system: they must. Either they depend on such systems for their livelihoods or the intricate fabric of "civilized" society would deteriorate without such devices. This lack of choice complicates acceptance enormously.

Why? First, because peoples' attitudes and perceptions are affected significantly when the obvious option — total rejection — is removed. When forced to use a tool that in their gut they would rather discard (like the Coke bottle), many people not only become less productive or effective: they seek means to vent their frustrations. They unconsciously, in many cases, decide to *make* the technology fail.

The study of technology acceptance is inherently multi-disciplinary because of its psychological, technical, and social context. Often people must adapt to technology rather than adapting it to their needs, including their cultural preferences. Culture is the ultimate context that molds peoples' expectations of technology and their behavior towards it. In a world intent on globalizing computer information systems, the culture brought to the technology experience by individuals inevitably takes on substantial importance.

The very nature of technology acceptance across cultures makes its analysis difficult. Concerns intrinsic to this area include:

- The appropriateness of technology.
- Cognitive decision-making styles versus those implemented by technology.
- Knowledge representation (e.g. semiotics and iconography).
- Social control (e.g. mechanisms and impact).
- Change management and adaptation.
- Globalization and localization standards.
- Technology transfer and diffusion of innovation.
- Interface design techniques or (more broadly) implementation for cross-cultural applications.

The appropriateness of technology is perhaps most critical, because change management and adaptation merely addresses the styles of implementation, once a suitable technology has been selected. Styles of implementation are a moving target, since technology mutates at an ever increasing rate. Principles of appropriateness, however, are based on cultural factors with a much longer timeline.

3. DIFFICULTIES IN ASSESSING CULTURAL ASPECTS OF TECHNOLOGY ACCEPTANCE

Studies of technological fit to task demands and decision making styles are difficult enough within single, somewhat heterogeneous cultures. However, when an attempt is made to evaluate technology acceptance across cultural boundaries, the task is complicated significantly (del Galdo and Nielsen, 1996). The difficulties, however, are mirrored by the potential payoffs, because of the imperative to manufacture goods for a global marketplace.

Given that technology acceptance is human-centered, the identification of representative people for evaluating acceptance factors and behavior is critical. Unfortunately, the modern world of satellite TV, global wire services and offshore manufacturing makes it very difficult to identify people who are largely of one culture versus another. In this work, the cross-national influence of disparate cultures is somewhat indelicately termed "contamination," only because the significant characteristics of one culture naturally meld with those of another, as soon as they come into contact. However, crisply distinct cultures are important methodologically to any research that claims cultural background plays a major role (e.g. in technology acceptance). But the availability of relatively distinct cultures in evaluating technology acceptance relies either on extensive (and expensive) international travel or painstaking segmentation of domestic populations along ethic boundaries.

Aside from subject selection, assessing cultural aspects of technology acceptance challenges the types of instruments typically used for such data collection. Most evaluation depends upon language — but language is a major variable between cultures. Therefore, a major issue arises as to whether instruments that seek to gauge technology acceptance should be in the
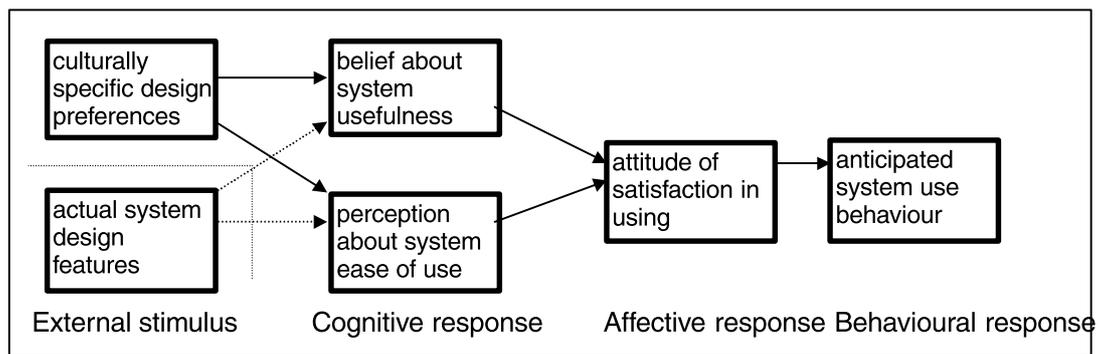
Figure 1. Modified technology acceptance model (Eves and Day 1997).

respondents' native tongues or in the vernacular of the assessor. Writing instruments in native language obviously increases understanding among respondents, but at the same time reduces methodological control by the assessor.

4. MODELS OF TECHNOLOGY ACCEPTANCE

Although significant work has been done in anthropology and other fields in regard to technology acceptance, its seminal treatment may be the series of articles in the late 1980s and early 1990s by Fred Davis. His work addressed the acceptance of computer systems.

Davis's Technology Acceptance Model (Davis, 1993) proposes that users' beliefs about system usefulness and their perceptions about system ease of use contribute substantially to their attitudes of satisfaction in using a technology, which in turn impacts actual system use behavior. "Usefulness" is essentially effectiveness, whereas "ease of use" is natural facility by the user in application of the technology. Cultural aspects added to Davis's approach in this chapter suggest that culturally specific user expectations (taken largely from anthropology and sociology) and specific styles of system design features (e.g. interface elements) precede and act upon beliefs about usefulness and ease of use. The combined effects are shown in Figure 1, which elaborates Davis's original model.

5. THE SUBSTANTIAL CONTRIBUTION OF ANTHROPOLOGY

"Culture" is a very contentious concept, because of its enormous breadth not only geographically but also over time. Although often considered esoteric and remote, anthropology contributes much to our understanding of man's responses to technologically sophisticated tools. Desmond Morris was not needed to remind societies that man has not changed much socially since soon after he stood upright, as the "naked ape." The basic instincts, interpersonal interactions, feelings and behaviors of men and women may change in application context, but in very fundamental terms the creature acts in accordance with its original design.

That design is the blueprint of culture (Hofstede, 1991), which is imposed in all societies soon after birth and which in some cases overwhelms the influences of family and physical environment. Culture put simply is a set of socially defined, shared values through which the individual's experiences are filtered. The filtering is subconscious in most people, but nonetheless powerful.

Technology acceptance studies have borrowed heavily from anthropology's explanations of man's reactions to his environment. For example, the notion of field dependence (also accepted in disciplines other than anthropology) can be ported to a technological context by anticipating the reactions of individuals to control features in computerized technology. Man inevitably anthropomorphizes his machines, especially those that seem to react to him in some of the same ways that other humans do. In cultures where the opinions of others are considered more important than individual judgement (i.e. field dependent societies), it is reasonable to assume that relatively high degrees of control and correspondingly limited freedom of action with the technology would be acceptable. On the other hand, in field independent societies — those in which individual volition is considered more important — technology users can be expected to resist control by their machines. The stimulus is computerized, but the context of the response is cultural.

Other factors from anthropology that figure significantly in technology acceptance begin with Geert Hofstede's universal cultural variables. These distillations of the human experience and context can be translated directly to anticipated technology acceptance behaviors — because in important ways technology is just another element in the environment to which individuals react. The cultural variables form a program for human behavior that transcends the character of the stimulus. Technology cannot expect to escape their influence.

Hofstede's variables, and others contributed by Edward Hall and Fons Trompenaars, include:
- Uncertainty avoidance: whether uncertainty is perceived as threatening.
- Preference for parallel versus sequential action: to do "more things at once" or "one thing at a time."
- Preference for diffusion versus specificity: whether business and personal relationships are separated clearly.
- Particularism versus universalism: whether behavior is rule-based versus relationship-based.
- Collectivism versus individualism: whether people like to work and make decisions in groups versus alone.
- High context versus low context: whether meaning draws in part from context or instead must be stated explicitly.
- Tendency toward transference: the extent to which a person blames others or one's self when a problem occurs.
- Power distance: the perceived degree of separation between an individual and figures of authority in his or her environment.

Each cultural variable influences the acceptance of technology, to varying degrees. For example, people from a culture that abhors uncertainty will be more comfortable with an unambiguous computer interface. And people from a culture that values group input to decision-making will be more influenced by co-workers in terms of accepting workplace technology. Of course, individuals' behaviors will vary from such generalities. The point is that cultural background predisposes large numbers of people toward widely held attitudes toward tools in the environment, technology included.

Base cultural variables may influence interface acceptance subtly, over an extended period. However, cultural reactions to technology *implementation* can be immediate. Work largely in industry has demonstrated that target cultures (as opposed to source cultures, where a technology is developed) take immediate notice of many interface features that impact computer use (Taylor, 1992; Hall and Hudson, 1997). These include sorting sequences; number, date, currency and time formats; the connotations of color usage; icon recognition and functionality, and reading direction (left-to-right, right-to-left, top-to-bottom).

6. ADVANCES IN ACCEPTANCE RESEARCH

The intercultural marketing of information technology products is part and parcel of the personal computing and telecommunications revolutions. Software in particular has become an ambassador of its culture of origin, thanks to the Internet and to worldwide standards in computing. This diaspora is accompanied by an accelerating attempt by multinational corporations to market wares in every culture where a buyer has the wherewithal to make purchases. So it is that "globalization" has become a driving concern of business everywhere, whether the product be motorcars, fax machines or computer software.

Software is a special case, however, due to its potential to be customized relatively easily to fit the needs of users — users with differing cultural expectations and experiences. In recognition of this potential, firms such as Intel, Claris, Sun and Microsoft have been evaluating cultural factors in interface acceptance for several years (Fernandes, 1995). Although many of their findings are considered proprietary and therefore not readily available to the public, enough has been published to form a basis for technology acceptance studies. A useful summary of work to date appeared in a review by Jane Carey of Arizona State University, in an early 1998 special issue of *Interacting with Computers* (Day, 1998). In reviewing a handful of books published in this area during the early 1990s, Carey provided a snapshot of efforts to internationalize and localize computer software.

Software that is "globalized" is merely packaged for use outside its culture of origin. Aspects unique to the culture of origin are removed from "internationalized" software, and changes are made to accommodate the specific cultural preferences of target cultures in software that has been "localized." Localization is facilitated by a movement within industry to engineer products in highly modular fashion, separating functionality from representational "resources." This makes it easier to substitute resources that are attuned to target cultures, in place of those used as defaults in cultures of origin.

Advances in technology acceptance research depend not only on product customization efforts by multinational business, but also on basic academic research. In one such study, Vanessa Evers and Donald Day found that the path to acceptance of interfaces in globalized software products varied significantly with users' home cultures (Figure 1).

7. APPROACHING THE BOTTOM LINE

At the beginning of this chapter it was noted that the peculiar perspectives brought to technology by local culture can profoundly affect the degree to which it is accepted — and therefore, how it is used. Users confronted by a technology (e.g. a computer interface) grounded in alien perspectives, with an implementation that is insensitive to local practice and values, inevitably will be less efficient and less effective in the use of that technology. Users certainly can learn to apply such technology, however awkwardly. But, as in all usability engineering, a key objective in developing software products for intercultural use is to minimize training while maximizing ease of use and usefulness. The flowering interest among multinational firms in the cultural preferences of overseas customers is founded upon very practical concerns: recent experience has shown that if they do not adapt interfaces to the needs of their other culture users, local firms rooted in target cultures will (Uren, Howard and Perinotti, 1993).

REFERENCES

DAVIS, F., 1993, User acceptance of information technology: system characteristics, user perceptions and behavioral impacts. *International Journal of Man–Machine Studies*, 38, 475–87.

DAY, D. (ed.), 1998, Shared values and shared interfaces: the role of culture in the globalisation of human–computer systems. *Interacting with Computers*, 9(3–4).

DEL GALDO, E. and NIELSEN, J. (eds), 1996, *International User Interfaces* (New York: Wiley).

FERNANDES, T., 1995, *Global Interface Design* (London: Academic Press).

HALL, P. and HUDSON, R. (eds), 1997, *Software Without Frontiers: A Multi-Platform, Multi-Cultural, Multi-Nation Approach* (Chichester: Wiley).

HOFSTEDE, G., 1991, *Cultures and Organisations: Software of the Mind* (New York: McGraw-Hill).

TAYLOR, D., 1992, *Global Software: Developing Applications for the International Market* (New York: Springer).

UREN, E., HOWARD, R. and PERINOTTI, T., 1993, *Software Internationalization and Localisation: An Introduction* (New York: van Nostrand Reinhold).

Describing and Predicting Human Interaction with Technology

N.A. Stanton
C. Baber

1. SUMMARY

In this article, the cognitive ergonomics of using public technology is investigated. A methodology for predicting human error with technology has been developed. Predictions from the method (combined with observation of user performance) form the foundation of the concept of "rewritable routines". This is in keeping with the tradition of building models of user cognition on the basis of observed and predicted errors. The concept is introduced and illustrated with examples. Implications for cognitive ergonomics are discussed.

2. INTRODUCTION

A method for predicting user error in transactions with simple machines has been developed (Baber and Stanton 1994; Stanton and Baber 1996) which is called Task Analysis for Error Identification (TAFEI). This method combines a description of human activity (using hierarchical task analysis) with a description of human–machine interaction (using state–space diagrams). The method has been used to describe and predict human interaction with devices for a variety of tasks such as boiling a kettle, withdrawing money from an automated teller machine, purchasing a rail ticket, recording a television program, and copying music from tape-to-tape. An example of the latter is illustrated in Figure 1, and further details may be found in Baber and Stanton (1994).

In Figure 1, the TAFEI diagram shows how task activity is mapped onto the device states for the goal of recording music
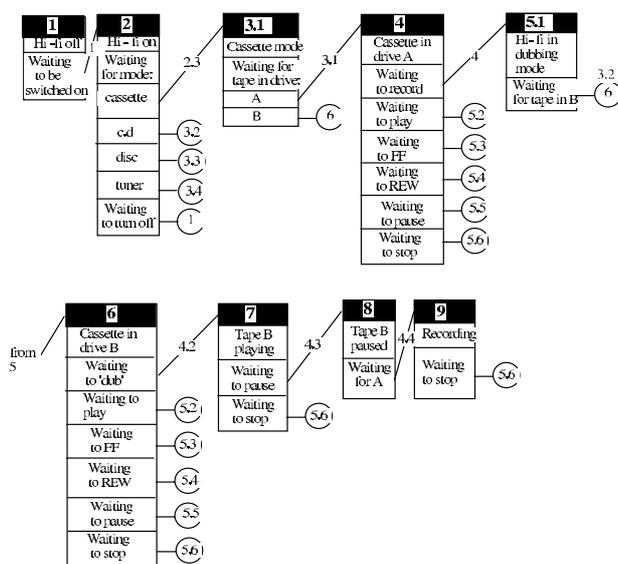
Figure 1.

tape-to-tape. As the diagram shows, the user does not explicitly press the "record button" to make a recording. This design is contrary to the task requirements for recording from other devices on the same music center. To record from the radio tuner, record player, or compact disc, the user is required to insert a tape into the cassette recorder and depress the play and record button together. The TAFEI approach makes inconsistencies in task structure explicit. This analysis can be useful to designers in helping them to understand why users may find it difficult to comprehend the task requirements. As the concept of recording from one device to another is similar across devices, the actions required of the user should also be similar. If they are different, as in this case, errors are likely to occur. To further understand the power of the analysis methodology, the theoretical basis underpinning user–device interaction are considered within a problem-solving framework.

3. PROBLEM SOLVING AND PUBLIC TECHNOLOGY

TAFEI, like other HCI methods, is based on Newell and Simon's (1972) ideas of movement through a problem space. As people move through the problem space, so they change their knowledge of the problem, acquiring new information and rejecting old hypotheses. Newell and Simon (1972) propose that this problem-solving behavior requires the person to hold current knowledge in short-term memory and goals in long-term memory. Problem solving then involves processes to reduce the gap between current knowledge and goals. This is a form of the means–ends analysis approach to Problem solving. People tend to employ means–ends analysis when they lack expertise to develop more sophisticated strategies. Interaction with public technology is performed on an intermittent basis and people are unlikely to try to reach "expert" level. Thus, we assume that means–ends analysis will be a common feature of such interaction. Further, we assume that while a goal can be reached via a limited number of relevant states, the interface of the machine could present states which appear relevant. In using public technology, the user will attempt to move from current state to the goal via relevant states. This is known in the problem-solving literature as "hill climbing". Actions are selected by comparing the current state with the goal, and the action which looks likely to lead to the goal will be chosen. This means that errors can arise when people reject "correct" actions which do not look as if they will reach the goal, or can perform actions which look as if they will reach the goal.

The notion that Problem solving involves movement through a defined problem space can be contrasted with that of Miller et al. (1960) who viewed Problem solving as a continual adaptation of behavior to a changing environment (in the test–operate–test–exit scheme). Other stances on Problem solving see the use of feedback as paramount in influencing activity, or view planning as largely dependent on environmental cues. Taking these views, we propose the second assumption in this article is that goals are decomposed into state-specific entities; Problem solving then involves generating a routine which will satisfy the immediate goal. TAFEI is based on the assumption that human–machine interaction follows a sequence of states, and at each state users will employ simple sequences of action to move them on to the next state. Thus, planning of action will be relevant only for each state. This assumption is shared with Simon and Young (1988: 591) who

propose "users engage only in simple forms of planning, [and] we can predict certain errors in user behavior due to the characteristics of a particular interface design". Suchman (1987) has also noted that much of human–machine interaction (HMI) is situated in the immediate context of the transaction with plans being developed on an ad hoc basis. It is not difficult to apply these assumptions to the description of HMI proposed by TAFEI. Thus, the first assumption in this article will be that people develop plans which are relevant to the current state in their interaction with a machine. In order for this to work, we need to have some means of defining the relationship between current state and goal state. We term this relationship "relevance".

4. RELEVANCE

The use of public technology is goal directed (Stanton and Baber 1996). This means that there will be a specific end-state of the interaction, e.g. obtaining a valid ticket from a ticket vending machine, withdrawing cash from an automated teller machine, etc. Possession of a goal does not imply possession of plans to achieve that goal. However, the range of possible actions will be constrained by the design of the machine and the user's interpretation of the machine. TAFEI assumes that for each state in human–machine interaction there will be a finite number of following states. The number of states will be limited by the design of the machine and the goals of the user. Furthermore, there will be several states from which a path to the goal can be seen (irrespective of whether such a path is "true"); these states will be termed relevant states.

We assume that users will interpret the relevance of actions in terms of the relationship between the current state and the goal. This is known in the problem-solving literature as "back chaining", where people create sub-goals which are achievable in the current state to allow progression to the goal. This means that the user has to work forward to the goal (to consider possible states) and backward from the goal (to determine relevant states). As Duncan (1993: 63) says, "Thus, the stimulus actually selected for control of behavior is not simply the best match to a pre-existing specification of what is needed in pursuit of a current goal. The goal itself may change; the selected stimulus is the one most relevant to the selected goal." We assume that the user will seek relevant information during the interaction, where relevance is determined by goal definition, machine state, and previous actions. The activity of seeking information is not intended to imply exhaustive search; rather the user will be defining the problem space at each state. Chi *et al*. (1981) demonstrate that problem classification can have a bearing on solution strategy and will vary between novice and expert. We assume that interaction with public technology involves a continuous process of problem classification.

5. REWRITABLE ROUTINES

In order to move from current to relevant states, eliminating other possible states, the user needs to retain some record of the interaction and to have some means of assigning relevance to states. At each state, this record will be modified. Thus, it will need to be rewritable. We assume that the record will be held in working memory, presumably in the articulatory loop which has a limited duration. This means that unless the record is updated, it will decay. As the record will also guide the next action, we see
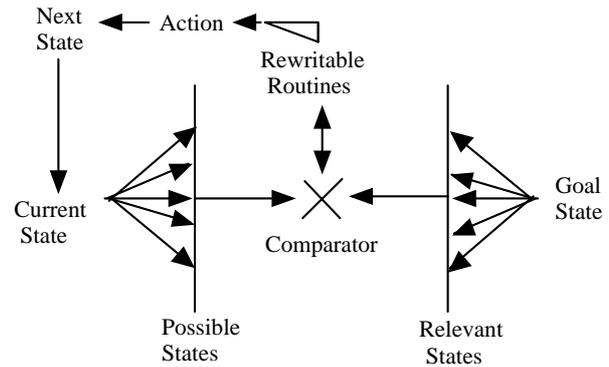
Figure 2.

this as a rewritable routine. To some extent this notion is similar to the 'partial provisional planning' hypothesis of Simon and Young (1988). Figure 2 presents a simple schematic of this process. The possible states (interpreted by the user from the machine) are compared against states which could lead to the goal. The comparator has a two-way connection to the rewritable routines (with the routines both influencing the comparator, i.e. by defining relevance, and taking the output to define action).

Example 1: A Rail Ticket Vending Machine

A number of errors which people make when using a ticket vending machine were observed. Around 80% of these errors could be predicted by TAFEI (Baber and Stanton 1996). The main errors were related to the mode of the machine, i.e "change given", "exact money only", "closed", "waiting". From the observations, people tended to use all machines as if they were in "change given" mode. This suggests that state indication on the machine was sufficiently poor to allow confusion, or that users were only employing some of the information. The machine studied had a wall-mounted LCD one meter above the machine to indicate mode, with a further display on an LED on the machine. From this we see that mode state is indicated. The fact that users missed these indications suggests that the comparator was operating on limited information, either because the users did not see the indicators or because the users took no notice of them. Not seeing the indicators would be a function of conspicuity (and this is certainly a possible explanation). Not noticing the indicators would be a function of relevance; users would hold a definition of relevance in their rewritable routines which would not require the mode indication, e.g. the routine could prompt users to find an available machine (i.e. one with no one waiting); by definition, this could take users to machines which were in states other than "change given" (there might be no one waiting at machines which were not "working" or which were "wait for assistance", and other users might have moved from "exact money" when they realized the mode).

Errors were observed in button use. This could be attributed to slips, with a large number of buttons in close proximity being likely to lead to the wrong button being pressed. While no data was collected on the buttons used, it seems unlikely that slips would account for all of these errors. A possible explanation lies in the fact that the button errors were often attributable to errors in sequence rather than operation. For example, from post-observation interviews, people in London reported having

difficulty in finding particular train and underground (subway) stations because these are listed alphabetically by type. This means that if a user is looking for "Euston", which is a train station, he is unlikely to find it by pressing an "E" key in the list of London Underground stations, and may well select "Embankment" instead. (Similarly, on machines in Birmingham, looking for "New Street" under "N" will not be successful because the station is listed as "Birmingham New Street".) This suggests that users work to a routine which differs from that devised for the machine.

The initial TAFEI study failed to predict a minority of the observed errors. These were "lose place in interaction", "lose ticket exit", "confusion due to options not available". It is proposed that each of these errors can be attributed to the description illustrated in figure 2. If the progression through states is not indicated, then the routine might not be updated appropriately. This would leave the user with a "gap" in the routine. "Lose ticket exit" referred to occasions when users were unable to locate the tray in which the ticket appeared. One explanation of this was that the ticket tray also served as the coin return tray; assuming that coin return and ticket dispensing were separate functions could lead to user seeking the ticket dispenser in another location. "Confused by operations not available" relates current state to users' interpretation of relevance, i.e. users could be seeking information which was either not present or was presented in another form.

Example 2: A Confectionery Vending Machine

In another study, we observed a number of errors which people make when using a confectionery vending machine. Around 70% of these errors could be predicted by novice users of TAFEI. Although purchasing confectionery from a vending machine is a relatively simple task, it is remarkably productive of error. Around 25% of the observed transactions resulted in errors of one form or another. The task required users to complete the following sequence of steps: (i) enter sufficient money to cover the purchase; (ii) check the nomenclature of the item of their choice; (iii) press the relevant number and character on a keypad; (iv) remove the item of confectionery; (v) turn the "change return" handle and collect the change.

The main errors people made were as follows: inserting the wrong money (the machine would only accept coins and would not accept pound coins), pressing the incorrect character or number, failing to turn the "coin return" handle and, consequently, failing to pick up their change. Most errors were associated with the latter two failures. We suspect that the delivery of the chosen item of confectionery, which satisfied the overall goal of the user, served to "close" the routine. If the change had been delivered before, or together with, the confectionery fewer errors of this type would have been observed.

Again, TAFEI failed to predict all the errors that occurred. Typically these errors occurred outside the state-based analysis of human–machine interaction. For example, TAFEI did not predict the occasion when a user attempted to insert a coin in an indent in the surface of the panel which resembled a coin slot. From the notion of relevance, this action could be interpreted as follows: the user sought some part of the machine with characteristics sufficient to satisfy the concept 'coin insertion place', and mistakenly selected an indent. This mistake could be compounded by the fact that a number of vending machines

keep the coin insertion slot covered until the user is intended to insert money, e.g. this is the case in the ticket vending machine.

DISCUSSION

In this article we demonstrate an initial specification for a theory of human-public technology interaction. The theory contains three main concepts:

1. Human–machine interaction with public technology is state-dependent. This means that the amount of preparation and planning conducted prior to the transaction will be limited. Hence, there is little need to read the instructions on machines. Observations have suggested that people tend not to read the instructions, or if they do, only to read part of the instructions. We suggest that the user will seek information regarding action to perform in the initial state, but will not seek information relating to subsequent states until those states are reached.

2. Human–machine interaction with public technology relies on rewritable routines. Users of public technology will plan actions which are relevant to the current state. When the next state is reached, the previous plans will be discarded. We believe that this is why even expert users of graphical-user interfaces are not able to recall the screen layout accurately; not necessarily because the users rely solely on information in the world, but because the planning of activity is focused on sequences of states. This would mean that users would recall only state relevant information (and not other details), and that users would primarily recall state exiting information (i.e. the details which take them from one screen to another). This issue is currently receiving attention in a series of studies.

3. Human–machine interaction with public technology should be designed to support opportunistic planning. Users interpret information from the machine in terms of relevance, and relevance is state-dependent. This means that presenting information too early will either confuse users or will lead to the information being ignored (as in the ticket vending machine example).

The approach taken in this article differs from previous research into public technology. Our concern has been to develop both a theory and method which can allow technology to be evaluated prior to its construction. This approach takes a model of human performance and applies this model through TAFEI. The aim is to have a means by which both user activity, cognition, and errors can be explicitly considered so that a theory of the use of the product can be developed during the early stages of design. This approach contrasts with other approaches which can be applied during the early stages of design, such as cognitive walkthrough. Whereas cognitive walkthrough, like other techniques in the verbal protocol family, requires people to describe and explain their activity; such approaches are not without their critics (Praetorius and Duncan 1988). While cognitive walkthrough may lead to people concentrating on issues of "relevance" (as discussed above), our concern is with the situations where relevance breaks down. This requires a form of investigation which allows us to capture the relationship between cognition and action in the context of human–machine interaction. Such a context is described in the TAFEI diagrams used in our work and is the focus of the analysis. Our research has been directed towards the

theory-led development of a methodology for human error identification. The theory of rewritable routines has helped to inform the development of TAFEI. Through this work gulfs between evaluation and execution may be bridged.

REFERENCES

BABER, C. and STANTON, N.A., 1994, Task analysis for error identification. *Ergonomics* 37(11): 1923–42.

BABER, C. and STANTON. N.A., 1996, Human error identification techniques applied to public technology: predictions compared with observed use. *Applied Ergonomics* 27: 119–31.

CHI, M.T.H., FELTOVICH, P.J., and GLASER, R., 1981, Categorization and representation of physics knowledge by experts and novices. *Cognitive Science* 5: 121–52.

DUNCAN, J., 1993, Selection of input and goal in the control of behavior. In A. Baddeley and L. Weiskrantz (eds) *Attention, Selection, Awareness and Control* (Oxford: Clarendon Press).

MILLER, G.A., GALANTER, E., and PRIBRAM, K.H., 1960, *Plans and the Structure of Behaviour* (London: Holt, Rinehart and Winston).

NEWELL, A. and SIMON, H.A., 1972, *Human Problem Solving* (Englewood Cliffs, NJ: Prentice Hall).

SIMON, T. and YOUNG, R.M., 1988, GOMS meets STRIPS: the integration of planning with skilled procedure execution in human–computer interaction. In D.M. Jones and R. Winder (eds) *People and Computers, IV* (Cambridge: Cambridge University Press) pp. 581–94.

STANTON, N.A. and BABER, C., 1996, A systems approach to human error identification. *Safety Science* 22: 215–28.

SUCHMAN, L.A., 1987, *Plans and Situated Actions* (Cambridge: Cambridge University Press).

Design of Menus: Choosing Names, Organizations and Selection Devices

K. R. Paap

Department of Psychology, New Mexico State University, Las Cruces, NM 88003, USA

1. INTRODUCTION

A menu can be defined as a set of options, displayed on a screen, where the selection and execution of one (or more) of the options results in a change in the state of the interface. The contents of a single screen will be referred to as a menu panel. Menu panels usually consist of a list of options. The options may consist of words or icons. The word or icon conveys some information about the consequences of selecting that option. When one of the options is selected and executed a system action occurs that usually results in a visual change on the screen. The complete set of options is usually distributed over many different menu panels. This allows the system to prompt the user with options that are likely to be useful and to hide options that are unlikely or illegal. However, layering the options across many menu panels also requires that the user be able to navigate between panels to find options that are not available on the panel currently in view. This brief description of a prototypical menu-driven interface highlights the fundamental issues of menu design that will be discussed. That is, how can designers choose good names for options, how should the options be organized within and between panels, and which selection technique should be implemented.

2. SELECTING NAMES FOR OPTIONS

2.1. Good Names

The name or phrase used to designate each option should be precise and permit the user to infer exactly those actions or objects that are controlled by this selection. It is easy for a designer to mistakenly believe that s/he can intuit the obvious name for a menu option. Furnas *et al.* (1984) determined that when people are asked to name things in a variety of domains (e.g. index words for cooking recipes) the probability of any two people generating the same name or description ranges from 8 to 18% across domains. Experts are no better than others at generating names with high intersubject agreement.

The goodness of a name or phrase is very much determined by the other names appearing on the menu panel. Error rates can increase by 45% when a *miscellaneous* option is added to a menu panel. A vague name like *miscellaneous* will poach selections that would otherwise be attracted to the correct option.

2.2. Adding Descriptors

Most errors in menu-driven systems occur because the meaning of the options is not clear to the user. One method of increasing the clarity is to append an expanded descriptor to each key word or phrase. Lee *et al.* (1984) report a direct comparison between the same set of options with and without descriptors. The descriptors were simply a list of the options that would appear next if a specific option were selected. The menus with descriptors were much preferred and had 82% fewer errors. The results clearly demonstrate that descriptors can be very effective when users have had limited experience with a menu panel that consists of options corresponding to fairly abstract categories. There are, of course, some costs associated with descriptors. Search times are probably somewhat longer and descriptors take up valuable screen space.

2.3. Using Icons

Options can also be represented or supplemented with icons. Icons afford two possible advantages over verbal options. First, if icons replace words as target alternatives, then there are situations in which the display can be searched in parallel and there is no cost associated with having a large number of options on a single panel. Second, even if users are considering one option at a time, categorizations of pictures can be faster than of words. This makes icons sound too good to be true, and owing to some inherent tradeoffs it is, indeed, impossible to reap all of the above benefits at the same time.

When the distinctiveness of an icon is enhanced by using simple figures that vary with respect to features like global shape, size, and orientation; it is likely that the simplification will make the icon more abstract and, thus, more error prone. On the other, more *representational* icons will be scanned sequentially, just like words. Relevant experiments reviewed by Paap and Cooke (1997) led them to conclude that supplementing category labels with icons is equivalent to adding a one-word example.

3. DESIGNING A SINGLE MENU PANEL

3.1. Three Types of Comparison Operations

The first stage of interaction with any menu panel involves the user forming a goal. The user directs his fixation to one of the options and evaluates the option in relation to the goal. If the option does not match the goal, then attention must be redirected and search continues. The best design depends on the type of search and comparison operations that are likely to occur as users consider the contents of the menu panel.

3.1.1. Identity matching

The fastest type of search occurs when users have generated a specific target that is literally displayed as one of the options. When this situation exists, the user can engage in identity matching, that is, the user compares each option to the specific goal held in memory to see if they are identical. This type of matching is fast because a holistic comparison can be made on low-level visual codes. Identity matching is likely to occur when the user knows precisely how the target is presented on the menu. For example, when a user knows that the target *delete* will be listed as *delete* and not as *erase*, *drop*, *zap* or anything else.

The following guidelines apply to the design of a menu panel that is likely to engage identity matching. If the options consist of a set of conventional names and the names also have a conventional order, then list the options in that order. If there is no conventional order and the set of options is small, then list them alphabetically. If the list is long and the list can be subdivided into distinctive categories, then present the options as alphabetized subsets.

3.1.2. Class-inclusion matching

Class-inclusion matching is likely to occur near the top of a hierarchical menu organization. The top panels frequently consist of a list of abstract categories. Users must make judgments about class inclusion, that is, whether the target is an instance of the category specified by an option. For example, is *delete* an instance of the category *editing commands*? Category decisions will be faulty if there is conceptual overlap between the categories and many targets might belong to two or more categories.

3.1.3. Equivalence search

Equivalence search occurs when the user knows what he wants, but does not know what it is called. Sometimes the target for an equivalence match will be formulated as a fairly abstract intention expressed as either a phrase or a possible synonym. Consequently the user cannot rely on a fast holistic identity search, but must engage in a slower semantic analysis of the relationship between the target and each option. The following guidelines apply to the design of a menu panel that is likely to engage equivalence matching. If some selections will be selected with far greater frequency than other selections, then list the options in descending order of frequency. If the options are equally likely to be selected, then list them in meaningful and distinctive categories.

3.2. Search Strategies

The preceding section on types of comparison operations focuses on the type of judgment the user makes when an option is evaluated. This section examines a related issue, namely, the strategies used to determine which options will be evaluated and the order of evaluation.

3.2.1. The search-time function

Search time surely will increase as a function of the total number of options that must be considered, but what is the shape of the function? Norman (1991) and other sources claim that the search time function will usually be logarithmic. The basis for the prediction is the assumption that the Hick–Hyman Law applies when users visually search a menu panel for a specific option. Paap and Cooke review the relevant data and conclude that: (1) one should *not* expect the Hick–Hyman Law to govern visual search for word targets and (2) empirically determined search time functions tend to be more linear than log-linear.

3.2.2. Stopping rule

A search is exhaustive if *all* the options are considered and, in contrast, is self-terminating if the user stops when an acceptable option is found. MacGregor *et al*. (1986) assume that users evaluate each option with respect to both a low- (L) and high-criterion (H). Any evaluation that yields a probability less than L will lead to the permanent rejection of that alternative and continuation of the search. An option that exceeds H will be immediately chosen and produce, by definition, a self-terminating search. Options that exceed L, but not H, are considered candidates. If the end of the list is reached the user will reexamine the candidate set. When the model was tested the mean number of options processed was approximated by multiplying the number of options on a test panel by the constant 1.2. Thus, users often examined all the options at least once and a small set twice. Self-terminating searches were rare in this environment, but they are likely to be more common if users can engage in identity matching, particularly on panels they are familiar with.

The dual-criterion model initially proposed by MacGregor *et al*. has been extended by Pierce *et al*. (1992). This work shows how characteristics of a menu design can influence the placement of the two criteria. In turn, the location of the criteria determines both the proportion of self-terminating, exhaustive, and redundant searches and the quality of system performance. One conclusion is that H becomes more lax as the number of options per page increases. In contrast, increasing the number of alternatives induces users to adopt a stricter L. Another way of summarizing these assumptions and their implications is to say that as the number of options increases users are less likely to tag promising options and more likely to gamble on the first attractive option they encounter.

4. ORGANIZATION AND NAVIGATION BETWEEN MENU PANELS

This section examines the factors important to deciding how the total set of options should be distributed across the individual menu panels. The distribution of options to panels will determine the navigation pathways, that is, the sequence of selections that will be required to get from one menu panel to another. Most menu systems are organized into a hierarchical tree, in which each node (menu panel) in the hierarchy can be reached only from a single superordinate node that lies directly above it in the hierarchy. Depth is usually defined as the number of levels in the hierarchy. Breadth is defined as the number of options per menu panel. A hierarchical structure with several levels requires the user to either recall or discover how to get from where s/he is to where s/he wants to go. The navigation problem (i.e. getting lost or using an inefficient pathway to the goal) becomes more and more treacherous as the depth of the hierarchy increases.

Despite the navigation problem there are three reasons for considering a system with greater depth: *crowding*, *insulation* and *funneling*. *Crowding* is the straightforward constraint imposed by the amount of available space on a panel. When the available space is exceeded, some depth must be introduced. *Insulation* refers to the opportunity for menu systems to prompt selections that are likely to be needed and hide those that are unlikely or illegal. *Funneling* refers to a reduction in the total number of options processed that is achieved by designing a system with more depth and less breadth. When greater depth is traded for less breadth, funneling can generate efficiency gains, particularly when the processing time per option is long.

There has been more theoretical and empirical work on the depth–breadth tradeoff than on any other issue related to the design of menus (for extended discussions, see Paap and Cooke 1997, Norman 1991). Much of the theoretical work uses mathematical models that compute the optimal organization given assumptions concerning various temporal parameters. Although interpretations remain debatable, it is safe to say that most designers will be tempted to use too much depth. In terms of both speed and accuracy optimal performance can often be achieved by presenting as many as 80 options on a single menu panel, providing that the options can be organized into distinctive categories.

Many factors conspire to make increasing depth a dangerous practice. Recall that adding depth provides the potential benefit

of funneling (i.e. fewer options need to be examined). However, this potential advantage is only important when the processing time per option is rather long. But, processing time per option is long when users need to make a slow semantic evaluation of each alternative. These types of decisions are much more error prone than quick identity matches. Furthermore, it has been empirically demonstrated that the top levels are of a hierarchy are more error prone than lower levels. When users make an error near the top of the hierarchy it takes longer to detect the error and then longer to recover from the error. Depth sinks efficient navigation.

The semantics of a menu system is probably much more important than the syntax (i.e., the depth–breadth structure). Let the natural subdivisions of the options dictate the organization. If some depth is required to avoid unacceptable crowding then user testing, at least at the top levels of the hierarchy, is essential. Modifications following a single phase of user testing have sometimes reduced the number of navigation errors by > 90%! Do not let an expert's intuition by your guide. Lee *et al*. had a group of experts design a menu. Another group of experts rated the menus for ease of use and the correlation between experts was virtually zero! In contrast, when the same set of menus were give to a representative sample of end users there was much higher agreement and, furthermore, the end-users preferences were highly correlated with actual performance.

Displaying a history of how the user arrived at the current panel may facilitate the navigation problem. A history helps the user consolidate a sequence of correct choices, enhancing the chances of being able to recall the correct pathway later. When an error is made, the history may help the user backtrack and find the correct path. Maps of the menu structure also aid navigation, but this is impossible for large systems. Map size can be compressed either by reducing the scope and showing only the area of current interest or by reducing the detail by showing only the major highways. The former should be better for helping users who are stuck in a specific place, but the latter may be better for improving the users' overall mental model of the system and, hence, their long-term performance. Furnas (1986) has been developing a *fisheye view* that blends both local detail and important distant landmarks. The contents of the current panel are based on a degree of interest function that takes into account the distance of each node in the system from the node currently in focus and the a priori importance of each node.

5. CHOOSING A SELECTION TECHNIQUE

There are a number of different techniques for selecting items from menus. Menu selection techniques all involve selection from a list of items that can be presented in a variety of formats (e.g. pull-down, pop-up, pie). By convention a horizontal layout that is always in view is often found at the top of the screen. In most other cases, however, menus items are arranged vertically and if not fixed can be opened and closed using pull-down or pop-up menus. Pull-down menus are menus that are selected and displayed in the same location each time they are selected, whereas pop-up menus appear at the current cursor location. Given a specific list format, the items in that list can be selected using a pointing device or an identifier.

5.1. Pointing
Menu items can be selected by moving a pointer to the location of the desired option and selecting that item. The location of an on-screen pointer or cursor can be manipulated with cursor keys, joysticks, trackballs, touch screens, pens and mice. Pointing with a sequence of cursor keystrokes is not very efficient, but it is difficult to rank the other alternatives because there is much variability in performance between alternatives of the same type.

5.2. Identifiers
Another family of techniques for selecting menu items involves direct entry of an identifier associated with each menu item. The identifier is typically a letter or digit, but sometimes a letter string or abbreviation. Perlman (1984) makes several cogent recommendations based on empirical results. If the menu panel is used in an application where users make frequent selections and if the set of options do not change over time, then it may be worthwhile to use letter identifiers that can be compatibly paired to each option (e.g. "d" for delete). Another advantage of compatible letter identifiers is that they will be easier to remember than number identifiers. If the list of options is dynamic rather than static, the use of digits is a good compromise. Incompatible letter identifiers should be avoided. If should be noted that the best case, compatible letters, turns into the worse case if the menu options are translated into another language.

5.3. Prompts and Feedback
Providing default options and appropriate feedback is likely to be an important part of the design of any selection mechanism. Default options are particularly desirable when one option is selected much more often than any of the others. Under these circumstances search and selection operations can be made more efficient by highlighting the default option or preselecting it by having its identifier automatically displayed. Regardless of the selection mechanism, the system should provide feedback indicating: (1) which options are selectable, (2) when an option is under the pointer and, therefore, can be selected (assuming a pointer is used), (3) which options have been selected so far and (4) the end of the selection process.

Two techniques are commonly used to prevent users from selecting inappropriate choices. One approach is to gray-out or otherwise mark those options that do not make sense in a given context. The unselectability of the faded option is reinforced when it is under the pointer since it will not become highlighted. The second approach is dynamically to alter the visibility of the options, i.e. inappropriate options simply disappear from the screen. It is difficult to offer guidelines with respect to these two techniques since their relative benefits and costs have not been explored empirically. Norman advocates graying-out because it provides the user with stable cues regarding presence and location of items. In addition the inaccessible though visible items can help the user to achieve a more complete model of the items that are available and how they are related to other items.

REFERENCES

FURNAS, G.W., 1986, Generalized fisheye views. *Proceedings of CHI '86*, pp. 16–23.

FURNAS, G.W., LANDAUER, T.K., GOMEZ, L.M. and DUMAIS, S.T., 1984, Statistical semantics: analysis of the potential performance of

keyword information systems. In Thomas, J. C. and Schneider, M. L. (eds) *Human Factors in Computer Systems* (Norwood: Ablex).

LEE, E., WHALEN, T., MCEWEN, S. and LATREMOUILLE, S., 1984, Optimizing the design of menu pages for information retrieval. *Ergonomics*, 27, 1051–69.

MACGREGOR, J., LEE, E. and LAM, N., 1986, Optimizing the structure of database menu indexes: A decision model of menu search. *Human Factors*, 28, 387–99.

NORMAN, K.L., 1991, *The Psychology of Menu Selection: Designing Cognitive Control at the Human/Computer Interface* (Norwood: Ablex).

PAAP, K.R. and COOKE, N.J., 1997, Design of menus. In Helander, M. G. Landauer, T. K. and Prabhu, P. V. (eds) *Handbook of Human–Computer Interaction* (2nd ed.) (Amsterdam: Elsevier).

PERLMAN, G., 1984, Making the right choices with menus. *Proceedings of INTERACT '84*, pp. 291–5.

PIERCE, B.J., SISSON, N. and PARKINSON, S.R., 1992, Menu search and selection processes: a quantitative performance model. *International Journal of Man–Machine Studies*, 37, 679–702.

Ecological Interface Design—Theory

G. L. Torenvliet and K. J. Vicente

Cognitive Engineering Laboratory, Department of Mechanical & Industrial Engineering, University of Toronto, Toronto, Ontario, Canada

1. INTRODUCTION

Ecological Interface Design (EID) is a theoretical framework for interface design that attempts to extend the benefits of traditional direct manipulation interfaces (DMI) to complex human–machine systems. While traditional approaches to the design of DMI may work well within some domains, they cannot address the unique challenges posed by complex work domains. EID has been developed specifically for complex domains and it uses the unique challenges of these domains as its starting point. Once the challenges of a specific domain have been characterized, EID addresses the domain-relevant capabilities and limitations of the human operator, and develops prescriptive interface design guidelines on this basis.

This article shows how the EID framework is a specific instantiation of the broader ecological approach to human factors described by Hajdukiewicz and Vicente (1999). The theory of EID can be captured in three broad principles, but the rationale behind them is important for their understanding. Accordingly, one will begin by structuring the interface design problem using two fundamental questions. These will lead to a discussion of two foundational concepts driving EID: the abstraction hierarchy (AH) and the skills–rules–knowledge (SRK) taxonomy. These concepts are then integrated into a statement of the central tenets of EID, in the form of three prescriptive guidelines. Finally, a few comments on the limitations and generalizability of EID are made. For a number of examples of how EID has been applied to various application domains, see Chery and Vicente (this volume).

2. ECOLOGY OF THE WORK DOMAIN: A CENTRAL ISSUE FOR INTERFACE DESIGNERS

The point of departure for EID is the unique set of challenges faced by operators of complex human–machine systems. In the course of their work, operators must respond to a range of events that can be categorized under three groupings:
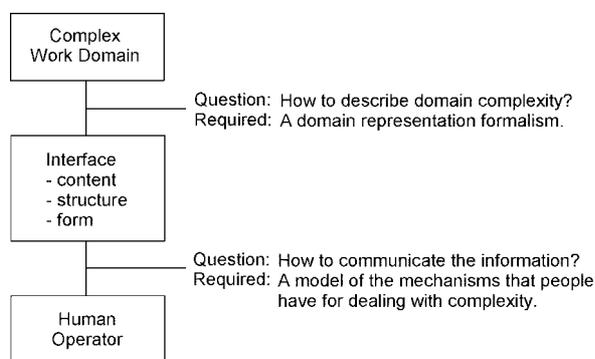
Figure 1. Basic structure of the interface design problem (adapted from Vicente and Rasmussen 1992).

- Most frequently, operators must respond to *familiar events* for which they have acquired the necessary skills through repeated exposure.
- Less frequently, operators must respond to *unfamiliar, but anticipated events*. While operators may not have developed the requisite skills that address this class of events directly, designers have anticipated these events. As a result, operators can rely on tools and aids provided by the designers to deal with this class of events.
- In rare cases, operators are called upon to respond to *unfamiliar and unanticipated events*. In addition to being unfamiliar, designers did not anticipate these events, and, thus, operators cannot rely on preplanned solutions in these situations. Instead, they must improvise a solution themselves through discretionary decision-making.

Traditional theories of interface design are generally well suited to helping operators cope with the first two classes of events, but are ill-suited to helping operators cope with unfamiliar, unanticipated events. While one valid solution might seem explicitly to design for a larger number of these rare events, this history of industrial accidents shows that all possible events cannot reasonably be enumerated. Unfamiliar and unanticipated events will always be a reality in the operation of complex systems. Thus, one needs a framework that can help to identify and provide the support that operators need to adapt to the unanticipated.

The first step in developing such a framework is to reconsider the most basic questions that should guide interface design so that the work domain is explicitly represented. As shown in Figure 1, the interface design problem includes at least three fundamental elements: the complex work domain; the human operator; and the interface that mediates between them. This structure naturally leads to two questions. The first can be generated from the relationship between the complex work domain and the interface: how should the designer describe the complexity of the work domain? The second follows from the relationship between the interface and the human operator: how should the complexity of the work domain be communicated to the operator? It is important to note that these questions capture the organism–environment reciprocity that is foundational to both ecological psychology and the ecological approach to human factors (Hajdukiewicz and Vicente 1999). It is for this reason that this framework is referred to as *ecological interface design*.

To answer the questions just described, two conceptual tools are needed. First, a representation formalism is needed to describe the content and structure of the work domain. Second, a model of the mechanisms that people have for dealing with the complexity of the work domain is needed to design the form in which to communicate this information to the operator.

3. DESCRIBING DOMAIN COMPLEXITY: THE ABSTRACTION HIERARCHY (AH)

A formal description of the work domain — independent of any particular worker, automation, event, task, goal or interface — drives EID. EID describes domain complexity using the AH (Rasmussen 1985). While the use of hierarchies is common in system design, AH has unique properties that sets it apart. It belongs to the class of stratified hierarchies (Mesarovic *et al.* 1970), and so has five general properties:

- Each level of the hierarchy describes the same system, but each different level provides a unique description (or model) of the system.
- Each level provides its own unique set of concepts, terms and principles.
- The selection of levels for describing a system depends on the observer and their knowledge and interest in describing the system. For many systems, however, there may be some levels that appear as natural or inherent.
- The requirements for proper system functioning at any level appear as constraints on the meaningful operation of lower levels, while the evolution of the state of the system is specified by the effect of the lower levels on the higher levels.
- Understanding of the system increases by crossing levels: by moving up the hierarchy, one obtains a deeper understanding of system significance with regard to the goals that are to be achieved, while in moving down the hierarchy, one obtains a more detailed explanation of the system's functioning in terms of how these goals can be carried out.

The structure of AH is further specified by a structural means–end relationship between levels (i.e. for each pair of adjacent levels, the upper level defines a set of ends, while the lower level decomposes these ends into structural means by which they can be accomplished). Thus, in contrast to other types of hierarchies, the elements included in an AH are explicitly related to work domain purposes. For an example of a completed AH, see Vicente and Rasmussen (1990).

Note that AH is not a specific representation, but rather a framework for developing representations for different work domains. The exact number of levels and their content will vary from one application domain to another as a function of the different types of constraints inherent in each domain. For instance, five levels of constraint have been found to be useful for describing process control work domains: the purposes for which the system was designed (functional purpose); the intended causal structure of the process in terms of mass, energy, information or value flows (abstract function); the basic functions that the plant is designed to achieve (generalized function); the characteristics of the components and the connections between them (physical function); and the appearance and spatial location of these components (physical form). Regardless of the application domain, however, the resulting hierarchy will have the properties described above.

The structure of the information contained in an AH makes it useful in helping operators to cope with unanticipated events. When operating in a 'normal' state, any work domain can be viewed as a set of constraints that describe the relationships between variables (e.g. at steady-state, the inflow into a reservoir is equal to the outflow). When unanticipated events occur, the structure and function of the work domain will change, and the constraints that usually hold under 'normal' circumstances will be broken (e.g. if a leak occurs in a reservoir, making the inflow and outflow equal will no longer lead to steady-state). Fault detection is thus equivalent to the detection of broken constraints. Because it is not possible to know in advance which constraint(s) will be violated, the complete set of goal-relevant constraints governing the work domain must be represented to permit operators to detect and diagnose broken constraints. An AH representation does just that by describing the physical and functional relations between variables. Representing these relations should make it possible for operators to cope with the entire range of operating demands, including unanticipated events.

In addition, there are also psychological justifications for adopting an AH representation. AH allows operators to manage the work domain at a variety of levels of detail, and so provides a mechanism for coping with the complexity of the domain. Operators can gain a bird's-eye view by using a high level representation or can drill down to a faulty object by moving to a lower level. This property of AH allows operators to cope with systems that would be unmanageable if they had to observe them in full detail and all at once.

While it can be argued that this property is shared by many hierarchies, AH is unique because it is explicitly goal-oriented. Each level in the hierarchy is connected to adjacent levels by structural means–ends relations. This type of functional structure deliberately supports goal-directed problem-solving. Consequently, workers can search the problem space in an economic fashion by evaluating the means that are relevant to the current set of goals.

Furthermore, research has shown that people find it natural to reason within an AH representation (Vicente and Rasmussen 1992). Researchers have had success in mapping the problem-solving protocols of world-class chess players, computer programmers, electronic trouble-shooters and nuclear power plant operators onto AH representations of their domains. Many of these studies have also shown that subjects' problem-solving trajectories begin at a high level of abstraction and then 'zoom in' to levels of more structural detail. Thus, there is evidence to indicate that the AH is a psychologically plausible problem-solving representation.

4. COMMUNICATING DOMAIN INFORMATION: THE SRK TAXONOMY

The next question to be addressed is: what mechanisms do people have for dealing with the complexity of a domain? Answering this will reveal effective ways to communicate the information in a work domain representation to the operator. As will be seen, the approach adopted by EID is to present information in a form so that operators can take advantage of the power of human perception. To do this, one needs to know something about the different mechanisms that people have for processing information, how they can be induced and their relative efficacy.

In the cognitive engineering community, the SRK taxonomy (Rasmussen 1983) is a widely accepted framework for describing these mechanisms. In brief, it asserts that information can be interpreted in three mutually exclusive ways — as signs, signals and symbols — and that the way in which information is interpreted determines which of the three levels of cognitive control is activated — skill-based behavior (SBB), rule-based behavior (RBB) or knowledge-based behavior (KBB) respectively. Thus, cognitive control may rely on a repertoire of automated behavioral patterns (SBB), a set of cue–action mappings (RBB) or problem-solving operations on a symbolic representation (RBB).

The three levels of cognitive control can be grouped together into two categories. SBB and RBB deal with perception and action, while KBB deals with analytical problem-solving. In general, perceptual processing is fast, effortless and can be

executed in parallel, while analytical problem-solving is slow, laborious and is only executed serially. Analytical problem-solving also tends to be more error-prone than perceptual processing. SBB and RBB, however, can only be activated in familiar situations because they require that the operator be attuned to the perceptual features of the environment. KBB, on the other hand, allows operators to cope with novelty. This description shows that there is a trade-off between cognitive economy and the ability to cope with novelty, so no one level is superior to any other.

How can the SRK framework be applied to design? Two characteristics of complex work domains make this possible. First, operators of these domains are highly skilled and have extensive experience in controlling the work domain. Second, interfaces for these domains are generally system-specific and generality is not required. Operators always deal with the same work domain, and thus transfer is not an issue. These two characteristics imply that complex work domains afford designing for SBB and RBB. Since operators will have extensive experience with the system, they will be able to attune themselves to the perceptual properties of the control room interface. Because operators always deal with the same work domain, if a way can be found to describe that work domain comprehensively, then the need for dealing with unanticipated situations can be minimized. This implies that reliance on KBB should be reduced.

The above assertions are backed by empirical evidence (for a review, see Vicente and Rasmussen 1992). Many researchers assert the immense power of perception, and other studies have found that perception is superior to analytical problem-solving in terms of errors and efficiency. Furthermore, research has shown that people naturally favor the less effortful route of perceptual processing over the more effortful path of analytical problem-solving, and that, given the proper conditions, a proficient level of performance can result from perceptual processing.

While people naturally prefer lower levels of cognitive control, presenting information in such a way that all task demands could be fulfilled by SBB alone does not imply that operators will not activate higher levels of cognitive control. Rather, the current demands of the task, the person's experience and the form in which information is presented together determine the level of cognitive control that is activated. Thus, the argument here is not that all interfaces should attempt to support all actions via SBB and RBB, but rather that all levels of cognitive control should be supported as operators tend to switch between levels freely. KBB needs support just as much as does SBB and RBB.

The discussion to this point has been somewhat of a simplification. The three levels of cognitive control cannot be treated independently, as a realistically complex task will usually require consideration of all three levels. The implication is that information should be presented to operators in a way to support all three levels of control. Taken as a whole, this discussion leads to a goal for interface design: to design interfaces in such a way as to not force cognitive control to a higher level than the demands of the task require, while at the same time providing the appropriate support for all three levels. The benefit of the SRK taxonomy is that it indicates the constraints on inducing each of the levels of cognitive control. For instance, SBB can only be activated when information is presented in the form of time–space signals, familiar perceptual forms (signs) trigger RBB and KBB is activated by meaningful relational structures (symbols).

5. ECOLOGICAL INTERFACE DESIGN

The preceding theoretical development is the foundation of the principles of EID, a set of prescriptive principles that allow designers to develop interfaces that satisfy the goal for interface design just presented (Vicente and Rasmussen 1992). This framework consists of three principles, each corresponding to one of the levels of cognitive control.

1. SBB — to support interaction via time–space signals, the operator should be able to act directly on the display, and the structure of the displayed information should be isomorphic to the part–whole structure of movements. This principle attempts to structure the interface to take advantage of SBB. Since the operator cannot usually act directly on work domain components, action at the SBB level is concerned with the manipulation of on-screen elements. The use of a mouse, trackball or some other direct manipulation device is preferred over command languages for this task because it maintains the communication of spatial–temporal aspects of the perception–action loop intact. Further, to aid skilled perceptual-motor performance in complex work domains, the interface should be designed in such a way that the aggregation of elementary movements into more complex routines corresponds with a concurrent chunking of visual features into higher level cues for these routines. In other words, the structure of the displayed information should be isomorphic to the part–whole structure of movements. This aim can be accomplished by revealing higher level information as an aggregation of lower level information. In this way, multiple levels are visible at the same time in the interface, and operators can guide their attention to the level of interest, depending on their level of expertise and the current demands.

2. RBB — provide a consistent one-to-one mapping between the work domain constraints and the cues or signs provided by the interface. This principle attempts to support RBB. At this level the interface should provide operators with signs that they can use to select appropriate actions. Conventional interfaces do not provide a consistent mapping between the perceptual cues they provide and the constraints that govern the work domain's behavior. This underspecification can lead to procedural traps, situations in which operators rely on their normal rule set but to little effect. EID attempts to overcome this problem by developing a unique and consistent mapping between the constraints that govern the behavior of the work domain and the cues provided by the interface. This approach should reduce the occurrence of errors, because cues for action will uniquely define the work domain state. Adherence to this guideline should make it possible for operators to often operate the plant by relying on perceptual cues instead of having to resort to KBB. There are two advantages to this approach. First, since RBB is less effortful than KBB, this approach is more economic. Second, because there is a 1:1 mapping between symbols and signs, operators can exhibit what looks like KBB by invoking RBB. The advantage of knowledge-based control is that it is not restricted to specific and/or frequently encountered situations like rule based control. Therefore, the second principle of EID allows operators to take advantage of the cognitive economy of RBB, while at the same time preserving the wide applicability of KBB.

3. Represent the work domain in the form of an AH to serve as an externalized mental model that will support knowledge based problem-solving. This final principle attempts to provide support for KBB. This is essential because KBB is usually an effortful and error-prone activity. The limitations of KBB are amplified when applied to complex systems that require problem-solving within a complex net of causal relations.

The approach of EID to this problem is to reveal the problem space, in the form of an AH representation, to the operator (for an example, see Vicente and Rasmussen 1990). This principle thus inherits all of the properties of the AH described above: it provides a psychologically relevant domain representation that contains the information that operators need to cope with unanticipated events. In addition, making the AH visible in the interface provides operators with a normative model of the work domain that can support thought experiments and other activities. In short, it relieves operators of some of the burden of keeping track of the causal net within which they are reasoning.

6. LIMITATIONS AND GENERALIZABILITY

While the EID framework enjoys considerable theoretical and empirical support, a number of limitations must be acknowledged. First, an AH representation can only be developed if the designer has an adequate understanding of the work domain. Thus, the EID framework may not be applicable to application domains that are not very well understood. Second, the robustness of an EID interface in the face of sensor noise and failure is not yet known. Empirical research is needed to address this issue. Third, in some cases, EID may require the measurement of some property that is not available via current sensor technology. While in some cases this problem can be overcome through programmed analytical routines, sensor technology may limit the applicability of EID.

Finally, although EID has been successfully applied to a broad range of domains (including process control, nuclear operations, aviation, medicine, and information retrieval; Chery and Vicente 1999), the generalizability of these ideas to a broader range of application domains needs to be assessed empirically. As the primary prerequisite for applying the framework is knowledge of the constraints inherent in a domain, there is good reason to be optimistic about EID's generalizability. In principle, it is irrelevant what these constraints are as long as they can be described in some way that can be mapped onto the perceptual features of a display. However, the generalizability of these claims needs to be continually tested empirically for new application domains.

REFERENCES

CHERY and VICENTE, K.J., 2000, (in this volume)

HAJDUKIEWICZ, J. and VICENTE, K. J., 1999, An ecological approach to human factors, 2000, (in this volume).

MESAROVIC, M.D., MACKO, D. and TAKAHARA, Y., 1970, *Theory of Hierarchical, Multilevel, Systems* (New York: Academic Press).

RASMUSSEN, J., 1983, Skills, rules, and knowledge; signals, signs, and symbols, and other distinctions in human performance models. *IEEE Transactions on Systems, Man, and Cybernetics*, SMC-13, 257–66.

RASMUSSEN, J., 1985, The role of hierarchical knowledge representation in decision making and system management. *IEEE Transactions on Systems, Man, and Cybernetics*, SMC-15, 234–43.

VICENTE, K.J. and RASMUSSEN, J., 1990, The ecology of human–machine systems II: Mediating 'direct perception' in complex work domains. *Ecological Psychology*, 2, 207–50.

VICENTE, K.J. and RASMUSSEN, J., 1992, Ecological interface design: theoretical foundations. *IEEE Transactions on Systems, Man, and Cybernetics*, SMC-22, 589–606.

Ergonomics of CAD Systems

H. Luczak, J. Springer and L. Schmidt

Institute of Industrial Engineering and Ergonomics, Aachen University of Technology, Aachen, Germany

1. INTRODUCTION

Design departments are usually equipped with highly complex CAD systems. Their application is supposed to increase the efficiency of the design process and to rationalize the product development process. However, more and more experiences of CAD users reveal a gap between the goals and the reality of CAD implementation. Also, experiments in different design departments show that users do not take full advantage of the functional capacities of CAD systems. This is one reason for poor performance in the production of technical drawings.

The characteristics of CAD systems that are most in need of improvement are their suitability for task requirements and their ability to adapt to different user qualifications. However, if ergonomic design rules are ignored, the consequence will be low design effectiveness, low efficiency and a higher strain level.

2. DESIGN WORK AND CAD SYSTEMS

2.1. Characteristics of Design Tasks

Any ergonomics approach to solving a design problem starts with an analytical procedure to identify and systematize the features and characteristics of the respective work or task (Luczak 1997). Hence, if the objective of this chapter is to describe and order the facts and regularities and to show how CAD systems interfere in design work, it is necessary to divide the task into a sequence of individual actions, so that the essentials of this (design) work become obvious.

Technically speaking, in the generation cycle of a product, the design department is the functional unit in a company where customer orders specified as customer demands or technical requirements are transformed into a mostly graphical model of the product. This model's work results consist of parts lists, parts drawings and composition drawings derived from the application of natural science and engineering knowledge about the product domain. The purpose of the CAD system in this context is to have the work result coded in a computer-compatible form, so that all functional units in the company can access the engineering data produced in the design department.

Organizationally speaking, the design process consists of phases in which a product order is subdivided into tasks and subtasks with a specified intermediate or final work result (Figure 1). Besides this phase-oriented sequential labor partition, a
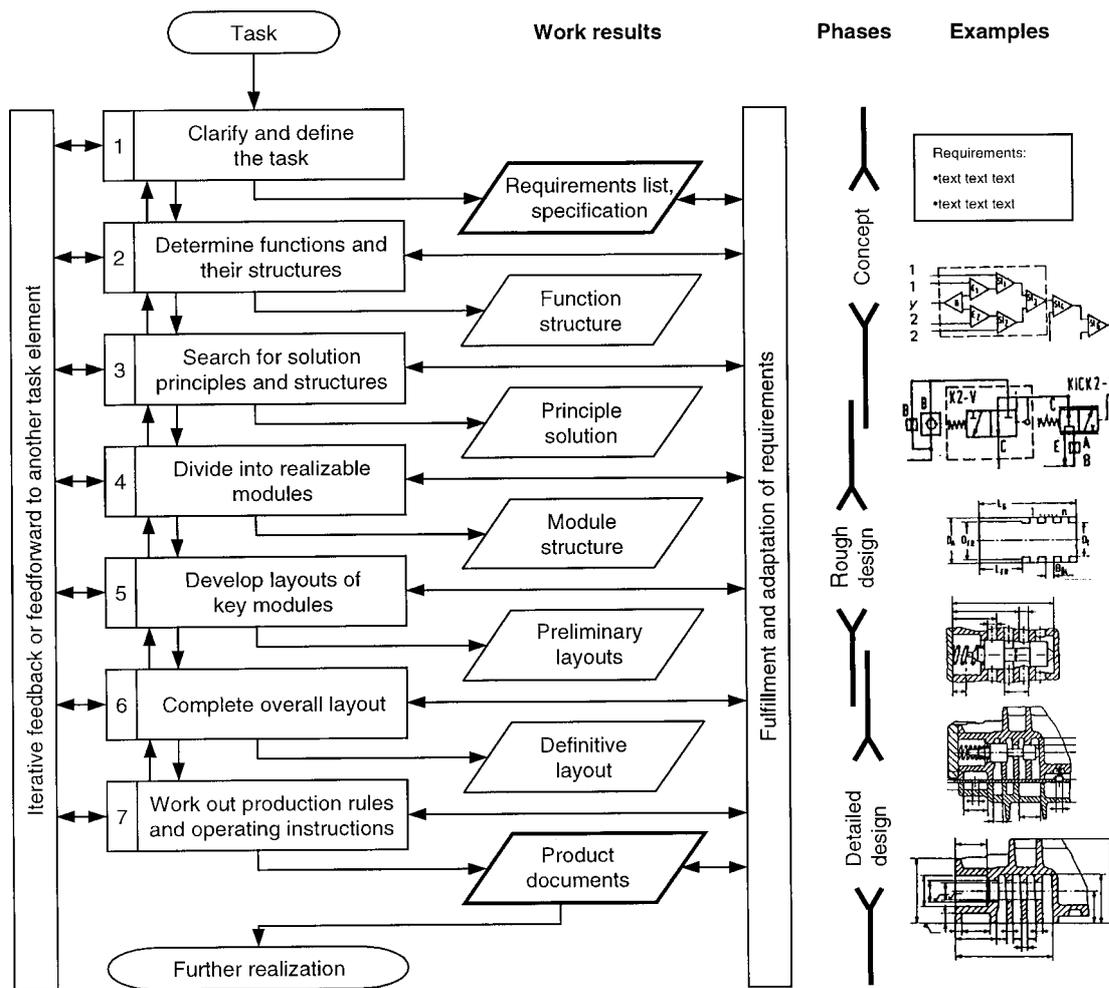
Figure 1. Steps and work results in the design process (according to VDI 1993, Pahl *et al.* 1996)

hierarchical partition from the academically qualified design engineer down to the draftsman can be found, as well as a competency oriented labor partition into mechanical, electrical, hydraulic, etc. problem-solving and integration procedures. The CAD system has the functions of rationalizing the operations management of information handling and of integrating different information chunks into a complete work result: the completely documented technical product.

Ergonomically speaking, design work is "informatory work" in that it consists of human information processing procedures only, but with different levels of cognitive control. According to Rasmussen (1986), three primary levels of cognitive control can be identified:

- Knowledge-based processes, such as defining objectives of a product, setting goals for development procedures, identifying problem domains, and initiating problem-solving where methods are unknown, or even developing the problem-solving methods themselves.
- Rule-based routine processes, such as task execution using prescribed methods and procedures with algorithmic and perhaps iterative characteristics, such as the application of iterative optimization tools or simulation procedures to test functional components of the product; mathematical model formulation and calculation; creation of parts lists or transforming sketches into drawings.
- Skill-based automated processes, such as data entry, data transfer or data recall without invoking conscious information processes.

2.2. Knowledge Representation for Problem-solving and Decision-making

To take these technical, organizational and ergonomic aspects of the product development process into account, it is necessary to adopt a complete and unambiguous structure for design work as a basis for CAD systems engineering. This structure forms a conceptual analytic framework for the task-specific representation of knowledge in a kind of navigation and solution space, according to the knowledge-based level of human information-processing procedures mentioned above. The framework is characterized by the compatible representation of the product developer's mental model, with the aim of supporting problem-solving and decision-making in every design process phase.

Fundamental definitions are made in the early, conceptual phases of the product development process when there is still little definite information available. Therefore, when developing CAD systems, special attention must be paid to these first phases: an abstract description of the design problem must be supported by suitable tools and an appropriate problem representation. During the whole design process a homogeneous structure should integrate various media (including data not digitally available) and access methods on information units. It should be possible to address information in its context using a uniform search strategy, as in a knowledge base.

A further important demand on the structure of a CAD system is its correspondence to an intuitive way of thinking, as well as to conscious systematic design methods. Design tasks can be performed more efficiently with computer-generated structures of knowledge representation and methods of interaction compatible with mental models, because unnecessary transformation effort is avoided. Together with the mentally compatible visualization of the product model structure as the carrier of semantic information, partial situation analyses then become possible. This is necessary to evaluate solution variants concerning different product and process versions to obtain construction alternatives.

Rasmussen (1986) described a normative structure for the development of design guidelines based on human information-processing abilities and limitations. His conceptual framework
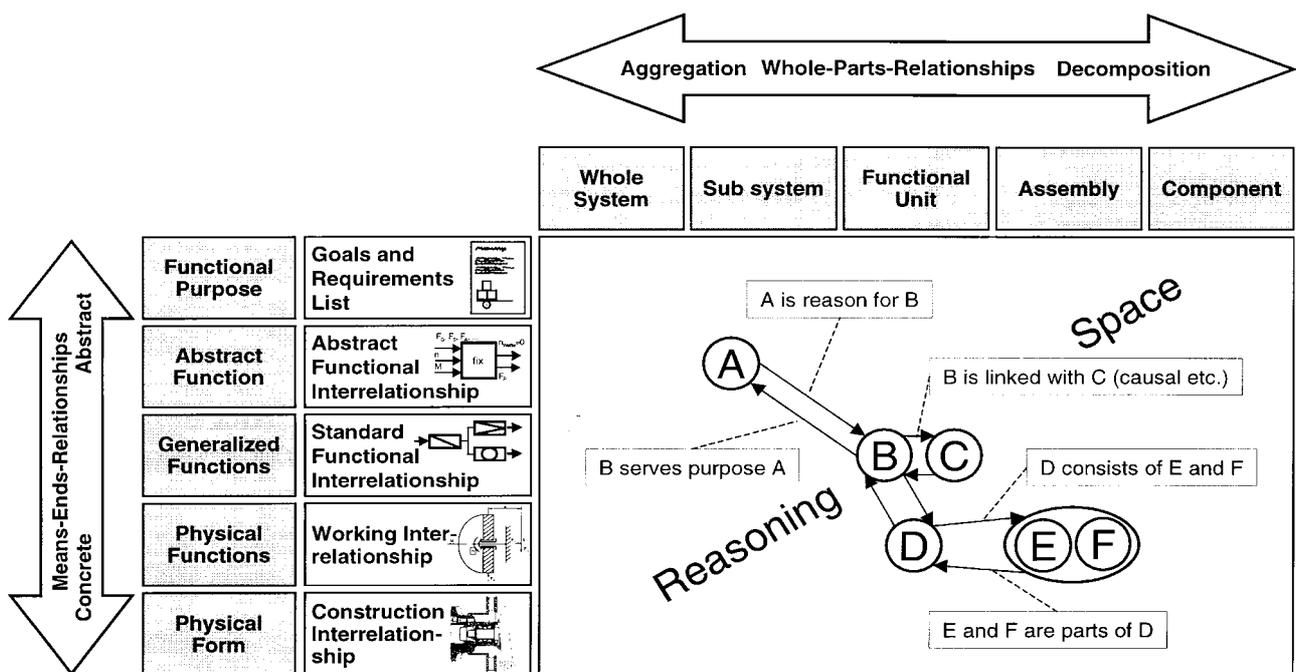

Figure 2. Reasoning space as reference structure for cognitive CAD systems engineering.

can be used to analyze cognitive tasks and subsequently to develop mentally compatible systems. A means-ends abstraction hierarchy represents functional properties of technical systems. This hierarchy describes bottom-up what components and functions can be used for, how they may serve higher level purposes, and, top-down, how these purposes can be implemented. Independently, and orthogonal to this functional dimension, various aggregation levels describe the whole-parts relationships of the system as a hierarchy of parts.

Concerning CAD work, a two-dimensional reasoning space can be the reference structure for cognitive CAD systems engineering, because design process phases can be assigned to different levels of abstraction (Figure 2). The designer's mental model of the task can compatibly be represented in this framework. For example, in the dimension of means-ends relationships, five abstraction levels can be separated.

The requirements list is a result of task definition and clarification to describe essential functional purposes at the most abstract level.

Identification of functional interrelationships in the next design phase refers to the level of abstract function, where conversions of energy, material and signals are defined.

The description of standard functional relationships (e.g. with the standard function items "transform," "modify," "link," "conduct" and "store") is the basis for principal solution variants.

To perform these generalized functions and working principles, physical effects and general material properties (like "solid," "liquid," "resistant to corrosion") are used.

Results of final design in the construction interrelationship are the complete set of drawings, parts lists and detailed production documents.

The hierarchy of parts shown in Figure 2 consists of five levels in this example, in structural analogy to the abstraction hierarchy concerning the whole-parts relationship of technical systems: "whole system," "subsystem," "functional unit," "assembly" and "component."

2.3. Structural Levels of Human–Computer Interaction

To obtain a model of human–machine systems design, it is necessary not only to think of appropriate knowledge representation, but also to focus on structural levels of interactions between people and machines. General guiding principles can be derived from the semiotic model — as first described by Morris (1946) — which has frequently been applied to human–computer interaction. It is capable of bridging the gap between a person- or work-centered view and a support-centered view based on tools or means and measurements. It may also be used to describe the interaction between a design engineer and his CAD system (Figure 3). This description emphasizes the tool-character of the CAD system for the individual work of the design engineer.

As outlined previously, the design engineer and his CAD system are bound to an organizational context by responsibilities for certain types of product orders. This may be a "manufacturing and assembly oriented design in layers," as shown in Figure 3. These orders are redefined as tasks or task elements on a pragmatic level. In a cyclic or stepwise process of order processing, task elements are assigned to the human designer or to the respective tool in the CAD system. In the above example, the designer must consider the physical principles of layer functioning and production, whereas during task execution, the task-representation and modeling in the CAD system leads to provision
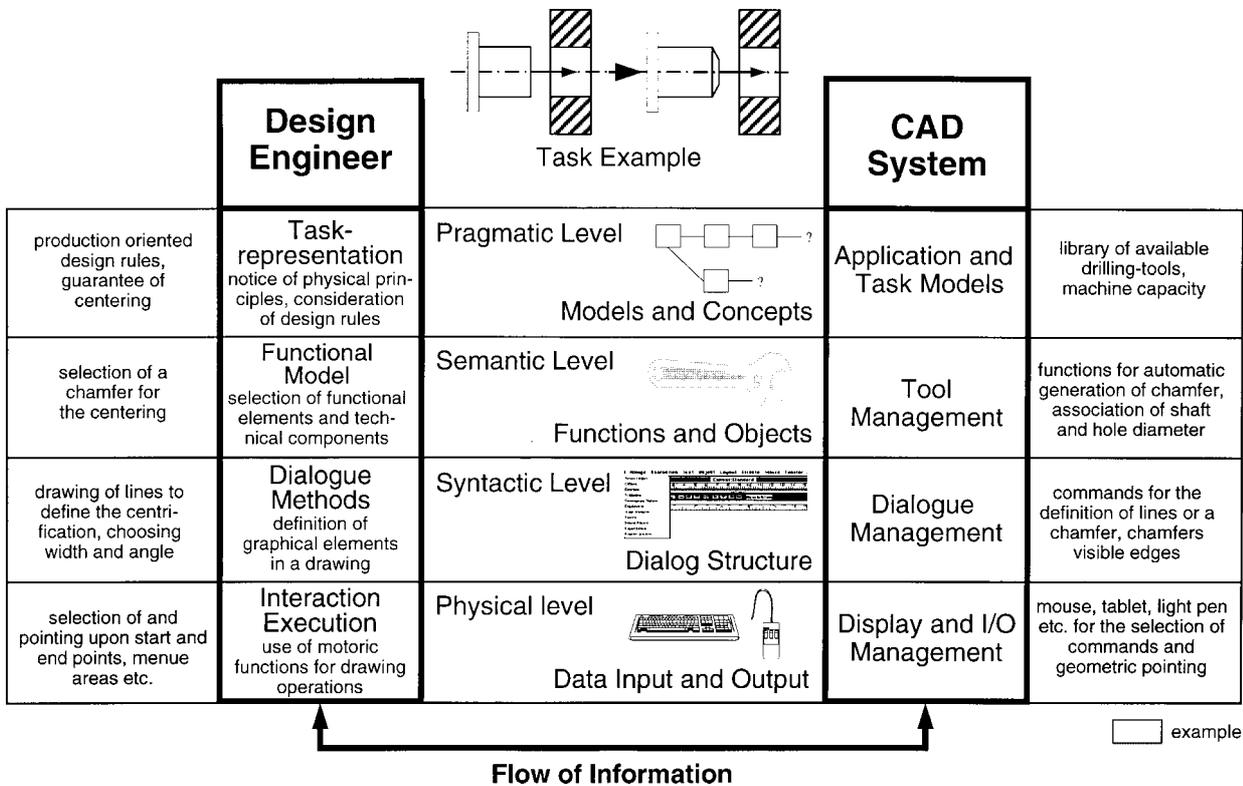
Figure 3. Example of a design task and its execution in the semiotic model of human–computer interaction (according to Luczak and Springer 1997).

and support with a "library of available drilling tools," a survey on "machine capacities for achieving a given surface quality," etc.

On the semantic level, functions of the CAD system are used to specify and generate functional elements and technical components of the artifact to be designed. For example, the selection of a chamfer for centering the shaft is supported by tools for object generation or manipulation in the CAD system. The chamfer is then automatically generated according to shaft diameter and the shaft and hole diameters are associated with a certain fit.

On the syntactic level, the definition of graphical elements in the drawing comes into focus, which means that the user selects dialogue methods to allow easy definition and arrangement of lines and other standardized graphical components of the drawing (center lines, visible edges, etc.).

Last but not least, the physical level also concerns input–output devices as shown in Figure 3. These ensure that dialogue steps are executed by physical operations and represented by perceivable representations on displays. This represents a technical information aspect: the signs must be expressed in a physical way, e.g. visual, auditory or kinesthetic information.

The semiotic model is not only useful for the overview and the example shown, but also it may serve as a structuring guideline for specifying the features of CAD systems used in industry today.

3. ERGONOMIC REQUIREMENTS FOR CAD SYSTEMS

The structure elaborated above can be used to identify the requirements of CAD systems in a task- and user-specific way. The task-appropriate representation of knowledge and the design of the user interface dependent on semantic and pragmatic aspects become the essential factors of success where an effective, efficient and accepted use of a CAD system is concerned.

3.1. Pragmatic Level

Requirements on the pragmatic level of the semiotic model are related to application models and concepts used in the design process:

- The designer should use a CAD internal model with characteristics compatible with his mental model, thus decreasing coding and decoding effort.
- The CAD system should have an easy-to-use functionality for changing the characteristics of models and for generating task- or user-specific models by the combination of different default models.
- The designer should describe characteristics using fuzzy data, especially in the early phases of the development process.
- Different models and different views of a model should be managed simultaneously and the designer should have the ability to switch between them, e.g. switching between global and detail problems (aggregation level) as well as between concrete and abstract definitions of model characteristics (abstraction level).
- Activities directly related to the design task should be aided by the following conceptual characteristics of a CAD system:
- Special functions should exist that allow the designer to mark solutions that have not reached their final state or for which better solutions might exist.
- Because the designer is using various abstraction levels for

the description of the solution, all descriptions must be related to each other. Hand sketch drawings (of functional structures as well as of geometrical layouts) play an important role. The path of a designer through the "reasoning and solution space" (as seen in Figure 3) is characterized by a large number of iterations between the various information resources. This requires parallel handling of the different resources as well as the connections between the representation modes.

- Different concepts should be incorporated for problem-solving or optimizing solutions (e.g. simulation, application programs like solution databases, advanced calculation methods like FEM — Finite Element Method).
- Different concepts for evaluation (e.g. use value analysis) should be integrated. The designer should thus use task or user specific evaluation parameters to verify the evaluation results.
- Furthermore, effective information management is necessary for an effective design process:
- The designer should access and use different information sources based on both alphanumeric and graphical information, such as standards, company standards, supplier information, similar designs created previously, handbooks or legal regulations. It should be possible to manage different information resources simultaneously.
- Information should be given in a consistent way to the designer.
- Functions for navigation, for activation and deactivation of constraints, for visualization and for the control of consistency should be integrated.
- Information must be actualized permanently.

3.2. Semantic Level

The activities of the designer are highly dependent on the availability and suitability of functions to operate on the various objects in a design. Therefore, the characteristics of functions and objects (the semantic level of the HCI model) of the CAD system are important for effective design. Requirements for functions based on the designer's activities are:

- Direct object manipulation should be used intensively in correspondence to the interactive tool metaphor (transparency of the result). Direct object manipulation supports generation of objects (rubber bendings to visualize the effects of generation), geometrical manipulations like stretching, bending, pressing, moving, rotating, etc., as well as deletion of complex objects by simply using virtual erasers as they are used in conventional drawing. Therefore, functionality must be integrated into the objects themselves, and must be visualized by dialogue metaphors as parts of the objects.
- Because most geometrical definitions must contain exact dimensions, numerical feedback must be given continuously during the design process. It can be implemented through written dimensions like those in technical drawings, through numerical displays, software rulers or snap modes based on user-defined grids. The user should be able to switch between scaled drawings and unscaled visualization, as often used in hand sketch drawings, e.g. to make a special functionality transparent (tolerances, functional measures like eccentricity, etc.).

- Hand sketch functionality should be integrated into the CAD systems, but the CAD system should allow the user to decide whether the hand sketch should be changed into the final geometry (CAD elements), or if the hand sketch is used to visualize the unfinished character of a solution.
- An UNDO function should exist to reverse an error and to aid an iterative design process. Mechanisms for storing a design history are also required, which means that copies of different design states must be stored in a user-defined way for documentation of the decision processes as well as for returning to previous design steps.
- The whole functionality of the CAD system must be adaptable dependent on the requirements of the task as well as the individual strategies of problem-solving.
- Objects of the CAD system are data, which are operated on by the functions of the CAD system. Because of the various demands on the product to be designed, the designer must handle a large number of different objects. Therefore, the type and structure of objects managed by a CAD system are important for an ergonomic system design:
- All types of geometric information should be integrated into a homogenous and consistent data structure. Constraints between geometry and related geometric objects like dimensions, crosshatches, etc. should be defined when the designer creates these related objects, and they should be managed by the CAD system automatically.
- The user needs continuous information about the characteristics of defined objects and their relations to other objects outside the CAD system (costs, technological data, etc.).

3.3. Syntactical and Physical Level

Requirements at the syntactical and physical level can be taken from International Standards Organisation (ISO) standards for CAD non-specific requirements, because in CAD offices these factors are basically the same as in normal offices. ISO has worked out a set of standards (ISO 9241) that provide a scientific basis for planning and evaluating workplaces with visual display units (VDU) based on ergonomic criteria. The standard currently comprises 17 parts: General introduction (part 1), task design (2), hardware and environmental factors (3–9), and software and usability (10–17). Parts 3–17 are implicitly adjusted to the semiotic model so that the reader can find the appropriate standard quickly.

For an improved design of CAD hardware and software, the needs of the designers must be recognized and formulated clearly for the system developers. With this requirement, the ISO standards as well as the "EU directive on the minimum health and safety requirements for work with display units" provide a general guideline, which must be specified to meet the needs of designers and the tasks and working conditions in the design department.

REFERENCES

LUCZAK, H., 1997, Task analysis. In G. Salvendy (ed.), *Handbook of Human Factors and Ergonomics*, 2nd edn (New York: Wiley), 340–416.

LUCZAK, H. and SPRINGER, J., 1997, Ergonomics of CAD systems. In M. Helander, Landauer and Prabhu (eds), *Handbook of Human–Computer Interaction*, 2nd edn (Amsterdam: Elsevier), 1349–94.

MORRIS, C.W., 1946, *Signs, Language and Behavior* (New York: Prentice-Hall).

PAHL, G., BEITZ, W. and WALLACE, K. (eds), 1996, *Engineering Design: A Systematic Approach,* 2nd edn (Berlin: Springer).

RASMUSSEN, J., 1986, *Information Processing and Human–Machine Interaction: An Approach to Cognitive Engineering* (New York: North-Holland).

VDI 2221, 1993, *Systematic Approach to the Development and Design of Technical Systems and Products* (Düsseldorf: VDI).

Functional Systems Design versus Interface Design

V. A. Riley

Honeywell Technology Center, 3660 Technology Drive, Minneapolis, MN 55418, USA

1. INTRODUCTION

In the traditional systems development process, a system is built to solve an operational problem, and the functions of the system (what the system does) and the operational logic of the system (how the system does what it does) are defined according to the engineering requirements. The user interface is typically defined toward the end of the development process. Because the functions and operational logic of the system have already been defined, the interface must then build a bridge between how the user thinks about the task and how the system was designed to operate. This requires the user to learn the system's functions and operational logic; when the user must perform a task using the system, he/she must first decide which system functions are appropriate to accomplish the task, then apply the appropriate procedures (based on the operational logic) to use those functions. For a complex system, this can require extensive training requirements and may cause errors during system operation.

In contrast, human-centered functionality seeks to define the functionality and operational logic of the system based on the user's task requirements and on how the user thinks about the operational environment. The interface is then a natural extension of how the user thinks about the task and environment, and the interface definition drives the definition of system functionality. This reverses the traditional process, by beginning with the interface requirements and ending with the engineering requirements. But because the final system works much more like the user thinks, extensive training is no longer required, and the user no longer has to grapple with error-prone complexity.

It should be noted that this approach is most effective when the system and the plant are different entities. For example, in aircraft operation, the operational environment and the aircraft exist separately from the systems that guide the aircraft, such as the flight management system; this provides an opportunity to employ the pilot's existing mental models of the environment and the aircraft in the design of the flight management system's functions, logic and interface. However, in process control, there may not be a similar external criterion; the plant may be the system that the user is operating. In the first case, the engineering solution to flight management system functions and operational logic may seem superfluous or arbitrary to the pilot because there is no apparent direct relationship to the characteristics of the aircraft or environment. Thus, learning to use the system requires learning arbitrary rules of operation that are specific to the system and are not directly relevant to flying the aircraft. In the second case, the process control plant establishes the operational criteria, so learning the system's operations is fundamental to using the plant. Consequently, human-centered functionality may be most valuable in those cases where the user has an existing mental model of plant operation that is separate from the operation of the system being designed, or where knowing the underlying operations of the system is not fundamental to operating the plant.

2. SYSTEM-CENTERED FUNCTIONALITY VERSUS HUMAN-CENTERED FUNCTIONALITY

The functionality of a system is the set of functions that the system can perform. The interface provides the means for the operator to control and monitor these functions. The operational logic is the set of procedures by which these functions are controlled and monitored. Typically, functions and operational logic are defined based on the engineering requirements, and the interface is added to the system as a means of representing the functions and operational logic to the operator.

An example of system-centered functionality is given by how the flight management system of a typical modern aircraft is used to set up a holding pattern. If the pilot is given a clearance such as "Hold at 20 miles south-west of Farmington", he/she must first determine what functions are provided by the system that would meet the requirements of the clearance. Once the pilot decides on the proper functions, he/she must remember how to access the functions through the interface. Finally, the pilot has to recall the procedure (operational logic) to be applied. In this case, the flight management system calculates trajectories in terms of waypoints (artificial locations on the ground), so the pilot must first define a waypoint at the place designated by the clearance (20 miles south-west of Farmington). The pilot must then insert that waypoint into the proper position in the flight plan, then define a holding pattern around it. This example demonstrates how system-centered functionality requires the user to think about the process in terms that are peripheral to the required operation but are dictated by the design of the system. In other words, the pilot must translate from the logic of the clearance to the logic of the system in order to tell the system how to comply with the clearance.

More human-centered functionality would remove the need for the pilot to perform this translation, instead enabling the pilot to enter the clearance requirements directly. For example, the system may require the pilot to enter "Hold 20 south-west Farmington", a sequence of steps that matches the logic of the clearance (Riley *et al.* 1998). In other words, the human-centered operational logic establishes system functionality that is compatible with task logic, so the input to the pilot and output from the pilot into the system are essentially the same.

3. FUNCTIONALITY VERSUS INTERFACE

It is important to note that the essential difference between a system-centered solution and a human-centered solution is at the level of system functionality instead of at the level of the interface. There is nothing about the functionality or operational logic that would necessarily dictate a textual or graphic interface, or any particular interface look and feel. While it is often tempting to apply graphic user interfaces to systems, the mere fact that the interface is graphic may not ensure, or even improve, usability.

The value of concentrating on functionality as the focus of usability efforts was demonstrated by researchers at NASA Langley Research Center (Abbott 1995), who applied a graphic user interface to an existing flight management system, while preserving the underlying functionality and operational logic of the existing system. Pilots who performed complex flight

procedures with the new, graphic interface showed no training or accuracy benefits when compared with pilots who performed the same procedures with the existing text-based interface.

4. FUNCTIONAL COMPLEXITY VERSUS CONCEPTUAL COMPLEXITY

Many users are concerned with the growing level of systems complexity, and some are calling for reduced complexity as a means to greater usability. However, many systems are complex because the operational environment and the tasks to be performed with the system are themselves complex; arbitrarily reducing system complexity may therefore make the system even less usable because its performance would be compromised.

One way of addressing this problem is to separate functional complexity from conceptual complexity. A good illustration of this distinction is provided by personal computers using the desktop interface; although these systems are far more complex (functionally) than the DOS machines that preceded them, users find them conceptually more simple. This is because the desktop interface translates the underlying functionality of the system into a conceptual world that the user already understands; file management functions, for example, are translated by the interface into operations involving virtual paper documents, file folders, rubbish bins or recycling bins, etc. To the extent the on-screen world works the way the real-world does, the user can perform tasks with little or no training.

This demonstrates how a suitable metaphor can make the underlying system-centered functionality appear to be human-centered. However, the metaphor is not a panacea; in the case of personal computing, the metaphor was imposed on the operating system after the essential functions of the system were already defined. This means that only a subset of system functions fall under the metaphor; secondary metaphors (such as the distinction between "radio buttons" for mutually exclusive options and "check boxes" for non-exclusive options) are applied where appropriate, but in the current desktop interfaces, many sophisticated system functions still require learning because the metaphor does not dictate their logical operation. The most powerful application of a metaphor, then, would be before system functions are defined, so the greatest amount of system functionality can fall under the logic suggested by the metaphor.

5. TECHNOLOGY-CENTERED VERSUS HUMAN-CENTERED SYSTEMS DEVELOPMENT

Truly human-centered functionality is defined based on how the user thinks about the operational environment and the logic of the tasks the user must perform, apart from the actual operation of the system. This suggests that the traditional system development process must be turn around. Figure 1 illustrates the current process, where operational requirements (the problem to be solved by the system) drives functional requirements (what the system will do to solve the problem), which in turn drives system requirements, which in turn drives the interface requirements. Figure 2, on the other hand, illustrates a more human-centered process, where task requirements drive interface requirements, which in turn drive system and functional requirements.
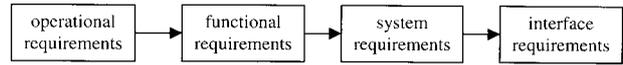
Figure 1. The traditional systems development process begins with engineering requirements; the interface is added after system functions and operational logic have been defined.
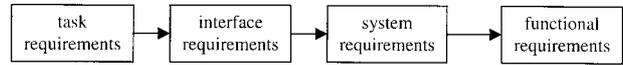

Figure 2. A more human-centered process begins with the user's task requirements and interface requirements; these, in turn, drive the engineering requirements.

Technology-centered system development has been appropriate where the capabilities of the technology limited the functionality of the system, and functionality had to be provided as efficiently (in terms of processing power) as possible. However, the current level of technology is making truly human-centered systems more possible. In the past, the user had to accommodate the system because the system was severely limited. In the future, more capable technology will enable the system to accommodate the user.

6. ROLE OF HUMAN FACTORS IN HUMAN-CENTERED SYSTEMS DEVELOPMENT PROCESS

All of this implies that the role of human factors in systems development processes will change. Since the beginning of human factors practice, practitioners have advocated an earlier role for human factors in the systems development process. One justification for this is that a better interface can be developed if the system's functions and operational logic are known earlier. Another justification is that early involvement by human factors experts can help influence the functionality of the system better to support the operator. A third justification, though, would represent the other extreme from the first, that human factors requirements drive the engineering requirements, rather than the other way around. This would require a much larger role for task analysis and other human factors methods at the beginning of system development, as a means to defining what the system will do rather than just how the system will be represented to the operator.

REFERENCES

ABBOTT, T. S., 1995, The evaluation of two CDU concepts and their effects on FMS training. *Proceedings of the 8th International Symposium on Aviation Psychology*, 233–8.

PALMER, M., ROGERS, W., PRESS, H., LATORELLA, K. and ABBOTT, T., 1995, *A Crew-centered Flight Deck Design Philosophy for High-Speed Civil Transport (HSCT) Aircraft*. NASA Technical Memorandum 109171 (Hampton: Langley Research Center).

RILEY, V., DEMERS, B., MISIAK, C. and SCHMALZ, B., 1998, The cockpit control language: a pilot-centered avionics interface. In *HCI-Aero 1998* (Montreal), pp. 35–9.

HCI Hypermedia: Usability Issues

A. Dillon

HCI Laboratory, Indiana University, USA

1. INTRODUCTION

hypermedia is a general term used to describe the presentation of graphical, textual, audio and video information in nodes (chunks) that can be linked together and accessed in a manner determined by the immediate interests of the user. Originating as an idea for mechanized information access and organization that better reflected the natural workings of the human mind by Vannevar Bush (1945), the potential of the computer to provide the best means of supporting hypermedia was recognized early on by thinkers such as Ted Nelson and Doug Engelbart (for a historical overview see Conklin 1987). The move toward hypermedia-based digital documents holds with it the promise of user-controlled, immediate access to the world of published information and stored data. While originally a specialist application domain, in the last few years the World Wide Web has brought to everyone's desktop the power and problems of hypermedia interaction. Yet from the outset, human factors researchers have noted a range of user issues that prevent the simple transition from analog to digital resources.

2. THE RANGE OF HUMAN FACTORS CONCERNED WITH HYPERMEDIA

Relevant research on the subject of interface design for information access in electronic domains began several decades ago, but it is only in the last 15 years or so that the technology existed to support empirical research into hypermedia interfaces. Since then, there has been an explosion of studies examining such standard user-performance issues as speed and accuracy of screen reading, the design of navigation aids for information spaces, the combinatorial value of images and text for information exploration, and the educational value of hypermedia information.

Dillon (1994) reviewed the ergonomic literature on digital documents and reported findings on the outcome and process differences between digital and analog documents at the physical, visual, and cognitive levels. Since then, interest in the social aspects of hypermedia creation and use has also blossomed and there now exists a large and sometimes contradictory research literature on the human factors of hypermedia design. Of particular interest to most researchers are the questions of what factors most affect user performance with digital documents and how best to exploit hypermedia for real-world tasks.

3. FUNDAMENTAL DIFFERENCES BETWEEN DIGITAL AND ANALOG MEDIA

Dillon (1994) outlined five outcome and three process variables that have been studied by researchers in hypermedia. The outcome variables are:

- Speed
- Accuracy
- Comprehension
- Fatigue
- Preference

In general, people have repeatedly been shown to read 20–30% slower from typical computer screens than paper (see, for example, Gould *et al.* 1987) and this has given rise to a concern with image quality and research at the perceptual level of use. "Accuracy" of use has proved more difficult to measure and many researchers have used highly constrained tasks such as error location in proof-reading or searching for target information in an information space as indices of performance accuracy or effectiveness. The results are highly task-dependent, but there are some suggestions that hypermedia is better suited to search and comparison tasks than complex learning tasks.

The effect of presentation medium on comprehension is particularly difficult to assess because of the lack of agreement on how comprehension can best be measured. However, in a detailed review of the learning outcomes of hypermedia and other instructional presentations, Dillon and Gabbard (1998) reported that the experimental findings show little support for the use of hypermedia over any other medium in educational settings.

There is also a popular belief that reading from screens leads to greater fatigue, but the results are inconclusive. The perception of fatigue or eye strain from screen reading is obviously linked with the user's preference for screen or paper-based reading, where it seems that paper retains popularity with many users, although this is highly contingent on task.

The major process differences that have been experimentally studied include:

eye movements

manipulations

navigation

Comparison of eye-movements with paper and hypermedia documents shows that there is little difference although screen readers may make slightly more regressions (Gould *et al.* 1987). Manipulation and navigation, however, are much more commonly assessed and the results indicate significant effects, with large hypermedia documents in particular producing navigation problems for users, leading to increased disorientation and poorer mental model formation of layout and narrative flow.

4. EXPLAINING THE DIFFERENCES

A range of potential explanations for these differences has emerged hand in hand with the findings of significant process and outcome differences between the media. Following Dillon's (1994) review, these are grouped by human factors under physical, perceptual, cognitive, and social issues.

4.1. Physical Sources of Difference

There are wide ranging physical differences between paper and hypermedia that affect the manner in which readers use them, such as:

orientation

aspect ratio

handling and manipulation

display size

While myths abound on the theme of display size or orientation, research indicates that the major problems occur in designing appropriate manipulation facilities for the digital medium where there are few standards beyond the use of a mouse

as a pointing and selection device. On balance, though, the empirical literature would seem to indicate that the primary causes of the reported performance differences between the media do not lie at the physical level of use.

4.2. Perceptual Sources of Difference

The user's visual processing of text and images is contingent on image quality which varies on computer screens as a function of the refresh rate and resolution. This has led to research into the relationship of image quality to user performance with digital documents. Some of the major areas of attention have been:

    flicker
    screen dynamics
    visual angle of view
    image polarity.
    anti-aliasing

The definitive work on perceptual variables in reading from computer screens was carried out by Gould *et al.* (1987) who empirically demonstrated that under the right conditions differences between the digital documents and paper disappear. Their results suggested that the performance deficit in speed was the product of an interaction between a number of individually non-significant effects: display polarity (dark characters on a light, whitish background), improved display resolution, and anti-aliasing. This image quality hypothesis has been replicated by Muter and Maurutto (1991). It is fair to conclude that high image quality is a necessary (but insufficient) precondition of efficient and effective use of any hypermedia.

4.3. Cognitive Sources of Difference

While the identification of the role of image quality was an important breakthrough for our understanding of human response to digital information, the potential of hypermedia to represent information in novel and configurable forms is what truly interests designers and users. This potential has served to emphasize the importance of cognitive issues in design. To date, researchers have focused on the following type of variables:

    Visual memory for location and layout
    Schematic representations of concepts and relationships
    Navigation patterns and mental map formation
    Comprehension of various information structures

Issues relating to re-structuring of information, together with the nature and amount of linking involved with hypermedia, have become central questions of human factors researchers (see, for example, Nielsen 1995). It is almost taken on faith by many that analog documents consist of a linear format which demands serial interactions, while hypermedia allows non-linear formats which offer more flexible and "natural" methods of use. Such a distinction would appear dubious and there is little evidence to support either assumption.

Systematic research on cognitive issues shows that users can become overwhelmed with highly interlinked nodes of information that fail to cohere or conform to expectations of structure. This has given impetus to studies of the strategies people adopt in navigating through information spaces, the best representations of location and order to provide users at the interface, and the best combination and organization of images, text, and sound to ensure learning. To date, few definitive answers

have emerged and it remains a major design challenge to produce a hypermedia document that outperforms its paper equivalent.

5. SOCIAL SOURCES OF DIFFERENCE

In recent years, interest in the social aspects of digital information creation and use has increased. As hypermedia becomes the universal interface style for so much of the information with which people interact, researchers have started to address the social and cultural variables that influence use and acceptance of digital documents. Issues that are seen as central at this level include:

    The sharing of meaningful representations
    Control of authorship
    Formation and application of document genres
    Collaborative hypermedia creation and use

It is clear that image quality is necessary but insufficient to ensure usability and that any hypermedia system invariably exists in a socio-cognitive environment of use that transcends physical space. Creators and users of hypermedia might be physically and temporally separated, of different language communities, and may create and shape personal hypermedia spaces that are unique. To date, most work on these issues has been discursive rather than empirical.

6. DESIGNING USABLE HYPERMEDIA: USERS, TASKS AND INFORMATION SPACES

The World Wide Web is essentially a hypermedia environment and its rapid growth has increased the need for understanding how best to design hypermedia. The lesson from research is that while it is important to take account of the visual ergonomic issues, any hypermedia application will only be of benefit to users if it supports their navigation, location, and sharing of information. Few computer-based innovations have proved as difficult to design so that users benefit significantly. Successful examples of hypermedia have only been developed through long-term iterative design processes involving frequent usability evaluations (see, for example, Landauer *et al.* 1993).

While following an iterative user-centered design process is usually beneficial, it can be slow and costly. In one respect, hypermedia applications share many of the standard interface design problems of any other application in terms of icon design, color coding, error message and documentation design, screen layout, etc., thus enabling designers to call on a large existing database of human factors research to guide parts of their design. What is specific to hypermedia, or at least what hypermedia places greater emphasis upon, is the flexibility afforded users to access any part of the information space they choose, the combinatorial form of the information types, and, with the WWW, the potential to traverse multiple sites which share nothing in common but a link.

Of necessity, any usable design must be grounded in the context of intended use which requires developers to analyze the users, their information tasks, and the environment in which they will work. Large individual difference effects have been observed and hypermedia presentations often seem to hinder rather than help the people at whom they are aimed (Dillon and Gabbard 1998). Clearly users with experience with any information type have expectations of form and organization that might be leveraged in hypermedia designs.

While it was recognized early in the research literature that not all information was well served by being presented in hypermedia form, few definitive answers have been provided concerning just which types of information work best. Shneiderman (1989) stated that the best hypermedia were information types that were naturally fragmented and that users only want to access in parts. This might suggest some natural limits on the type of information that users wish to explore in hypermedia form, but as yet there appears to be no slowing of the push to digitize any and all types of information. Furthermore, such a perspective is limited to imagining hypermedia versions of existing forms of information; as hypermedia evolves we should anticipate the emergence of new document genres and types that have no analog equivalent.

Any analysis of information type cannot be divorced from task requirements users have for it. For example, the complete works of Shakespeare may not seem an ideal candidate for fragmentation and hyperlinking, yet a digital version would prove highly useful to a scholar searching for quotations or seeking parallels across plays or poems. Certainly, where location is supported by search facilities, hypermedia offers tremendous potential to users exploring large documents. However, once located, many users seem to prefer viewing a printed version of the target information. This suggests that thinking about hypermedia should move forward from comparisons with analog documents toward an appreciation of what contexts of use it best serves and how it might supplement and be supplemented by existing media to support any given users and tasks.

7. RECOMMENDATIONS

hypermedia offers a powerful means of presenting large amounts of rapidly accessible information to a user that can be explored as needed. However, hypermedia can also overwhelm and disorient users if it is not well-designed. The concept of hypermedia is sufficiently wide-ranging to prevent simplistic design guidelines on link density or screen size to be specified, although there is an ever-expanding list of design recommendations being pushed by web-designers. However, to maximize the advantages and minimize the potential problems, the following general recommendations have emerged from the human factors research on hypermedia:

1   Provide high image quality screens to improve readability.
2   Use colors sparingly; black text on white backgrounds is optimal for reading lengthy text.
3   Support simple manipulation of screen contents with a mouse or equivalent input device; minimize repetitive motion such as scrolling through use of directly accessible navigation controls.
4   Links should clearly signify destination and be retraceable; increasing link density can increase the cognitive load on users.
5   Structure the information to conform to user expectations, or provide structural cues such as maps or overviews to introduce new structures. Determine new structures by considering the tasks being performed.
6   Use animation only to attract attention or convey a process.
7   Minimize load time. Users strongly dislike waiting for screens to load.
8   Add value by offering facilities to perform desirable or advantageous tasks that are impossible, difficult or time-consuming with analog media.

However, all the above recommendations are context-dependent and are less critical to the success of any hypermedia application than the central HCI recommendation of testing any design on real users.

REFERENCES

BUSH, V., 1945, As we may think. *Atlantic Monthly*, 176(1), July, 101–8.

CONKLIN, J., 1987, Hypertext: an introduction and survey. *Computer*, September, 17–41.

DILLON, A., 1994, *Designing Usable Electronic Text: Ergonomic aspects of human information usage*. (Bristol PA: Taylor & Francis).

DILLON, A. and GABBARD, R., 1998, hypermedia as an educational technology: Review of the quantitative research literature on learner comphrension, control and style. *Review of Educational Research*, 68(3) 322–49.

GOULD, J., ALFARO, L., FINN, R., HAUPT, B., and MINUTO, A., 1987, Reading from CRT displays can be as fast as reading from paper. *Human Factors*, 26(5) 497–517.

LANDAUER, T., EGAN, D. REMDE, J., LESK, M., LOCHBAUM, C., and KETCHUM, D., 1993, Enhancing the usability of text through computer delivery and formative evaluation: The SuperBook Project. In: C. McKnight, A. Dillon, and J. Richardson (eds) *Hypertext: A Psychological Perspective*. (Chichester: Ellis Horwood), pp.71–136.

MUTER, P. and MAURUTTO, P., 1991, Reading and skimming from computer screens and books: the paperless office revisited? *Behaviour and Information Technology*, 10(4) 257–66.

NIELSEN, J., 1995, *Multimedia and Hypertext: the Internet and Beyond.* (Boston: AP Professional).

SHNEIDERMAN, B. (1989) Reflections on authoring editing and managing hypertext. In E. Barratt., (ed.) *The Society of Text*, (Cambridge MA: MIT Press), pp.115–31.

Human Acceptance of Information Technology

A. Dillon

SLIS, Indiana University, Bloomington, IN, USA

1. INTRODUCTION

Despite significant investments in information technology in developed nations over recent decades, concern exists over the extent to which such expenditures have produced the intended benefits. At least part of this concern is based around the issue of whether any information technology is accepted by its intended users. Human factors professionals are interested in understanding the determinants of acceptance and ensuring new designs are built and implemented so as to minimize resistance. This concern has extended the traditional ergonomic concern with usability, or ability to use, to cover acceptance, or willingness to use.

2. DEFINING AND STUDYING ACCEPTANCE

User acceptance can be defined as the demonstrable willingness within a user group to employ information technology for the tasks it is designed to support. Thus, acceptance theorists are less concerned with unintended uses or non-discretionary use of technologies and more interested in understanding the factors influencing the adoption of technologies as planned by users who have some degree of choice. By developing and testing models of the forces shaping user acceptance, human factors researchers seek to influence the process of design and implementation in a manner that will minimize the risk of resistance or rejection by users.

The scientific concern with user acceptance is comparatively recent, since traditionally, developers and procurers of new technology could rely on authority to ensure that technology was used, at least in many industrial/organizational contexts. However, current working practices, as well as the large market for leisure and educational applications of information technology have enabled greater discretion among users thus increasing the need to determine the dynamics of acceptance.

The literature on acceptance is broad, ranging from case studies of accepted technologies, to the individual psychological characteristics of acceptors or resistors (Dillon and Morris 1996 has a detailed review of various theories and models of user acceptance). Each facet of the literature can provide some understanding of what makes users accept or reject a system but, since the issue is complex, it is unlikely that a single-variable explanation can be derived of the level of acceptance any information technology will receive among its intended users. For present purposes it makes sense to consider the evidence on the characteristics of the accepted (or rejected) technology separately from the evidence on the characteristics of the accepting (or rejecting) user, before reviewing the interplay of both these factors in current models of acceptance.

3. CHARACTERISTICS OF ACCEPTABLE TECHNOLOGY

According to Rogers' (1995) innovation diffusion theory, five characteristics of a technology determine its acceptance:

- Relative advantage (the extent to which it offers improvements over available tools).
- Compatibility (its consistency with social practices and norms among its users).
- Complexity (its ease of use or learning).
- Trialability (the opportunity to try an innovation before committing to use it).
- Observability (the extent to which the technology's gains are clear to see).

Numerous diffusion studies have demonstrated that innovations affording relative advantages, compatibility with existing practices and beliefs, low complexity, potential trialability, and observability, will be more extensively and rapidly accepted than an innovation with the opposite characteristics. In particular, three of these characteristics seem to have the greatest influence: relative advantage, compatibility and lack of complexity. While the diffusion model has broad appeal, there are concerns that the characteristics Rogers lists are too loosely defined to provide a sound basis for a complete theory.

The importance of complexity and trialability have long been raised in the human–computer interaction (HCI) literature where these concepts find resonance in the literature on usability and user evaluations (e.g. Nielsen 1993). Usability is frequently linked to certain qualities of the user interface that are under the control of the designer and HCI professionals place great emphasis on ensuring, through systematic usability evaluations, that users can operate a technology effectively, efficiently and satisfactorily. However, HCI research has concentrated less directly on the concept of acceptability or adoption of new technology, making the plausible assumption that usability is a prerequisite of acceptance.

Shackel (1991) is one of the few HCI researchers to make explicit the link between usability and acceptability. According to his formulation, an acceptable system is one that appropriately satisfies the requirements of its users for utility, usability and cost. These attributes can be easily linked with Rogers' five characteristics showing a close overlap between two distinct perspectives. However, while ability to use any technology is obviously necessary, it is not sufficient to ensure acceptability, and many technologies that are demonstrably usable are never accepted by the target users.

4. CHARACTERISTICS OF ACCEPTING USERS

Many researchers have attempted to identify psychological variables that distinguish users who accept or reject technologies. In a meta-analysis of research, Alavi and Joachimsthaler (1992) suggest that the most relevant user factors determining technology acceptance are cognitive style, personality, demographics and user-situational variables.

Cognitive style refers to the characteristic ways in which individuals process and use information and can be seen in information processing terms as a stable pattern of handling incoming stimuli and formulating responses. More than 100 different dimensions can be found in the literature, although a core cluster accounts for the majority of the work on this topic. To date, however, few cognitive style dimensions have been shown to predict user behavior with technology reliably.

Personality traits such as need for achievement, degree of defensiveness, locus of control and risk-taking propensity are

frequently proposed as important predictors of acceptance. The literature on this topic tends to blur the distinction between personality and cognitive style and the results of studies into such traits have equally failed to yield significant insights.

Among the demographic variables that have been studied, age and education have been shown to influence system use in some contexts. As expected, higher educational attainment and lower age both seem to influence use positively, but the relationship is weak. Coupling demographic variables with contextual knowledge improves matters substantially and variables such a training, experience, and user involvement, correlate well with acceptance of new technology. Alavi and Joachimsthaler (1992) found that the broad group of user-situational factors was more important than individual difference variables.

Innovation diffusion theory also suggests that factors at the level of the individual user are important. Rogers (1995) divides technology or innovation adopters into five categories depending on their speed of uptake: innovators, early adopters, early majority, late majority and laggards. Rogers plots these categories over a normal distribution where the division between early and late majority is viewed as the mean, and thus laggards and late adopters constitute 50% of the population. Rogers estimates that early adopters and innovators (~16% of the population according to his theory) are more likely to manifest risk-taking, adventure seeking personalities as well as being wealthier and more educated than the norm.

Thus, there is some agreement on the individual and situational factors influence the acceptance of new technologies but the weight of evidence suggests context might be more important than personality or individual psychological factors alone.

5. MODELING THE PROCESS OF ACCEPTANCE

While the identification of core technological and psychological variables underlying acceptance has provided some insight, few human factors researchers have attempted to link both sets of variables explicitly into a unified theory for design and implementation purposes. The most important theoretical work in this area however has involved socio-cognitive analyses of the dynamics of user action. Models of acceptance have emerged from this work which place emphasis on the attitude of users, and in so doing, seek to predict long-term user acceptance by measuring early affective responses to any new technology.

Of these models, the most widely cited is Davis *et al.* (1989) Technology Acceptance Model (TAM). TAM predicts user acceptance of any technology is determined by two factors: perceived usefulness and perceived ease of use. Perceived usefulness is defined as the degree to which a user believes that using the system will enhance his or her performance. Perceived ease of use is defined as the degree to which the user believes that using the system will be free from effort. According to TAM, both perceived usefulness and perceived ease of use have a significant impact on a user's attitude toward using the system.

Davis' research shows that TAM can explain ~50% of the variance in acceptance levels for many routine office applications, and the results from several studies of TAM indicate that usefulness is the most important predictor of use, explaining significantly more variance than ease of use ratings by users.

Impressively, TAM has been widely applied across different application types with consistent results. Furthermore, it is easy to administer since it involves little more than asking users to provide ratings of agreement/disagreement to a series of short statements such as "Learning to use this application would be easy for me." However, research on TAM is typically based on a single time period when users are exposed to a ready-made system. This makes it useful for choosing between competing technologies at the implementation stage but less applicable in the early stages of design where designers are trying to determine how best to design a technology so that it will prove acceptable.

6. DESIGNING AND IMPLEMENTING ACCEPTABLE TECHNOLOGIES

While significant inroads have been made into the determinants of user acceptance, there remains the tricky issue of applying these insights in the design process to ensure the resulting technologies are likely to be accepted. This poses the joint challenge of determining acceptance before any technology is fully developed and then implementing it in a manner that supports uptake.

One cannot simply rely on models such as TAM since these require the user to experience the technology in order to formulate their perceptions. To date, little work has been carried out on the reliability and validity of TAM scores for early prototypes. Furthermore, it is a fundamental tenet of most organizational theories that the collective response to a new system in any working or social environment is likely to be determined by more than the isolated, individual ratings of the members.

Within the user-centered systems design tradition of HCI emphasis is placed on the early and continuous involvement of users. Primarily this is to serve as testers of prototypes and to provide insights into task and work practices that need to be supported by any technology. Even if a technology is engineered to be highly usable, and be shown to be so through formal testing, there exists no guarantee that this will lead to acceptance.

Theorists from the socio-technical systems tradition such as Eason (1988) argue that information technologies are embedded in working practices and that these practices manifest a network of social relationships such as cooperation among users, management relationships and so forth. Accordingly, any technology cannot be fully analyzed or understood in terms of usability where this is conceptualized in isolation of the goal-oriented organization or work context it is intended to support. To optimize jointly both the social and technical attributes of any organization, allowance must be taken at the design level of the social dynamics of any organization or group within it

Socio-technical systems theory has given birth to a framework for technology design that emphasizes the analysis of all stakeholders, not just the direct users of a technology, the formation of planning groups to oversee the design, the performance of prototyping exercises, and the analysis of likely impact the technology will have on the organization. The intention of such a design process is to avoid unpleasant side effects in working practices (which would lead to resistance) and to ensure as much a social solution as a technical solution to the computing needs of an organization.

Eason (1988) views acceptance in terms of two competing forces: control and enhancement. Control factors are those that

impose rules or structures upon the users, thereby removing autonomy (control over their own actions) from them. According to socio-technical thinkers, working group autonomy is to be encouraged since it is considered to increase satisfaction and long-term performance. Among the control issues raised with respect to technology design are: access, reliability, confidentiality, monitoring, pacing, stress, social contact. Low or high presence of certain factors (e.g. low reliability, high pacing) with the introduction of a new technology is likely to reduce the users' perception of control and thus increase the risk of resistance.

Enhancement factors include sense of mastery, growth of knowledge, discretion, ability to act informally, requirement for certain skills, and enabling worker cooperation. These factors should all be maximized as appropriate for the context (though skill requirements are not to be inherent for certain situations). A technology that is designed to support such factors is likely to increase user acceptance in an organization.

To date there has been little controlled study of the importance of such control and enhancement variables. Socio-technical researchers tend toward case studies of designs and their implementation rather than controlled experiments rendering specific and individual weighting of control and enhancement factors problematic. However, this perspective offers insights that might prove amenable to further research that moves us beyond the search for single technological or user variables.

7. SUMMARY

Determining user acceptance of a system is a difficult but important part of human factors research and application. While there is currently no complete theory or model that explains and predicts acceptance, there is an emerging understanding of the key variables in the technology, the user and the implementation process that affect acceptability. To be accepted, a technology must satisfy basic usability requirements and be perceived as useful by its intended user community. User experience and training will impact acceptance levels as will the manner in which the technology is implemented to contribute to organizational goals and working practices.

REFERENCES

ALAVI, M. and JOACHIMSTHALER, E.A., 1992, Revisiting DSS implementation research: a meta-analysis of the literature and suggestions for researchers. *MIS Quarterly*, 16, 95–116.

DAVIS, F., BAGOZZI, R. and WARSHAW, P., 1989, User acceptance of computer technology: a comparison of two theoretical models. *Management Science*, 35, 982–1003.

DILLON, A. and MORRIS, M., 1996, User acceptance of information technology: theories and models. In M. Williams (ed.), *Annual Review of Information Science and Technology*, vol. 31 (Medford: Information Today).

EASON, K., 1988, *Information Technology and Organizational Change* (Washington, DC: Taylor & Francis).

NIELSEN, J., 1993, *Usability Engineering* (Boston: AP Professional).

ROGERS, E., 1995, *Diffusion of Innovations* (New York: Free Press).

SHACKEL, B., 1991, Usability context, framework, definition, design, and evaluation. In B. Shackel and S. Richardson (eds), *Human Factors for Informatics Usability* (Cambridge: Cambridge University Press).

Human–Computer Interaction (HCI) Standards

T. Stewart and T.F.M. Stewart

System Concepts Ltd, 2 Savoy Court, The Strand, London WC2R 0EZ, UK

1. INTRODUCTION

Although standards can be unwelcome constraints, they can also be useful tools, saving designers from repeating others' mistakes or "re-inventing the wheel" and helping to ensure that human–computer interaction (HCI) issues are taken seriously in the design process. The paper therefore starts with a brief discussion of the importance of standards followed by an overview of selected standards activities. It ends with a review of where different standards fit into the system development process.

2. WHY STANDARDS ARE IMPORTANT IN IMPROVING USABILITY

Standards promote consistency, good practice, common understanding and an appropriate prioritization of user interface issues.

2.1. Consistency

Most users have horror stories about inconsistency between, and even within, systems. Standards provide a consistent reference across design teams or across time to help avoid confusion. In other fields, consistency, for example, between components that should interconnect, is the prime motivation for standards. It is certainly a worthwhile target for user interface standards.

2.2. Good Practice

In most fields, standards provide definitive statements of good practice. In user interface design there are many conflicting viewpoints about good practice and standards, especially international ones, can provide independent and authoritative guidance. International standards are developed slowly, by consensus, using extensive consultation and development processes. This has its disadvantages in such a fast-moving field as user interface design and some have criticized any attempts at standardization as premature. However, there are areas where a great deal is known that can be made accessible to designers through appropriate standards and there are approaches to user interface standardization, based on human characteristics, that are relatively independent of specific technologies.

2.3. Common Understanding

Standards do not guarantee good design but they do provide a means for different parties to share a common understanding when specifying interface quality in design, procurement and use.

- For users: standards allow them to set appropriate procurement requirements and to evaluate competing supplier's offerings.
- For suppliers: standards allow them to check their products during design and manufacture and to provide a basis for making claims about the quality of their products.
- For regulators: standards allow them to assess quality and provide a basis for testing products.

2.4. Appropriate Prioritization of User Interface Issues

One of the most significant benefits of standardization is that it places user interface issues squarely on the agenda. Standards are serious business and whereas many organizations pay little regard to research findings, few organizations can afford to ignore standards. Indeed, in Europe, and increasingly in other parts of the world, compliance with relevant standards is a mandatory requirement in major contracts.

3. CURRENT USER INTERFACE STANDARDIZATION INITIATIVES

The International Organisation for Standardisation (ISO) is the worldwide standardization organization responsible for developing international standards in the field of mechanical standardization.

ISO is organized into technical committees (TC) and the ergonomics committee responsible for HCI standards is ISO TC159 SC4. Table 1 shows the main stages of ISO standards development.

International standards are important, partly because the major manufacturers are international and partly because the regional European Standardisation Organisation (CEN) has a strategy of adopting ISO standards wherever appropriate as part of the creation of the single market. The activities of the relevant CEN working group (CEN TC122 WG 5) are also described.

3.1. ISO TC159 SC4: Ergonomics of Human– System Interaction

This committee is responsible for ergonomics standards in the field of human–system interaction. The primary technical work takes place in working groups (WG) and each of these has responsibility for different work items. WG 2–5 are responsible for ISO 9241 (Table 2). WG 5 is also working on ISO 14915: *Multimedia User Interface Design — Ergonomics Requirements for human–computer multimedia interfaces*. This four-part standard will establish software ergonomics requirements and recommendations for interactive multimedia user interfaces that integrate and will synchronize different media (static-like text, graphics, images and dynamic media like audio, animation and

Table 1. Main stages in the development of an international standard

WI	Work Item — an approved and recognized topic for a working group to be addressing that should lead to one or more published standards
WD	Working Draft — a partial or complete first draft of the text of the proposed standard
CD	Committee Draft — a document circulated for comment and approval within the committee working on it and the national mirror committees. Voting and approval is required for the document to reach the next stage.
DIS	Draft International Standard — a draft standard circulated widely for public comment via national standards bodies. Voting and approval is required for the draft to reach the final stage.
FDIS	Final Draft International Standard — the final draft is circulated for formal voting for adoption as an international standard.
IS	International Standard — the final published standard

Note: documents may be reissued as further CDs and DISs

Table 2. Parts and status of ISO 9241

ISO 9241: Ergonomics Requirements for Office Work with Visual Display Terminals (VDTs)	Responsible working group	Status end 1998
Part 1. General Introduction	WG6	IS
Part 2. Guidance on task requirements	WG4 (finished)	IS
Part 3. Visual display requirements	WG2	IS
Part 4. Keyboard requirements	WG3	IS
Part 5. Workstation layout and postural requirements	WG3	IS
Part 6. Enviromental requirements	WG3	FDIS
Part 7. Display requirements with reflections	WG2	IS
Part 8. Requirements for displayed colors	WG2	IS
Part 9. Requirements for non-keyboard input devices	WG3	FDIS
Part 10. Dialogue principles	WG5	IS
Part 11. Guidance on usability	WG5	IS
Part 12. Presentation of information	WG5	IS awaiting publication
Part 13. User guidance	WG5	IS
Part 14. Menu dialogues	WG5	IS
Part 15. Command dialogues	WG5	IS
Part 16. Direct manipulation dialogues	WG5	IS
Part 17. Form-filling dialogues	WG5	IS

Table 3. Parts and status of ISO 11064

ISO 11064: Ergonomic Design of Control Centers	Responsible working group	Status end 1998
Part 1. Principles for the design of control centers	WG8	DIS
Part 2. Principles of control suite arrangement	WG8	DIS
Part 3. Control room layout	WG8	FDIS
Part 4. Workstation layout and dimensions	WG8	WI
Part 5. Displays and controls	WG8	WI
Part 6. Enviromental requirements for control rooms	WG8	WI
Part 7. Principles for the evaluation of control centers	WG8	WI
Part 8. Ergonomics requirements for specific applications	WG8	WI

video). WG 6 is mainly concerned with ISO 13407: *Human-centred Design of Interactive Systems*. WG 8 is concerned with ISO 11064: *Ergonomics Design of Control Centres* (Table 3).

3.2. European Committee for Standardisation (CEN)

CEN is the European equivalent of ISO. Its members are the national standards bodies of the countries in the European Union (EU) and in the European Free Trade Association (EFTA).

The European Commission (EC) declared that 1992 represented the end of trade barriers in Europe and that from then there would be a single European market for goods and services. Different national standards are at least potential barriers to trade and so part of this initiative involved the replacement of national standards with European ones.

There was therefore a flurry of activity to create CEN standards where none existed before. The chosen strategy was to adopt international standards, where appropriate, and short circuit the standards-making process. The CEN Ergonomics Committee TC122, through its WG 5 (Ergonomics of VDTs), was given the brief to facilitate and support the ISO TC159 SC4 user interface work with a view to adopting the International Standard ISO 9241. Since then, the other standards under development by ISO TC159 SC4 have been included in the work program of CEN TC122 WG 5.

4. USING STANDARDS IN SYSTEM DEVELOPMENT

The contents of the main user interface standards, including those under development in the context of how these standards might be used during a number of typical HCI activities are described briefly (Table 4).

The activities follow an approximate time sequence with some additional activities added towards the end of the list. They are not intended to represent the "right way" to approach user interface design nor are all the activities necessary in every project.

4.1. Feasibility Studies and Market Research

Standards are not primary sources for designers looking for new ideas and approaches or inspiration for innovative designs but

Table 4. Typical design activities and issues

Feasibility studies and market research
Requirements analysis including business analysis/task analysis
Interface specification and task specification
Initial design, simulation, prototyping and modeling
Design and build   Dialogue design, interface navigation, display design
Keyboard/input device
Workplace and console
User support, documentation, manuals and training
Safety critical systems
User-centered design methods
Health and safety issues
Job design, group working organization
Users with special needs, children, the disabled and the aged

they often represent important constraints on the potential design. During the feasibility stage it is important to identify which standards if any are relevant. For published standards, the primary source is the national standards organization, e.g. the American National Standard Institute (ANSI) in the USA, the British Standard Institute (BSI) in the UK. Standards under development are more problematical and although much information is potentially public, knowing where to look can be difficult. In the first instance, use the standards numbers and the committee references and contact your national standards organization. They should put you in touch with those involved from your country. Most of the ISO member bodies operate mirror national committees that not only comment on drafts as they are developed but also can provide up-to-date information about the work-in-progress.

4.2. Requirements Analysis Including Business Analysis and Task Analysis

At present, there is little help from standards for this stage although a Technical Report on Human-centred Methods is under development. Engineering usability into products requires a commitment to usability requirements as well as functionality requirements. However, setting usability objectives on these requirements only make sense if there is an agreed way of specifying usability. ISO 9241-11: 1998: *Guidance on Usability* was developed to provide a framework for such a statement of usability. It provides guidance on usability specification that includes descriptions of the context of use, the evaluation procedures to be carried out and the criterion measures to be satisfied when the usability of the system is to be evaluated. There are various situations in which usability may be evaluated, for example in product development, in procurement or in product certification. The common framework presented in this section should be useful in all of these situations.

4.3. Interface Specification and Task Specification, Initial Design, Simulation, Prototyping and Modeling, Design and Build

Specification, design and build involve different HCI activities, but the same standards are relevant. Standards offer specific guidance on well-established design requirements, e.g. the pressure required to operate a key, or the contrast required for an image to be distinct from its background.

4.3.1. Dialogue Design and Interface Navigation

ISO 9241 10: 1996: *Dialogue Principles* presents high-level ergonomic principles that apply to the design of dialogues between humans and information systems. These include suitability for the task, controllability and error tolerance, among others. The principles are supported by scenarios that indicate the relative priorities and importance of the different principles in practical applications.

ISO 9241-14: 1997: *Menu Dialogues* provides recommendations on menu structure, navigation, option selection and execution, and menu presentation (by various techniques including windowing, panels, buttons, fields, etc.).

ISO 9241-15: 1998: *Command Dialogues* provides recommendations on command language structure and syntax, command representations, input and output considerations, feedback and help.

ISO FDIS 9241-16: 1998: *Direct Manipulation Dialogues* provides recommendations on the manipulation of objects, and the design of metaphors, objects and attributes. It covers those aspects of "graphical user interfaces" that are directly manipulated and not covered by other parts of ISO 9241.

ISO 9241-17: 1998: *Form-filling Dialogues* provides recommendations on form structure and output considerations, input considerations and form navigation.

4.3.2. Display Design

"Display design" refers both to the design of display hardware and to the presentation of information on the display.

Display hardware specification and design for office VDTs is covered in ISO 9241-3: 1992: *Display Requirements*. This part deals with the design of screen hardware for visual display terminals. In addition to design specifications, this part also contains a proposed user performance test as an alternative route to conformance. (Note that the ergonomics requirements for flat panels are dealt with in ISO FDIS 13406-2: 1998.)

ISO 9241-7: 1998: *Display Requirements with Reflection* deals with the ergonomic requirements for, and details of, methods of measurement of reflections from the surface of display screens, including those with surface treatments. ISO 9241-8: 1997: *Requirements for Displayed Colours* deals with the ergonomic requirements for multicolor displays that supplement the monochrome requirements in Part 3. Displays for control rooms are dealt with separately in ISO 11064.

Software aspects of display design are covered in ISO FDIS 9241-12: 1998: *Presentation of Information*. This part deals with the specific ergonomics issues involved in representing and presenting information in visual form. It includes guidance on ways of representing complex information, screen layout and design, as well as the use of windows.

There is already a substantial body of material available in guidelines and recommendations and this part represents a distillation of the most useful and relevant ones.

4.3.3. Keyboard and Input Device Design

Keyboard specification and design (in terms of the operation of the keys and its ergonomic qualities) is covered in ISO 9241-4: 1998: *Keyboard Requirements*. This deals with alphanumeric keyboard design. In addition to design specifications, this part also contains a proposed user performance test as an alternative route to conformance. It deals with the ergonomic aspects of the keyboard, not the layout that is specified in ISO 9995: *Keyboard Layouts for Text Office Systems*.

Non-keyboard input devices are becoming increasingly popular and ISO FDIS 9241-9: 1998: *Requirements for Non-keyboard Input Devices* deals with the ergonomic requirements for pointing devices including the mouse, tracker ball, etc. that can be used in conjunction with a visual display terminal.

4.3.4. Workplace and Console Design

Office workplaces incorporating VDTs are covered in some detail in ISO 9241-5: 1998: *Workstation Layout and Postural Requirements*. This part deals with the ergonomic requirements for a visual display terminal workstation that will allow the user

to adopt a comfortable and efficient posture. Workplaces in control rooms are dealt with separately in ISO DIS 11064-3: 1998: *Control Room Layout*. Environmental considerations (visual, acoustic, thermal) are covered in ISO 9241-6: 1998: *Environment Requirements*.

4.4. User Support, Documentation, Manuals and Training

In assessing the usability of a product in practice, real users take account of the documentation, manuals and training received as well as the specific characteristics of the product. ISO 9241-13: 1998: *User Guidance* covers some of these aspects and provides recommendations for the design and evaluation of user guidance attributes of software user interfaces including prompts, feedback, status, on-line help and error management.

4.5. Safety Critical Systems

Although many user interface standards are generic and apply to a wide range of different task situations, safety critical systems pose special risks and may demand special standards. For example, the size, spacing, force and travel of keys on a keyboard to be used in an office environment may not be achievable when the same keyboard has to be used in a control room where protection from sparks or use by operators wearing gloves may override normal requirements.

ISO 11064 Parts 4–8 will address a number of HCI issues in safety critical systems.

4.6. User–centered Design Methods

Although there may be disagreement about what user-centered design means in detail, most HCI specialists would argue that it is fundamental to the practice of human–computer systems. Indeed, much of ISO 9241 has an implicit user-centered design philosophy behind it.

ISO FDIS 13407: 1998: *Human-centered Design Process for Interactive Systems* describes the ergonomic process to be applied within the design process to ensure that proper attention is paid to the human and user issues. Coverage includes usability design and evaluation, user-centered design methods, the use of ergonomics standards in design activities and evaluation methods. The standard is aimed at those responsible for managing design processes and it presents a high level overview of what activities are recommended for human centered design.

4.7. Health and Safety Issues

Health and safety concerns were the starting point for a number of the standardization and regulatory measures. The Council of the European Economic Community (as it then was called) published a Directive on 29 May 1990 on the minimum safety and health requirements for work with display screen equipment (90/270/EEC). This Directive comes under Article 118a and is concerned to ensure that there is a minimum level of health and safety in member states and is aimed at employers. The directive does not therefore refer to the product standards (e.g. ISO 9241) even though they deal with much of the same technical content. With the exception of the UK, they are unlikely to be used formally in national implementations of the Display Screen Directive.

In practice, the European Standards EN 29241 developed by TC122 (based on ISO 9241) can be used by suppliers to demonstrate compliance with the ergonomics state-of-the-art.

4.8. Job Design, Group Working and Organizational Issues

In developing ISO 9241, there was a clear recognition that many of the problems often attributed to poor equipment or workplace design may in fact stem from poor job design. Thus ISO 9241-2: 1992: *Guidance on Task Requirements* provides guidance on the design of display screen tasks based on nearly half a century of research and organizational practice in socio-technical systems.

4.9. Users with Special Needs, Children, the Disabled and the Aged

Many users of information technology feel that little account is taken of the requirements of so-called "normal" users. Users with special needs — whether transitory or permanent — are even less well catered for. One of the major benefits of information technology is its potential to extend human capacity and to complement human skill. A new work item has recently been started on accessibility for users with special needs.

5. CONTACTS FOR FURTHER INFORMATION

For the latest information about the status of standards, contact your national standards body or the regional organization in Europe (CEN) or the ISO:

- Comite Europeen de Normalisation (CEN), rue Brederode 2, Boite 5, B-1000 Brussels, Belgium. E-mail: infodesk@cenclcbel.be
- International Organisation for Standardisation (ISO), Central Secretariat, 1, rue de Varembe, Case postale 56, CH-1211 Geneva 20, Switzerland. Url: http://www.iso.ch/

Human Factors and Digital Networks: Intranets, Extranets and the Internet

C. Forsythe and E. Grose

Statistics and Human Factors, Sandia National Laboratories, Albuquerque, NM 87185–0829, USA

1. INTRODUCTION

Forget the fanciful props of popular science fiction. Instead, extrapolate from current technology trends to envision the world of 2005, 2010 and beyond. It seems reasonable to suppose that most electronics will be linked via digital networks. Your wearable computer will provide access to a current inventory of items on your kitchen shelves. Likewise, it will let you start your dishwasher or adjust your home climate control after you have left the house. There will be a fusion of technologies. Your television, stereo and home computer will be merged and you will make on-demand selections of programming from libraries containing an endless breadth of recorded media. This will require a simple download that will occur within seconds, if not fractions of a second. Here is only a sampling. The intent has been to portray a world where nearly every facet of our lives has been interconnected through digital networks.

Now, put aside the promise imaginable with digital networks. Continuing to extrapolate from current trends, there is a downside. Imagine that the 1960s sitcom you downloaded the previous evening was infected with a tenacious virus. This virus quickly spread to corrupt the operating system software linking each of your home electronics and appliances, as well as software running locally on several devices. Over the course of a few hours virtually every electronic device in your home becomes completely unusable. Your initial inconvenience might only last the day or so needed to eradicate the virus and reinstall software that had been infected. However, to your chagrin, you may find that there are unanticipated problems. Over the past few years there have been software upgrades. In reinstalling the operating system software and various other software applications and peripheral devices, you encounter incompatibilities between software applications. Consequently, some devices are not recognized by the operating system and others do not function properly. You expend countless hours attempting to fix the problems. Technicians install and reinstall software. And in the end, you may attain the same degree of functionality you had prior to contracting the virus.

These scenarios are mentioned as a means of drawing attention to the more dramatic problems faced with digital networks. For the past three decades, Human Factors has focused tremendous effort on identifying aspects of user interface design that affect the moment-to-moment efficiency and satisfaction of users. These efforts have been justified through time savings, reduction of human error and general customer approval. However, for the typical user, these gains might seem incremental.

It is our belief that there are human factors issues associated with digital networks, and associated advances in technology, that dwarf the issues that have been the primary focus of Human Factors' past efforts. As users become increasingly reliant on digital networks, they also become increasingly vulnerable to malevolent acts and technical failures that have the potential to leave them sidelined for hours, if not days. The remainder of this chapter will address these issues, and others, discussing the unique problems the Human Factors profession must address as digital network technology continues its forward march.

2. VULNERABILITY TO MALEVOLENT ACTS

One must be concerned with two general types of threats. First, there are direct threats whereby hackers gain access to computer systems for purposes of theft, espionage or sabotage. The second type of threat is indirect and involves the dissemination of viruses that spread uncontrolled, infecting mostly unknown victims. In either case, the outcome can be costly, whether the victim is an individual or an organization. Interestingly, the mainstream media often presents a somewhat inaccurate representation of the security risks. Almost daily reports of obscure vulnerabilities in popular commercial software create an impression that the main impetus of most hackers is to discover and exploit such shortcomings. In reality, hackers primarily take advantage of human nature. For example, lists of the most commonly used passwords are readily available. Likewise, users rarely change their passwords, they leave slips of paper with their passwords in the open, and if asked over the telephone, they will often divulge their passwords to a stranger under marginally credible pretenses.

Two Human Factors issues arise. First, we are currently witnessing a proliferation of passwords, Personnel Identification Numbers (PIN) and other mechanisms for authenticating one's identity. With the technology commonly in use today, users cannot be expected to adhere to the most minimal security provisions (i.e. create a unique password, memorize passwords so they are not written anywhere and change passwords on a regular basis). This dilemma will only worsen as the Internet enables increasing access to remote resources. Being knowledgeable of memory processes, Human Factors experts should be uniquely capable of devising concepts that enable reliable verification, without the drawbacks currently faced with simple password-based concepts, or the intrusive and sometimes questionable reliability of biometric devices.

Second, for most organizations, protection of their computer network resources depends on a relatively small number of personnel assigned to computer security and/or network administration. Human error by these personnel, be they errors of omission or commission, may leave network resources vulnerable to intruders. Previous analysis has identified that human system components are one of the weakest links in networked systems (Forsythe 1997). Furthermore, techniques are available for identifying failure modes associated with human error and assessing the adequacy of existing and proposed measures to preclude these failures (Forsythe and Grose 1997). This is a role for which Human Factors is uniquely qualified due to our familiarity with Human Reliability Analysis and other methodologies for assessing and preventing human error.

Avoidance of viruses presents an especially difficult challenge. The current model is one wherein virus detection software is used to scan files to determine if they have been infected by any of the thousands of known viruses. Because a virus must be known and a means developed to defeat the virus, at best, this solution

provides only partial assurance. An ingenious hacker can easily introduce a new virus to millions before it is detected and virus-scanning software updated. And this does not consider the delinquency of most end users with respect to updating their virus scanning software, another human vulnerability to exploit.

A second solution requires users to rely on certification sites that provide some level of guarantee that the files they certify are virus-free. Certification can be easily forged. Furthermore, the presence of such organizations offers an invitation for malevolents to engage in games of cat and mouse wherein the objective is to induce mistaken certification of infected files. With virus-resistant software unlikely, the remaining option is to modify user behavior. Warnings and cautions are of limited effectiveness. A more promising approach would be to develop interfaces that unobtrusively alert users to the risk associated with specific on-line behaviors. There has been extensive research concerning the perception of risk and devising means to heighten the awareness of risk. Facilitating objective assessments of risk is a job well suited to Human Factors.

3. COMPLEXITY OF COMPUTER SOFTWARE AND HARDWARE

With each new version, there is a tendency for popular commercial software products to add "Internet-ready," and other features. These additional features introduce increased complexity to the software. This trend is evident with all types of software: operating system, server, browser and office products such as word processing and spreadsheets. The problem is that each type of software must interact and greater complexity results in increased potential for incompatibilities. These incompatibilities may have minor consequences (e.g. unavailability of software features and system errors) or more significant consequences (e.g. system crashes with subsequent data loss or complete system inoperability).

It may appear that these problems only involve software reliability and may be corrected through more rigorous testing and evaluation. However, there is a proactive course of action. It should not be forgotten that human actions are the immediate cause of software incompatibilities. Users upgrade software following irregular schedules. They install and uninstall software and hardware. And they move and delete systems files. Consequently, software must operate in an environment that often exhibits far greater heterogeneity than intended by software developers. In organizations, standardization has often been proposed as a solution. This has been based on a belief that problems with software incompatibility may be alleviated by restricting the ability of users to diverge from standard hardware/software configurations. However, it has been commonly observed that users resent and resist such attempts to impose conformity and will inevitably find means to deviate from the standard.

Heterogeneous computing environments are partially a product of the variability in computer users. For many users, it represents a means of personal expression. In contrast with efforts aimed at stifling heterogeneity through standardization, a characterization of variability in the behavior of computer users would allow the design of products that accommodate the resulting variability in computing environments. The characterization and accommodation of user populations is a familiar problem for Human Factors. In fact, it is the basis for anthropo-

metrics. The same strategies used to characterize the physical and cognitive dimensions of user populations may also be applied to describe user-generated variability in computing environments. With such knowledge, levels of accommodation based on divergence from a population mean could be incorporated into software requirements (e.g. accommodate 98th percentile user). These requirements could address concepts such as the percentage of overall storage space that is utilized, the frequency of software upgrades and the tendency to simultaneously run multiple applications. Furthermore, concepts may be applied that involve definition and qualification of software products with regard to a parameter space that incorporates different dimensions of user-generated heterogeneity in computing environments (Forsythe *et al.* 1995).

4. OPPORTUNITIES FOR HUMAN FACTORS

The most important aspect to understand about digital networks, such as the Internet, is the wealth of opportunities that they enable. As a profession, Human Factors has been routinely relegated to a subservient role within the product development process. Even so, the profession has certainly flourished over the past decades. However, the profession has also been frustrated and left to regret what might have been possible if usability had received fair consideration.

Within product development environments, power tends to be associated with the ability to control funding and to control design. Traditional businesses and organizations offer very limited opportunities for Human Factors to wield influence over either means of exerting power. Consequently, we argue that the potential for Human Factors has barely been realized and as a direct product, we have yet to witness the true potential afforded by computing technology.

In human history, there probably has never been a more significant challenge to traditional business and economic models than that posed by the emerging "digital economy" (Kelly 1997). The ability is unprecedented for ordinary individuals, operating from a $1000 home computer, to communicate and interact with large numbers of people. Consider publishing. Traditionally, there have been limited channels by which writers could publish their work with hopes of high volume distribution. Those controlling these limited channels possessed enormous power to censor available information, promote certain perspectives and determine the success or failure of talented individuals. With the Internet, ordinary home computer users have been given a low-cost means of readily broadcasting to millions of potential consumers. Publishing is only one of countless examples. It is this freedom to operate outside of established business models that creates an extraordinary opportunity for innovation and provides avenues to success that have never before existed.

The emergence of digital networks has created an atmosphere ripe with opportunities for Human Factors to realize its true potential to positively shape everyday life. Likewise, our profession is held back by nothing more than the willingness of individuals to pursue their ideas. Human behavior permeates all levels of interaction with digital networks and by exploiting our unique knowledge of human behavior, Human Factors has an opportunity to step forward with solutions that will make a genuine difference.

Consider a simplistic example. Keyboard errors are well

understood and data is available for the most common errors. Spelling errors are similarly well understood. When entering an Internet address, it is reasonable to expect that one in every hundred entries, or so, will contain either a keyboard error or misspelling. Knowing these highly predictable occurrences of human error and aware that certain Internet sites receive thousands, if not millions, of visitors every day, an opportunity was realized. Entrepreneurs purchased the domain names similar to those of popular Internet sites, such as Microsoft Corporation, but with minor errors in the Internet address (e.g. microsof.com) and then, for a modest fee, offered to redirect misdirected Internet traffic to the desired destination. In this case, a business has emerged on the basis of knowledge of the simplest characteristics of human behavior.

After an hour using their computer to access the Internet, reasonably astute Human Factors professionals applying their expertise should be able to identify similar, although likely more sophisticated, means of improving the user's experience with this technology. This is the opportunity that the rampant spread of digital networks has extended to the Human Factors profession.

5. CONCLUSION

The Internet of the future will be quite different from the one to which we have now become accustomed. Ascending technologies such as wave division multiplexing, fiber optics, satellite, asymmetric digital subscriber lines, and wireless ethernet are laying the foundations for tremendous increases in bandwidth. This promises an explosion of electronic activity on the global network. The popular media have warned that the impending "information explosion" will somehow leave users paralyzed. In reality, the Internet, in its current form, demonstrates that human beings are superbly equipped to filter unwanted data. Far from being paralyzed by "information overload," users have shown a capacity to function effectively despite a substantial increase in day-to-day information exchanges.

However, with the information explosion, or actually the media explosion, there is an increasing awareness that time is a precious commodity, and perhaps more valuable than money. Young, old, rich or poor, there is a growing resentment toward anything that steals or squanders our time. In this regard, it becomes unacceptable for computers to add tasks to our already busy schedules. This not only drives the need for more usable products, but also creates the need to be able to measure usage

patterns in order to focus developers on why products succeed or fail. With its connectedness, the Internet offers an unparalleled opportunity to collect such usage data. In this regard, the Internet offers a unique capacity for self-correction.

To take full advantage of this capacity for self-correction, it is necessary for someone to ask the appropriate questions, and no discipline is better positioned than Human Factors. The vast majority of web products fail to attract the level of usage initially expected by their developers (Grose *et al.* 1998). Usage statistics reveal that a relatively small number of factors are needed to predict the usage of web products. Products that are relevant to a large population, have relatively little competition, and satisfy the needs of users will see extensive usage. Similarly, more complex questions may also be answered. To realize the potential of Human Factors to impact the Internet, it is necessary to ask and answer questions regarding the interactions of humans with digital networks, and in the answers, there resides the opportunity for the Human Factors profession to realize its potential.

ACKNOWLEDGEMENTS

REFERENCES

FORSYTHE, C., 1997, Human factors in agile manufacturing: a brief overview with emphasis on communications and information infrastructure. *Human Factors and Ergonomics in Manufacturing*, 7, 3–10.

FORSYTHE, C., ASHBY, M.R., DIEGERT, K.V., PARRATT, S.W., BENAVIDES, G.L., JONES, R.E. and LONGCOPE, D.B., 1995, *A Process for the Agile Product Realization of Electro-Mechanical Devices*. SAND95-0536 (Albuquerque: Sandia National Laboratories).

FORSYTHE, C. and GROSE, E., 1997, Human factors in high consequence manufacturing systems. *Society of Manufacturing Engineers*, MM98-105, 1–8.

FORSYTHE, C., GROSE, E. and RATNER, J., 1998, *Human Factors and Web Development* (Mahwah: Lawrence Erlbaum Associates).

GROSE, E., VARNER, L. and FORSYTHE, C., 1998, Predicting penetration of Intranet content and applications. In *Proceedings of the Human Factors and Ergonomics Society*, 5–9 October, Chicago, IL (Santa Monica: Human Factors and Ergonomics Society).

KEVIN, K., 1997, New rules for the new economy: twelve dependable principles for thriving in a turbulent world. *Wired*, 5.09, 140–97.

Human Speech Digitization and Compression

P. R. Michaelis

Lucent Technologies, Bell Laboratories, 11900 N. Pecos Street, Denver, CO 80234, USA

1. INTRODUCTION

Nearly all modern speech-based computer and communication systems transmit, route or store speech digitally. One obvious advantage of digital techniques over analogue is the ability to provide superior audio quality (e.g. compact discs versus phonograph records, or digital cellular telephones versus analogue). Other advantages include the ability to send many more simultaneous transmissions over a single communications channel, route speech communication through computer-based switching systems and store speech on computer disks and in solid-state memory devices. This paper describes techniques that reduce the amount of data required to digitize speech.

2. SPEECH DIGITIZATION

The simplest way to encode speech digitally is to generate a sequence of numbers that, in essence, trace the 'ups and downs' of the original speech waveform. For example, if one wished to digitize a waveform in which all of the important acoustic information is < 4000 Hz (4000 cycles s^{-1}), the basic steps of this analogue-to-digital conversion would include:

1. Filter from the original signal all information > 4000 Hz.
2. Divide the original signal into 8000 segments s^{-1}.
3. Go through the segments in order, measuring and recording the average amplitude of the waveform within each segment.

The purpose of the first step is to prevent 'aliasing' — the creation of false artifacts caused by the undesired interaction of the sampling rate with the frequency of the observed events. The phenomenon in motion pictures, where the spokes of a rapidly rotating wheel may appear to be standing still or even moving backwards, is an example of aliasing.

The second step, sampling at twice the frequency of the highest-frequency sine wave, is necessary to capture both the peaks and the valleys of the wave.

To envision the third step more easily, imagine that the original waveform is drawn on a sheet of paper. Within every segment, each of which represents 1/8000 s, the height of the waveform is measured with a ruler. The sequence of numbers obtained in this manner constitutes a digital representation of the original waveform.

Regarding the "ruler" used to measure within-segment speech amplitudes, speech quality comparable with that of a modern telephone requires 12 bits per segment, 8000 segments s^{-1}. (As a point of comparison, audio compact discs use 16 bits per segment, with 44 100 segments s^{-1}.) The resulting data rate of 96 000 bits s^{-1} means that a typical 1.44 MB floppy diskette can hold only ~2 min of telephone-quality speech.

Modest reductions in the data rate can be achieved by using logarithmic amplitude encoding schemes. These techniques, which represent small amplitudes with greater accuracy than large amplitudes, achieve voice quality equivalent to a standard 12-bit system with as few as 8 bits per segment. Examples include the m-law (pronounced "myoo law") coding found on many US digital telephones, and the A-law coding commonly used in Europe.

For many applications in which the cost of transmission or the cost of storage is important, such as wireless telephony or voice mail systems, the data rate reductions achieved with simple m- and A-law encodings are inadequate. One way to achieve significant reductions in the data rate is to extract and digitize the frequency content of the waveform (rather than simply to digitize the shape of the waveform).

Many coders that work in this manner have software components that map to physical components of the human vocal mechanism. They reduce the data rate by encoding only the parameters that control the changeable components of the speech production model; for example, the parameter that controls overall amplitude and the parameter that adjusts the fundamental pitch of the electronic "vocal cords".

3. THE HUMAN SPEECH PRODUCTION MECHANISM

Given that many components in these coders have physiological counterparts, it is helpful to understand the human vocal mechanism prior to examining the coders.

The major physical components of the human speech mechanism include the lungs, the vocal cords and the vocal cavity. When a person speaks, the lungs force air past the vocal cords and through the vocal cavity. The pressure with which the air is exhaled determines the final amplitude or "loudness" of the speech. The action of the vocal cords on the breath stream determines whether the speech sound will be voiced or unvoiced.

Voiced speech sounds (e.g. the "v" sound in "voice") are produced by tensing the vocal cords while exhaling. The tensed vocal cords briefly interrupt the flow of air, releasing it in short, periodic bursts. The greater the frequency with which the bursts are released, the higher the pitch.

Unvoiced sounds (e.g. the final "s" sound in "voice") are produced when air is forced past relaxed vocal cords. The relaxed cords do not interrupt the airflow; the sound is instead generated by audible turbulence in the vocal tract. A simple demonstration of the role of the vocal cords in producing voiced and unvoiced sounds can be had by placing one's fingers lightly on the larynx, or voice box, while slowly saying "voice"; the vocal cords will be felt to vibrate for the "v" sound and for the double vowel (or diphthong) "oi" but not for the final "s" sound.

The mechanisms described above produce what is called the excitation signal for speech. Many properties of the excitation signal will differ when comparing one person to another. However, when examining an individual only three parameters in the excitation signal will vary as they speak: amplitude of the sound, proportion of the sound that is voiced or unvoiced and fundamental pitch. This can be demonstrated easily. If one were to hold one's mouth wide open, without any movement of the jaw, tongue or lips, the only remaining changeable characteristics of sound generated by the vocal system are the above three parameters.

At any given time, excitation signals contain sounds at many

different frequencies. A voiced excitation signal is periodic. The energy in its frequency spectrum lies at multiples of the fundamental pitch, which is equal to the frequency with which the vocal cords are vibrating. An unvoiced excitation signal contains a random mixture of frequencies similar to what is generally called white noise.

The vocal cavity "shapes" the excitation signal into recognizable speech sounds by attenuating certain frequencies in the signal while amplifying others. The vocal cavity can accomplish this spectral shaping because it resonates at frequencies that vary depending on the positions of the jaw, tongue and lips. Frequencies in the excitation signal are suppressed if they are not near a vocal cavity resonance. However, vocal cavity resonances tend to amplify, or make louder, sounds of the same frequency in the excitation signal. The resulting spectral peaks in the speech sounds are called formants. Typically, only the three or four lowest-frequency formants will be < 5000 Hz. These are the formants most important for intelligibility. (The upper frequency limit for many audio communication systems, including the public telephone system in the USA, is ~3400 Hz. This is why speech sounds that differ chiefly in their upper-frequency formant structure, such as "f" and "s", tend to be hard to distinguish on these systems.)

For spoken English, a simple classification of speech sounds according to manner of formation would include vowel, nasal, fricative and plosive sounds. In the formation of vowels, such as the "ee" sound in "speech" and the diphthong "oi" in "voice", the breath stream passes relatively unhindered through the pharynx and the open mouth. In nasal sounds, such as the "m" and "n" in "man", the breath stream passes through the nose. Fricative sounds are produced by forcing air from the lungs through a constriction in the vocal tract so that audible turbulence results. Examples of fricatives include the "s" and "ch" sounds in "speech". Plosive sounds are created by the sudden release of built-up air pressure in the vocal tract, following the complete closure of the tract with the lips or tongue. "Talk" contains the plosive sounds "t" and "k". Except when whispering, the vowel and nasal sounds of spoken English are voiced. Fricative and plosive sounds may be voiced (as in "vast" or "den") or unvoiced (as in "fast" or "ten").

4. SPEECH COMPRESSION

The parameters computed by coders that follow this vocal tract model fall into two categories: those that control the generation of the excitation signal and those that control the filtering of the excitation signal.

Two different signal-generating mechanisms are required to produce a human-like excitation signal. One mechanism generates a periodic signal that simulates the sound produced by vibrating human vocal cords. The other produces a random signal, similar to white noise, suitable for modeling unvoiced sounds. Thus, when a voiced sound must be produced, such as the "ee" in "speech", the output from the periodic signal generator is used; for the unvoiced "sp" and "ch" sounds in "speech", the random output from the other generator is used.

In some systems, a weighted combination of the random and periodic excitation is used. This can be helpful in modeling voiced fricative sounds, such as the "z" sound in "zoo". However, many coders restrict the excitation so that it is modeled entirely by either the voiced or unvoiced excitation source. In these coders,

selection of the excitation is controlled by a two-valued voicing parameter, typically referred to as the voiced/unvoiced decision.

In addition to the voiced/unvoiced decision, the excitation function is scaled by an amplitude parameter, which adjusts its loudness. Finally, if the system is to generate something other than a monotone, it is necessary for the period of the voiced excitation source to be variable. The parameter that controls this is called the pitch parameter. In summary, three parameters are sufficient to control a simple excitation model (i.e. a model that does not take into account vocal tract differences among people): an amplitude parameter; a voiced/unvoiced parameter; and, if voiced, a pitch parameter that specifies the fundamental periodicity of the speech signal.

Various techniques have been used to simulate the manner in which the human vocal cavity imposes a particular spectral shape on the excitation signal. One of the first techniques developed uses a bank of band-pass filters, similar in many respects to the adjustable multiband 'graphic equalizers' found on some high-end stereo systems. The center frequencies of these filters are fixed; an adjustment in the gain of each filter or channel allows the desired spectrum to be approximated, in much the same way that adjusting the tone controls may vary the spectral characteristics of a stereo system.

The chief drawback to this approach is the large number of filters required. The number of filters can be reduced if it is possible to control their center frequencies. Specifically, by matching the center frequencies of filters to the desired formant frequencies, one can encode speech with only three or four tuneable band-pass filters. The important point is that even though the center frequencies of the filters must now be encoded along with the gains of the filters, the total number of parameters required for accurate shaping of the excitation signal is reduced greatly.

Although early speech synthesis systems relied on analogue mechanisms to filter and shape the excitation signal, modern speech compression systems rely entirely on digital filtering techniques. With these systems the decoded speech signal heard at the receiving end is the output of a digitally controlled filter that has as its input the appropriate excitation sequence. Digital control of the filter is accomplished through the use of a mathematical model — in essence, an equation with constants and variables in which the desired spectral filtering is specified by setting the appropriate values for the variables. Great reductions in the data transmission rate are achievable with this approach because the same mathematical model is preloaded into both the encoder and decoder. Therefore, the only data that must be transmitted are the relatively small number of variables that control the model.

A good example is the technique known as linear prediction, in which speech samples are generated as a weighted linear combination of previous output samples and the present value of the filter input. This yields the following expression for each output sample ($S[i]$) as a function of previous samples ($S[i-1]$, $S[i-2]$, …, $S[i-n]$), the prediction weights ($A[1]$, $A[2]$, …, $A[n]$) and the filter input ($U[i]$):

$$S[i] = A[1]S[i-1] + A[2]S[i-2] +… + A[n]S[i-n] + U[i].$$

The filter input in this equation ($U[i]$) is the product of the amplitude parameter and the excitation sequence. The total

number of coefficients in the equation (*n*) determines how many spectral peaks, or formants, may be approximated.

Once the complete set of parameters (amplitude, voicing, pitch, spectral parameters) has been specified, a speech decoder can produce a constant speech-like sound. To generate intelligible natural-sounding speech, the model parameters need to be updated as often as 40–50 times s^{-1}. To envision this process, it is helpful to recall how motion pictures work: apparent motion — in this case, a smoothly varying speech sound, rather than a smoothly varying image — is achieved by updating with sufficient frequency what are, in fact, still images. (Some systems that store speech in this format, such as the Lucent Technologies Intuity™ AUDIX®, voice messaging system, allow users to adjust the playback rate without the shift in tone that would accompany, for example, playing a 33 RPM phonograph record at 45. This is accomplished by adjusting how long each set of speech production parameters stays "in the gate" before being updated, in much the same way that "slow motion" is achieved with motion pictures.)

One of the first products to incorporate this style of speech compression was a children's learning aid introduced by Texas Instruments in 1978, the Speak & Spell,. It used 10-coefficient linear predictive coding (LPC-10) to model speech. The data rate for this LPC-10 model was 2400 bits s^{-1}. (The actual data rate in the Speak & Spell is considerably less than 2400 bits s^{-1} because a one-bit repeat code was used when adjacent parameters were judged sufficiently similar.) This low data rate was achieved, in part, by "hard-wiring" the excitation parameters that tend to vary from person to person. This meant that if people's vocal tract characteristics differed from those that had been built into the speech production model, their voices could not be reproduced without distortion.

The ability to model a wide variety of voices accurately is achieved by systems in which the excitation function is not hard-wired, but is instead under software control. A good example is the Intuity AUDIX voice messaging system, which uses code-excited linear prediction (CELP) to model speech. The data rate for typical CELP-based systems ranges from 4800 to 16 000 bits s^{-1}. (The higher data rates are seen more frequently in systems where it is important to maximize the speech quality or to reduce the computational complexity of the coder.) Compared with similar-quality uncompressed digitized speech, these techniques yield data rate reductions of at least 6:1, and as high as 20:1.

REFERENCES

COX, R. V., KROON, P., CHEN, J.-W., O'DELL, K. M. and ISENBERG, D. S., 1995, Speech coders: from idea to product. *AT&T Technical Journal*, 74, 14–22.

JAYANT, N. S. and CHEN, E. Y., 1995, Audio compression: technology and applications. *AT&T Technical Journal*, 74, 23–34.

MICHAELIS, P. R. and WIGGINS, R. H., 1982, A human factors engineer's introduction to speech synthesizers. In A. Badre and B. Shneiderman (eds) *Directions in Human–Computer Interaction* (Norwood: Ablex).

TETSCHNER, W., 1991, *Voice Processing* (Boston: Artech House).

Human Ecology: Developing Ecological Auditory Warnings

N. Stanton
Department of Design, Brunel University, UK
J. Edworthy
Department of Psychology, University of Plymouth., Plymouth PL4 8AA, UK

1. INTRODUCTION

Auditory alarms take many forms. Despite more advanced ergonomic warnings (Patterson 1982) many warnings consist of horns, klaxons, whistles, sirens, bells, buzzers, chimes, and gongs. There are various characteristics and features associated with the different technologies of auditory alarms, such as intensity, frequency, conspicuity, and noise penetration ability. Although much research has already been conducted in this area and improvements suggested and made as a result (Stanton 1994; Stanton and Edworthy 1999), this does not mean that all the problems have been solved, or that the information which already exists has filtered its way into current design practice. Case studies in coronary care, intensive care, aviation, manufacturing, and power generation show that although auditory signals may be successful in attracting attention, they can at times be disruptive distractions carrying little useful information about the problems of which they are informing. Further, a study by Patterson (1982) suggests that between four and seven warnings can be acquired reasonably quickly, thereafter performance slows down dramatically. Up to seven warnings can be retained, even after one week of absence, and this figure could be up to nine if the warnings were presented regularly. However, it is not just the sheer number of alarms, which cause problems for the user; in many cases the alarm sounds themselves are confusing because they are similar to one another either in sound quality or in meaning. There are also some problems in extrapolating from laboratory and in situ testing to the way in which alarms are most likely to be heard in the actual work environment. For example, Weirs and Kershner (1984) suggest that although twelve auditory alarms may be discernible on a relative basis (in tests at nuclear power plants), if absolute identification is required, this number is likely to be halved. There are at least two fundamental design problems. The first is how to design auditory alarms so that their meanings are conveyed to the listener in the most effective way. The second is the question of how many auditory alarms can be learnt and recalled comfortably. The two problems are interrelated, and one way to tackle them is to capitalize upon what the users understand about sound (including alarms and warnings) and its meaning.

The amount of information that can be transmitted by auditory tones is clearly very limited although it has also been shown that more advanced ergonomic alarms can convey larger amounts of information. The idea of transmitting some information about the nature of the problem through the auditory medium is an attractive prospect. There are a number of research projects which suggest how sound might be used to enrich the working environment that could be applied to the design of warnings and monitoring sounds. There are also methods from other areas of warning and information design research which might fruitfully be applied to auditory warning evaluation (Zwaga and Easterby 1984).

2. THE METHOD

Blattner, Sumikawa, and Greenberg (1989) suggest that there may be some principles common to visual symbols (icons) and auditory messages (earcons) in transmitting information to the user of a computer system. Earcons have been divided into two classes, in the same way as icons, known as representational and abstract. Representational earcons are perhaps the easiest to design in that they sound like the thing they represent to some degree. This requires that the representational object has a sound, and that the sound will be recognized and easily interpreted by the human observer. Gaver (1986) investigated symbolic (e.g. applause for approval) nomic (e.g. the sound of a closing metal cabinet for the closing of a file) and metaphoric (e.g. falling pitch for a falling object) mappings for representational earcons. He suggested that representations need not be realistic, but they should capture the essential features of the thing they represent if they are to be successful. Other research shows that the essential psychological features of events are conveyed through particular acoustic aspects of those events. Abstract earcons, on the other hand, require the development of a distinctive audio pattern which appears to be very like the construction of traditional auditory alarms. There are problems with the use of abstract sounds, because they can often be associated with several different types of meaning, which could result in confusion. However, some traditional sounds may already contain meaning, e.g. the symbolism of the fire bell. Gaver's classification (i.e. symbolic, nomic, and metaphoric) enables us to reinterpret sounds and understand the type of information they contain.

Development and construction of auditory symbols could follow a procedure similar to the one used in the development of visual symbols — for example, the method for the development of public information symbols (ISO TC 145/SCI DIS 7001) as proposed by Zwaga and Easterby (1984). Figure 1 illustrates our first iteration of an alarm construction and evaluation methodology adapted from this method. The figure distinguishes between the process (those activities required by the method, i.e. design, testing, and modification) and the output (the results produced by the aforementioned activities). The principal parts of the process are dealt with in more detail by the following sections. The activities may be likened to a funneling process: many sounds are entered into the testing procedure, but only a small subset of refined and modified sounds are left at the end.

2.1. Generate Trial Sounds

First of all, the number of potential referents needs to be identified. Some care needs to be taken not to identify too many of these as learning problems will be encountered later on. Rationalization may therefore be necessary at this stage. As well as any existing sound that may already be available (e.g. from the equipment currently being used), designers may wish to generate new sounds. This can be done by recording natural sounds (e.g. the sound of rapid heartbeat) and using synthesizers to create sounds (e.g. falling pitch for a falling blood pressure). Quite sophisticated mechanisms are available at relative low cost for sampling, mixing, and editing sounds which offer the alarm designer almost infinite

possibilities. The extensive range of options available leaves the designer in a quandary regarding which sound to choose for each function and this procedure would go some way in alleviating this problem. Part of the design process could be allocated to the user. For example, structured interviews or a free-association paradigm could allow potential users of the system to generate their own ideas about which sounds would be most appropriate. These ideas could then be modified by the designer.

2.2. Appropriateness–ranking Test

The appropriateness-ranking test involves initial screening of the potentially large number of sounds from the initial set. Sounds are presented to respondents together with their referent function. Respondents are requested to rank order the sounds for each function until the entire set has been presented. The purpose of this phase is to select a subset of highly ranked sounds for further testing. For each referent function two or three potentially useful sounds may be taken forward as potential warning sounds for inclusion in the final warning set. If a referent function has no appropriate sound, some modification of the sound set may be necessary, otherwise that referent may have to be returned to the initial design stage.

2.3. Learning/Confusion Test

A paired-associate learning task is then invoked to train participants in the referent functions for the set of sounds. After learning the sound set, participants would be tested immediately and again after a one-week interval. The dependent variables might be the time taken to learn the function of each sound and the ability to recall the function when the sound is played immediately after the learning phase and after a one-week interval. Confusions are also noted in order to determine which sounds are mistaken for other functions. The confusion matrix is an important part of this process. A whole sound set should be presented which corresponds with all of the sounds that will be presented in situ in order to identify any overlaps which may exist. Some confusion between sounds may occur because of a similarity in function rather than in acoustic properties. Whilst this is a theoretically interesting issue, which needs to be explored, it presents something of a problem in the initial identification of referents that should be borne in mind at that time.

2.4. Urgency Mapping Test

The urgency mapping test requires respondents to rate the urgency of each sound as it is played to them. Previous studies by Edworthy et al. (1991) have shown consistent and predictable effects through adjustments to certain characteristics of the sound. These findings may be used to carry out modifications where inconsistencies are found between subjective ratings of urgency for a sound and the objective urgency requirement of the function. The purpose of this exercise is to obtain ranking data on the relative acoustic urgencies of the warnings without knowledge of the objective urgencies of the referents. The relative urgencies of the referents should be rank-ordered if possible — although this would be difficult if the urgency of the referent was context-specific — and then correlated with the acoustic urgencies already obtained. If the matching between warning and referent is poor, the sounds can be modified in order to increase, or decrease, their urgency appropriately.

2.5. Recognition/Matching Test

The matching test requires the participant to assign a sound to one of the referent functions from a list they are given. They may assign more than one sound to a function if they see fit. This information is used for a preliminary modification of the sounds where required. Sounds that are successfully matched to functions at this stage proceed for further testing.

In the recognition test, respondents are required to recognize the situations to which the sounds refer. Sounds are presented to respondents who are required to write down the sound reference number and what they think the sound represents until all of the sounds have been presented. Some development of the most pragmatic procedure will be necessary for auditory warnings, as opposed to public information symbols, because of the demands that listening to the complete set of warnings will make on the respondent. (It is relatively easy for the respondent to scan a set of symbols in a few seconds — this is not the case for auditory warning sounds.) It is important that the respondent is aware of the nature of all the sounds which are later to be linked to referents before the procedure begins. An iterative procedure, such as response surface methodology, is envisaged as being appropriate. Respondents should be free to write more than one sound reference number against a particular function if they wish. These responses are then evaluated against the referent function that they were intended to convey. Again, if any function fails to reach an acceptable criterion level (e.g. ISO suggest 66% recognition accuracy as a minimum for visual symbols) then it is modified or referred back to the initial design stage.

2.6. Generate Standardized Verbal Descriptions

The verbal descriptions of the sounds serve as identifiers only and are given to label sounds with referent functions. In a similar way to public information symbols, these verbal descriptions would allow some flexibility in the application of the designs while retaining the essence of the sound.

2.7. Operational Test

An operational test may be carried out in a simulated or real environment. While this sort of testing often lacks the controlled rigor of experimental laboratory studies, it may provide the users with the ecological validity necessary to gain acceptance. At this stage, it is important to carry out an audibility test, especially if the warning set would typically be heard under noisy conditions. The operational test also has the advantage of testing sounds within the environmental context in which they are to be used. To do this effectively would require an observer (or observers) to collect data on the responses of the participant(s) to a set of alarms as they occurred. The appropriateness of these responses could then be evaluated during a debriefing session with the participant(s). This evaluation would indicate the success with which the alarm system was able communicate the problem to the participant(s).

3. CONCLUSIONS

Further refinement of the methodology is likely to include developing viable methods for carrying out the generation of trial sounds by respondents, the recognition and matching tests, and the degree to which urgency mapping is a necessary (or indeed

feasible) feature of the particular warning set being designed. However, the design of a representational auditory alarm system could communicate the nature of the fault to the human observer (e.g. pilot, nurse, or central control room operator). This type of display has the potential to bring the human observer 'closer' to the system (e.g. aircraft, patient, or power generation equipment) being monitored.

Representational auditory alarm displays could mean an increase in the amount of information that can be conveyed by a non-speech medium. The procedure would also allow already learned associations between 'traditional' alarms and their meanings to be capitalized upon, because the production is a two-way exchange between designer and potential user. Auditory displays of the nature discussed in this article may also benefit by the fact that they could be universally understood by all users, independently of spoken language. However, it is essential that they are first tested on a representative sample of the user population, and that once standards have been established they are adhered to. By ensuring that a multiplicity of alarms is not used for the same function, this would alleviate some of the problems associated with current systems.

REFERENCES

BLATTNER, M.M., SUMIKAWA, D.A. and GREENBERG, R.M., 1989, Earcons and icons: their structure and common design principles. *Human Computer Interaction* 4: 11–44.

EDWORTHY, J., LOXLEY, S., and DENNIS, I., 1991, Improving auditory warning design: relationship between warning sound parameters and perceived urgency. *Human Factors* 33(2): 205–31.

GAVER, W.W., 1986, Auditory icons: using sound in computer interfaces. *Human–Computer Interaction*. 2: 167–77.

PATTERSON, R.D., 1982, Guidelines for auditory warning systems on civil aircraft. *CAA Paper 82071*. (London: CAA).

STANTON, N.A., 1994, *Human Factors in Alarm Design*. (London: Taylor & Francis).

STANTON, N.A. and EDWORTHY, J., 1999, *Human Factors in Auditory Warnings*. (Aldershot UK: Ashgate).

WEISS, C. and KERSHNER, R., 1984, A case study: nuclear power plant alarm system human factors review. In Attwood, D. and McCann, C. (eds) *Proceedings of the International Conference on Occupational Ergonomics*, pp. 81–5. (Rexdale, Ont.: Human Factors Association of Canada).

ZWAGA, H. and EASTERBY, R., 1984, Developing effective symbols for public information. In Easterby, R. and ZWAGA, H. (eds) *Information Design* (London: Taylor & Francis).

Human–Computer Interaction (HCI)

G. Salvendy

School of Industrial Engineering, 1287 Grissom Hall, Purdue University, West Lafayette, IN 47907, USA

1. WHAT IS HCI?

Human–Computer Interaction (HCI) deals with the design of computing systems for human use. In accomplishing its objective, HCI design considers the limitation and availability of technology, the specific population that will use the computing systems and the tasks and functions for which it will be utilized. In pursuit of its objective, HCI researchers and designers consider the physical, cognitive, perceptual and social attributes and limitations of the human. Thus, HCI arrives at designing systems which all users, independent of skill level, cultural background or functional limitations resulting from disability, aging or circumstance, can effectively use.

The discipline of Human Factors and Ergonomics, since its inception more than 50 years ago, has been the leading force in designing machines for human use. HCI design is a specific case in Human Factors and Ergonomics, requiring the design of a system which is currently utilized by > one billion people. No other system or machine has ever existed with the physical, social and cognitive complexity so widely utilized, as the computer. For this reason, HCI evolved some 20 years ago as a separate professional and scientific discipline.

There are currently > 30 000 HCI professionals worldwide for whom three major international conferences specifically on HCI are held (ACM's CHI annual conference, Interacting with Computers biennial conference and the International Conference on Human–Computer Interaction biennial conference). In addition to these, there are national conferences on the subject-matter.

The HCI discipline is very well researched, having > 20 000 publications, which are disseminated via conference proceedings in five specifically dedicated journals on the subject matter (*Behaviour and Information Technology*, *International Journal of Human–Computer Interaction* [published in cooperation with the International Ergonomics Association and the Human Factors and Ergonomics Society], *International journal of Human–Computer Studies*, *Human–Computer Interaction* and *Interacting with Computers*) and books. Despite these significant activities in HCI, the description suffers from the following shortcomings:

- Less than 1% of corporate research and development in computer technology is dedicated to HCI.
- HCI lacks its own scientific base, and is neither respected by industry or academia.
- HCI is pulled by technology rather than pushing technology.
- HCI is primitive in its comparison to human–human communication.

2. TAXONOMY AND FRAMEWORK

HCI does not yet have a scientifically based and operationally oriented taxonomy which could provide a basis to pool together the large quantity of diversified research in a unified conceptual and operationally model that could be utilized by designers to further improve HCI and by researchers more effectively to target their research studies. This is partly because HCI professionals predominantly come either from computer science or human factors and ergonomics. The computer scientists do not get enough (and frequently get no) background in behavior science; the ergonomists do not get enough (and frequently get no) background in computer science. Yet, the HCI discipline is the interaction between the two.

Because of this lack of taxonomy in HCI, researchers and designers have developed a framework for capitalizing in designing HCI on representing the mental model of a user associated with a specific task performance. The original Macintosh computer had nicely taken advantage of this by, for example, requiring discarded materials to be placed in a wastebasket by using the mouse to drag the item to the wastebasket icon.

Another framework for HCI research and design is to minimize the complexity of task performance. This is achieved by chunking items of similarity, reducing the number of alternatives presented to the user, providing decision support to the user to reduce the mental load associated with task performance, designing an intelligent software to do a part of the task which originally was schedule to be performed by humans or, finally, by the use of visualization and multimedia enablers, in certain situations, for the presentation of information in a significantly less complex form than would be required in the presentation of similar information for decision making.

The emphasis by HCI researchers and professions to simplify the computer interface design is contrary to the finding by industrial psychologists who advocate enriching rather than simplifying, since the former results in both higher job satisfaction and performance, lower error rate, absenteeism and lower labor turnover. If so, why are the HCI professionals not enriching HCI design? The reason is presumably that HCI professional view their task as an interface design, when in effect it determines the structure, content and nature of the task performed by the user. This lack of understanding on the part of the HCI professional creates major social and psychological problems associated with computer based work. These, among others, include increased stress, anxiety, depression, and social isolation, and decreased job satisfaction in the workplace.

3. DESIGNING AND USING COMPUTERS

Nine key attributes need to be considered when designing and using computers. Each is presented along with a discussion of the key considerations for each attribute:

3.1. Design and Development of Software Systems

In designing and developing a software system, the following two items are most important: that it provides adaptability of use to fulfill each individual's needs and that it is enabled easily to correct errors created by the user. The software should enable the user to perform the task in a logical manner in which the user would like to perform the task.

3.2. User Interface Design

In this attribute, the most important two considerations are to minimize mental workload and ambiguity, and to maximize task strength.

3.3. Evaluation of HCI

HCI should be evaluated for consistency in design, extent of usability of the interface, and for the design of satisfaction associated in using the interface. For each of these three attributes standardized checklists and questionnaires are available.

The ultimate objective for evaluating HCI is to derive an evaluation index that would provide quantitative evaluation of the effectiveness of each HCI. This could be tested by an independent laboratory (such as an independent laboratory in the USA used by hardware products). By so doing, the consumer would be well informed regarding which products may be the best for their use.

3.4. Individual and Cultural Differences

Regarding individual and cultural differences, there are two main considerations: the spectrum novice versus the expert. As an example, experts perform better when using commands whereas novices perform better when using menus. In cultural differences, both format and structure are important. For example, Americans perform better when menus are presented horizontally, whereas people in Mainland China perform better when menus are presented vertically.

3.5. Multimedia and Visualization

The term "multimedia" implies the use of more than one of such items as voice communication, graphics, color and alphanumeric presentation. Multimedia is most helpful in either presenting complex problems or if one wishes to present both a trend and an exact presentation. An example of the latter is well illustrated in watch design by presenting the time digitally (which gives exact time but takes longer to read than analogue presentation) and with the dial by analogue presentation.

Visualization is most effective for presenting abstract concepts such as the sum of squares in a statistics procedure of analysis of variance.

3.6. Intelligent Interface and Systems

Computerized systems that the human cannot perform very effectively or systems for which intelligent interfaces can be designed economically such as the tuning of electronic tuning machines are prime candidates for intelligent system design. In this case, the procedure and thought processes involved in the tuning were documented, and based on this information a fuzzy logic-based combined neuronet and knowledge-based system was developed for intelligently (without human intervention). The machines can now be tuned at a fraction of the time and with greater accuracy than was possible with human tuning.

Intelligent interfaces can also act as decision supports in HCI operations. For example, in supervisory control of flexible manufacturing systems, a real-time look-ahead capability is provided to the computer interface supervisor in which information is used by the supervisor better to schedule the machine utilization and, thus, minimize throughput time.

3.7. Input Devices and Design of Work Stations

The physical part of HCI is concerned with input devices and workstation design. The effective design and implementation of these in the workplace will result in reduced fatigue and lower muscle skeletal discomfort than what otherwise would be possible. Three-way adjustable chairs with good back support and arm rests are as essential as wrist and foot rest devices and the use of input devices which require the use of minimum force.

Carpal tunnel syndrome frequently results from prolonged use of computers. This effect can be reduced by using open ergonomically designed input devices and workstations, and by providing rest periods during the workday. However, the most critical factor in determining the occurrence of carpal tunnel syndrome in computerized work is the length of time one works on the computer. Hence, job rotation between computerized and non-computerized work is an essential element for reducing the probability of carpal tunnel syndrome occurrence.

3.8. Computer–supported Collaborative Work and Organizational Issues

The key issues here are how organizational learning can be accelerated by utilizing and improving the intranet, and how the web can facilitate collaborative publishing, manufacturing and commerce. The key requirement for achieving these objectives is the design of adaptive interfaces which accommodates for differences in design requirements for different collaborative partners.

The more enriched (as opposed to simplified) the task of each collaborator, the better the understanding of cognitive style and cognitive aspects of their collaborator which results in increased effectiveness of the collaborative work.

3.9. Standardization

There are national and international standards for the design and use of computer related hardware, software and system operations. These standards should be consulted before designing and using HCI. These standards are good, but they frequently provide the minimum, rather than optimum or maximum, of requirements to achieve a most effectively functioning system. Additionally, these standards are for interface design and, as such, do not consider the psychological aspects of task content which are associated with the task performance. As such, these standards do not provide guidelines on the design of HCI jobs, break schedules, requirements, job rotation needs or derivation of acceptable work performance.

REFERENCES

HELANDER, M.G., LANDAUER, T.K. and PRABHU, P.V. (eds), 1997, *Handbook of Human–Computer Interaction* (2nd ed.) (Amsterdam: Elsevier).

SALVENDY, G. (ed.), 1997, *Handbook of Human Factors* (2nd ed.) (New York: Wiley).

SALVENDY, G. (ed.), 1999, *Handbook of Industrial Engineering* (3rd ed.) (New York: Wiley).

Hypertext and Hypermedia

P. R. Vora

US WEST 1515 Arapahoe Street, Denver, CO 80202, USA

> "The human mind does not work that way. It operates by association. With one item in its grasp, it snaps instantly to the next that is suggested by the association of thoughts ...."
>
> Vannevar Bush (1945)

1. INTRODUCTION

Hypertext is intended to overcome the artificiality of index-based systems of storage and retrieval by providing computer-supported *links* between related pieces of information (*nodes*) (Figure 1). By doing so, hypertext permits non-sequential (or non-linear) associative mode of information access, which is both more direct and more immediate than possible in a conventional paper-based system of information storage.

In addition to being a non-linear mode of information access, hypertext may refer to one or several of the following:

- Information creation: hypertext is described as "non-sequential writing" — a *liberating* medium, where readers can annotate and make their own links making different paths through the text.
- Information storage and management: hpertext is also an approach to information management in which data are stored in a network of nodes connected by links.
- Information presentation and access: Nielsen (1995) characterized hypertext more by its "look and feel" rather than the technology; hypertext users should *feel* that they can move freely through the information according to their own needs. This ability to navigate non-linearly using links is probably hypertext's most important characteristic.

Hypertext and hypermedia are often distinguished, with the former referring to text-only systems and the latter referring to systems that support multiple media. This distinction is not made in this chapter, however; the term "hypertext" is used generically to refer to both text-only and multimedia systems.
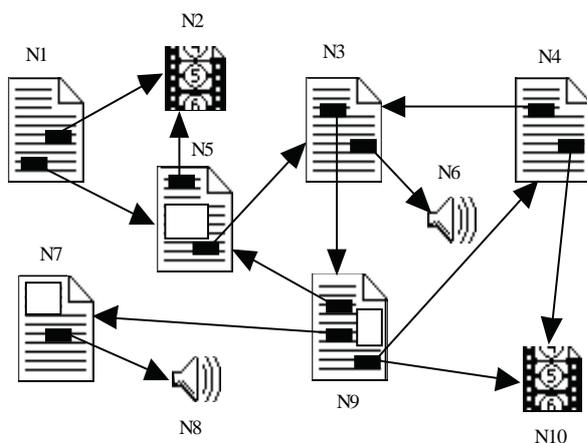
Figure 1. A simplified representation of hypertext. N1–10 are nodes; the arrows between them are links. N2 and N6 are nodes with video content, whereas N6 and N8 are nodes with audio content.

2. A BRIEF HISTORY OF HYPERTEXT

The concept of hypertext is credited to Bush (1945) who envisioned "memex" to support intellectual work using the complex interlinkage of accumulated knowledge. However, it was T. H. Nelson who coined the term "hypertext" and offered his vision of "Docuverse," which would contain the entire world's literature, accessible and interlinked from any point in the global network. And it was Doug Engelbart (1963), who emphasized the need to augment human intellect through cooperative work environments via shared electronic spaces, where groups of workers could manage and exchange concepts and ideas in the form of linked structures.

Despite the differences in their visions, they described hypertext as a means for enhancing idea processing by allowing the freedom to read non-sequentially, annotate freely and to collaborate on writing of complex documents.

3. REALITIES OF HYPERTEXT USE

The main justification for hypertext is the claim that useful information retrieval is much easier through associations, rather than through indexing. The ensuing freedom to browse is the single most important advantage of hypertext technology. However, the freedom and flexibility offered by hypertext often accompany the well-known problems of *cognitive overhead* and *disorientation* (Conklin 1987).

Cognitive overhead applies to both readers of hypertext as well as its authors. For readers it is the overhead of making decisions as to which links to follow and which to abandon, given a large number of choices. Whereas for authors it is the additional mental overhead to create and keep track of nodes and links. Disorientation, on the other hand, is the tendency to lose one's sense of location and direction in a non-linear document — the "lost in hyperspace" problem.

To mitigate these problems associated with information seeking in hypertext, users are offered several navigation support features, to help answer their questions such as: Where am I? Where to go next? How do I get there? How do I return?

4. INFORMATION SEEKING AND NAVIGATION SUPPORT

Fundamentally, there are two forms of information seeking: *navigation* and *browsing*; information retrieval (IR), yet another form of information seeking, is discussed below. The main difference between navigation and browsing is based on user goals. In navigation, users information seeking behavior is goal- or purpose-oriented. Whereas, in browsing users explore the available hypertext to get a general idea about one or several topics.

Hypertext designers use one or more of the following techniques to help users navigate in hypertext.

4.1. Paths, Guided Tours and Tabletops

Paths or *guided tours* are pre-authored linear paths through the document. They are comparable to the notion of "trails" suggested by Bush (1945), where the authors ("trailblazers") suggest a useful path through the information. This "ordered sequence of nodes" does not mean restricting the presentation to only one node at a time. Trigg (1988) suggested using *tabletops*, where a screenful of nodes is presented.

Although paths and guided tours violate the fundamental notion of non-linearity and browsing freedom in hypertext, they may facilitate navigation for novice users.

4.2. Backtrack and History Lists

Probably the most important navigation facility in hypertext is allowing the users to *backtrack* — that is, to return to the nodes they have visited before. Unlike backtrack, which allows users to return to previous nodes one at a time in a reverse temporal sequence, *history lists* allow users to jump directly to any of the previously visited nodes. History lists may be presented textually by listing the node names, graphically by showing visited nodes in an iconic form, or by a combination of both.

4.3. Bookmarks

Some hypertext systems allow users to bookmark nodes so that they can return to them later. Bookmarks are particularly useful for browsing large hypertext systems such as the World-Wide Web, where users stumble across interesting information that they may wish to peruse later.

4.4. Landmarks

Like both natural and man-made landmarks, hypertext landmarks help users to identify their current position in the hypertext network and help in orientation. One approach to identifying landmarks is based on connectivity. The basic idea is that important topics tend to be connected either to many other topics directly or to topics that are in turn well connected, e.g. Home Page on a Web site.

4.5. Overview Diagrams or Maps

The approaches to support navigation discussed so far help readers locate pertinent information quickly. *Overview diagrams* (or *maps*), on the other hand, support navigation by helping readers understand the overall structure of hypertext. Similar to table of contents in linear texts, overviews or maps attempt to delineate hypertext's scope and organization. However, in many situations, the design problem is constructing a usable map-like representation. For even a medium-sized hypertext with moderate connectivity, the node-link structure may appear to be a "visual spaghetti" (Conklin 1987). Proposed solutions, such as those listed below, allow users to examine local detail while still maintaining a global perspective — also referred to as "focus + context" methods.

4.5.1. Fisheye views

A fisheye view displays in detail the area of interest, while presenting distant areas in successively less detail (Furnas 1986). Fisheye views have been found to significantly improve user performance when compared with the traditional zooming on a hierarchically structured network.

4.5.2. Three-dimensional (3D) representations

Some of the 3D visualization techniques, developed under the *Information Visualizer* project at Xerox PARC, include *Cone Trees* for visualizing hierarchies and *perspective walls* for visualizing linear structures (Card 1996). Cone trees are hierarchies laid out uniformly in three dimensions, either vertically or horizontally. Whereas, perspective walls represent linear structures having wide aspect ratios that are difficult to accommodate in a single view.

Usability of the 3D representations is often enhanced by the use of color, lighting, shadow, transparency, hidden surface occlusion, continuous transformation, smooth interactive animation, motion cues, and 3D perspective. However, an important usability problem with 3D representation still remains — the choice of input device for manipulation of 3D objects on two-dimensional display surfaces.

4.5.3. Multi-level overviews

Another solution to the problem of representing large information spaces may be to abstract information at several levels to provide progressive disclosure of the content. These overviews are often presented as hierarchical structures with various levels of detail. This approach can be useful not only for navigation, but also for improving comprehension.

4.6. Filtering Hypertext Network

Filtered views of hypertext network are possible when hypertext systems classify (or *name*) nodes and links. Users can then specify their interests in terms of node and link types and the system can generate a filtered hypertext. Furthermore, some hypertext systems may lend themselves to "structure search," where a user could ask for a diagram of all subnetworks that match a given pattern of node and link types.

4.7. Adaptive/Intelligent/Personalized Navigation

Intelligent navigation support mechanisms help users select a path through the hypertext that is tailored (or personalized) for a particular purpose. For example, one may generate hypertext by computing the degree of relatedness between member nodes of a hypertext system rather than using explicitly hard coded links. Most adaptive hypertext systems consider either the user goals or user characteristics to provide intelligent navigation support.

5. INFORMATION RETRIEVAL AND INFORMATION FILTERING

Finding information in hypertext simply by traversing links, especially in large hypertexts, can become cumbersome and compromise usability. Therefore, most reasonably sized hypertexts provide additional ways of accessing information via *information retrieval* (search or query) and/or *information filtering*. In *Information retrieval* (IR), users *search* for specific keywords by querying the content of the hypertext system. Whereas, in information filtering (IF), the system presents only those documents to the users that match their profiles (or interests). For example, information retrieval might be to find the name of the companies that make supercomputers; whereas, information filtering might be used to inform a user every time a new supercomputer is introduced in the market.

A caveat is in order. Neither IR nor IF guarantee increase in hypertext usability. A good user interface is important to ensure that information is used effectively and that the users can quickly get to the relevant information.

5.1. Integrated Information Retrieval and Browsing

To access information from large databases, information retrieval and browsing are likely to be integrated in future hypertext

interface. One example is *Starfield display* by Ahlberg and Shneiderman (1994). In this approach, users perform information retrieval (or query) simply by dragging a set of sliders representing different characteristics of the hypertext system. The results of queries are presented continuously on a spatial display, the "starfield" display, which is a two-dimensional scatterplot with additional features to support selection and zooming. The user can progressively refine the query using sliders or select any object (displayed as a dot or a "star") in the scatterplot.

5.2. Improving Hypertext Comprehension

An implicit assumption in hypertext's interaction model is that the readers are capable of creating coherent cognitive maps of underlying hypertext structures by selective browsing through nodes. Reality, however, is that unless the structure of hypertext is made comprehensible to the user, the user will find hypertext fragmented and incoherent. Many researchers believe that it is incomprehensibility of hypertext that causes disorientation, not non-linearity *per se* (Vora 1994, Thüring *et al.* 1995)

There are two ways of improving comprehensibility of hypertext: by using appropriate metaphors and by improving design of hypertext documents.

5.3. Use of Metaphors

By using familiar metaphors, the structure of hypertext can be made intuitively recognizable; users can transfer knowledge and skills from a familiar domain to another less familiar area and can predict the behavior of objects. Several types of metaphors have been used in hypertext systems: Cards, books, encyclopedias, libraries, cities and maps. Metaphors should be selected carefully, however. Badly chosen metaphors are counter-productive. Often, a non-metaphorical interface, with well-designed features, works better.

5.4. Design of Coherent Hypertexts

Design of coherent hypertext systems requires that necessary cues are provided at both the individual *node level* and at the organizational or the *structural net level* (Vora 1994, Thüring *et al.* 1995). This is similar to writing coherent linear text, where comprehension is required at the microstructure, macrostructure and superstrucutre level (van Dijk and Kintsch 1983). For example, one may comprehend the words, but not the meaning of a sentence, or one may understand sentences and still be confused by the overall organization of the text.

In hypertext, to improve comprehension at the node level, one can use the traditional reading models to improve comprehension. And at the net level, the designers should attempt to limit the "fragmentation" so users can understand how the text is distributed over several nodes and what they have to do with each other. A coherent hypertext could then be achieved by the following measures:

- Labeling the links: link labels make explicit the semantic relationships between the nodes and enable readers to identify the links of interest and make hypertext easier to read and navigate. Vora (1994) found that graphical representation of information with labeled links helped in comprehension and improved navigation and search performance.
- Providing continuity through context: conveying a sense of continuity across nodes is important for comprehension as it helps readers form semantic relationships among nodes in a hypertext application.
- Aggregating the information into higher order units: *aggregate* or *composite* nodes help users to identify important themes (or functional categories) of hypertext.
- Providing a graphical overview: graphical representation or maps can convey the structure to the user. If the target user population is diverse, it may help to offer multiple overview diagrams or access paths through the hypertext (Vora 1994).

6. HYPERTEXT USABILITY

Although, like software interfaces, hypertext interfaces can be evaluated for usability along the usability dimensions of *Learnability*, *Efficiency*, *Memorability*, *Errors* and *satisfaction* (Nielsen 1995), it is crucial that they be evaluated on two additional usability dimensions: *disorientation* and *comprehension*.

6.1. Disorientation

Four kinds of disorientation are particularly relevant when evaluating the usability of hypertext systems: not knowing where to go next; not knowing where one is; not knowing how one arrived at a particular node; and knowing where the information is, but not knowing how to get there. In addition to commonly used measures of disorientation (e.g. number of "backtracks" to previously visited nodes, restarts from the "home" node and subjective feelings of disorientation), a useful measure of hypertext usability and disorientation is user perception of the size of hypertext. This is important for hypertext because, unlike linear text, which present clear boundaries, hypertext readers have no cues available for its size and scope.

6.2. Hypertext Comprehension

Although the term comprehension includes understanding of the content of hypertext, hypertext comprehension as described here means understanding hypertext's structure in terms of relationships among the nodes and links. Therefore, useful indicators of comprehension are users' mental models of the information organization in hypertext. This can be achieved directly by asking the users' to construct the maps of information in hypertext or indirectly by extracting the structure from their verbal protocols (requiring participants to think aloud about what they are doing, how they are doing it and why they are doing it).

Yet another alternative is to use a *teaching method*, where after a users have learned the system by browsing or retrieving information using pre-determined tasks, they are asked to *teach* another participant (a confederate) how the system is organized (Vora 1994).

7. APPLICATIONS OF HYPERTEXT

Although hypertext has been used in several computer, business, intellectual, educational, and entertainment and leisure applications, probably the most important application of hypertext is the World-Wide Web (or W3 or the Web). Since its introduction at the Hypertext '91 conference, the growth of the Web has been astronomical; Web servers on the Internet has risen from 130 in December 1993 to > 4 million sites in January 1999 (www.netcraft.com) (see the chapter on the Internet and the World-Wide Web in this volume).

8. FUTURE OF HYPERTEXT

Undoubtedly, most of the hypertext-related research in the future will involve the Web. Therefore, the following predictions of the future of hypertext include the technological advances made on the Web.

8.1. Content–based Linking and Retrieval and Integration of Virtual Reality

In the future, linking and retrieval of information will be based not only on text or simple graphics, but will also use images, sound and video. Probably the biggest advances will be made in linking among temporal media such as sound and video, where links are not just present on a page, but become spatial and temporal opportunities for hypertext readers. Inclusion of temporal media and interactivity will finally merge with virtual reality to create hyperworlds offering not just richer interactive hypertext environments, but immersive experiences as well.

8.2. Personalized Hyperspaces

Considering the growth of the Web and the vast amount of, importance of personalized hyperspaces can not be overemphasized. Creation of such personalized hyperspaces will be driven by intelligent agents, which will not only gather information in its raw form, but also interpret information to create personalized hyperspaces and personalize to fit user tasks and needs. They will aid not only in information access and organization but serve as negotiators for its owners for transactions for exchanging information that may or may not involve money.

For the agent technology to be successful, however, it is necessary that the information objects are richer and carry extensive metadata and there is an agreed upon standard way to describe different types of information objects.

Without a richer, expressive, and standardized language to describe the information, retrieval and discovery of information and structure will not be possible and the agents' technology will remain in a nascent stage.

8.3. Cooperative and Ubiquitous Hypertext Systems

Cooperative hyperspaces will become a reality through improved methods of collaboration and improved mediums of communication. Collaboration will be both asynchronous and synchronous. It will support "nomadic" environments (mobile participants) and create what Streitz (1996) called "Ubiquitous Collaboration" which supports *polyphasic activities*. In other words, incompatibilities due to differences in presentation format and media will disappear.

8.4. Open and Interoperable Hypertext Systems

To enable creation of personal and collaborative hyperspaces, interoperability and availability of open hypermedia systems (OHS) will become an absolute necessity. Traditional hypertext systems have often used private formats to store both links and data, thus limiting interaction. Open hypertext systems will also require increasing emphasis on standards and interoperability. This will eventually make Neslon's vision of "Docuverse" a reality; it would contain the entire world's literature, accessible and interlinked from any point in the global network.

Finally, as all ubiquitous and usable technologies, hypertext will become a *transparent* technology — it will become *the way* to interact with information on computers.

REFERENCES

AHLBERG, C. and SHNEIDERMAN, B., 1994, Visual information seeking: tight coupling of dynamic query filters with starfield displays. *CHI '94 Conference Proceedings*, 313–17.

BUSH, V., 1945, As we may think. *Atlantic Monthly*, 176, 101–8.

CARD, S., 1996, Visualizing retrieved information: a survey. *IEEE Computer Graphics and Applications*, 16, 63–7.

CONKLIN, J., 1987, Hypertext: an introduction and survey. *IEEE Computer*, 20, 17–41.

ENGELBART, D.A., 1963, A conceptual framework for the augmentation of man's intellect. In Howerton and Weeks (eds) *Vistas in Information Handling* (London: VI Spartan), 163–74.

FURNAS, G.W., 1986, Generalized fish-eye views. *Proceedings of CHI '86*, 16–23.

NELSON, T.H., 1981, *Literary Machines* (Swarthmore: author).

NIELSEN, J., 1995, *Multimedia and Hypertext: The Internet and Beyond*. (New York: Academic Press).

STREITZ, N.A., 1996, Plenary Session, Opening Keynote: Ubiquitous collaboration and polyphasic activities: from desktops and meeting rooms to cooperative buildings and virtual organizations. *Hypertext '96*, ACM Conference on Hypertext, Washington, DC.

THÜRING, M., HANNEMANN, J. and HAAKE, J.M., 1995, Hypermedia and cognition: Designing for comprehension. *Communications of ACM*, 8, 57–66.

TRIGG, R.H., 1988, Guided tours and tabletops: tools for communicating in a hypertext environment. *ACM Transactions of Office Information Systems*, 6, 398–414.

VAN DIJK, T.A. and KINTSCH, W., 1983, *Strategies of Discourse Comprehension* (London: Academic Press).

VORA, P.R., 1994, Evaluating interface styles and multiple access paths in hypertext. Unpublished PhD dissertation, Department of Industrial Engineering, State University of New York at Buffalo.

Information Design: Warning Signs and Labels

D. R. Clark† **and M. R. Lehto‡**

†Department of Industrial & Manufacturing Systems Engineering, Kettering University, Flint, MI 48504–4898, USA

‡School of Industrial Engineering, Purdue University, West Lafayette, IN 47907–1287, USA

1. WARNINGS IN THE HIERARCHY OF HAZARD CONTROL

There is a number of approaches to the control of safety hazards that can generally be hierarchically arranged, in order of preference, as (1) design-in, (2) safeguarding and (3) administrative controls. Design-in safety relies on creating systems that have inherent safety due to hazards being either lacking or at a level, from the user's perspective, insufficient to harm. In many cases, attempting to achieve this level of inherent safety can result in the loss of the original utility of the system. More often, one sees a spectrum of engineering safeguards applied, ranging from those that are tightly integrated to the design to those that are added in the field. The effectiveness of safeguards depends, in large measure, on how independent their performance is from the actions of the people they are design to safeguard. At the extreme, a transition occurs to administrative controls that depend on proper procedures, adequate information, and effective action of people. The application of warnings as a form of hazard control is usually placed here.

Warning signs and labels necessarily must rely on people to read, understand, and follow any explicit or implicit information given. However, overwarning can cause habituation that can result in reduced overall effectiveness. Therefore, warnings should be considered a supplement to engineering of the system itself. The purpose of a warning is to alert people to the presence of a hazard, in hopes of avoiding unsafe or risky behavior.

The errors that people make can be classified into three broad categories: (1) errors that occur when performing routine behavior, (2) when performing non-routine behavior and (3) intentional violations. Warnings that interrupt performance of a task can successfully remind people to not perform incorrect behavior before it occurs. As well, warnings can stimulate recovery from errors. Warnings, though, should be used with care, especially when considering the level of one's experience, as it is found that experienced users, familiar with the task and the product, often fail to read warnings. One potential reason may be that the user is highly focused on task-related goals, only looking for information that will satisfy the end-goal of task completion. Interrupting the performance of a task may also take the user out of an automatic form of behavior and encourage them to enter a less routine information-seeking mode of behavior.

2. GENERAL GUIDELINES FOR DESIGN AND APPLICATION OF WARNINGS

Many general guidelines can be inferred from warnings research. These include: (1) conforming to standards and population stereotypes when possible, (2) making warnings conspicuous and each component legible, (3) simplifying the syntax of text and combinations of symbols, (4) making symbols and text as concrete as possible, (5) making sure that the cost of compliance is within a reasonable level, (6) being selective, (7) integrating warnings into the task and hazard related context, and (8) matching warnings to the level of performance at which critical errors occur for the relevant population. These general guidelines can be expanded upon and classified into three broad categories; format, content and mode of presentation.

2.1. Format

The format of a warning refers to its physical aspects, which include text, iconic and/or graphic information, both in individual and in collective form, and the size of the label, methods of color-coding, and the arrangement of the components. Twelve specific guidelines (F1–12) have surfaced from the literature:

- (F1) Provide a signal word or words in capital letters at the top of a warning sign or label to indicate the sign or label is a warning.
- (F2) Use lowercase for text within a warning sign or label.
- (F3) For text within smaller labels, a visual angle of at least 10 min of arc for the intended viewing distance should be provided. For text within signs, a visual angle of a minimum of 25 min of arc should be present for the intended viewing distance.
- (F4) For short messages or signal words, text characters should be sans serif.
- (F5) Text characters should be dark against a light background.
- (F6) The stroke width-to-height ratio of text characters should be between 1:6 and 1:10.
- (F7) The width-to-height ratio of alphanumeric characters should be between 1:1 and 1:3.5.
- (F8) Provide for a brightness contrast of at least 50% between the warning text and its background.
- (F9) Consider color-coding schemes consistent with ANSI Z35.1/Z535.1.
- (F10) Avoid crowding of components on the sign or label, ensuring that each component is legible.
- (F11) The presence of adverse viewing environments, such as the presence of dirt, grease and other contaminants, can degrade the legibility of warning components.
- (F12) Consider the use of symbols/pictographs when users are performing routine behavior.

Consider using text when employees need to make decisions or are performing non-routine behavior.

2.2. Content

The content of the warning refers to the message itself, its level of abstraction, and its syntactic structure. Some specific guidelines (C1–17) that can be derived from these generic guidelines and from the research follow:

- (C1) Messages should focus on critical errors that cause a significant safety problem. When choosing warnings, one should be selective to avoid habituation and warning overload, and thus reduce the potential for ignoring the warning. Therefore, avoid long lists of hazards and messages that describe trivial hazards or hazards that are obvious to the intended audience.
- (C2) Focus on developing messages for the following two types of error situations: (1) forgetting to perform an action

ordinarily performed (i.e. the sign or label reminds), and (2) not knowing the consequences of performing or failing to perform some action.

- (C3) When a user's performance is skill-based, meaning they are performing an automated set of procedures, and they commit an error based on their failure to perceive a condition or motor variability, provide a warning signal and consider training.
- (C4) When performance is rule-based, meaning the user is following a set of rules and the behavior is not yet automated, and the error is caused by a incorrect or inadequate rule, determine whether the rule was originally developed on the basis of knowledge or judgment-based behavior. If it seems to be judgment-based, such as when people speed on the highway, focus on enforcement. If it seems to be knowledge-based, determine whether a warning sign or label can be used to interrupt the task (i.e. to place its message into short-term memory at the time it is relevant).
- (C5) When performance is knowledge-based, meaning the user is problem-solving, and the error is caused by inadequate knowledge, the amount of knowledge necessary to prevent the error should be determined. If the knowledge can be described with a small number of rules, consider a warning sign or label containing these rules in the form of step-by-step instructions. Otherwise, focus on training, instruction manuals, or other forms of education.
- (C6) When performance is judgment-based, meaning the user is experiencing an affective reaction of some sort, and the error is caused by inappropriate priorities, evaluate the user's behavior pattern. If the undesired behavior pattern appears to have significant value to the user (i.e. pleasure, comfort, convenience, etc.) or is likely to be entrenched, focus on enforcement through supervision.
- (C7) Regardless of the level of a user's performance, consider messages that minimize the cost and increase the benefits of compliance.
- (C8) If a large number of potential warnings is present after applying these guidelines, increasing the probability of overloading the user, other means of providing the information (such as instruction manuals or training courses) should be considered to reduce warning overload.
- (C9) When subjects are inexperienced, consider pictographs (having a more detailed design) instead of symbols.
- (C10) When performance is at a skill- or rule-based level, consider brief messages that describe conditions or actions. Also, consider symbols or pictographs instead of text, since these make the warning more memorable.
- (C11) When performance is at a knowledge-based level, whereas the user needs to understand the reasoning "why" an action or behavior needs to be performed, consider more detailed messages that describe both conditions and actions.
- (C12) When performance is at a judgment-based level, in that the incoming information is processed as values that evoke an affective reaction, consider messages that describe the hazard and the benefits of compliance. Also, consider citing highly credible sources.
- (C13) When the hazard is complex or occurs in different manifestations, consider abstract text, which better covers hazard contingencies, rather than a long list of concrete examples or symbols.
- (C14) When knowledge or understanding of a product or task is low, consider concrete text instead of abstract symbols or pictographs. Concrete text is easier to comprehend and the interpretation of symbols has been found to vary across different cultures.
- (C15) Use text and symbols that people in the intended user population can comprehend. Consider language, reading level and cultural effects.
- (C16) Use short, simple sentences; complex conditional sentences, particularly those containing negations, should be avoided. Longer messages are not necessarily better comprehended.
- (C17) Symbolic signs or labels should focus on describing conditions (i.e. flammable). With few exceptions (i.e. a slash/bar to indicate negation) they should not combine multiple meanings or be used to describe complicated sequences of actions.

2.3. Mode of Presentation

The mode of presentation of a warning refers to the location and task-specific timing of one's contact with a warning. Specifically, the following guidelines (M1–5) should be followed:

- (M1) The warning sign or label should be presented at a location and time in which the danger is still avoidable.
- (M2) The location and timing of presentation should minimize the cost or difficulty of compliance.
- (M3) Make an attempt to present the warning sign or label at a time when the person has available attentional capacity.
- (M4) When performance is skill-based, determine if the task can be interrupted to bring attention to the label. If this can be done, consider a warning sign or label that describes the condition and prescribes an action. If this cannot be done, consider providing a warning signal, training or modifications of the product.
- (M5) Avoid imbedding sign or label in a cluttered background.

3. ASSESSMENT OF WARNING LABELS AND SIGNS

3.1. Assessment of Effectiveness

The steps that must occur before a warning can prevent an accident include: (1) attention, (2) comprehension, (3) decision-making, (4) action-taking and (5) sufficiency of the action to avoid the accident. The probability of avoiding an accident can never be greater than the probability of successfully completing a single step.

The dependent variables related to the effectiveness of warnings can be grouped into three areas: perceptual factors, comprehension levels and those factors associated with behavior patterns. An assessment of the perception of warnings should give insight as to the conspicuity, or attention attracting ability of the sign. Variables related to the perception of warnings include reaction time, accuracy of task completion, attention to different elements of the warning, the use of tachistoscopic procedures and measures of legibility distance. When assessing the comprehensibility of a warning, four variables are often used: symbol recognition or matching, message recall, psychometric (rating)

scales and readability indexes. Assessing behavior patterns also offers some insight about the effectiveness of warnings under realistic conditions. The most common approach is to conduct an experiment in which two groups of subjects are tested on the same task, one group is the control without the warning, and the other group receives the warning. Field observations are also useful in determining warning effectiveness.

3.2. Role of Standards in Assessment

Published US standards helpful in the design or assessment of warnings labels and signs include:

- *ANSI Z535 Standards*. This series of standards replaced the ANSI Z35 and ANSI Z53 standards and includes: (1) ANSI Z535.1, Safety Color Code; (2) ANSI Z535.2, Environmental and Facility Safety Signs [typical sign consists of multiple panels for signal words, messages, and symbols]; (3) ANSI Z535.3, Criteria for Safety Symbols [methodology for the development of understandable symbols]; (4) ANSI Z535.4, Product Safety Signs and Labels [similar to ANSI Z535.2, except generally smaller]; and (5) ANSI Z535.5, Accident Prevention Tags [for temporary hazards].

- *ANSI Z129.1, Hazardous Industrial Chemicals – Precautionary Labeling*. This standard is applicable to chemical products and mixtures and is complementary in scope to ANSI Z535.2 in the environmental or facility area, and to ANSI Z535.4 in the product area.

- *OSHA 29 CFR 1910 General Industry Standards*. Three sections of these standards are particularly applicable: (1) 29 CFR 1910.145, Specification for Accident Prevention Signs and Tags; (2) 29 CFR 1910.144, Safety Color Code for Marking Physical Hazards; and (3) 29 CFR 1910.1200 Hazard Communication Standard. The first two sections are very similar to ANSI Z35.1/Z53.1/Z535.1. The last, also known as the Right-to-Know Standard, requires employers to provide specific information to employees about the hazardous chemicals they are or may be exposed to through on-container labels and Material Safety Data Sheets (MSDS).

- *DOT 49 CFR 170 Hazardous Materials Transportation Regulations*. These regulations address the "diamond" signs on the outside of trucks, whose primary purpose is to alert emergency response personnel to potential hazards, and are comprised of nine classes of hazards, with corresponding colors, symbols, and text.

- *NFPA 704, HMIS®, and ASTM D 4257*. The Standard System for the Identification of Fire Hazards (NFPA 704), the Hazardous Material Identification System® (National Paint and Coatings Association) and the Safety Alert System (ASTM D 4257) are all designed to aid workers and/or emergency response personnel in the identification of hazards and the appropriate procedural response. They address fire, health, reactivity levels and specific hazard concerns.

4. THE CONTINUING EVOLUTION OF WARNINGS

As knowledge accumulates regarding the factors that impact the effectiveness of warnings, criteria for the design and assessment of warning signs and labels found within both mandatory and voluntary standards evolve accordingly. Since most standards are revised every 5 years or less, one should always try to obtain the latest versions of standards referenced here.

Finally, bear in mind that in a court of law compliance with standards is generally not a sufficient sole defense as standards are often viewed as minimum requirements (often due to a "one size fits all" strategy in their development). Therefore, it is necessary fully to understand the potential hazards of any system being analyzed (i.e. environment, facility, product, etc.), apply the hierarchy of hazard control to achieve the most effective protection for those who use or are exposed to the hazards, and apply warnings as a sole remedy sparingly.

REFERENCES

AMERICAN NATIONAL STANDARDS INSTITUTE/CHEMICAL MANUFACTURERS ASSOCIATION. *American National Standard for Hazardous Industrial Chemicals – Precautionary Labeling*, ANSI Z129.1; *American National Standard for Material Safety Data Sheets*, ANSI Z400.1.

AMERICAN NATIONAL STANDARDS INSTITUTE/NATIONAL ELECTRICAL MANUFACTURERS ASSOCIATION. *American National Standard Safety Color Code*, ANSI Z535.1; *American National Standard for Environmental and Facility Safety Signs*, ANSI Z535.2; *American National Standard for Criteria for Safety Symbols*, ANSI Z535.3; *American National Standard for Product Safety Signs and Labels*, ANSI Z535.4; *American National Standard for Accident Prevention Tags (for Temporary Hazards)*, ANSI Z535.5.

EDWORTHY, J. and ADAMS, A., 1996. *Warning Design: A Research Perspective* (London: Taylor & Francis).

HADDEN, S.G., 1986. *Read the Label* (Boulder: Westview).

LEHTO, M.R. and MILLER, J.M., 1986. *Warnings*, vol. I: *Fundamentals, Design, and Evaluation Methodologies* (Ann Arbor: Fuller Technical).

MILLER, J.M., LEHTO, M.R. and FRANTZ, J.P., 1994, *Warnings & Safety Instructions* (Ann Arbor: Fuller Technical).

O'CONNOR, C.J. and LIRTZMAN, S.I., 1984, *Handbook of Chemical Industry Labeling* (Park Ridge: Noyes).

US DEPARTMENT OF LABOR, OCCUPATIONAL HEALTH AND SAFETY ADMINISTRATION. 29 CFR 1910 *General Industry Standards*; 29 CFR 1926 *Construction Standards*.

US DEPARTMENT OF TRANSPORTATION. "Hazardous Materials Transportation Regulations", 49 CFR 170 *et seq*.

Interactive Speech Technology

C. Baber

University of Birmingham, UK

1. INTRODUCTION

Interactive speech technology is that collection of technologies which permit humans and computers to communicate through speech. In this entry, attention will be given to speech recognition systems (see also speech-based displays). This area has been growing apace for many decades and has a reached a stage where it has confidently moved from academia to industry (this is not to deny that industry has often a played crucial role in developing speech technology, but to point to the fact that interactive speech technology is now commercially viable). For ergonomists, the issues relating to speaking to a computer, as opposed to using some other form of control device, are only now beginning to surface. As we see more speech-based applications on the market, it is likely that the question of how best to design interactive speech technology for human use will become paramount.

2. TECHNOLOGY AND APPLICATIONS

Over the past few years, the technology used in speech recognition systems has improved dramatically. In 1991, most commercial products were speaker-dependent, i.e. you had to undertake rather laborious training of the speech recognizer to get it to recognize your way of speaking and had to train all the words you wanted to use before you could use it. In 1998, most commercial products are speaker-independent, or at least speaker-adaptive. This means that the speech recognizer comes with a vocabulary already in place and with patterns of speech which it can try to match against your speech. The speaker-adaptive systems modify stored patterns to get a better fit with your speech as you continue to use the recognizer. Not only has this aspect of the technology changed, but also the manner in which patterns are created and stored has developed to a point where most systems use some form of statistical pattern matching algorithm, rather than simple template matching. The upshot of this development is that speech recognizers are able to spot boundaries between words and recognize individual words from a stream of speech; of course, users still need to speak clearly and not mumble, but they no longer need to carefully enunciate each word and pause before they say the next. Some researchers have suggested that the capability of speech recognition algorithms are beginning to reach a plateau.

Typically, speech recognizers are given a score out of a hundred to indicate their performance. This score relates the number of words which have been correctly recognized out of each hundred words spoken. Thus, a score of 95% indicates an error rate of 5%. Generally, speech recognition systems continue to work most effectively with phonetically balanced vocabularies. A further measure to note at this point is perplexity, i.e. the number of possible vocabulary items at any given point in the dialogue. The greater the number of alternative words which can be spoken, the higher the perplexity, and the greater the possibility of misrecognition. Various types of syntax have been applied to limit perplexity in vocabularies.

We have also seen changes in the proposed uses of the technology over the past decade. In 1991, there was much interest in using speech for industrial data entry or military command and control applications. While these applications are still receiving interest, they have been superseded by the rapid expansion in telephone-based speech systems. Early telephone-based systems relied on very small vocabularies, such as "yes" and "no", and tortuous dialogues that involved listening to a long string of menu options before emphatically declaring "yes" to the required option. In contemporary systems, the vocabulary is gradually expanding and the dialogues are becoming more human-like (although whether one wants a human-like conversation with a computer remains a moot point).

3. FEEDBACK

Feedback can be thought of as reactive, instrumental, and operational. Reactive feedback refers to the feedback a person receives when operating a control, i.e. when a button is pressed, it moves and makes a clicking sound. Unfortunately, there is no reactive feedback inherent in speaking into a microphone. How is one to know that the microphone is "live", or that one is speaking loudly enough, or that there is no noise or distortion in the connection between microphone and speech recognizer? This means that some form of artificial feedback is required. For example, auditory beeps were used to pace speakers using speaker-dependent systems in the 1980s. The beep would indicate that a word had been recognized and the next word could be spoken; the general consensus was that this helped first-time users but very quickly became irritating. With the development of connected-word recognizers, the pacing beep is now obsolete. However, the problem of how to provide users with reactive feedback remains. Typically, this problem is circumvented by using instrumental feedback as reactive feedback, i.e. reactive feedback is provided by the fact the recognizer is working, which can be confusing to the user and can make it difficult to determine the cause of false rejection errors.

Instrumental feedback concerns the performance of the recognizer itself and allows the user to answer the question, has the word I have just said been recognized? Instrumental feedback can be provided to users of a speech recognizer at various points in the interaction: the most obvious point will be after each individual subunit of dialogue (word, digit, letter) or after each complete dialogue unit (command, phrase, digit string, etc.). Designers faced with limited screen space, or with a purely audio channel, might be tempted to provide feedback on a word-for-word basis. However, this can introduce a very high working memory load for the user and will require recognition errors to be corrected as soon as they occur, which will disrupt the production of a command and hence be undesirable. It has been suggested that instrumental feedback should be provided at the end of a complete dialogue unit (Hapeshi and Jones 1989), but this introduces a further problem for error handling: if the user is forced to wait until the end of a dialogue unit before receiving any feedback, then it is difficult to decide which of the words has been misrecognized, especially if a substitution error has occurred. A solution to these problems is to provide a text window for the complete dialogue unit, but to display each subunit as it is

recognized. This is found to be the most acceptable form of textual feedback, and the one which gives the best user performance (Schurick, Williges, and Maynard 1985).

Operational feedback relates to system functioning; for example, if one speaks the command "Close file" and the file on the screen closes, then the system has correctly recognized and responded to the command. If the wrong operation is performed, then an error has occurred. However, this makes error correction somewhat laborious; the wrong operation has to be undone and the correct action repeated. Baber *et al.* (1992) demonstrate that operational feedback can also take the form of a symbol, representing the command, appearing adjacent to the object being controlled, e.g. an arrow icon adjacent to a valve on a process control display. If the symbol fails to appear, if it appears elsewhere on the display, or if the wrong symbol appears, then a recognition error can be identified.

4. ERROR CORRECTION

People assume that human speech is error-free, and are not particularly forgiving of errors in speech recognition. However, it is doubtful whether human speech really is free of error so much as overlain with a well-established set of rules and procedures of correcting and otherwise handling error. It is proposed that a peculiar problem with interactive speech systems is that the user must follow quite formal and rigid rules of dialogue that do not currently have the flexibility of human conversation, and which can become frustrating if these "rules" appear to be broken by the computer. For example, having to speak words from a fairly restricted vocabulary set at a slower than normal pace requires a different style of speech than conversation, and there does not appear to be any recognized "rules" for handling errors in this style of speech. Some systems employ specific error correction dialogues, such as "comma" (a different word is recognized) — "No", "Back", "comma" — while other systems require the user to repeat the misrecognized word (although this can prove difficult in connected word recognizers). Baber and Hone (1993) demonstrate that selection of error correction dialogue can be related the recognition accuracy, i.e. if accuracy exceeds 95%, then simple repetition will suffice, but if accuracy falls below 90% then some form of error correction dialogue might be required.

It has become customary to divide recognition errors into three types: rejection, insertion, and substitution. Rejection errors occur when a legal vocabulary item is spoken by the user and the ASR device does not respond. This suggests a problem in the communication between user and device. Brown and Vosburgh (1989) found that rejection errors accounted for between 2–3% of recognition errors in their studies. Thus, while rejection errors can occur, they are rare and can often be dealt with by improving communications. The next type of error, insertion, occurs when spurious noise, originating from either the user or the environment, is recognized as a legal vocabulary item. Brown and Vosburgh (1989) found that insertion errors accounted for between 5–6% of recognition errors. This problem could be reduced using a number of techniques, such as altering microphone position, providing some means of masking for ambient noise, performing enrolment in similar ambient noise levels to task performance, etc. The final type of error, substitution, occurs when a legal vocabulary item is spoken by the user and

the device "recognizes" a different item. This type of error accounts for over 90% of recognition errors (Brown and Vosburgh 1989), and is the main focus of efforts to develop error correction techniques. As the most common form of recognition error will be substitution error, there has been much interest in the design of dialogues which will either reduce the likelihood of substitution — for instance, by employing a vocabulary set of phonetically distinct words or through the inclusion of error correction commands.

5. VOCABULARY

The issue of perplexity was discussed in relation to technology. One approach to dealing with perplexity is simply to limit the number of words that a speaker can use. A feature of early applications was the limited choice of words offered to users, e.g. users might be restricted to using the digits 0–9 and "yes" and "no". While this limitation can allow some operations, it necessarily restricts functionality and leads to devices operating in what might be termed "Spanish Inquisition" mode with users having to answer a long string of questions before finding the information they require. It is apparent that, even with questions designed to elicit "yes" or "no", some users will reply with other words (Baber et al., 1997). Thus, using a restricted vocabulary with apparently unambiguous questions need not eliminate out-task vocabulary. An alternative approach is to develop an exhaustive vocabulary, containing all the words which people are likely to use. However, increasing vocabulary could return us to the problem of perplexity (without a sound syntax to support such a vocabulary). Rather than increase vocabulary size, one could employ a form of "word-spotting", i.e. the speech recognizer is "tuned" to search for specific words in the incoming speech. The implications of this approach are that appropriate, relatively unambiguous keywords can be defined and that people can be encouraged to limit their extraneous speech. A recent trend in large vocabulary systems has been to move from word spotting to topic spotting, i.e. rather than simply "listen" for particular words, the recognizer is able to detect keywords associated with a given topic and can call upon appropriate subsets of vocabulary pertaining to that topic.

6. WORKLOAD AND STRESS

One of the key issues for applications of automatic speech recognition is the effects that different environments may have on human speech production. As the technology becomes efficient enough to be considered for deployment in situations which were previously considered "adverse", then it becomes important to consider how workload and stress could influence performance. Baber (1995) reported a series of studies investigating the effects of workload and stress on speech production and proposed the following problems. Given the tendency to revert to well-learned behaviors under stress, people may attempt to structure their utterances without due regard for the syntax. This problem could also manifest itself in speech errors, which involve violation of vocabulary constraints in using speech recognition systems, where users revert to more familiar words and phrases. If the quantity of information in feedback increases, people will be more likely to treat this in general rather than specific terms. This will lead to an increase in the likelihood of user error (i.e. not responding to erroneous feedback). People vary their response strategy in

terms of task load, and if the number of tasks a person performs increases, there will be a tendency to narrow attention on specific cues, which may mean that feedback is treated in general rather than specific terms. This problem will be exacerbated if feedback competes with the primary task modality. In terms of speech production, increased demands lead to an increase in speech rate. Furthermore, as demand increases, people will reduce the amount of speech they utter, either by saying nothing or "fitting" their speech to the demands of the situation.

Many of these problems are typical of other forms of human performance under stress, but serve to highlight some of the problems peculiar to speech. It is proposed that, while some of the problems can be solved through technological innovation, many of the fundamental problems relating to speech under stress require ergonomic solution.

REFERENCES

BABER, C., 1995, The effects of workload on the use of speech recognition systems, *Proceedings NATO/ESCA Workshop on Speech under Stress,* Aix-en-Provence: ESCA, 75–82.

BABER, C., USHER, D.M., STAMMERS, R.B. and TAYLOR, R.G., 1992, Feedback requirements for ASR in the process control room. *International Journal of Man–Machine Studies,* 37, 703–19.

BABER, C. and HONE, K.S., 1993, Modelling error recovery and repair in automatic speech recognition. *International Journal of Man-Machine Studies*, 39, 495–515.

BABER, C., JOHNSON, G.I. and CLEAVER, D., 1997, Factors affecting users' choice of words in speech-based interaction with public technology. *International Journal of Speech Technology*, 2, 45–60.

BROWN, N.R. and VOSBURGH, A.M., 1989, Evaluating the accuracy of a large vocabulary speech recognition system. *Proceedings of the 33rd Annual Meeting of the Human Factors Society*, Santa Monica, CA: Human Factors Society.

DILLON, T.W., NORCIO, A.F., and DeHAEMER, M.J., 1993, Spoken language interaction: Effects of vocabulary size and experience of user efficiency and acceptability. In G. Salvendy and M.J. Smith (eds) *Human–Computer Interaction: Software and Hardware Interfaces* (Amsterdam: Elsevier), pp. 140–5.

HAPESHI, K. and JONES, D.M., 1989, The ergonomics of automatic speech recognition. *International Review of Ergonomics*, 2, 251–90.

SCHURICK, J.M., WILLIGES, B.H., and MAYNARD, J.F., 1985, User feedback requirements in automatic speech recognition, *Ergonomics*, 28, 1534–55.

Internet and the World Wide Web

P. R. Vora

US WEST, 1515 Arapahoe Street, Denver, CO 80202, USA

1. HISTORY

The Internet — "network of networks" — grew out of an initiative of the US Department of Defense in the late 1960s. The motivation for the creation of the network (originally named ARPAnet after the government's Advanced Research Projects Agency) was to help scientists and researchers share scarce and physically remote computing resources and information.

The Internet gave rise to several applications to help users communicate with each other, to find information and to share resources on the Internet. These include e-mail, FTP (File Transfer Protocol), Telnet, Usenet, archie, Gopher and several others. Excluding e-mail, despite their usefulness, most other applications were not very "usable"; they required at least some working understanding of UNIX –the most common operating system on the Internet. However, with Gopher, developed at the University of Minnesota and named after its mascot, which provided a menu-based interface to organize and access information, the Internet became a little more usable.

Although Gopher made information access easier, several problems remained with the use of the Internet because of its size and diversity:

- users needed to use several different applications to access the Internet resources;
- users could not see the content before downloading a file to their local computer and viewing it; and
- there was no simple way to get to the resources related to the current piece of information.

The World-Wide Web (or the WWW or the Web), a hypertext-based application, offered solutions to all of these problems and made itself a "killer app" for the Internet.

Work on the Web began in March 1989, when Tim Berners-Lee of CERN, the European Laboratory of Particle Physics, proposed a project to be used as a means of exchanging information and transporting research and ideas throughout the organization. CERN and MIT (Massachusetts Institute of Technology) now share the Web project.

A common misconception is that the Internet and the Web are the same. It is not so. The Internet is the infrastructure — it is the physical medium used to transport data. The Web is a collection of protocols and standards used to access the data available on the Internet. Specifically, the Web is defined by three standards:

- URI (Uniform Resource Identifiers): a method to specify the location of a resource on the Internet;
- HTTP (HyperText Transport Protocol): a protocol for transferring hypertext-based information over the Internet; and
- HTML (HyperText Markup Language): a language of interchange for hypertext on the Web.

2. REASONS FOR THE WEB'S SUCCESS

Although standards and protocols existed before the Web, they were designed as "closed" systems; they did not interact with other Internet applications and protocols in a seamless way. In addition to this interoperability, the following characteristics contributed to the success of the Web:

- Open and extensible architecture: the Web's open architecture is easily extensible and can accept advances in technology, including new protocols, object types, and data formats. This allows the Web to evolve with users' needs.
- Consistent access interface: the Web, through its simple hypertext link-based navigational model, provides users a consistent means to access a variety of media in a simplified fashion. And all documents, whether real or virtual (i.e. dynamically generated), look similar to the user.
- Universal readership and platform independence: information on the Web is accessible from any type of computer, anywhere, as long as it is connected to the Internet and has a Web browser. This also makes it easy and almost instantaneous to distribute information on the Web.
- Ease of publishing and universal authorship: because of the simplicity of writing HTML and relatively low cost of having a Web presence, almost any person can publish information on the Web.

3. USABILITY ISSUES WITH THE WEB

Despite the seductive benefits of the Web outlined above, many Web sites have failed to realize their potential. Some of the common problems that compromise usability on the Web are discussed below.

3.1. Slow Response Time or Download Speed

In the GVU's Ninth WWW User Survey (GVU 1998), across all groups of users, taking too long to download pages (i.e. "speed") was the most commonly experienced problem with the Web (64.8% of respondents). In the same survey, ~53% of respondents reported that they had left a Web site simply because the site was too slow. The percentage of respondents reporting the problem of "speed" has been consistent over the last two surveys — 63% in the Eighth and 66% in the Seventh survey respectively.

3.2. Linkrot or Dangling Links

The next most frequently cited problem on the Web is a growing one — dangling links (or *linkrot*). Although solutions for dealing with broken links are well-known to Web designers, i.e. automated link-checking, redirecting URL, comprehensible error pages, redirecting broken links to a search page and so forth — most sites do not seem to employ these techniques. And the users are still presented with the "404-File Not Found" errors. This problem can be attributed to the embedded nature of links (as HTML markup) on the Web pages. Moving (or deleting) a page involves checking all references to this document within all other documents. For the Web this would involve checking all other documents in the world which, obviously, is not feasible. The result is linkrot.

3.3. Inefficient and Ineffective Navigation

The links provided on the Web are often cumbersome, irrelevant, or trivial. This problem is compounded by inconsistencies in

indicating link anchors: what aspects of words or phrases should be marked as links have not been addressed. There are several occasions, where instead of linking the words describing the destination Web page, hypertext links are made to the action of clicking (e.g. "Click here for Information on Usability" instead of "Information on Usability"). Another common problem is use of the links from graphic images, which do not afford the action of clicking and readers are often unsure as to which graphics are linked on a Web page.

Navigation problems on the Web are also caused by the single-window design of Web browsers, which preclude the preservation of context that could be achieved if both anchor and destination were displayed. Although *frames*, which permit partitioning a Web browser window into independent sections, have provided a partial solution to this problem, they have caused problems related to bookmarking, printing and accessibility.

3.4. Difficulty in Finding Information

Granted, the link-based navigational model of the Web is easy to understand. However, finding relevant information on the Web is neither easy nor effective by simply browsing the pre-authored hypertext links. To aid in finding information on the Web, several Web sites now maintain indexes that categorize and organize information by subject (e.g. Yahoo!, http://www.yahoo.com). Although the categorization, done by humans for such indexes, is helpful, their scope is limited considering the volume of data that is made available daily on the Web. Search engines, on the other hand, are more exhaustive in their coverage (e.g. Alta Vista, www.altavista.digital.com; Lycos, www. lycos.com; InfoSeek, www.infoseek.com). However, Web users who type in a search request are often presented with thousands of "hits," many of which contain references to irrelevant Web sites while leaving out others that hold important material.

3.5. Poor Content Quality

The Web's strength of universal authorship is also its weakness. There are no intermediaries on the Web that can validate the usefulness, correctness and quality of the content on the Web. Furthermore, a considerable amount of information on the Web is either outdated or not updated; such Web sites are commonly referred to as "cobweb" sites. Without a clear indication of the life of the content or a well-defined expiration date, the useful content gets lost and becomes more difficult to identify.

4. ACCESSIBILITY

Although universal readership is an important benefit of the Web, a significant portion of the users are unable to access all the information available on the Web — in particular, those with visual, hearing, and physical impairments. In this respect, the Web has tended to become exclusive or excluding. As the Web has embraced "new" technologies, it has pushed assistive/adaptive devices beyond their current capability. And it has made it difficult for people with disabilities to participate as equals in work, education and leisure.

5. CAUSES OF USABILITY PROBLEMS ON THE WEB

5.1. Perceived Need to Have a Web Presence

Although there are several reasons for unusable designs on the Web, many usability problems on the Web can be attributed to "Web-site Envy" or the perceived need to have an immediate Web presence. Many organizations rush to be on the Web without spending sufficient time to understand the needs of potential users or to analyze or judge the scope of the content to be put on the Web. This rush to have a Web presence causes them to disregard any methodology or guidelines for designing and organizing information on the Web.

5.2. Use, Activity and Poor Design

Both the use and the activity on the Internet have led to severe degradation in response times. According to the Internet Traffic Report (www.internettrafficreport.com) on a Sunday 14 February 1999 at 18:45 GMT, the Global Internet Traffic Index was 52 (on a scale of 0–100, with 100 being the best response time); this is usually much lower during the week. Not all the blame can be placed on the traffic on the Internet, however. Compared with the text-heavy pages on the Web in its infancy, Web pages have become bandwidth-hungry with inclusion of images, animations, multimedia and scripts, which take extra time to download.

5.3. No Metadata: Impoverished Content

In contrast to human indexers, automated programs used by search engines have difficulty identifying characteristics of a document such as its overall theme or its genre–whether it is a poem or a play, or even an advertisement. The Web, moreover, lacks standards that would facilitate automated indexing. As a result, documents on the Web are not structured so as to allow programs to reliably extract the metadata information such as author, date of publication, length of text, and subject matter that a human indexer might find simply through a cursory inspection.

5.4. Usability of Web Browsers: Poor Support of Standards

Although graphical Web browsers have made the Web more accessible, several usability issues need to be addressed in their design — in particular, their non-compliance with the standards. With the Web browsers competing for the market share, very little attention is paid to standards. Even dominant browsers, like Netscape Navigator and Internet Explorer, which claim to support the new HTML 4.0 standard in their recent releases, do not support them completely and reliably. The presentation of Web pages is, therefore, unpredictable and requires Web page designers to either spend more time testing their Web pages with a variety of browsers, design their Web pages for more than one type of browser, or design for the lowest common denominator.

5.5. Usability of Web Authoring Tools

The current crop of WYSIWYG Web page authoring tools is also responsible for poor design. Many WYSIWYG tools of today not only generate excessively redundant and unnecessary code, but also place more emphasis on the presentation and consequently add unnecessary graphics to remain true to the user-generated layout. Both these practices add to the download time for Web

pages (Vora 1998b). Finally, emphasizing presentation, a majority of them do not offer good support for designing accessible Web pages.

5.6. Use of Opaque Technologies

The technology is no longer transparent to users who now not only need to understand terms such as plug-ins, helper applications, Shockwave, PDF, JavaScript, Java, ActiveX and so forth, but also are able to install appropriate applications on their machines to take full advantage of the Web.

6. SOLUTIONS TO THE USABILITY PROBLEMS ON THE WEB

6.1. Follow a Human Factors Methodology

Although the Web technology makes evolution, enhancement or changes to content relatively easy, Web site design should not proceed in a haphazard manner. A human factors methodology for designing Web that includes *planning, analysis, design and development, testing, implementation* and *maintenance* should be followed for designing usable Web sites (Vora 1998).

A particularly important stage is usability testing, where the design assumptions and decisions are tested with representative. Relative simplicity of the Web technology also makes it possible to conduct remote usability evaluations with a distributed population of users.

Another critical stage is maintenance, where the challenge is to provide current and updated content while maintaining the integrity of the Web site. Maintenance also requires that Web site usage trends are analyzed and changes are made to accommodate users' needs. Finally, it is important to keep up with the advances in technology, and implement them if it improves usability and offers a richer interaction environment for the users.

6.2. Understand the Needs of Content Owners and Authors

When designing Web sites the focus should not be exclusively on the users; the needs of authors and content owners should be addressed as well. This is necessary because the authors and owners are responsible for ensuring ongoing usability of a Web site after its deployment. Knowing who the authors are, how knowledgeable they are about computers and Web technology, and where they are located (at one location or at different locations) are all important to determine whether any training is necessary, and what approach is best for managing Web sites.

6.3. Separate Structure, Presentation and Semantics of the Web Content

Although originally intended as a simple mark-up language, HTML has become a mish-mash of markup and presentation language. Changes are happening in the Web world in the form of Cascading Style Sheets (CSS) and Extensible Markup Language (XML) to complement HTML, which not only address the current concerns with HTML, but also help enhance the user experience when interacting with the Web.

CSS is a language for describing layout and graphical style. CSS can be used to control colors, white space, multicolumn layouts, fonts and more, allowing designers to leverage document styles without modifying each page. CSS is uniquely powerful because it defines the look-and-feel for a document by annotating its semantic structure.

XML, on the other hand, allows creation of documents using tags beyond those specified in the HTML standard. This allows development of tag-sets that are more meaningful in the context of the subject-matter and with an associated style sheet language (i.e. Extensible Style Language [XSL]) offer better control over presentation. In the future, XML is likely to be the language of choice for building other languages so as to make interoperability and information sharing possible. In fact, several XML-based applications have already emerged to offer a variety of functionality including:

- Multimedia functionality, SMIL and HTML+TIME.
- Describing specialized languages, Mathematical Markup Language (MathML), Chemical Markup Language (CML).
- Distributing and updating software, Open Software Description (OSD).
- Electronic Commerce, OBI (Open Buying on the Internet).

However, CSS and XML are not universally and reliably implemented in the browsers. It will be at least a few years when their use becomes common.

7. FUTURE OF THE INTERNET AND THE WEB AND ASSOCIATED HUMAN FACTORS ISSUES

Undoubtedly, the Internet and the Web will continue its exponential growth with the potential of exacerbating the current human factors issues related to responsiveness, interoperability, and interactivity. Although some of the future technologies will address these concerns, the applications of the Web in different domains will give rise to different kinds of usability problems. Some of these upcoming technologies and associated human factors issues are discussed below.

7.1. Higher Access Speeds: Still Higher Bandwidth Demands

With the technologies such as cable modems, Digital Subscriber Line (DSL) and satellite-based network delivery, the speeds that the users will have to access the Internet will increase. However, user experience is likely to remain the same because the content providers will demand still higher bandwidth by packaging more multimedia content.

7.2. Narrowcasting or "Personalizing" the Web Content

It is almost a cliché to talk about "information glut" on the Web. To manage access to information and to help users "filter" information, an emerging solution is narrowcasting or personalizing the content on the Web. Rather than providing the information, which appears the same to all the users, the users will be increasingly offered the option of customizing the content to meet their unique needs and interests. The challenge will be to determine what the users want — that is, identifying the "user profile." Currently, two approaches are used to solve the problem: by asking the users to specify their interests and use that information to customize the content and by "watching" their interaction with site and extracting a profile from the gathered information. In the future, personalization will become essential for usability and the burden will be placed on the information

providers to separate relevant and irrelevant information for the users.

7.3. E-Commerce: Addressing Security, Privacy and Trust

Considering the growth estimates of electronic commerce by Forrester Research (1998) — from $48 billion in 1998 to $1.3 trillion by 2003 in the USA and $3.2 trillion globally — the distinguishing characteristics of successful presence for business on the Internet will be ease of use, customer service and efficiency.

A successful user experience for electronic commerce will make it imperative that the user concerns about security, privacy and trust on the Internet are mitigated. In the GVU's 9th survey, the reasons for not making purchases over the Web were:

- not trusting that credit card number will be secure (39.1%);
- not being able to judge the quality of the product (39.3%); and
- not trusting that personal information will be kept private (26.9%).

The World-Wide Web Consortium (W3C), the industry, and the government are working together to address these issues with minimal legislation. In May 1998, the W3C released a public working draft of the Platform for Privacy Preferences (P3P) to provide a framework for informed online interactions (http://www.w3.org/P3P/). P3P allows Web sites to disclose their privacy practices by outlining which data they collect from users, what they use the data for, whether the data will be shared with other parties, and if the data will be shared in an identifiable manner. At the same time, it allows the users to express their own privacy preferences by outlining what information they would like to be made available and in what context. When a disagreement arises between a site's privacy practices and the user's privacy preferences, user is prompted to resolve the negotiation and offered other options to proceed.

Although P3P allows a mutually convenient way for the companies and the consumers to transact business over the Web, the issue of trust, however, still remains. That is, how do users know that a site is not misrepresenting its privacy practices? A number of services such as TRUSTe (http://www.truste.org) and BBBOnline (http://www.bbbonline.org) will provide the assurance that Web site's policies accurately reflect their practices. These assurances are manifested on the site as visual labels ("seal of trusts") on Web sites.

As more privacy safeguards are deployed, the consumer confidence in the Internet should increase. However, in addition to the technology, the design question about "trustworthiness" of a site is still open: What design elements make one site more "trustworthy" than the other?

REFERENCES

GVU, 1998, Graphic, visualization, and usability center (GVU), Georgia Institute of Technology. *GVU's 9th WWW User Survey*. [http://www.gvu.gatech.edu/user_surveys/survey-1998-04/]

VORA, P.R., 1998a. Human Factors methodology for designing Web sites. In C. Forsythe, E. Grose and J. Ratner (eds) *Human Factors and Web Development* (Malwah: Lawrence Erlbaum Associates), 153–72.

VORA, P.R., 1998b. Designing for the Web. A survey. *Interactions*, 3, 13–30.

WORLD-WIDE WEB CONSORTIUM (W3C) http://www.w3.org

Knowledge management in HCI design

C. Stephanidis and D. Akoumianakis

Institute of Computer Science, Foundation for Research and Technology-Hellas, Science and Technology Park of Crete, Heraklion, Crete, GR-71110, Greece

1. INTRODUCTION

In the relatively short history of Human–Computer Interaction (HCI), several approaches have been developed to account for the design of computer-based interactive artifacts. The most prominent have been the paradigms of Human Factors evaluation, cognitive science, and more recently, user-centered design (UCD). (It should be noted that in recent years there has been a growing interest in more developmental approaches to HCI design (Carroll 1991), such as activity theory, distributed cognition, situated action models, etc. As these approaches still seek their ultimate formulation in the field of HCI, they will not be addressed here. However, their potential value as new prescriptive frameworks for designing computer-mediated human activities in an Information Society is fully acknowledged.) Human Factors evaluation and cognitive science have been criticized with respect to the underpinning scientific ground and the means (e.g. instruments, methods and tools) for achieving a specific set of objectives. The literature in the field of HCI reports many on-going debates (e.g. Nardi 1996), and offers an insight into the diversity that characterizes the field. The essence of the critique is that, on the one hand, traditional large-scale user testing (e.g. the Human Factors evaluation paradigm) is both costly and suboptimal, while on the other hand any attempt for a generalized model or theory (e.g. the information processing psychology approach of cognitive science) for designing interactions between humans and machines is simply not feasible.

Instead, techniques are needed that focus on the requirements of end users and provide early feedback to design, to reduce the cost of designing defects and to meet specific usability objectives. The term that was coined to this approach was user-centered design and the first comprehensive collection of papers on the topic appeared in Norman and Draper (1986). The normative perspective of UCD is to fulfil the need for usability by providing techniques that foster tight design–evaluation feedback loops, iterative prototyping, early design input, end-user feedback, etc. In recent years, due to the compelling need to cost-justify usability throughout a product's lifecycle, the field moved towards a variety of UCD techniques (Figure 1), and the consolidated experience gave rise to several computer-based tools to facilitate the process.

Here the major categories of these tools are described to identify their underpinning assumptions, to assess their contribution to UCD in HCI and to elaborate some of the requirements that future HCI design support environments need to address and satisfy. The following section classifies the state of the art in computer-aided HCI design into six categories of tools. These categories are representative of the three prominent HCI design perspectives already mentioned. Then, a retrospective
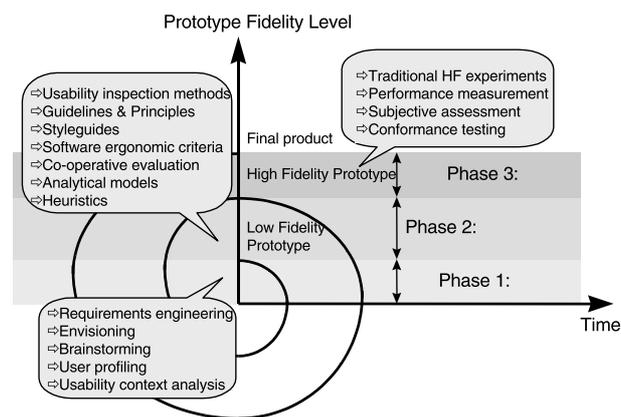
Figure 1. User-centered design — techniques, methods and tools.

account of these tools is provided to pinpoint shortcomings and limitations. The next section formulates the proposal of knowledge-intensive design environments (KIDE) as a new category of tool aiming to address the requirements of interaction design in the emerging Information Society. Subsequently, a brief case study is presented as an example of a design environment exhibiting the properties of KIDE.

2. COMPUTATIONAL SUPPORT FOR HCI DESIGN

The efforts to provide computational support for HCI design have emerged from the compelling need for cost-justified usability and more effective means to an end. This section focuses on how the consolidated wisdom has been embedded into computational tools for design support. In this context, we will be concerned with the categories of computer-supported design environments that have emerged to provide support for the various stages of the user interface software design process. (It is important to mention that this account will cover only software systems that have been developed specifically for user interface design support tasks, i.e. tools supporting prototyping and/or development of user interfaces, or design as applied in software; other engineering disciplines (e.g. architecture, civil engineering) will not be addressed.) It should be noted that the prime intention is to depict a road map for user interface design support software, rather than to provide an exhaustive description of specific systems.

2.1. Categories of Tools

Figure 2 depicts schematically the major categories of tools that are currently available. (For a more elaborate account of each class of tool, see Stephanidis and Akoumianakis 1999.) The first category comprises *tools for working with guidelines*. Such tools are primarily intended to substitute the paper-based medium for propagating guidelines into design, by tools which either provide on-line hypertext access to guideline reference manuals; or, integrate a subset of relevant guidelines into knowledge bases that could subsequently augment the design phase; or, automate the evaluation of certain components of a user interface according to recommendations resulting from general, or context specific guidelines. A more recent development in the above line of work, is the effort to advance *tools for experience-based usability guidelines*. This approach extends the scope of tools for working with
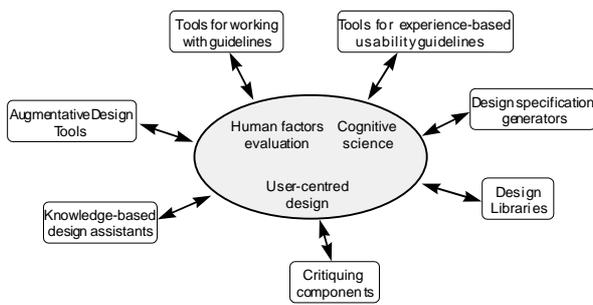
Figure 2. Prominent categories of user interface design support tools.

guidelines to facilitate depositing and retrieval of design experiences and the construction of "living" design repositories.

Critiquing systems and components aim to empower designers when designing low level details of a user interface. In user interface design, critiquing has been employed to facilitate an exchange of reasoned opinions about a product, or action, which triggers further reflection on, or changes to, the artifact being designed (Fisher *et al*. 1991). In the recent literature, this notion of critique has given rise to dedicated and domain-specific implementations of critiquing components embedded into design environments (e.g. Fisher *et al*. 1991).

Knowledge-based design assistants aim to bring Human Factors knowledge closer to the development process of a user interface, by integrating a collection of ergonomic design rules with user interface management systems. In such efforts the content and scope of the Human Factors design input provided is limited to heuristic or style rules. Thus, the level of design assistance offered is limited to identifying, and sometimes automatically correcting, faults in the design that can be traced through the available rules.

Design libraries have emerged from the effort to provide support for design re-use. They constitute a relatively recent development in the area of HCI design and have been closely related with work in the area of design augmentation and rationale. Despite the recognized potential of this approach, the relevant literature reports only very few efforts dedicated explicitly to this line of activity (e.g. Grammenos *et al*. 1999).

Design specification generators have their roots in the model-based paradigm for user interface development (Szekely 1996). The common characteristic of such tools is their reliance upon models of design knowledge (e.g. user, task, environment) to derive specifications and some times generate the code of the interactive components of an application. Existing systems, typically offer some form of modeling tools to assist designers or developers in building the models. These tools range in scope, and include text editors for building textual specifications of models, form-based tools to create and edit model elements, specialized graphical editors, etc.

The categories of tools reviewed so far, are commonly characterized by the fact that they are artifact oriented. In other words, they encapsulate design knowledge (in the form of rules, heuristics or alternative representations) that relates to a particular artifact. As a result, they offer little support for assessing the underlying reasoning, or the process that gave rise to a given artifact. In addition, they offer limited support for collaborative design. To account for these shortcomings, but also as a result of

a more fundamental shift towards analytical design support, recent efforts have been exploring *argumentative design tools* towards design rationale capture and retrieval (Carroll and Moran 1996). Design rationale in the context of HCI refers to a cluster of techniques aiming to provide rationalization for design. In the recent past, several efforts have been undertaken to support design rationale capture and retrieval in HCI (Carroll *et al*. 1996).

2.2. Retrospective

The current generation of design support tools exhibits several shortcomings that impede their wider use and adoption by practitioners. In certain cases, these shortcomings are related to the HCI design strand into which the tool is rooted and therefore they cannot be easily remedied. In other cases, remedy can be attained by extending the capabilities of the tool to interoperate with other tools or design environments. In any case, the critique that follows does not intend to delimit the usefulness of the tools, but to illustrate the shortcomings of providing non-integrated and partial support for HCI design.

Scope of design input. A well-known and documented shortcoming of the current generation of design support tools is their insufficiency to capture and reason about context-oriented parameters. It is frequently found that the recommendation offered by a tool remains the same, irrespective of the context of use. A typical example of such tools is the systems available for working with guidelines. Their underlying model for managing the guidelines does not reflect any context-oriented attributes, thus making customization impossible. Recent developments attempt to provide a remedy for this shortcoming, by following alternative pathways. One approach is that reported in (Grammenos *et al*. 1999) where the tool allows the designer to encode the reasoning behind a particular decision as well as the constraints pertaining to a particular design case, that led to that decision. Another approach is based on compiling usability cases and presenting recommendations by correlating context-specific attributes.

Loose coupling/integration with user interface development systems. In the majority of the cases, design support tools do not interoperate with popular user interface development environments. The primary reason for this is that developers of such systems to not wish their systems to be tied to a particular user interface development environment, as this entails frequent updates to reflect versioning. On the other hand, the problem that is brought out is that the design knowledge encoded in design support tools can rarely be automatically articulated into prototypes that can be experienced by end users. As a result, design outcomes are loosely accounted during development.

Extensibility, maintenance and versioning of encoded knowledge. In currently available tools, design knowledge is typically encoded as collections of ergonomic design rules. However, there is no practical (e.g. non-programming) support for: (1) maintaining the rule base (e.g. supporting updates, identifying competing rules, coping with conflicting recommendations); (2) extending the scope of the rule base (e.g. dedicated programming functions for implementing new rules); (3) versioning of rule bases, so as to depict specific requirements of particular design cases. Once again, more recent efforts such as that described in (Grammenos *et al*. 1999) explicitly account for these shortcomings by offering techniques to automatically identify and resolve conflicts, as well

as extensible Application Programming Interfaces for the transparent integration of new rules once they have been encoded into the rule base.

Corporate support. Another important shortcoming of existing tools, is their lack of supporting corporate practices. This does not only relate to customizing a guidelines reference manual, but also to developing domain-specific styleguides and offering organization-wide support for appropriating the recommendations of these styleguides. In other words, it is not possible for an organization to encode a corporate style guide into the representation supported by an existing tool and subsequently provide this representation as an internal company standard to be observed by different business units and development sections. As a result, it is practically impossible to support consistency and persistency in the use and application of Human Factors knowledge.

Reporting. Reporting design defects and alternative solutions is another important issue that needs to be supported by design support tools, if they are to provide an effective and efficient medium for integrating good design practices, into software design and management. To this end, designers need to be able to effectively document and report the results of their assessments so that they can be communicated to developers, management and other stakeholders.

3. PERSISTENT HUMAN FACTORS KNOWLEDGE MANAGEMENT IN HCI DESIGN

To overcome some of the above shortcomings and to meet the emerging requirements of the Information Society, it is necessary to reconsider several aspects of HCI design and the tools through which it is conducted and practiced. To this end, this section describes some of the critical requirements that should characterize the next generation of design support tools, and provides highlights of some of our efforts towards attaining these requirements, by means of a case study.

3.1. Requirements for Computational Environments

To accomplish their tasks effectively, designers need support, which is not uniform, throughout the various phases of design. Consequently, there is no single best category of design support tool that is preferable to others. Instead, what is needed is a conscious and purposeful effort towards *knowledgeable environments* for *informed* and *rationalized* HCI design. The basic objective of such environments (Figure 3) is to provide computer-mediated instruments to bridge the gap and preserve interoperability across design activities (as performed by human designers), design perspectives (as reflected in the various approaches to design such as analytical/prescriptive, predictive, process versus artifact orientation) and the corresponding design support tools (e.g. engineering models, rational capture and retrieval, critiquing components, specifications generators, prototyping tools). Such environments should exhibit several characteristics, which are briefly described below.

First of all, they should be integrated systems of interoperable design components to provide a greater scope for design input. In this context, an integrated system implies a computational tool that bridges across alternative design regimes (e.g. descriptive/predictive, augmentative/reflective, proactive/reactive), the tools
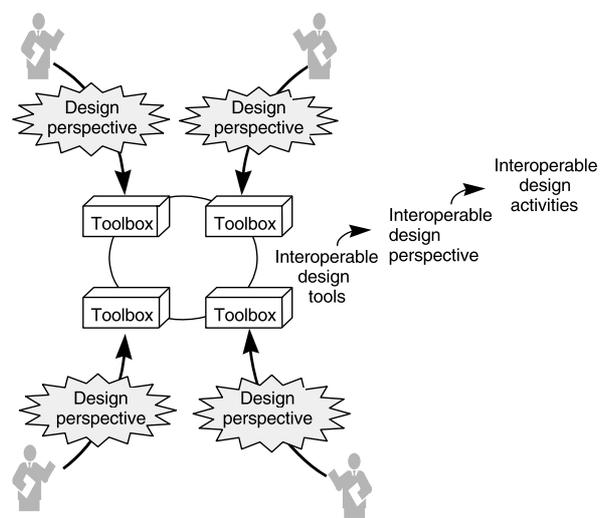

Figure 3. Bridging across design activities, perspectives and corresponding tools.

that serve them, and the humans that employ them. On the other hand, the notion of interoperable design components means that different tools (e.g. engineering models, design rationale editors, critiquing components, specification generators) facilitating alternative design perspectives (e.g. reflection, argumentation, critiquing and automation respectively), can expose data to, and receive data from, one another, so as to support the broad variety of design activities through an evolutionary and persistent computational protocol. In this manner, KIDE can provide a corporate means to effective design processes.

Second, the scope of such computational environments should not be limited to mere propagation of design knowledge, but should also facilitate design knowledge construction as related to the new virtual spaces likely to emerge and the new range of end user experiences likely to be encountered. To facilitate this task, design environments should allow for knowledge to flow and be articulated in generative manners. As information technology is used pervasively within modern organizations, it qualifies as a medium to facilitate both the flow and the generation of knowledge. A related aspect is that the type of designs that should be supported by the new generation of these tools should not be constrained to traditional interaction metaphors, such as the visual desktop. In particular, the scope of design support to be offered by computational environments should be extended beyond the conventional desktop, as in the future, it is expected that new means for interacting with computers will emerge to facilitate the novel virtualities which will increasingly predominate computer-mediated human activities.

Third, due to the iterative and evolutionary nature of HCI design, focus on the user interface artifact being designed is not sufficient. Equally important is the process through which the artifact evolves, the design rationale behind it and the collective wisdom that has been progressively accumulated and can be reused to inform future design cases. It is, therefore, important that emerging design environments provide a more analytic support for unfolding alternatives, building scenarios to envision and understand the design options, as well as for sharing and reusing past design experience and best practice. Meeting this

target, also bypasses the shortcoming on current tools to facilitate extensibility, maintenance and versioning of the encoded knowledge.

Finally, the above characteristic properties of knowledgeable design environments for HCI should be conceived within a broader organizational shift in perspective whereby the currently prevailing paper-based medium for documenting and propagating design knowledge is replaced by "living design memories." The term living design memories is used to refer to the use of technology as a novel, interactive medium for knowledge management in HCI design. To this end, knowledgeable computational environments for HCI should provide reusable, extensible and evolutionary repositories of encoded knowledge which can be used to shape and inform future designs, as well as to facilitate knowledge persistency within and across organizational units.

3.2. Case Study Overview

The above have motivated the development of a KIDE to demonstrate the applicability of the above mentioned four principles. The target application domain covered is the area of accessibility of user interfaces by disabled and elderly people.

The conceptual overview of the KIDE is depicted in Figure 4. The accessibility KIDE integrates four distinctive software tools, each offering a particular service. Thus, for instance, DESIGN-AID follows the design argumentation paradigm for analytical HCI design to allow the designer to articulate elements of a design language and accordingly augment the interaction capabilities of a toolkit. Additionally, the system provides tools and facilities for encoding and retrieving rationale for specific interaction elements, constructing scenarios, encoding argumentation, as well as for reusing past experiences and practice. ELICIT is a user modeling environment which elicits user characteristics and compiles rules regarding suitable interaction elements. ADAPT is a user interface adaptation environment which comprises several tools to facilitate the automatic compilation of a lexical specification scenario for an accessible and user-adapted interface.

Finally, CRITIC-AID embodies a theory of "error and repair" common to critiquing systems. The tool cooperates with ADAPT to check and verify correctness and consistency of adaptation decisions. This cooperation is implemented as a salient-critiquing component that is activated, as soon as a particular test is invalidated. CRITIC-AID encapsulates and implements a three-phased critiquing process. In the first phase, the tool identifies the problem. Then it compiles a representation of the problem that explains to the designer how the problem was encountered, and how it may be addressed. Finally, it repairs the problem upon the designer's instruction.

What is important to note in relation to these developments are: (1) each tool is a separate and autonomous entity; (2) the design perspectives adopted by the tools (figure 4) and their respective contributions to the design activity are radically different; (3) these tools are intended for different people (e.g. the user modeling analyst, the designer, the assistive technology expert). Despite such variation in scope and intended objectives, all tools interoperate to constitute the integrated KIDE. In this case, the basis of interoperation is design abstractions in the form of unified interaction object classes (or abstract interaction object classes). These are platform and metaphor independent interaction constructs capable of polymorphic instantiation to facilitate alternative accessible faces for different user groups (Stephanidis *et al*. 1997).

A second important property of the KIDE, as depicted in Figure 4, is its interoperation with a user interface development environment. Interoperation in this case is based on well-founded semantic protocol, making use of the unified object classes; such a protocol can be implemented through simple message-passing. (It is important to mention that the unified object classes constitute the common and shared vocabulary between the KIDE components, as well as the KIDE and the user interface development environment.) An example of the type of information that is exchanged between the accessibility KIDE and the user interface development toolkit is depicted in Figure 5. It illustrates the contents of the accessibility recommendations server which can be interrogated either prior to user interface implementation (Figure 5) to review the type of recommendations available, or alternatively, during user interface implementation, provided that the toolkit provides the required Application Programming Interface functionality.

As shown, the accessibility server compiles recommendations per attribute of interaction object classes. Additionally, it provides information on how the recommendation was compiled (e.g. which design criterion was assigned and what preference and indifference expressions gave rise to the recommendation). In this manner the designer may review the recommendation for a particular attribute, or all recommendations per dialogue state, or all recommendations which comply to the same design criterion.

From the above, it follows that the accessibility KIDE unifies not only alternative design tools (Figure 4), intended for different users and embodying different design models, but also the different perspectives on design possessed by the human users of these tools. Thus, analytical and process-oriented instruments, such as DESIGN-AID, are "glued" together with artifact-oriented design tools, such as ADAPT, and critiquing components, such as CRITIC-AID. The basis for such a unification is a common vocabulary (e.g. unified interaction object classes) and well-defined communication protocols between the tools and across the KIDE and the user interface development environment.

4. SUMMARY AND CONCLUSIONS

This article has provided an outline of prominent HCI design traditions and identified the type and scope of computational
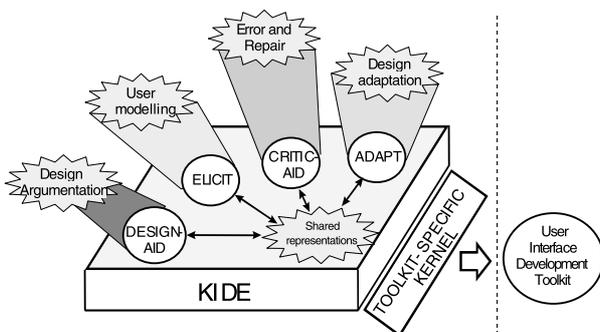

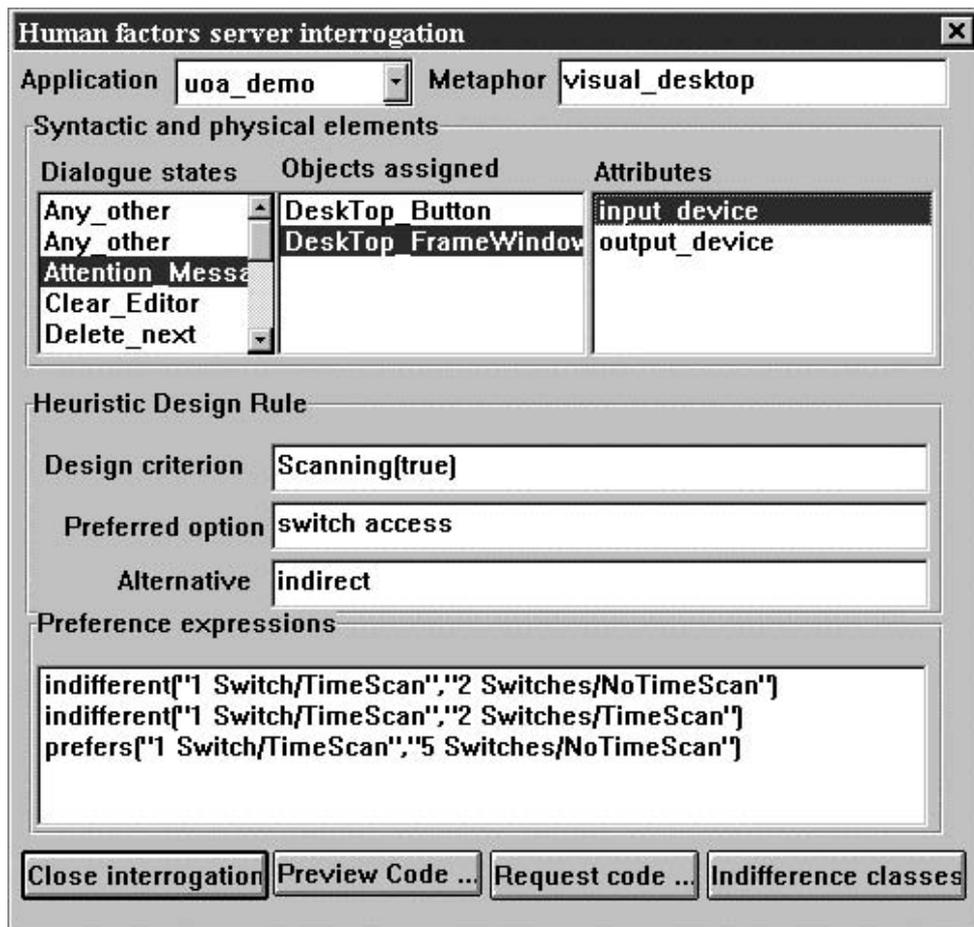**Figure 4. Conceptual overview of the accessibility KIDE.**

Figure 5. Interrogation of the accessibility recommendations server.

tools which have risen to support design activities. Additionally, a proposal for a new generation of design-support tools in HCI was advanced to facilitate the increasingly complex nature of user interface design in the context of the emerging Information Society. From the account and arguments presented in this article, we can draw a number of conclusions. First of all, it was pointed out that design support is not uniform throughout the various phases. Second, there is no specific category of tool that is best suited for the broad range of design tasks; instead there are different tools that can be used to facilitate different tasks. Third, designers will increasingly need support that exceeds the functional scope of any particular category of tool, thus the requirement for supporting environments of integrated and interoperable design components.

In conclusion, it should be pointed out that, design support tools are not a panacea for the designer, as their capability to automate parts of the design activity is constrained by the inherent complexity of the design activity. Consequently, in the future, the focus should be on computational design support tools that: (1) combine design functionality (e.g. specifications generation, exploration, reflection on past experiences, automation, reuse, and critiquing) in an integrated manner; (2) augment the capabilities of the human designer by offering suggestions, and context bound assistance; and (3) facilitate collaboration between members of the design team through tight design-evaluation feedback loops, so as to assure timely and effective design input.

REFERENCES

CARROLL, J. (ed.), 1991. *Designing Interaction* (London: Cambridge University Press).

CARROLL, J. and MORAN, T. (eds), 1996, *Design Rationale: Concepts, Techniques and Tools* (Mahwah: Lawrence Erlbaum Associates).

FISCHER, G., LEMKE, A. C., MCCALL, R. and MORCH, A. I., 1991, Making argumentation serve design. *Human–Computer Interaction*, 6, 393–419.

GRAMMENOS, D. AKOUMIANAKIS, D. and STEPHANIDIS, C., 1999, Integrated support for working with guidelines: the Sherlock Guideline Management System. *International Journal of Interacting with Computers*, special issue on "Tools for Working with Guidelines," 11.

KENT, A. and WILLIAMS, J.G. (eds), *The Encyclopaedia of Microcomputers* (New York: Marcel Dekker).

NARDI, B. (ed.), 1996, *Context and Consciousness: Activity Theory and Human Computer Interaction* (Cambridge, MA: MIT Press).

NORMAN, D. and DRAPER, P., 1986, *User-centered Design: New Perspective in Human Computer Interaction* (Hillsdale: Lawrence Erlbaum Associates).

STEPHANIDIS, C. and AKOUMIANAKIS, D., 1999, Knowledge-based interaction design. In A. Kent and J.G. Williams (eds), *The Encyclopaedia of Computer Science and Technology* (New York: Marcel Dekker, Inc.).

STEPHANIDIS, C., SAVIDIS, A. and AKOUMIANAKIS, D., 1997, Tutorial on unified interface development: tools for constructing accessible

and usable user interfaces [Tutorial no. 13]. In *International Conference on Human–Computer Interaction (HCI International '97)*. [http://www.ics.forth.gr/proj/at-hci/html/tutorials.html]

SZEKELY, P., 1996, Retrospective and challenges for the model-based interface development. In F. Bodard and J. Vanderdonckt (eds) *Conference Proceedings of 3rd Eurographics Workshop on the Design, Specification and Validation of Interactive Systems (DSV-IS'96)* (Namur: Springer).

Knowledge-based Man-modeling: Job Design Procedure (Man-modeling for Job Design)

K. Landau

Darmstadt University of Technology, Department of Ergonomics, Petersenstrasse 30, 64287 Darmstadt, Germany

1. THE NEED FOR CAD PROCEDURES IN JOB DESIGN

As technology at the workplace becomes increasingly complex, many designers of work processes are sorely tempted to regard the skills and abilities of the actual worker as being of secondary importance or even to ignore them completely. It is, however, a well-known fact that the benefits of ergonomically correct job design and job environment include not only improved worker performance and well-being, but also, over the longer term, the prevention of job-related disease.

Extensive ergonomic job design input is an essential factor in the prevention of job-related and occupational diseases. For some time now, insurance companies in the USA have postulated that ergonomic job design could reduce the incidence of musculoskeletal diseases by one-third (Snook 1978). More recent data indicate that roughly 30% of all diseases and injuries are attributable to overloading and that overloading accounts for something like 22% of all traumata in the dorsal region (Putz-Andersen and Waters 1991). Data collected by the Federal German Statistics Office also reveal that industrial accidents and occupational diseases are responsible for only a very small proportion of the health erosion attributable to work activities. The available statistics on rehabilitation measures show that patients suffering from occupational diseases and recovering from industrial accidents (not including accidents occurring on the way to and from work) accounted for only 6.5% of total rehabilitation procedures. In the majority of the remaining cases (89% of total rehabilitation procedures), it is mainly chronic diseases that are involved, e.g.

- musculoskeletal disease;
- cardiovascular disease;
- respiratory disease; and
- gastrointestinal disease.

Work stresses are, at the very least, a contributory cause to these diseases and they should therefore be classified as "job-related diseases." In the sphere of physical work it is mainly the handling of loads that can lead to increased disease susceptibility and, consequently, to frequent, long absences from work.

These data demonstrate clearly that practice-oriented methods of evaluating physical work, in particular load handling, are needed to eliminate or reduce short- and long-term health risks. It must be remembered that simple stimulus–response models (if–then relationship between occupation, job and disease) are of only limited value because they fail to take account of numerous relevant factors from both the work and the private spheres. The resulting standard deviation in the data obtained is so large that no meaningful juxtaposition of work stress and worker health is possible. To reflect the real conditions prevailing in a person's working life, it is necessary to use a multi-causal, retrospective approach that makes allowance for the individual's constitutional makeup and also for extra-occupational stresses to which he is exposed.

This means that detailed knowledge of the design conditions prevailing at the workplace and of the tools and equipment used there must be known and that any evaluation of the job design must be based on valid ergonomic knowledge. A given workplace design will thus give rise to stresses of a given intensity which are *potentially* capable of causing a given symptom or disease with a predetermined percentage probability. Unfortunately, scientifically validated data from which "maximum permissible stress" tolerance limits can be derived are frequently lacking and it is consequently necessary when preparing job design proposals to cobble together data obtained from investigations performed with different evaluation procedures. Software tools can be of considerable assistance here.

The state-of-the-art is considerably more advanced in the design of jobs involving physical work than is the case with information-intensive work. It is thus not surprising that the majority of the software tools on man modeling are for the design of jobs involving physical work.

All the methods described naturally have their specific fields of application, advantages and disadvantages, varying degrees of both accuracy and cost, and focus on different points in the ergonomic design process.

2. HISTORICAL REVIEW

The European Union's Guidelines on Work Safety and Machine Design will exert a lasting influence on the work of job designers and production planners and also on industrial medicine and work safety regulations. They also raise the question as to whether those responsible for implementing ergonomic workplace and job design requirements in industry have at their disposal the tools needed for the analysis, evaluation and design of the man–machine interface.

The former two-dimensional (2D) diagrams of man at the workplace have now been largely replaced in industry by the computer-aided, man-modeling, anthropometric procedures of the 1980s and 1990s. Mutually competitive procedural approaches now being used were frequently developed for a specific purpose, e.g. for the automobile industry, the aircraft manufacturing industry or for other products where anthropologic, ergonomic design is a key factor.

The scientists developing these procedures seldom exchanged ideas and it is consequently safe to assume that a lot of work was duplicated and that the potential advantages of synergy were largely ignored.

Nearly 100 different man-modeling software packages are available on the market at the time of writing. The fields of application of the first packages were workplace design, especially in aircraft and motor vehicle cockpits, and also in industry. Their aim was to portray the human form to scale and in as realistic a way as possible, so that the postures imposed on the worker by tools and equipment and by the work object could be simulated. Later refinements made it possible to portray visual and reach

distances and to perform collision investigations. All in all they were useful tools to replace tedious former procedures using body outline stencils.

Further refinements subsequently built into these software packages gradually transformed them into multivalent instruments for the ergonomist. Attempts are currently being made to improve their modeling accuracy, i.e. to make them produce realistic postures automatically instead of the user having to go through a complicated dialog procedure. Simulation of routine body movements and of the spinal column and the hand–arm system, etc. has improved substantially and work is now proceeding on making the programs capable of showing computer algorithms for applied forces, maximum permissible loads and many other data relating to the worker and the workplace.

3. DEFINITIONS OF AND REQUIREMENTS TO BE FULFILLED BY MAN–MODELING PROCEDURES (MMP)

MMP can be defined as graphic and alphanumeric models using available knowledge on human physiology and on technical aspects of body movements and information reception and processing to produce a model of man at the workplace that is as realistic as possible.

The illustration must not be static. Appropriate simulation techniques must be used to portray the work sequences at the workplace during the course of a shift.

MMP are intended to aid the performance of the fundamental tasks involved in ergonomics — analysis, measurement, evaluation and design. In the field of *analysis* they should facilitate definition of the work system, its elements and its structure. In *measurement,* they should provide support in determining body angles, forces and movements in typical working postures. *Evaluation* is the area where the most extensive requirements exist. Here they must be able to assess energy turnover, force limits and areas of reach and vision, etc. They must be capable of creating stress profiles and of predicting degree of worker fatigue or reliability of a given work system. The evaluation must not be restricted to feasibility and tolerability of the work. It must also indicate whether the work can reasonably be expected of a worker and what discomfort parameters will apply.

The actual purpose of MMP is, however, to provide support to designers and production planners in their *design work.* Their role here is to visualize the man–machine interface, to animate movement sequences, e.g. for workers of different shapes and sizes, and finally to integrate the MMP results into other planning tools, e.g. the planning of plant layouts, facilities management systems, etc.

4. CURRENT RESEARCH STATUS

Only very few of the MMP currently available contain mathematical processes with which the program can process predetermined target points to produce the resultant postures. Sophisticated solutions capable of doing this are a significant breakthrough because they can take into account factors like angle limitations of joints, interdependent factors involved in movements of arm–shoulder system or the pelvispinal region, etc. and thus produce typical human postures that obey the laws of gravity. The body movements in most of the available man models are also still highly simplified and cannot be generated

by the computer internally. One very important development is the use of closed movement cycles that are obtained by video analysis of typical movement sequences.

Another significant development in recent years is body scanning, the collection of anthropometric data by opto-electronic methods. Although the initial accent here was on the sitting position, mainly for research on vehicle design, body scanning has now been extended to cover a wide variety of postures and fields of application. Figure 1 shows the VITUS 3D Body Scanner and the RAMSIS software package as an example of this method.

Over the past 10 years there has been significant progress in the evaluation of physical work involving load handling. It would be logical to integrate this newly acquired knowledge in MMP.

Four types of models — physiological, psycho-physical, biomechanical and epidemiological — (cf. Ayoub 1993, for example) have now been developed for the evaluation of load handling procedures.

The physiological models are based mainly on knowledge relating to cardiovascular and respiratory strains. They are suitable primarily for assessing the long-term effects of a given type of work against the criterion of tolerability. They are not capable of predicting traumata caused by isolated peak stresses.

Psychophysical models establish the relationship between the objective intensity of a physical stimulus and its subjective perception by the worker against the criteria of whether a type of work can reasonably be demanded from either individual workers or groups of workers.

Epidemiological models provide information on the covariance of occupational stresses and job-related or occupational diseases against the criterion of tolerability.

Biomechanical models have a special role to play. This is to provide data on the interaction of work content-related, physical factors and the living body. Mechanical analyses and cover obtain these data:

- body stability;
- maximal physical strength;
- movement sequences;
- mechanical strains affecting individual organs; and
- the mechanical work performed.

Segment models in which the human body is portrayed in the form of linked segments are widely used. The individual segments are rigid, have defined physical properties (mass,

Figure 1. Collection of anthropometric data with the VITUS 3D Body Scanner (source: Tecmath, TUM).

dimensions, focal point coordinates, etc.) and linked together flexibly. Segment models can be static or dynamic and either 2D or 3D. Whereas static models ignore the inertial effects of both the loads being handled and the parts of the body involved, quasi-dynamic models take account of the inertial effects and moments of the external loads only and dynamic models take account of the inertial effects of both the loads and the body parts.

Two-dimensional models are symmetrically constructed in the sagittal plane and cannot register asymmetrical movements or, for example, twisted body postures. This is unfortunate because it is precisely this type of factor that can have injurious effects.

Single Equivalent models illustrate the reactions between muscles and ligaments in a single force vector. Optimization and EMG-controlled models can handle multiple force vectors.

Some of these different types of models are contained together with graphic man models in the software tools RAMSIS, AnySIM or ErgoMan (see below and Landau 1999). There is nowadays an increasing tendency to supplement the graphic portrayal of a load handling operation with an evaluation of the data obtained from a biomechanical model.

The integration of the MMP into the workplace design software is a key feature of several program systems (Figures 2 and 3).

This can offer substantial productivity benefits to industrial users by making it possible to prepare complete plans of both micro- and macro-work systems, starting with the choice of materials and equipment and proceeding via posture optimization and visual geometry to timing calculations and forecasts of personnel requirements. Given the sophistication of these systems, it is not surprising that a considerable amount of development and also ergonomic validation still remains to be done.

It is at present impossible to be sure that the designers and production planners are able to recognize critical body postures and movements, to predict the physiological cost of various types
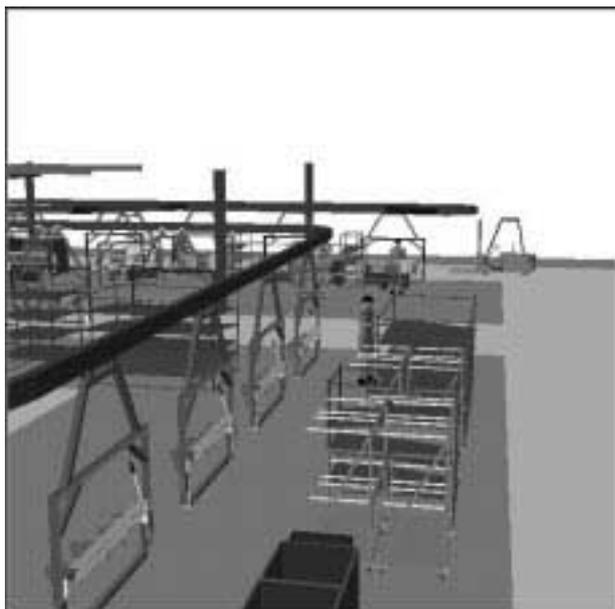
Figure 2. An extract from the Ergoman planning software (source: Delta).

Figure 3. Extract from the AnySIM planning software (source: AnySIM).

Figure 4. Anthropos as an example ov Virtual Ergonomic Prototyping (source: IST).

of physical activities or to appreciate the effects of stress accumulation during a shift or from daily repetition of given postures and movements.

One development following a course of its own is Virtual Ergonomic Prototyping that establishes a link between genuine man and virtual man. An example of this is CAVE visualization systems which check questions of assembly feasibility under critical spatial conditions (Figure 4).

5. CRITICISM OF MMP

Although man-modeling clearly offers major advantages, the risks inherent in a perfectionistic procedure should not be forgotten:

- Could simulation software one day start to dictate the individual actions to be performed by people at work and turn human beings into flesh-and-blood robots?
- Could it happen in a globalized labor market that a preference develops for workers from some specific corner of the world because their body measurements, strength and psycho-mental characteristics correspond best to the computer simulation sequence?
- To what extent will the workers be able to become involved in the design of their own jobs?
- Will man modeling, initially developed to portray and understand the human being, be used to manipulate or

replace him? This is not the vision of human work activity currently prevailing among international ergonomists.

Although these programs certainly offer substantial benefits, those using them should never lose sight of the basic ethical tenet of ergonomics — adaptation of work to the human being for the benefit of mankind.

REFERENCES

AYOUB, M.M., 1993, Modeling in manual materials handling [seminar handout].

LANDAU, K. (ed.), 1999, *Man-Modeling Job Description Procedures* (Stuttgart: Ergon).

PUTZ-ANDERSON, V. and WATERS, T.R., 1991, Revisions in NIOSH Guide to Manual Lifting. Paper presented at the national conference "A national strategy for occupational musculoskeletal injury prevention – Implementation issues and research needs", University of Michigan. Ann Arbor, Michigan.

SNOOK, S.H., 1978, The design of manual handling tasks. *Ergonomics*, 21, 963–85.

SOURCES

IST GmbH, Waldstrasse 13, D-68649 Gross-Rohrheim, Germany; Tel.: +49 6301/606 300; Fax.: +49 6301/606 309; Internet: www.tecmath.de

AnySIM Simulationssysteme GmbH, Heisenbergbogen 1, D-85609 Aschheim-Dornach, Germany; Tel.: +49 89/994025 0; Fax.: +49 89/994025 11; Internet: www.anysim.de

DELTA Industrie Informatik GmbH, Schaflandstraße 2, D-70736 Fellbach, Germany; Tel.: +49 711 57151 0; Fax +49 711 57151 99; Internet: www.delta-ii.de

TUM-Munich University of Technology, Ergonomics Department, Boltzmannstrasse 15, D-85748 Garching; Tel.: +49 89 289 15388; Fax.: +49 89 289 15389; Internet: www.lfe.mw.tum.de

TECMATH GmbH & Co. KG, Sauerwiesen 2, D-67661 Kaiserslautern, Germany; Tel.: + 49 6301 606 0; Fax.: + 49 6301 606 109; Internet: www.tecmath.de

Models of Graphical Perception

G. L. Lohse

The Wharton School, University of Pennsylvania, Philadelphia, PA 19104-6371, USA

1. AN OVERVIEW OF THE GRAPHICS LITERATURE

1.2. Use of Graphics

Graphics are a powerful tool for the discovery of ideas (Lohse and Walker 1994). Exploratory data analysis helps identify relevant information from a set of data as well as gain new insight or understanding of a problem. Often verbal and numeric cues are insufficient to trigger some insight into the specific nature of the problem (Larkin and Simon 1987). Drawing a picture or a graph often helps break a mental block by providing a memory structure to aid the decomposition of information in the problem. Graphics also allow users to digest a lot of information quickly. Decision accuracy and decision speed have always been important reasons to use graphics. Graphs aid data reduction, data summary, improve information search (Robertson *et al.* 1993) and facilitate computation. Graphics, diagrams and pictures transcend language barriers. Icons and international signs are examples of symbols used to communicate across language barriers. Cognitive psychologists found that visual information can benefit observable behaviors such as recall, comprehension and retention of information (Arnheim 1969, Umanath and Scamell 1988). Graphics enhance persuasion (Vogel *et al.* 1986). Vogel *et al.* also found that color overheads increased comprehension and retention of information when compared with black and white.

1.2. A Historical Perspective of the Evolution of Graphs and Charts

A historical perspective of the evolution of graphics provides insights about contemporary graphics usage (Daru 1989). Playfair is considered the founding father of quantitative graphics. His intent was to improve the general understanding of facts normally shown in tables. It took over a half century after Playfair's death before graphs gained acceptance among even a small group of economists and statisticians. An additional 20 years passed before governments used graphs in widely circulated publications. Since the turn of the century, many experiments examined the effectiveness of different types of graphs and tables for displaying data (Eells 1926, Croxton 1927, Croxton and Stryker 1927, von Huhn 1927, Washburne 1927). Businesses increased the use of visual means for advertising and presenting their products. Graphics also gained wide acceptance in the mass media.

As early as 1915, the American Society for Mechanical Engineers set standards for graphic presentation (JASA 1915). From the 1940s to the 1960s, scholars compared the relative effectiveness of presenting numerical data by the use of graphs and tables (Carter 1947, Feliciano *et al.* 1963, Schutz 1961a, b, Vernon 1950, 1953, 1962).

Graphics emerged as a specialized field called drafting or graphics design. Computers returned graphics to the masses by automating the manual drafting components. Unfortunately, many people are unable to produce pleasing, ergonomically correct, substantially true and statistically valid graphs (Johnson *et al.* 1980). Indeed, Cleveland (1984) found that 30% of the graphs reported in the 1980 (volume 207) of *Science* contained one or more errors.

In summary, three major points arise from this historical review of graphics. First, graphs are a tool for reasoning about information. At some point in the evolution of science, the transformation of knowledge into visual and graphical representations ought to be considered an advancement of the thinking tools of the time. Second, people invented graphs on an *ad hoc* basis as the need arose. The final point is that despite a relatively long history, there is not a common theory for testing the effectiveness of graphic data displays. While it is easy to invent a new graph type, it is hard to invent one that works well.

2. DESIGNING GRAPHS AND CHARTS

Bertin (1983) provides a structural taxonomy of graphs, network charts and cartograms (graphical maps). All other things being equal, if one graph requires a shorter viewing time than another, Bertin hypothesized that it is more efficient for that particular question. Ideally, the answer to a question posed to a graph should be perceivable instantly. One should not need to engage in an arduous deciphering task to decode the answer from the graph.

Tufte (1983, 1990) presents a fascinating collection of illustrations and graphics. He discusses graphic design quality and develops empirical measures of graphic performance. His central tenet was that well-designed, powerful graphics are the simplest. His novel ideas for well-designed graphics are supported by excellent examples; however, Tufte does not offer any empirical support for his design guidelines.

Most "how-to" graphics handbooks rely solely on the author's intuition and experience. Although some handbooks are based on empirical research (Cleveland 1985, Kosslyn 1989), recent research in visual psychophysics discredits some of the sweeping generalizations suggested by other graphics handbooks (Legge *et al.* 1989, Spence 1990, Spence and Lewandowsky 1991).

The literature is rife with experimental data assessing differences in graphic design (e.g. tables, line graphs, grouped bar graphs, etc.). Experimental variables that influence performance include: characteristics of the task, graphic display design variables and expertise of the decision-maker (Ives 1982, DeSanctis 1984, Jarvenpaa and Dickson 1988, Jarvenpaa 1989, Kleinmuntz and Schkade 1993). Without a theory for predicting results, it should come as no surprise that this literature is fraught with conflicting outcomes and a paucity of robust findings.

3. MODELS OF GRAPHICAL PERCEPTION

This section describes research that facilitates the evaluation of graphic displays (Gillan and Lewis 1994), automates the evaluation of graphic displays (Tullis 1983, 1988, Lohse 1991, 1993, 1997, Kieras and Meyer 1995), or automates the design of a graphic (Mackinlay 1986, Casner 1991).

3.1. MA–P – Gillan and Lewis

Gillan and Lewis (1994) developed a componential model of human interaction with graphs called the mixed arithmetic–perceptual model (MA-P). MA-P derives a model of the user from task analyses of users performing graphical perception tasks. The

task analysis used in MA-P identified five component processes: search, encode, arithmetic operations, spatial comparisons and respond. MA-P assumes (1) subjects execute component processes in sequence, (2) each component process requires the same amount of time to execute and (3) total time to complete the task is the sum of the number of component processes. MA-P predictions do not reflect WM capacity constraints.

An empirical study compared actual performance to MA-P predictions. Overall, a simple linear regression model explains 90% of the variance in mean reaction times (n = 24). However, like GOMS, MA-P yields an identical prediction for color or monochrome displays with or without grid lines. Thus, MA-P would need more detailed process operators for detection, discrimination, and perceptual grouping to explain performance for subtle differences in graphical presentation formats.

3.2. DAP – Tullis

Tullis (1983, 1988) identified six parameters for objectively evaluating screen formats. Based on a usability analysis of these six parameters from 520 displays, Tullis developed regression equations to predict average search time and display quality. His program analyzes screen displays, makes an assessment of usability, recommends improvements in the screen layout, calculates an average visual search time and rates overall quality.

Predictions generated by the program rely heavily on the Gestalt principle of proximity. Although the predictions generated by DAP are quite good (r = 0.71), DAP only analyzes monochrome, alphanumeric tables whereas graphical displays afford more determinants of grouping in addition to spatial proximity.

These determinants include color, graphical borders, highlighting, typefaces and multiple fonts. Items coded with the same color are a perceptual group. Rectangular boundaries also help form groups, although not as effectively as color since containment is detected serially. Highlighting, using reverse video, increased brightness, flashing, underlining or color, can be very effective in attracting user attention. For graphical displays, DAP needs to be extended to include these other display features.

3.3. UCIE Lohse

UCIE (Lohse 1991, 1993) simulates how people answer certain questions posed to bar graphs, line graphs and tables by predicting a sequence of eye fixations that contains the information needed to answer the query (Figure 1). Within each fixation, UCIE identifies the perceptual and cognitive subtasks. Next, UCIE assigns each component task a known cognitive engineering time parameter (e.g. 300 ms to process a six-letter word; John and Newell 1990). UCIE predicts total task execution time by summing the time parameters over all the component tasks. UCIE time predictions reflect subtle differences in similarity, proximity, color and overlap among adjacent non-target objects in each fixation. UCIE also simulates the flow of information through WM (Kosslyn 1985).

An empirical study compared actual performance to UCIE predictions for 576 combinations of presentation formats and question types from 28 subjects (n = 16,128). For conditions that predominantly involve serial processing, zero-parameter predictions from UCIE explained 60% of the variance in reaction times. While UCIE time predictions are good, additional research is needed to explain individual differences in performance.
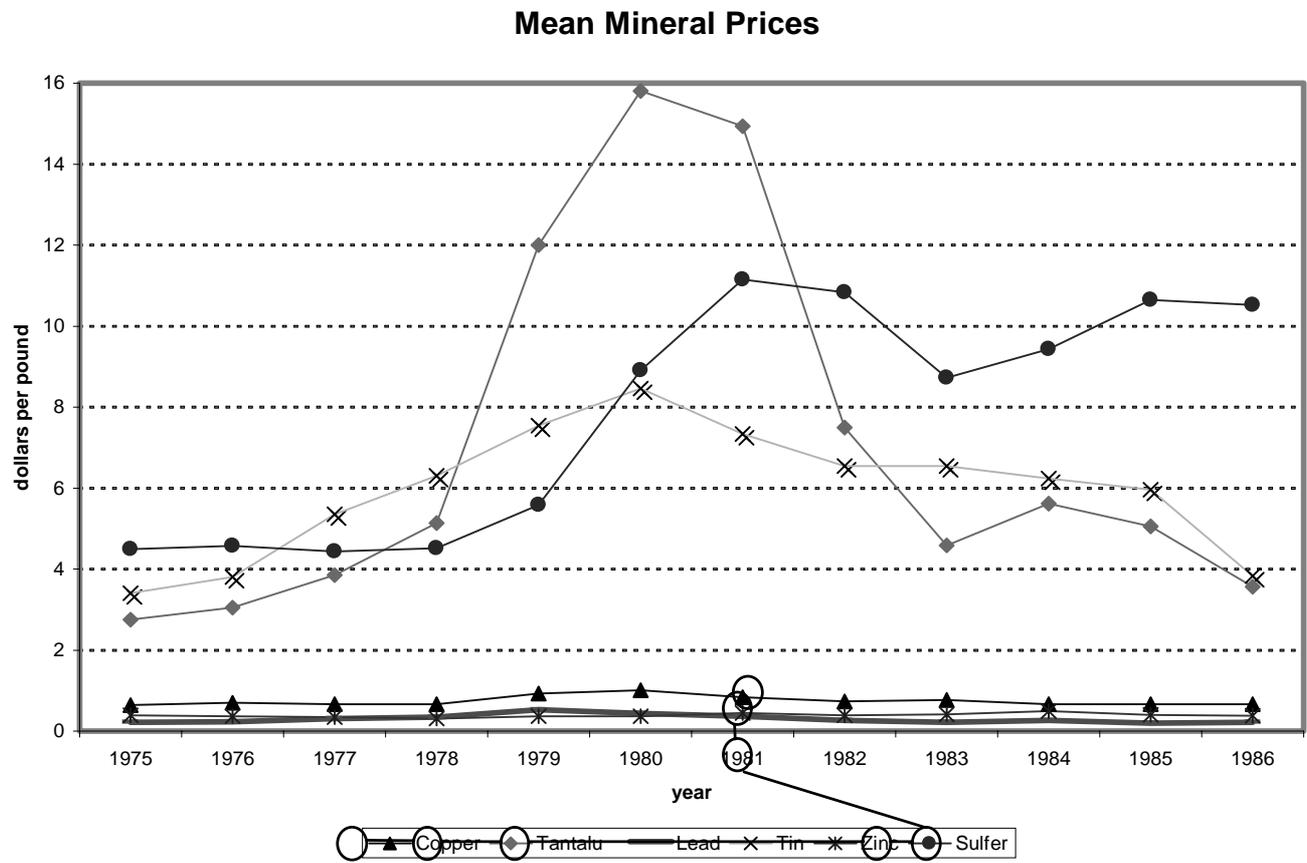
Mean Mineral Prices

Figure 1. Sequence of eye movements to answer the question "In 1981, did zinc cost less than copper?" posed to a line graph.

3.4. APT – Mackinlay

Mackinlay (1986) developed a compositional algebra for systematically generating a wide variety of graphics from a small set of presentation elements. For example, length encodes quantitative relations (e.g. largest) whereas color does not express quantity well. In contrast, color facilitates perceptual grouping among nominal data categories whereas length does not readily promote perceptual grouping. APT provides two important advances for automating graphic design. First, nearly every graphic presentation tool designed after APT, including BOZ (Casner 1991) uses APT's formal graphical language. Second, APT designs graphics with a minimal amount of input from the graphic designer.

APT has some limitations. The rankings of visual primitives for expressiveness of ordinal and nominal data are based on intuition and conjecture. APT also lacks knowledge about choosing font sizes, selecting line widths, and positioning objects for rendering the graphic design. Finally, APT does not consider the end user task of answering some question posed to the data. This prevents APT from generating different graphic presentations of the same data to support different questions.

3.5. BOZ – Casner

BOZ renders graphs from raw data and a task analysis of a specific question posed to the data (Casner 1991). It replaces demanding logical operators in the codified task description with less-demanding visual operators that reduce visual search. For example, BOZ might replace a comparison of numeric values with a perceptual judgment comparing the heights of two bars. BOZ renders multiple presentations of the same data customized to different task requirements.

Using BOZ, Casner and Larkin (1989) create four displays of one set of complex airline reservation data and predict the efficacy of the four displays using a LISP. Except for display 4, each visual operator reduced visual search time in the manner predicted by the simulation. The major contribution of BOZ is that it creates graphic designs that support a specific task. BOZ does have several limitations. It does not consider fonts, typeface, or color. The graphical presentation language does not compute task completion time. BOZ also does not consider general conventions (left to right reading order, spatial numbering patterns, etc.).

3.6. GOMS – John

GOMS models have four components: (1) a set of goals, (2) a set of operators to perform actions, (3) a set of methods to achieve the goals and (4) a set of selection rules for choosing among competing methods for goals (Card *et al.* 1983). GOMS assumes that routine cognitive skills are a serial sequence of perceptual, cognitive, and motor activities. Each of these actions has a time parameter.

Chuah *et al.* (1995) used GOMS to model the airline flight reservation tasks used by Casner (1991). They used one information search algorithm for each display to make precise predictions about the number of eye movements required to complete the information acquisition task. Chuah *et al.* compared zero-parameter predictions from the GOMS model to the empirical data collected by Casner and Larkin (1989). The average absolute error was 8%. These early results are very encouraging. They demonstrate that GOMS models provide accurate measures of performance for a relatively complex graphical perception task.

3.7. EPIC – Kieras and Meyers

The EPIC architecture is a computational modeling environment for simulating a range of human perceptual-motor and cognitive tasks (Kieras and Meyers 1995). EPIC requires, as input, a set of production rules that specify what actions must be performed to do the task. The model generates a specific pattern of activities necessary to perform the task. Model simulations explain specific patterns of effects in quantitative detail (Meyers and Kieras 1995).

Unlike other production rule systems, EPIC executes multiple actions in parallel. At the start of each cognitive cycle, EPIC updates the contents of WM with new inputs from perceptual and motor processes. WM does not include any mechanism for rehearsal, decay or overflow. The production rule interpreter tests the contents of WM against all conditions of rules stored in the production-rule memory. At the end of the cycle, EPIC executes associated actions in parallel. Thus, there are no central-processing bottlenecks in EPIC. Extensions of the visual processor will allow EPIC to handle such tasks as answering a question posed to a line graph.

Table 1. A comparison of features supported by models' graphical perception.

Model feature	MA-P Gillan Tullis	DAP Lohse	UCIE Mackinlay	APT Casner	BOZ John	GOMS Keiras	EPIC
Based on task analysis	✔	✔	✔		✔	✔	✔
Time predictions	✔	✔	✔			✔	✔
Screen layout and tables		✔	✔	✔	✔	✔	✔
General purpose tasks						✔	✔
Automatic designs				✔	✔		
Color			✔	✔			✔
Individual difference			✔				✔
WM limitations			✔				✔
Parallel tasks						✔	✔
Font size							✔
Errors							
Typeface							
Animation							
Practice/skill effects							
Learning new schemata							

4. FUTURE DIRECTIONS

Table 1 compares some of features supported by the graphical perception models reviewed here. All models, except APT, stem from detailed analyses of the user's task. All models, except APT and BOZ, predict task completion time. Time predictions using MA-P can be invariant to the graph type. For example, MA-P lacks detailed operators to predict processing time differences between two bar graphs of the same data with and without grid lines. Tullis designed DAP for analyzing screen designs. DAP does not handle charts and graphs. UCIE only models information processing from bar graphs, line graphs or tables. EPIC incorporates a model of WM, although the WM component of the EPIC architecture is in the early stages of development. CPM-GOMS and EPIC are designed specifically for parallel tasks.

None of the models fully examine individual differences in user capabilities. Individual differences include levels of skill, practice acquiring information from particular graph formats, knowledge of the graphical perception tasks, learning to process information from new or novel graph formats, age, and culture (e.g. most Asian reading patterns are not left to right). In addition, all models only consider skilled, error-free behavior. All of the models function in the domain of static, two-dimensional, graphic displays. Animation and dynamic displays are beyond the scope of current models. Such extensions would be difficult for two reasons. First, little is known about the perception and cognition of animated displays. Second, models would need to incorporate episodic memory.

This chapter motivates a computational cognitive engineering approach for quantifying the effectiveness, efficiency and quality of graphic designs. Unlike empirical graphics research, a quantitative cognitive modeling approach can provide robust objective predictions using an algorithm that could be incorporated into software for automated graphic design. Quantitative models provide a falsifiable theory of graphical perception and cognition that can be tested empirically by comparing decision time predictions from the model with actual decision times. The next generation of graphics research not only should enhance our understanding of graphical perception and cognition but also provide quantitative predictions that facilitate the objective evaluation of the effectiveness, efficiency and quality of graphic displays. Clearly, models of graphical perception are still in their infancy. Much work remains in extending the basic foundation to an understanding on all of the underlying graphical perception phenomena.

REFERENCES

ARNHEIM, R., 1969, *Visual Thinking* (Berkeley: University of California Press).

BERTIN, J., 1967, *Sémiologie Graphiques* (2nd ed.) (Paris: Gauthier-Villars); English translation: BERG, W.J., 1983, *Semiology of Graphics* (Madison: University of Wisconsin Press).

CARD, S.K., MORAN, T.P. and NEWELL, A., 1983, *The Psychology of Human–Computer Interaction* (Hillsdale: Lawrence Erlbaum).

CARTER, L.F., 1947, An experiment of the design of tables and graphs used for presenting numerical data. *Journal of Applied Psychology*, 31, 640–50.

CASNER, S., 1991, A task-analytic approach to the automated design of graphic presentations. *ACM Transactions on Graphics*, 10, 111–51.

CASNER, S. and LARKIN, J.H., 1989, Cognitive efficiency considerations for good graphic design. In 11th Annual Conference of the Cognitive Science Society, 16–19 August 1989, University of Michigan, Ann Arbor, MI (Hillsdale: Lawrence Erlbaum), pp. 275–82.

CHUAH, M.C., JOHN, B.E. and PANE, J., 1994, Analyzing graphic and textual layouts with GOMS: results of a preliminary analysis. In Plaisant, C. (ed) *CHI '94 Proceedings of the ACM Conference on Human Factors in Computing*, Boston, MA, 7–11 May 1994 (New York: ACM Press), pp. 323–4.

CLEVELAND, W.S., 1984, Graphs in scientific publications. *American Statistician*, 38, 261–9.

CLEVELAND, W.S., 1985, *The Elements of Graphing Data* (Monterey: Wadsworth).

CROXTON, F.E., 1927, Further studies in the use of circles and bars. *Journal of the American Statistical Association*, 22, 36–9.

CROXTON, F.E. and STRYKER, R.E., 1927, Bar charts versus circle diagrams. *Journal of the American Statistical Association*, 22, 473–82.

DARU, M., 1989, The culture of quantitative graphicacy. *Information Design Journal*, 5, 191–208.

DE SANCTIS, G., 1984, Computer graphics as decision aids: directions for research. *Decision Sciences*, 5, 463–87.

EELLS, W.C., 1926, The relative merits of circles and bars for representing component parts. *Journal of the American Statistical Association*, 21, 119–32.

FELICIANO, G.D., POWERS, R.D. and KEARL, B.E., 1963, The presentation of statistical information. *AV Communication Review*, 11, 32–9.

GILLAN, D.J. and LEWIS, R., 1994, A componential model of human interaction with graphs: I. Linear regression modeling. *Human Factors*, 36, 419–40.

IVES, B., 1982, Graphical user interfaces for business information systems. *MIS Quarterly*, special issue, 15–47.

JARVENPAA, S.L., 1989, The effect of task demands and graphic format on information processing strategies. *Management Science*, 35, 285–303.

JARVENPAA, S.L. and DICKSON, G.W., 1988, Graphics and managerial decision making: research based guidelines. *Communications of the ACM*, 31, 764–74.

JOHNSON, J.R., RICE, R.R. and ROEMMICH, R.A., 1980, Pictures that lie: the abuse of graphs in annual reports. *Management Accounting*, 62, 50–6.

JOINT COMMITTEE ON STANDARDS FOR GRAPHIC PRESENTATION, 1915, Preliminary report. *Journal of the American Statistical Association*, 14, 790–7.

KIERAS, D.E., WOOD, S. and MEYER, D.E., 1995, Predictive engineering models using the EPIC architecture for a high-performance task. In Katz, I.R., Mack R. and Marks, L. (eds) *CHI '95 Proceedings of the ACM Conference on Human Factors in Computing*, Denver, CO, 7–11 May 1995 (New York: ACM Press), pp. 11–18.

KLEINMUNTZ, D.N. and SCHKADE, D.A., 1993, Information displays and decision processes. *Psychological Science*, 4, 221–7.

KOSSLYN, S.M., 1985, Graphics and human information processing. *Journal of the American Statistical Association*, 80, 499–512.

KOSSLYN, S.M., 1989, Understanding charts and graphs. *Applied Cognitive Psychology*, 3, 185–225.

LARKIN, J.H. and SIMON, H.A., 1987, Why a diagram is (sometimes) worth ten thousand words. *Cognitive Science*, 11, 65–99.

LEGGE, G.E., GU, Y. and LUEBKER, A., 1989, Efficiency of graphical perception. *Perception and Psychophysics*. 46, 365–74.

LOHSE, G.L., 1991, A cognitive model for the perception and understanding of graphs. In Robertson, S.P., Olson G.M. and Olson J.S.

(eds) *CHI '91 Conference Proceedings*, New Orleans, LA, 28 April–2 May 1991 (Los Alamitos: ACM Press), pp. 131–8.

LOHSE, G.L., 1993, A cognitive model for understanding graphical perception, *Human–Computer Interaction*, 8, 353–88.

LOHSE, G.L., 1997, Models of graphical perception. In Helander, M., Landauer, T.K., and Prabhu, P. (eds) *Handbook of Human–Computer Interaction* (2nd ed.) (Amsterdam: North Holland/Elsevier), 107–35.

LOHSE, G.L. and WALKER, N., 1994, Classifying graphical knowledge. In Kent, A. and Williams, J.A. (eds) *Encyclopedia of Computer Science and Technology*, vol. 30 (New York: Marcel Dekker).

MACKINLAY, J.D., 1986, Automating the design of graphical presentations of relational information. *ACM Transactions on Graphics*, 5, 110–41.

MEYER, D.E. and KIERAS, D.E., 1995, A computational theory of human multiple-task performance: the EPIC information-processing architecture and strategic response-deferent model. *Psychological Review*.

ROBERTSON. G.G., CARD, S.K. and MACKINLAY, J.D., 1993, Information visualization using 3D interactive animation. *Communication of the ACM*, 36, 57–71.

SCHUTZ, H.G., 1961a, An evaluation of formats for graphic trend displays. *Human Factors*, 3, 99–107.

SCHUTZ, H.G., 1961b, An evaluation of methods for presentation of graphic multiple trends. *Human Factors*, 3, 108–19.

SPENCE, I., 1990, Visual psychophysics of simple graphical elements. *Journal of Experimental Psychology: Human Perception and Performance*, 16, 683–92.

SPENCE, I. and LEWANDOWSKY, S., 1991, Displaying proportions and percentages. *Applied Cognitive Psychology*, 5, 61–77.

TUFTE, E.R., 1983, *The Visual Display of Quantitative Information* (Cheshire: Graphics Press).

TUFTE, E.R., 1990, *Envisioning Information* (Cheshire: Graphics Press).

TULLIS, T.S., 1983, The formatting of alphanumeric displays: a review and analysis. *Human Factors*, 25, 657–82.

TULLIS, T.S., 1988, Screen design. In Helander, M. (ed.) *Handbook of Human–Computer Interaction* (Amsterdam: North Holland/Elsevier), pp. 377–411.

UMANATH, N.S. and SCAMELL, R.W., 1988, An experimental investigation of the impact of data display format on recall performance. *Communications of the ACM*, 31, 562–70.

VERNON, M.D., 1950, The visual presentation of factual material. *British Journal of Educational Psychology*, 3, 19–23.

VERNON, M.D., 1953, Presenting information in diagrams. *AV Communication Review*, 1, 147–58.

VERNON, M.D., 1962, *The Psychology of Perception* (London: Penguin).

VOGEL, D., LEHMAN, J. and DICKSON, G., 1986, The impact of graphical displays on persuasion: an empirical study. In *Proceedings of the 7th International Conference on Information Systems*, San Diego, CA, pp. 240–54.

VON HUHN, 1927, Further studies in graphic use of circles and bars: a discussion of the Eells experiment. *Journal of the American Statistical Association*, 22, 31–6.

WASHBURNE, J.N., 1927, An experimental study of various graphic, tabular, and textual methods of presenting quantitative material. *Journal of Education Psychology*, 18, 361–76.

Multimedia Production

H. M. Khalid

Institute of Design and Ergonomics Application, Universiti Malaysia
Sarawak, 94300 Kota Samarahan, Sarawak, Malaysia

1. INTRODUCTION

The convergence of computing and media presentation technologies has led to the coining of *multimedia*, a term that encompasses a variety of computer-controlled *interactive* media. "Media" refers to various forms of presentation that can be captured and processed using a computer, such as video, audio, text, graphics, images; while "multi" indicates that several of these media elements are used in the same application. The Interactive Multimedia Association defines multimedia as delivery of information, usually via a personal computer, that combines different content formats (text, graphics, audio, still images, animation, motion video) and/or storage media (magnetic disc, optical disc, video/audio tape, RAM).

Multimedia thus implies that although the presentations are designed and authored by one or several experts ahead of time, the information presentation is controlled by the user. Interactivity, therefore, is essential for multimedia. A multimodal interface may exploit different user senses, such as touch, hearing and seeing. Multimedia communication, then, aims at providing a rich environment of images, graphics, sound, text and interaction, which supports a variety of human activities from decision-making to emotional involvement. Once a buzzword, today multimedia is ubiquitous in most forms of information presentation and documentation; the World Wide Web is a good example.
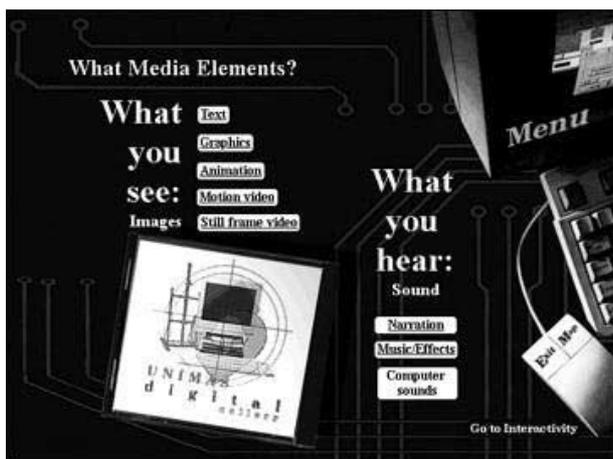
Figure 1. Unimas Digital Gallery presentation (Khalid 1997) is a display of various artworks illustrated in text, graphics, video, sound and animation. By clicking on keyword-buttons, users can view directly illustrations of media elements and return to the main screen via the "menu" button. The interface is easy to use with a clearly marked exit and a map to support user navigation.

Multimedia has various user contexts: information kiosks and games for home users, presentations and desktop conferencing for office workers, training and education for students and teachers in the classroom, data visualizations for scientists, and user manuals for engineers. The challenge for the ergonomist, then, is how to blend and integrate the media elements so as to improve human–computer communication. The purpose of this article is to highlight ergonomic aspects of multimedia production, involving the skillful use of different modalities and forms of presentation to communicate essential ideas and concepts. Focus will be on ergonomics guidelines for multimedia design and multimedia usability. Additional information on human factors aspects of multimedia interaction are in Waterworth (1992), and Chignell and Waterworth (1997). Research and theories on multimedia interface design are discussed in Blattner and Dannenberg (1992), the technical aspects of multimedia technology are described in Vaughn (1995), while Furht (1996) presents various tools and applications in multimedia.

1.1. Ergonomics and Multimedia

The design of an informative interface that provides a rich sensory experience of multimedia poses a considerable challenge, and ergonomics' tools can be used to exploit multimedia to its advantage. The highly engaging and interactive style of multimedia communication has the potential of providing effective human–machine interactions. However, multimedia can also be overwhelming, disconnected and difficult to understand; thereby, effective multimedia requires a conscientious design effort. Generally, multimedia design involves both micro- and macro- issues. At the micro- level the concern is with local coherence of information content and presentation that does not result in overcrowding of screen output. Principles of ergonomic screen design may be applied, such as highlighting important information, using appropriate fonts and colors, and organizing blocks of information within screen layouts. At the macro- level the concern is with connectivity, navigability and global coherence of a set of individual presentations that constitute the multimedia application. Landmarks, that is, superimposed visual artifacts, may be used to maintain the orientation and can aid in navigation. Therefore, general principles of ergonomics design in Human–Computer Interaction (HCI) also apply to multimedia design (Helander *et al.* 1997).

2. MULTIMEDIA USES AND APPLICATIONS

The goals of multimedia are to create a user interface that is interactive, pleasing to the senses, and enables the assimilation of information. To this end multimedia should be used when rich media, hyperlinking and interactivity can add significant value to an information presentation. For example, multimedia can enrich and enhance traditional conference presentation, such as in Figure 1 — a creative mix of content comprising text, graphics, video, music and voice narration. This creates an aura of superimposed expressions that, when crafted artfully, increases the interest of the viewing audience. At the same time it allows for multiple channels of information which reinforce the message and integrate the information presentation.

There are many applications of multimedia:

- *Training and education* — the ability to provide access to

remote multimedia stored in databases and to interact with a remote instructor.

- *Distributed work groups* — may benefit from multimodal information presentations, such as in engineering design.
- *Remote experts* — providing images from inaccessible locations, e.g. oil rigs, to experts who can communicate advice, or medical doctors for diagnosis and treatment.
- *Information/sales terminals* — located in shops, departure lounges, public places, providing information on products or services, with videotelephone access to a salesperson.
- *Interactive television* — entertainment, shopping, education and information services utilizing domestic television or Internet.
- *Computer-based information services* — built around personal computer access to information services, e.g. Internet.

Multimedia technology has become important to governments in several countries as a vehicle to enhance information exchange and control. The Information Super Highway in the USA, once a government initiative, later became a private industry initiative. In Asia, Malaysia's Multimedia Super Corridor provides high-capacity broadband network with the purpose to serve industry in the design and production of multimedia products and services. In Singapore the IT2000 initiative ensures that all households are networked using broadband connections. At present this is primarily used for entertainment, while professional applications are being developed.

2.1. Multimedia and Learning

The development of computer-aided instruction and audiovisual media has created much interest in applying multimedia in education and training. Although research findings are not entirely conclusive, using multimedia material people may remember ~10% of what they read, 20% of what they hear and 30% of what they see (Khalid 1998). Most studies in multimedia-based training have focused on learner satisfaction rather than on what has been learnt. Concerns such as whether learning occurs, whether reflective thought is involved or merely sensory experience, has revived some interest in certain theoretical perspectives. Dual Coding Theory, when applied to design of multimedia material, can enhance learning by the simultaneous presentation in visual and auditory modalities. Constructivist Theory, on the other hand, can support learners to develop cognitively flexible processing skills and to acquire knowledge structures that are coherent in terms of content. But these concepts still remain to be proven in research.

3. MULTIMEDIA DEVELOPMENT

An ergonomics framework for multimedia development is summarized in Figure 2. Generally the design evolves through different developmental phases: formulation of purpose, needs analysis, information design, usability features, production planning and evaluation.

Basic design issues that must be addressed early in the multimedia development process include:

- Inputs to the multimedia design process:
- What is the objective of the multimedia application?
- Who is the audience?
- What is the nature of the content?
- What is the delivery platform?
- What are the available development resources?
- What are the user's expectations?
- Multimedia design outputs:
- What combination of media?
- Where and when will various media be used?
- What is the overall quality of production elements?
- Will media elements be synchronized or user invoked?
- What primary branching structure?
- What user navigation features?
- What type of interactions and how frequent?

3.1. Multimedia Metaphor

There are various approaches to multimedia, each is based on different metaphors. Chignell and Waterworth (1997) differentiated between performance, presentation and document metaphors. *Performance* multimedia uses an orchestration of actors and events to create a sensory experience like a stage play. Timing is very important to synchronize the media elements. The media elements must then be coordinated with the user so that it is possible for the user to interact. This type of presentation can be created using Macromedia Director, and it is frequently used in edutainment.

Presentation multimedia uses slides or snapshots to present a set of ideas that form the focus and intention of the presentation. To detour from the linear presentation sequence, links can be offered between slides. This type can be created using Microsoft PowerPoint and it is often used in business or conference presentations.

Document multimedia is concerned with detailed text and ideas similar to a book. It consists of pages or nodes that are organized around a textual framework. This type can be created using Authorware Professional or ToolBook.

The development of HTML and the Web has given a boost to document multimedia. Web documents combining text, pictures, sound and video can be accessed on the Internet. Today, performance multimedia is typically distributed on CD-ROM, document multimedia on the Web, and presentation multimedia is delivered directly to an audience, typically using a notebook computer, and a data projector for projection on a large screen.

3.2. Multimedia Authoring

Authoring is the task of creating or developing multimedia. In multimedia development the concept is much more complex than writing a book given the numerous steps involved in the authoring process. Authoring tools are used for designing interactivity and user interface, and for assembling multimedia elements into a single cohesive application. Some useful tools include: card/page-based tools (e.g. Hypercard, Toolbook), icon-based (e.g. Authorware Professional, IconAuthor), and time-based authoring tools (e.g. Macromedia Director, Passport Producer).

3.3. Navigation in Hyperspace

Multimedia typically provides many alternative links and pathways, and users are often "lost in hyperspace." Thereby navigation becomes an important issue. The user must understand how to get back to the origin. A navigation map is helpful. It outlines the links or connections between various areas and helps the user to organize the content, and plan how to move around. There are many types of navigation structures:
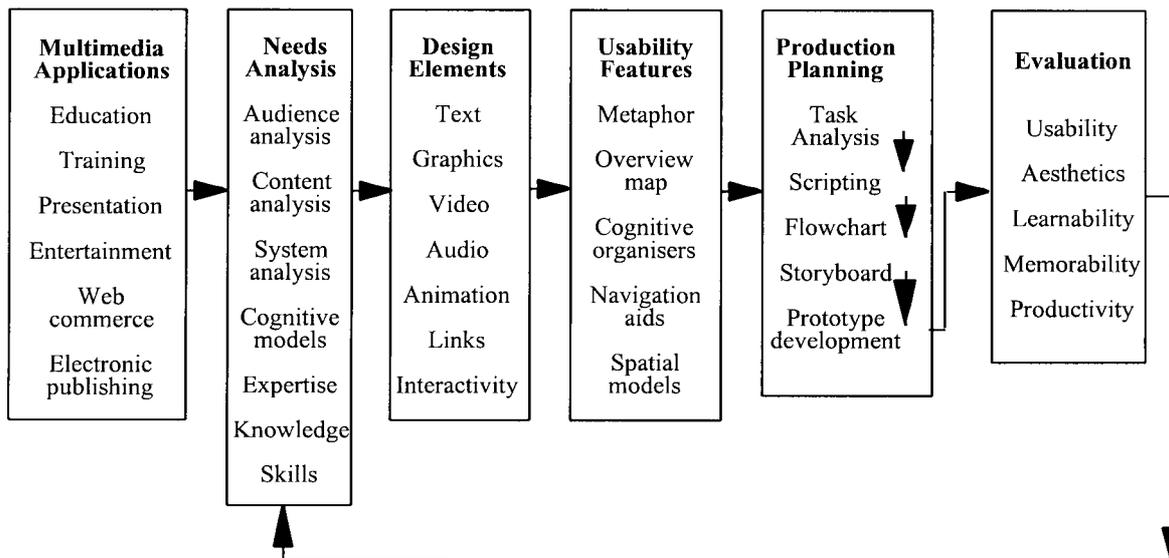
Figure 2. Ergonomics framework for multimedia development.

- *Linear* — users navigate sequentially from one frame to another.
- *Hierarchical* — users navigate along the branches of a tree structure that is shaped by the natural logic of the content.
- *Nonlinear* — users navigate freely through the content of the project, unbound by predetermined routes.
- *Composite* — users may navigate freely (nonlinear) but are occasionally forced to attend to linear presentations of critical information.

Navigation can also be improved by organization of information and by spatial representation:

- *Organization of information.* Global coherence in multimedia applications is an important ergonomic issue. The use of cognitive organizer tools, such as overview, maps, table of contents, and linking paths, can support organization and coherence of the information structure. Another issue is when and how to use hyperlinking. It is often claimed that nonlinearity of hypertext is especially suited to effective browsing, suggesting that hyperlinking should be used whenever possible.
- *Spatial representation.* The use of mnemonic systems, and a familiar spatial structure (e.g. the interior of a house) may increase the memorability of multimedia. They help to organize arbitrary sets of objects, concepts or events that can then be remembered easily. The Book House project, for example, is an interesting spatial illustration for finding books in a library (Pejtersen and Rasmussen 1997). Different rooms are used for different types of books, and by clicking on them, the library reference is presented. Providing spatial models of information structure can function as cognitive "prostheses" to multimedia users.

4. GUIDELINES FOR DESIGN OF MULTIMEDIA INTERFACES

Just as there is no standard definition of multimedia, little is also known about how to design usable multimedia interfaces. This is a difficult task since the time-varying nature of multimedia

interfaces demands new design principles. While not much empirical research exists that can support the many design guidelines found in the literature, the need to apply a certain structure seems inevitable for the design of multimedia.

To date, there are relatively few guidelines that can be shared among multimedia developers. Various ergonomic strategies can be used to address particular properties of multimedia. One strategy is to use a design team comprising engineers, computer scientists, ergonomists, industrial designers, graphics artist and application specialists. The use of design teams will see methodologies emerge specifically geared towards multimedia interface design, using the talents of all disciplines.

To help enhance the usability and ergonomics of multimedia, authoring and design guidelines have expressed the need to consider guidelines at several levels: conceptual (e.g. developing good metaphors), linking (e.g. creating dynamic links), prostethics (e.g. using navigational tools such as overviews and organizers), and structural (e.g. specifying appropriate network and spatial structures). Some design guidelines express the need to provide information in small chunks, reveal increasing levels of detail, not inundate the user but to focus on substance, provide an overview of the contents, and index the information. At the current state of knowledge about interface guidelines for multimedia, developers are cautioned to treat guidelines with a grain of salt, to realize that effectiveness may be "hit" or "miss," and to test the interface guidelines thoroughly with the help of the test audience.

Navigation is an important and difficult issue. The central problem is to let the user freely explore the data space, yet to provide guidance in the process. Overview diagrams, fisheye views, embedded menus, guides and a visual cache of previously examined nodes will all help in navigation. Users have to make decisions as to which links to follow and which to abandon. Given a large number of choices, links therefore induce a cognitive overhead.

The design of the multimedia interface must support flexible interaction. This requires the ability to structure material

hierarchically. Thereby users will understand how the system is designed and they will be free to navigate either in a depth-first or breadth-first strategy. Designers can also provide annotation techniques (e.g. keywords) to assist user navigation, browsing and retrieval of information.

Another type of navigation simulates the spatial properties of the world around us, in particular navigating and manipulating in 3D multimedia world. The "walking" metaphor is used to simulate the motion of a human body through 3D scenes. However, it may not be a desirable metaphor because of tethers, obstructions, and the slowness of physical movement (Blattner and Dannenberg 1992).

A common difficulty experienced by multimedia users is disorientation. The problem arises from the need to know where one is in the network, where one came from, and how to get to another place in the network. A similar problem is multithreaded navigation. Each navigation thread requires its own sense of orientation, and the user's overall sense of disorientation is multiplied unless effective orienting support is provided. This type of navigation pushes the limits on working memory and attention.

To reduce disorientation, Nielsen (1990) proposed to:
- create a straightforward and usable organization of the content material;
- make the structure apparent;
- provide maps and overviews;
- use annotation to increase rhetorical structure;
- create paths and hotlists; and
- add search tools to provide alternate means of "getting around."

Guidelines on multimedia design have focused on ways to facilitate learning, including the need to organize and display databases of various types; and to provide links to information sources in a logical and consistent manner. Interface design heuristics, commonly used for HCI design, can be applied to multimedia, such as: use simple and natural dialogue; speak the user's language; minimize the memory load; be consistent; provide feedback; provide clearly marked exits; provide shortcuts; prevent errors; provide good error messages (Nielsen 1990).

5. MULTIMEDIA USABILITY

There is no standard method for evaluating multimedia usability. Its evaluation is inherently more difficult than evaluation of software usability because there is hardly a notion of a standard task in multimedia applications, nor are there obvious measures of task performance using multimedia (Khalid 1998). Instead, one has to rely on subjective ratings of quality and usability of the multimedia, along with tests of comprehension of its content after the multimedia has been used. Global measures obtained from questionnaires and other subjective ratings must be supplemented with diagnostic measures. In addition to these evaluations one could also evaluate the overall information design — that it does not confuse, distract or annoy a significant number of users. The learnability and usability of navigation tools should also be tested, particularly those that are innovative or unusual.

Chignell and Waterworth (1997) proposed four empirical measures of usability that may be diagnostic of structural properties of usable multimedia: node accessibility, link recognizability, landmark recognition and convergence of conceptual structure. In usable multimedia linking paths should be clear and available for selection, and links for each node should be recognizable and make sense to the user. Also, the user's model of the structure should converge toward the system's representation of structure after a certain period of usage.

Nielsen (1990) identified four categories for the evaluation of hypertext documents which can also be applied to multimedia documents. Utility — a measure of whether the document actually helps a user perform the intended task; integrity — a measure of the completeness of the document; aesthetics — a measure of how pleasing the system is to the user; usability — a measure of the effectiveness of communication on five usability parameters: learnability (of the interface), efficiency, memorability, errors and satisfaction. The first experience most people have with a new system is learning to use it. Learnability is, therefore, considered the most fundamental usability attribute. A memorable system enables users to remember how to use and navigate in multimedia after some period of not having used it.

In multimedia, there are two common types of errors: failure to complete the navigation or search task, and deviations from the ideal path, such as when users have to backtrack. Both of these errors are indications that users have difficulties finding what they are looking for, and the designer should consider redesigning the links or improving the navigation by providing appropriate navigational aids.

6. FUTURE RESEARCH

We need human factors knowledge of *users* (e.g. psychology of multimedia users), *tasks* (e.g. standard models of multimedia for different tasks), *technology* (e.g. tools for constructing multimedia), and *environment* (e.g. multimedia in virtual environments). Multimedia has for many years been technology driven. There is a complex array of technological artifacts and computers. To design an educational courseware or entertainment, or an amalgam of both — "edutainment" — there are numerous options for selecting media, mixing and timing them, and for providing user feedback. Hardly any of these issues has been subjected to rigorous research in ergonomics. Perhaps the lack of interest has been because multimedia at present time is viewed more as an art than as a science.

We need new models of browsing that allow constrained exploration within a topic, and usability studies to evaluate the various maps and overviews that have been developed for visualizing multimedia structure.

We also need guidelines for adopting common rules in movie making (e.g. camera angle, panning, zooming) to multimedia. Experience rather than research guided these rules. In addition, we need guidelines to enhance spatial models of information to make them more usable and effective.

7. CONCLUSION

Recent developments in Web presentation technology imply that multimedia will continue to expand in importance. New media concepts, such as streaming of audio and video, demand new approaches to multimedia development; how they are conceived, designed, and implemented. This requires the skills of usability experts and designers who can develop tools and techniques that can assist in the coordination and presentation of media.

Ergonomics should support multimedia designers — the real users of human factors data, principles and tools — by packaging information and guidelines in a usable and meaningful form for direct application. To this end, much research will be necessary to support multimedia production.

REFERENCES

BLATTNER, M.M. and DANNENBERG, R.B. (eds), 1992, *Multimedia Interface Design* (New York: ACM Press).

CHIGNELL, M. and WATERWORTH, J., 1997, Multimedia. In G. Salvendy (ed.), *Handbook of Human Factors and Ergonomics* (2nd ed.) (New York: John Wiley), 1809–61.

FURHT, B., 1996, *Multimedia Tools and Applications* (Boston: Kluwer).

HELANDER, M., LANDAUER, T. and PRABHU, P. (eds), 1997, *Handbook of Human–Computer Interaction* (2nd ed.) (Amsterdam: Elsevier).

KHALID, H.M., 1997, Human factors in multimedia production. In *Proceedings of the 13th Triennial IEA Congress, IEA'97* (Tampere: Finnish Institute of Occupational Health), 17–22.

KHALID, H.M., 1998, Multimedia usability and its implications for ergonomics. In P. Scott, R. Bridger and J. Charteris (eds), *Global Ergonomics* (Amsterdam: Elsevier), 581–4.

NIELSEN, J., 1990, *Hypertext and Hypermedia* (New York: Academic Press).

PEJTERSEN, A.M. and RASMUSSEN, J., 1997, Ecological information systems and support of learning: coupling work domain information to user characteristics. In M. Helander, T. Landauer and P. Prabhu (eds), *Handbook of Human–Computer Interaction* (2nd ed.) (Amsterdam: Elsevier), 315–46.

VAUGHN, T., 1995, *Multimedia: Making It Work* (New York: Osbourne McGraw-Hill).

WATERWORTH, J. A., 1992, *Multimedia Interaction: Human Factors Aspects* (Chichester: Simon & Schuster).

Natural Language Communication

V. Raskin

Department of English and Interdepartmental Program in Linguistics, 324 Heavilon Hall, Purdue University, West Lafayette, IN 47907-1356, USA

1. INTRODUCTION

"Natural language" is the term used to denote all human languages that have evolved in the course of human history and been used monolingually to serve fully the needs of the communities that employ them. There are ~5600 extant natural languages, e.g. live languages and dead languages, for which there are records and/or descriptions (e.g. Sanskrit, Latin, Old English).

Natural languages are distinct from such artificial languages as specially designed international languages for human use (e.g. Esperanto, Ido, Volapük) or programming languages for computer use (e.g. Fortran, C++, Java) in many ways but, most significantly, in the manner they are acquired rather than learned by the users. A live natural language is seemingly effortlessly acquired by the native members of its speech community, often after a pretty minimal exposure to it, by the age of ~30 months. This natural language acquisition faculty remains active for a few years and starts fading to disappear entirely in most individuals by the end of their pre-teen years. Artificial languages, along with foreign natural languages to which an individual may be exposed later in life, have to be learned at a significant cognitive and temporal expenditure; in the case of foreign language, a foreign accent is a sure give-away sign of the language having been learned rather than natively acquired.

Natural language is the basis of all human activity and, therefore, an important component of many fields of study. The main discipline, however, which studies natural language as a universal human faculty and the most observable function of the human mind is linguistics.

2. LINGUISTICS

One of the most ancient academic disciplines, dating back several millennia and associated closely, in antiquity, with the study of foreign languages and preservation of dying languages and sacred texts in them, linguistics finally won its independence from such related fields as philosophy, theology, and philology in the early 19th century by developing a particular historical domain that was of little interest to other fields, namely the study of language families and, later in the century, reconstruction of ancestor languages, such as proto-Info-European, with the help of a well-defined methodology based on a systematic comparison of germane words in descendant languages.

This emphasis on precise methodologies served linguistics well in the 20th century when a momentum for an algorithmic, mathematicalized study of language developed in the 1920s, with the advent of structuralism and the subsequent emphasis on synchronic description of language, largely at the expense of its historical exploration. The American branch of structural linguistics, Leonard Bloomfield's descriptive linguistics, became most influential in the 1940–50s after introducing a near-algorithmic procedure for a field description of an unknown language, starting with the collection of a corpus and continuing with the recursive application of the segmentation and distribution procedures from the lower to higher levels of language structure.

These levels define the central subdisciplines of linguistics. Phonetics and phonology study the sound of language: the former empirically so and the latter theoretically. Morphology studies such parts of the words as the root, prefix, suffix and infix, as well as the words themselves and their groupings into parts of speech, such as nouns, verbs, prepositions, etc. Syntax is concerned with the complex organization of words into phrases, clauses, and sentences. Semantics deals with the meaning in natural language; pragmatics with meaning in context. The so-called hyphenated disciplines, socio-linguistics, psycho-linguistics and neuro-linguistics deal with the functioning of language in society, the relations of language to human psychology, and the place of language in the study of the brain respectively.

The mathematicalization, or more accurately, formalization of linguistics was institutionalized by Noam Chomsky in the late 1950s, whose twofold contribution included the development of the theory of formal grammars as a branch of mathematical logic and of transformational generative grammar as an implementation of a formal grammar. A finite set of several types of grammatical rules is postulated as the syntax of a language, and the syntactic structure of each sentence in the language is a formal derivation resulting from a consecutive application of a subset of these rules to the initial symbol S (sentence). The phonological and semantic components, developed a little later, complement the dominant syntactic module of the Chomskian edifice for linguistics.

The conception of language which has emerged from this approach and dominated linguistics ever since, even though most contemporary linguists deviate from some of Chomsky's more specific claims, is that of an algorithmic rule-governed mechanism, internalized in the minds of the native speakers. Many linguists agree with Chomsky that there exists an innate predisposition for natural language acquisition, which is what makes native language acquisition so ostensibly easy and painless. Because such a predisposition cannot be designed for any specific natural language (because, of course, a baby may be exposed to any natural language and acquire it the same as it would any other language), it is concluded that such a faculty contains the rules of "universal grammar," i.e., the foundational features that are common to all natural languages. The search for language universals and their principles and parameters has been the declared goal of contemporary linguistics, even though, in practice, as in many other disciplines, most work is done on the technical aspects of the proposed formalisms and their numerous modifications.

3. NATURAL LANGUAGE AND ERGONOMICS
3.1. Human–Human Communication

All human activity is based on language, most of it on the natural language of the community or the chosen common natural language of a multilingual community, increasingly so English,

in the last decades. All activities, such as manufacturing, services, communication, etc. that are served by ergonomics are no exception, so natural language, its nature and rules, and the body of knowledge accumulated about natural language in linguistics and its applications are all essential for ergonomics.

The primary function of natural language is communication among humans. There are several different modes of human–human communication. It is customary, after Paul Grice's influential work in the 1950–70s, to think of the fact-conveying, *bona fide* mode of communication as primary. In this mode, the speaker and hearer are committed to the literal truth of what is said. A slight and necessary extension of the mode allows such non-literal devices as metaphors and implicatures as long as the speaker makes it clear to the hearer how the utterance is to be understood. This mode is based on complete linguistic cooperation — and trust — between the speaker and hearer. Other, non-*bona fide* modes of communication, such as humor, play-acting, advertising, propaganda and lying, have their own principles of cooperation but none of the modes implies a commitment to the truth of what is being said.

All of these modes have to contend with two essential characteristics of natural language that complicate human–human communication, namely, underspecification, sometimes rather misleadingly referred to as "vagueness," and ambiguity. Underspecification of reality by language means simply that no utterance is capable of containing all the details of the situation it attends to describe. Thus, when the speaker says, "John was late for the calculus class this morning," those hearers who know who John is and what calculus class is meant will have a feeling of complete understanding. The sentence, however, leaves an infinite number of questions unanswered.

Some of them are obvious and may actually be asked by some hearers as a follow-up of their understanding of the sentences. These are such questions as, Why was he late? or How late was he? Others are much less obvious and, correspondingly, rather unlikely to be asked: What was he wearing? Which of the two doors did he use? Did he have a large red plastic diamond-shaped box in his hands? What had he eaten before coming to class and when? Does he know Amy? Has he ever heard of Jack London? These examples are suggested in the increasing order of remoteness and even absurdity but each of the answers may become imminently important a second later if the conversation goes that way.

Underspecification is accepted by native speakers (and hearers) as a fact of life and is rarely commented upon or much researched. (The computer, however, cannot, accepted it.) Ambiguity is a different matter: much of what native speakers do with language is a pretty sophisticated and almost entirely unconscious procedure of disambiguation. The fact is that just about every word in a natural language has multiple meaning, many syntactic structures can be analyzed in two or more ways, and as an obvious result of that, sentences, which are made up of words put together by syntactic structures, tend to be at least potentially many ways ambiguous. Thus, the much-analyzed sentence "The man hit the colorful ball" is believed to be four-way ambiguous; the specially concocted of hackneyed semantic examples "The paralyzed bachelor hit the colorful ball" is 25-way ambiguous.

Native speakers negotiate this potentially disastrous situation skillfully by using sentences in disambiguating contexts and by providing to the hearers the extra information needed to understand each sentence in the intended meaning. Sure enough, this sometimes falls through and misunderstandings occur, but because these are unacceptable to speakers, every effort is made, in the process of Gricean cooperation, to prevent ambiguity from affecting comprehension. (Ambiguity is allowed and even expected in humor, poetry, fiction, advertising, and some religious practices, and the speakers and hearers handle it according to a different principle of cooperation.)

Obviously, then, underspecification and ambiguity in human–human communication in natural language has to be taken into account in ergonomic studies. The efficiency of human–human communication affects all aspects of productivity. The most standard commands, requests, and other constantly recurring messages must be kept as short, precise, and contrastive with regard to each other, as possible: a good example of an ergonomic change to increase contrastivity was the decision to introduce the non-existent word *zwo* instead of the regular *zwei* for "two" into the German dialect of soccer stadium announcers because, in the noisy channel of stadium public address communication, *zwei* could be mistaken for the rhyming *drei* "three," and a dangerous confusion about the score of the game might ensue.

Standard situations require standard contexts, and the effects of underspecification and ambiguity are skillfully minimized by such practices, whether the originators of the established procedure are aware of these language phenomena or not. Non-standard emergency situations have been known to produce uncontrolled and unintended ambiguity resulting in disasters. Rule changes often have the same effect. This affects public address announcements, posters, the adoption of new formulaic commands, messages, and exchanges over communication media, etc. There is a growing realization of these dangers in the areas of ergonomics and industrial engineering and an increasing number of linguistically trained experts.

3.2. Human–Computer Communication

Much, if not already most human productive activity is conducted now with the help of computers. Human–computer interaction and the human factors aspect of computer development and use have, therefore, achieved paramount importance. Linguistics has contributed crucially to the study of these phenomena through its application, computational linguistics, or natural language processing (NLP), often seen also as the natural-language branch of artificial intelligence (AI).

NLP designs and implements automatic systems that, typically, take text in a natural language as input, process it according to the predefined tasks, and then generate output, which may be in the same or another natural language or in some other stipulated format, such as, for instance, a database report or a chart.

Historically, the first task, for which NLP was used in the early 1950s, was machine translation (MT) between pairs of such best-known and -described languages as English and Russian or English and French. While realizing the impossibility of simple word-for-word translation and devising an ingenious system of overcoming this difficulty syntactically, MT soon ran into the "semantic barrier": it became clear to the MT pioneers that no high-quality fully automatic translation was possible without

building the understanding of the text into the system, and this was not considered feasible at the time — to some extent, because of ambiguity and underspecification. As a result, MT degraded to human-assisted post-editor based systems, some of which were found to be practically useful enough to last for decades, and even to machine-aided translation systems, in which humans performed all the intellectual tasks and computers performed simple but time-consuming tasks, such as dictionary look-ups. This temporary collapse of MT as a naive ambitious goal of replacing a huge army of expensive human translators and interpreters with machines gave both MT and NLP a bad name in engineering fields in the 1960–70s, which persisted longer than it should have in spite of the crucial changes in NLP.

In the mid-1980s, MT started making a spectacular comeback as a knowledge-based, meaning-oriented application. This was due partly to progress in general AI and partly to the increasing role of computational semantics, which has moved from the fledgling efforts of its pioneers in the 1970s to the breakthrough developments of the early 1990s in handling ambiguity and underspecification and on to its currently dominant position in NLP. Contrary to the prevalent NLP ideology of the late 1960s–late 1980s, when an enormous amount of talent, effort, and funds were invested in avoiding computational semantics, usually through increasingly detailed syntactic parsing, the current state of the art easily assumes the necessity for the computer to understand the meaning of the text it processes. This makes it possible to diversify NLP applications to intelligent information retrieval and Web searches, automatic abstracting, summarization, etc. In computational semantics, input text is converted into its meaning representation at the required level of granularity, the system manipulates these representations according to the prescribed task and then converts them into natural language text or any other format.

Throughout the brief history of NLP, this representation-based approach, whether syntax- or semantics-based, has had competition in the statistics-based approach, involving no representation of the text in any form or shape but rather trying to achieve the system task on the basis of word lists, frequency counts, etc. Conceptually simpler and lower in cost, the statistical approach has a lower limit on accuracy and on the range of possible applications but it has frequently been opted for commercially if accuracy is less important than cost. There is a healthy division of labor between the representative and statistical approaches, however, with some simpler tasks delegated to the latter and the more intellectually demanding tasks handled by the former, with the statistical output often providing valuable help.

One obvious ergonomic application of NLP has been in the development of user-friendly computer interfaces. A low-cost and accurate NLP interface would allow the human user to interact with the computer in the native natural language without any training in the computer's operating system or a programming language. An additional speech recognition module, translating sounds into spelling, would make the typing of input optional, and these modules have been steadily improving in quality and decreasing in cost and training time as well as making a natural pace of speech more acceptable to them. Optimum interface solutions are apparently of hybrid types, combining the most effective GUI and menu-based interfaces with the more accessible and lower-cost NLP ones.

REFERENCES

AGOU, S., RASKIN, V. and SALVENDY, G., 1993, Combining natural language with direct manipulation: the conceptual framework for a hybrid human–computer interface. *Behavior and Information Technology*, 12, 48–53.

HUTCHINS, W.J. and DOMERS, H.L., 1992. *An Introduction to Machine Translation* (London: Academic Press).

CHOMSKY, N., 1965. *Aspects of the Theory of Syntax* (Cambridge, MA: MIT Press).

MANNING, C.D. and SCHÜTZE, H., 1999. *Foundations of Statistical Natural Language Processing* (Cambridge, MA: MIT Press).

NIRENBURG, S. and RASKIN, V., 2000. *Principles of Ontological Semantics* (Cambridge, MA: MIT Press).

Product Sensorial Quality

L. Bandini Buti

Società di Ergonomia Applicata, Milan, Italy

1. INTRODUCTION

Usability is defined by the International Standard Organisation as "the effectiveness, efficiency, and *satisfaction* with which specific users obtain specific results in specific environments" (ISO DIS 9241-11). It introduces the concept of "satisfaction," which refers to the sphere of sensations. The sensorial quality of a product is, therefore, a conceptual category linked to the concept of usability.

Jordan (1997) suggests that products can be classified with respect to the user on three levels. The first is that of functionality. A product will result as useless if it does not have an adequate functionality. A product cannot be used or useful if it does not possess the functions necessary to the performance of the tasks for which it was designed. If a product does not possess the correct functionality it will render the user unsatisfied. To satisfy the needs of the user at this level, the ergonomic specialist must understand the destination of use of the product and the context and environment in which it will be used.

The second level is that of usability. When users are used to an adequate functionality they will want a product that is easy to use. An appropriate functionality is a prerequisite of usability, but does not guarantee it.

The third level is that of pleasure. After becoming used to a product, users will quite quickly want something more, a product that provides not only functional benefits, but also emotional benefits. The agreeability of a product represents a new challenge for ergonomics, a challenge that requires a comprehension of the intended user not only as a physical and cognitive processor, but also as a rational and emotional being, with values, tastes, hopes and fears. It is a challenge that requires a deep understanding of the product–person relationship: what will the properties of a product if it is to arouse and provoke certain emotional responses in the user? And how can a product convey a particular group of values? This is a challenge that requires the capture of the ephemeral — formulating methods and systems of investigation for the study and quantification of emotional responses.

2. SENSORIAL QUALITIES

Schopenhaur says (1819), "Primitive matter had to go through a decidedly long series of transformations before the first eye opened … this eye is the indispensable intermediary of consciousness, for which and in which only, the world can have a reality, and without which it is absurd to even think of such a thing, since the world is nothing other than representation, and as such needs a conscious subject to sustain its existence."

We can, therefore, consider the world as a provider of stimuli that, by means of the sensory canals (principally sight, hearing, touch), allow us to perceive the world and make us aware of our surrounding environment. This does not as yet regard the qualities that the stimuli possess. In fact, perception can be activated with equal intensity by effective but unpleasant stimuli or by equally effective but pleasant stimuli. This is the aspect that concerns the sensation of well-being.

Studies of safety, well-being, and comfort do not allow us to evaluate the sensations experienced by an individual when confronted with an object, a situation or an environment. In other words, subjects express satisfaction and pleasure for various solutions in relation to the sensations that the object transmits to them through their senses as much as in relation to efficiency and effectiveness. This leads us to think that the pleasantness transmitted to an individual by an object is, together with comfort, an aspect relevant to quality research. This is valid for every object we come into contact with. When examining a product for the first time, we unconsciously effect a synthesis of the sensations the product communicates to us and which manifest themselves as a sense of pleasure, indifference or repulsion. When one then experiments by touching the product, feeling its weight, using it, other stimuli are received and the sensations become more precise. Before being able to judge the product on its real characteristics, a judgement of "pleasantness" has already been made, and this will heavily condition any opinion of the product.

The responsibility for reaching high levels of aesthetic quality in products has traditionally been assigned to sensibilities of the designer, as decreed by the current success of the work of creative people such as fashion stylists and industrial designers. But such intuition is no longer sufficient for the introduction of high levels of quality in complex, mass-produced, industrial products, which require, as well as creative work, specific methods for the analysis and application of the sensorial qualities.

The market for durable goods, in particular for electro-domestics, other electrical appliances, and motor vehicles, is becoming ever more interested in quality with the aim of providing an answer to buyers who demand comfort and usability from systems, and comprehensibility of informative means and instructions. One needs not only to identify the sensations perceived as agreeable by the user, but to be able to evaluate them rigorously, propose them in the context of design projects, describe them in the specifications, reproduce them in the product, and to hone them by comparing proposals with results. Sensorial qualities (such as, for example, the tactility of a control unit and the sensory characteristics of its functioning) cannot, in fact, be developed by basing the relevant work exclusively on the individual capabilities and sensibilities of the designer. If they themselves are the only instrument for the evaluation of quality, they become a limiting factor, as they are the only "measurer" of the phenomenon.

2.1. Subjective Sensations

Some product characteristics (material, dimensional aspects, efforts required in use, contact surface characteristics, etc.) can be measured with instruments, and qualified numerically with values valid for time and space (objective data), while the
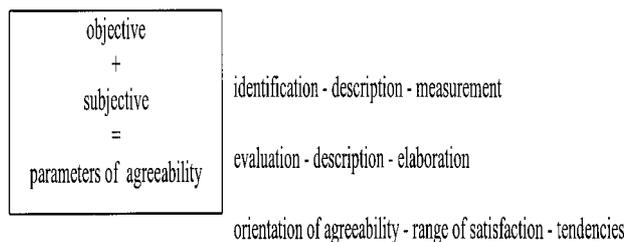

objective
+
subjective
=
parameters of agreeability

identification - description - measurement

evaluation - description - elaboration

orientation of agreeability - range of satisfaction - tendencies

Figure 1. Parameters of agreeability.

sensations they produce in those who enter into contact with them (tactile sensations, perceptions concerning ease of use, significance of colors, etc.) are linked to the complex interaction between product, use, and individual (subjective data).

The analysis of product agreeability must, therefore, regard both the measurable aspects of the product (objective parameters) and individual reactions to use (subjective sensations).

The objective of the analysis of sensorial characteristics is the evaluation of aspects of product agreeability, in order to be able to reproduce them and make them the object of specification, not differently from how the technical performance of materials and artifacts (design specifications, contract and supply specifications) are defined.

The subjective sensations felt by humans when they come into contact with objects are linked to the senses:

- tactile sensations: the form, dimensions, and weight of the product;
- prehensile sensations: the quality, softness, grip of surfaces;
- functional sensations: modes of use and operation;
- thermal sensations: conductibility and thermal capacity;
- chromatic sensations: the colors and finishes of the surfaces and the chromatism of the product; and
- acoustic sensations: the resonance of the material, the action, alarms and acoustic feedbacks.

There are also sensations linked to taste and smell, which, however, are to be considered as secondary in the evaluation of the sensorial characteristics of the product.

2.2. Sensorial Variables

Subjective "agreeability" or "pleasantness" is strongly linked to variable individual, cultural, sensorial, and temporal aspects:

- the *social or ethnic group culture* to which the subject belongs, is a cause of individual differences in the perception of quality in a product, as this depends on the history and experiences of the relevant group (e.g. in many parts of the Orient, white is the color of mourning, while in the West it is black), on modes of living (living in a town or city rather than in the country, etc.), on one's occupation, and so on.
- the *sensorial aspects* represent an element of differentiation in the perception of quality, since each individual has his or her own characteristics, which can derive from congenitive or acquired functional reductions (e.g. linked to age), or which derive from activities the subject carries out, such as the difference in tactile perception between a subject who performs heavy manual tasks (e.g. a builder) and another who, on the other hand, performs refined activities (e.g. a computer operator).
- the *temporal aspects* regard the alteration of the evaluation of sensations perceived over time. Certain color combinations that were judged by our fathers as vulgar and gaudy, are now normally appreciated; this is not to mention the evolution in the acceptance of different types of music in recent generations, etc.

3. AGREEABILITY

The evaluation of the emotional aspects of products requires specific study methodologies with the collection of objective data on products and the evaluation of sensations experienced by representative subjects in the direct contact of using products.

The evaluation of emotional aspects has required the development of a specific methodology for the study and introduction of the concepts of agreeability into product conception. The SEQUAM methodology (Sensorial Quality Assessment Method) was applied for the first time in 1992, to the internal components of motor vehicles. (The applications, beginning in 1992, have been effected by the Società di Ergonomia Applicata di Milano (Italy) on behalf of Fiat Auto (Turin), on those motor vehicle components in most direct contact with users.) It has as its objectives:

- The creation of scales of product satisfaction that, when associated with objective aspects, can highlight ranges of satisfaction (a satisfaction index of the products compared).
- The formulation of project specifications for the agreeability of the various product aspects (the most suitable and acceptable characteristics for surfaces, handle forms, softness, noise, etc.).
- The definition of orientations of agreeability that must guide designers.
- The effecting of predictive evaluations (that is, the projection of evaluations in the future).
- The elaboration of transmissible and comparable data on solid bases which may be applied during projection.

The theme for study can be one product or a collection of products dedicated to a specific function (e.g. a steering wheel, an environmental thermostat, a CD player, etc.). Experience tells us that studies of generic products, homogenous only in their functional typology (e.g. levers or push-buttons, leaving their function out of consideration) are to be avoided, as agreeability can never be disassociated from the modality and purpose of contact. Studies of products dedicated to the same function but with differing modes of operation/use (e.g. manual and automatic gear-boxes, or driver's and front passenger's seats in motor vehicles) are also to be avoided.

3.1. Methods of Testing

The evaluation of emotional aspects is divided into tests that allow us to evaluate the sensation of agreeability perceived by the individual:

1. *Laboratory tests* with products isolated from their context of use (e.g. steering wheels, gear-knobs, washing-machine controls presented on panels). Carrying out a test in a laboratory with an isolated product, allows one to use a relatively high number of samples, and to render the tests quicker with optimum time management, but it does not allow one to evaluate sensations linked to use.

2. *Simulation tests in a "natural" context* (electro-domestics fitted in experimental kitchens, motor vehicle parts fitted to simulatory models). These tests are carried using conditions that are close to those of real use, even if bringing subjects to the place of testing already creates conditions with a certain degree of artificiality. Compared with laboratory tests these are more realistic, but they are also more demanding in terms of time and cost.

3. *Tests in real conditions* are the closest to reality and consist of observations of individuals' behavior in their usual place of work/activity. They are onerous as it is the laboratory that comes to the subject and they can be limited by the fact that often it is only possible to test products that the subject

actually owns (e.g. in the study of electro-domestics in the home).

3.2. Operative Flow–chart

Studies carried out with subjects are very effective at describing how much is accepted and requested in a specific moment, but are, on the other hand, very ineffective in the evaluation of predictive situations. They give a good picture of today, not of tomorrow. But any action that has planning as its objective must first of all be concerned with tomorrow, since that which is analyzed has without doubt been designed a few years before, and that which is being designed will be on the market in a few years' time (with a gap that can be as large as 5 + 5 years in the case of motor vehicles). As a result of this, a design methodology that is content with studying the present runs the risk of staying in and consolidating the past, of becoming, in effect, an obstacle to evolution. For these reasons the SEQUAM methodology foresees that the operative plan is repeated during three sequential moments of study:

- study of the present;
- study of innovation; and
- verification with prototypes as a support to the development phase.

3.2.1. First phase – study of the present

This study is executed on products selected from among those on the market, chosen as being particularly interesting in relation to the various aspects of agreeability in such a way as to dispose of the widest possible range of components of the stimuli. The aim of this first phase is to obtain quantitative evaluations of the appreciation of the components of the stimuli and to orientate the design of mock-ups, necessary for the systematic study of agreeability. The study of the present develops through:

- the definition of the sensorial parameters that characterize

the product under examination (e.g. touch, rotation, opening, closing, etc.) and the identification of the qualities one wishes and is able to study (e.g. dimensions, weight, finishes of surfaces, etc.);

- the selection of products already manufactured that possess various gradations of the qualities one wishes to study. Every product will be defined by its measurable characteristics (dimensions, hardness, weight, resonance, range of movement delta, etc.), or its qualitative characteristics (genre of form, set of forms, surface characteristics, chromatic aspects, characteristics of operation, noise characteristics, etc.);

- the choice of subjects representative of the end-user. Subjects must not be specialists in the sector to which the product belongs (such as testers, maintenance workers, employees of the producing company), as they might be conditioned by familiarity with the product or by company loyalties. Subjects must be able to simulate the experience of using the product in real-life even though they are performing under laboratory conditions which differ from those of operating in real-life;

- the execution of analyses for the characterization of the parameters that have been the object of subjective analysis with numerical values;

- the execution of the study in a laboratory where conditions as close to real conditions of use as possible have been reconstructed. The test foresees that subjects will be observed (filmed) while they see/try products and they respond to open or specific (multiple choice etc.) questions; and

- the elaboration of data and the correlation of subjective data (scales of satisfaction) with objective data (measurement of parameters of agreeability of products) in order to define the satisfaction value of the various parameters tested. Data
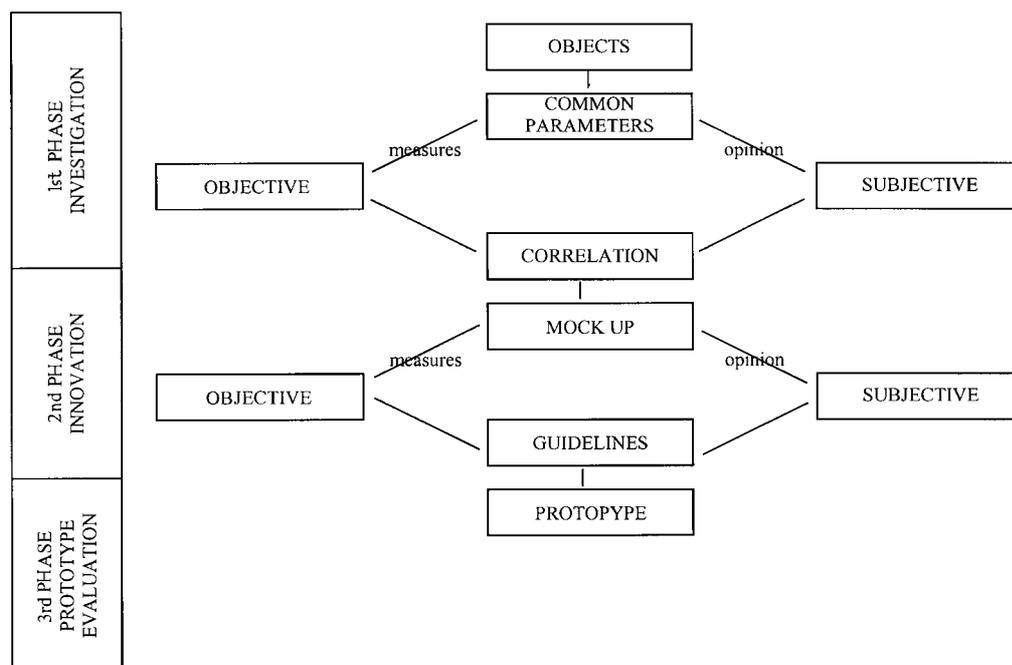

Figure 2. SEQUAM methodology foresees that the operative plan is repeated during three sequential phases of study.

will be continuously confirmed and enriched by the impressions collected, by suggestions, and by analysis of the filming of the tests.

3.2.2. Second phase — study of innovation

The study of the present gives an indication of the agreeability of those products already in production and, therefore, whilst significant for today, it tells us little about innovative trends. Thus, it is necessary to introduce into the study, a phase in which products created specifically for research purposes are analyzed. This phase foresees that the planning and realization of homogenous series of mock-ups, on which to conduct a study of the subjective response to the varying of individual components in situations of "low noise," that is, not 'polluted' by previous preconceptions. Models must allow us to:

- study parameters that in the first phase showed tendencies worthy of further investigation;
- analyze products that respond completely to the preferences emerging from the preceding phase;
- analyze the tendencies emerging from the study but not verified because of a lack of suitable products; and
- study innovative tendencies and trends.

The study methodology is the same as that applied to the first phase and, generally speaking, is executed with the same subjects. The questions are derived from those asked in the first phase, but are enriched by developments made during the planning phase and by the adaptations required by the technical characteristics of the models.

The results of these first two phases consist of project guidelines concerning the agreeability of the typologies studied. They must be considered as valid here and now, and thus to be re-verified in the course of time with the evolution of expectations of use and its characteristics.

3.2.3. Third phase — verification with prototype

Projects elaborated during research must not be intended as design projects in as much as they have not been conceived as integrated in an organic planning context, but, rather, have been elaborated for the specific ends of research and are often proposed for the study of limits of condition (e.g. grips or handles that are excessively large or small). Planning results obtained in the first two phases must be considered as a stimulus and means of orientation for the whole design process and as a means of rigorously evaluating aspects of agreeability and of rendering them easily transmissible.

The first two phases allow one to obtain the maximum amount of information within a reasonable period of time, as required by large industries. These conditions must be defined as "laboratory" conditions even in the most favorable of cases, in which functional products are available, as the studies take place in conditions more or less distant from real conditions of use. In fact, under experimental conditions, subjects tend to pay attention to the product itself, whereas in everyday use they are interested only in the results of their actions.

The final verification of the characteristics of a product's agreeability requires the carrying out of research with functional prototypes that possess all the characteristics of the finished product and that have been installed in coherent surroundings, both from the functional and formal point of view, and that can be used under real conditions by the generality of potential users.

When the development of a project allows for the production of prototypes that are, in terms of finish and function, fairly close to the finished product, the results of the first two phases may be the object of verification under real conditions of use on the part of a large sample of subjects.

During the project development of a product, in which ergonomics must form part of the planning group, verifications can be carried out on the various stages of the product comparing tests with the data collected in the two phases.

4. CONCLUSIONS

Studies of sensorial quality provide objective data (scales of values), subjective data on products (scales of satisfaction) and the comparison of objective and subjective data. The data as a whole enables us to formulate planning specifications for the agreeability of products, expressed as:

- optimum and/or maximum and/or minimum numerical values;
- trends and tendencies;
- values or tendencies that have specific validity for determined categories of products and/or users (e.g., sports cars, older users, etc.); and
- arguments that are not easily definable or that are heavily influenced by the immediate surroundings, but which must be taken into consideration and evaluated during the evolution of planning.

Studies do not produce design plans, but rather specifications for agreeability planning. In order to be effective they must be diffused as a systematic and transmissible approach to sensorial aspects and as a means of orientation that does not, however, limit the creative freedom of planners and designers. The application of the SEQUAM method permits dialogue on sensorial quality between planners, designers, and suppliers, based on objective data.

In fact, industries that produce on a large scale are characterized by an ever more complex organizational system, huge investments, and the pressing demands of the market, for which reason they need the concepts of ergonomy and agreeability to be able to enter into the normal flow of procedures so as to be then introduced into product development with the smallest margin of uncertainty.

In companies, studies carried out for the introduction of ergonomics into design projects can be directed towards a specific product in the course of its development or the generality of present and future models. When the objective is the development of a specific product, a work group is formed in which ergonomics researchers maintain a close contact with managers in charge of the planning and formal development of the model, with marketing services, suppliers, and purchasing departments. This renders the communication of results from studies easier, even if they are not expressed in a strictly formal way.

Formality, however, becomes indispensable when research studies have to be able to make use of the generality of present and future products, for which reason, the results of the studies, the principals and techniques of ergonomics and agreeability must be rapidly and effectively diffused at all levels of the company concerned. In other words, one must stimulate an orientation towards these themes of the structures that go to make up the complex system of planning, and which have to be able to gain

access to information in real time and with ease, in as much as few potential users (of whom there are a great number and who cannot really be predicted) have been able to have direct experience of research.

Data extrapolated form studies must be produced in a format that can be entered onto an Intranet site and available for all company operations interested in planning and design.

The system must enable:

- the transmission of data in a synthetic form that is easy for users with differing backgrounds to read (technicians, designers, stylists and administrators);
- the use of a graphic form adequate for rapid consultation; and
- the obtaining of more levels of reading with different degrees

of definition, facilitating reference to research literature in its entirety for the acquisition of details of research carried out and the justification of results obtained.

REFERENCES

BANDINI BUTI, L., 1998, *Ergonomia e Progetto* (Milan: Maggioli).

BONAPACE, L. and BANDINI BUTI, L., 1997, L'oggetto gradevole. *Ergonomia*, no. 10.

JORDAN, P.W., 1997, The four pleasures. In *Proceedings of the IEA 13th Congress*, Helsinki.

NAGAMACHI, M., 1989, *Kansei Engineering* (Tokyo: Kaibundo).

NORMAN, D.A., 1988, *The Psychology of Everyday Things* (New York: Basic Books).

SCHOPENHAUR, A., 1819, *Die Welt als Wille and Vorstellung* (Lipsia).

Speech-based Alarm displays

N. A. Stanton

Department of Psychology, University of Southampton, SO17 1BJ, UK

1. INTRODUCTION

Speech synthesis is receiving increasing attention as a potentially valuable medium for a variety of applications of information display. This paper addresses the use of speech as a means of displaying alarm information in "process control" type tasks. Typically the operator of a control desk will monitor the state of a process from a room that is sited remotely from the plant. Therefore, the operator's main source of information about the plant status is via process page displays presented on Visual Display Units (VDUs), which may contain up to 800 pages of information. Further, the process plant is controlled by a team of operators and a supervisor who are in contact, either face-to-face or via telephone, with plant engineers.

A speech-based display medium might have potential benefits to offer control room operation. Benefits that are often associated with speech displays include: breaking through attention, eyes-free—hands-free operation, omnidirectionality, no learning required, reduction in visual clutter, and public display. It has also been suggested that the auditory channel is particularly well suited to the transmission of warnings (Stokes, Wickens, and Kite 1990). Given the demands placed upon the operator in the control room, communication of alarm information using the auditory channel might therefore present a way in which better use might be made of their limited attentional resources.

The use of auditory displays in control rooms is not a new idea; in fact most control rooms employ non-speech auditory displays in conjunction with visual displays for conveying alarm information. However, non-speech warnings are clearly limited in terms of the amount of information that can be transmitted and the number of different types of signal that a human can discern (Stanton 1994). It is relatively easy to discriminate between up to seven different tone-based auditory warnings, but beyond this number it becomes much more difficult. Thus speech might be a more flexible and informative medium than tone warnings as this can be used to alert the operators to the problem, inform them of the nature of the problem, and cue the required response. However, Baber (1991) warns that although synthesized speech appears to be an attractive display medium for the presentation of warnings in control rooms, one needs to consider the appropriateness of speech for the tasks performed in the application domain before it is recommended. Baber (1991) also presents some design considerations regarding warning systems — for example, he suggests that the warning should sound distinct from human speech and that the message should be worded as a short phase containing a minimum of five syllables. These recommendations are intended to increase intelligibility and inform the operator that the message is from the machine and not from another human operator.

2. HUMAN VERSUS SYNTHESIZED SPEECH

In an experimental study reported by Baber, Stanton, and Stockley (1992), it was suggested that there might be a place for speech-based alarm displays in control room activities. In particular, they pointed out that a major benefit of speech in information-processing terms is that it can be used to relieve the visual channel. This was illustrated by introducing Wickens' (1984) model of information processing which suggests that visual–spatial information and auditory–verbal information draw on separate "pools" of attentional resources. Therefore transferring the alarm system's information from the visual display to an auditory channel might provide the human operator with greater capacity to deal with the incoming information and handle it appropriately. This could be seen as a possible solution in attempting to spread the mental workload demands of the task. Thus, in information processing terms, speech-based alarm systems might reduce the attentional demands on the visual channel. By comparing human and synthesized speech-based alarms, Baber *et al.* (1992) proposed that speech-based alarm displays need to be considered in terms of the tasks that operators are required to perform. For this purpose a model of alarm handling (Stanton, Booth, and Stammers 1992) was taken into the experimental laboratory so that the medium could be investigated in a more rigorous manner. From this study it was proposed that, in order to maintain a distinct difference between information communicated from the displays and information communicated between operators, synthesized speech was more appropriate than "human-like" speech. It was also suggested that speech might be suited to only a limited range of display information.

3. SYNTHESISED SPEECH VERSUS TEXT

Stanton and Baber (1997) investigated the operational performance differences in an experimental "process control" task by comparing speech and traditional scrolling text alarm displays. It was proposed that speech-based displays would be superior for attracting attention and single fault management (i.e. one alarm connected to one event), whereas scrolling text displays would be superior for more complex multi-alarm problems (i.e. many alarms connected to one event). The combination of speech and text should therefore lead to superior performance overall. Synthesis-by-rule was chosen over prerecorded messages for two main reasons. First, the alarm messages should be distinct from human speech. Second, in a real human supervisory control task there could be up to 20,000 alarms (such as in a nuclear power station). It would be a daunting task to prerecord all 20,000 messages, whereas synthesis-by-rule offers a practical solution. The surprise recall task used a pencil and paper so that participants could record the alarms they recalled.

Briefly, the experimental task required participants to conduct a planned activity in a simulated "process" plant. They had control over valves (open or close) and boiler heating (off, low, medium, high, or very high). They could also inspect the status of the plant elements — for example, tank levels, valve positions, and boiler temperature. Using this information, they were required to heat up a liquid until it was within predefined thresholds,

then they had to condense it whilst it was inside these thresholds. Several elements of the process could go wrong. For example, the source liquid could run out, the supply pipe could crack, the temperature of the boiler could be too hot or too cold, or the coolant tank could run out. Each of these problems had an associated alarm. The subject's goal was to "process" as much liquid as possible to achieve the maximum "output". In addition, the participant was requested to attend to a spatial secondary task when the workload on the primary "process" task permitted. The results generally suggested that performance was better in the text and speech and text (SandT) conditions when compared with the speech only condition. No statistical differences were found between the text and speech and text conditions. Interestingly, participants' performance in the speech only condition was generally significantly worse than that of participants in the other two conditions. This effect appeared to be reduced substantially when speech was combined with a scrolling text alarm display. Thus, one must consider why speech-based alarm displays appear to be detrimental to performance in a "process control" task.

4. DISCUSSION AND CONCLUSIONS

It has been suggested that the presentation of verbal information is inappropriate for fault diagnosis (Robinson and Eberts 1987). The suggestion is that the operator isolates the plant component in terms of its spatial reference rather than its verbal reference. Alarm information in the experimental task described within the second study was verbal in all three conditions, but the duration of the information was much shorter in the speech-based alarm condition. This is a more likely explanation of the findings, as all conditions were equally disadvantaged in "spatial reference" terms, i.e. in all of the conditions alarm information was presented verbally, as speech, scrolling text, or both. Baber (1991) also pointed out that poor quality synthesized speech and the resulting high memory demands this makes on operators could degrade performance. The latter of these probably had the greatest effect.

The results from the surprise recall task show quite clearly that participants in the speech-only condition were unable to recall as many alarm messages as participants in the other two conditions. This supports the findings of the first study that report recall performance for a synthesized speech condition was very poor. This suggests that synthesized speech is processed at a surface level, rather than at a semantic level, and could account for the poor recall performance. Thus, one could surmise that synthesized speech is inappropriate for tasks that involve a memory component.

Speech also has a "durational" component, which may lock people out of the interaction until the message is complete. This would occur even if the speech is of a high quality, whereas visual displays can be sampled at the operator's pace and are permanently displayed. Speech presents a paradox of operation: participants are required to respond immediately or the message will be lost, but have to wait until the message has ended before responding.

Stokes *et al.* (1990) suggest that speech-based warnings serve as both an attention attractor and as a channel for transmitting information about the failure. Although the synthesized speech used in both studies was distinct from human speech, participants in the speech-only condition were significantly slower to accept alarms than participants in the other two conditions. Participants

in the speech-only condition must have either ignored the message or failed to realize that the spoken alarm required an immediate manual response. Following a series of experiments examining fault management in process control environments, Moray and Rotenberg (1989) claimed that operators might observe abnormal values, but fail to act because they are already busy dealing with another fault and may wish to finish that problem before turning to a new one. They term this phenomenon "cognitive lock-up", and note that human operators have a preference for serial, rather than concurrent, fault management. Therefore, new alarm information is likely to be ignored until the operator is free to deal with it. If this alarm information is presented through a transitory medium, then it will be lost. This explanation of fault management suggests that synthesized speech-based alarm display systems are not appropriate for process control tasks. There is some evidence to suggest that irrelevant speech may have adverse effects upon performance (Smith 1990). The effects appear to occur independently of intensity. This could also lend some explanation to the study in this paper. Consider the participant working to maintain the process whilst speech alarms are being presented: the alarm information may be described as "irrelevant" if it does not relate to the particular task in hand, and performance is disrupted.

In conclusion, it is suggested that speech alone as a medium for alarm displays cannot be recommended for tasks where there is: (a) a memory component; (b) likely to be some delay before the fault is attended to; and (c) likely to be more than one alarm presented at a time, and the operator is required to assimilate information from a variety of sources using spatial reference. If speech is to be incorporated into the alarm system for "process control" tasks, it is recommended that it is paired with other media such as a text-based display. However, speech-based alarms might be appropriate for tasks where: an immediate response is required; the "operator" is away from the control desk; the situation is typically one-alarm to one-event and fault management is serial in nature.

REFERENCES

BABER, C., 1991, *Speech Technology in Control Room Systems: a human factors perspective* (Ellis Horwood: Chichester).

BABER, C., STANTON, N.A. and Stockley, A., 1992, Can speech be used for alarm displays in 'process control' tasks? *Behaviour and Information Technology* 11(4): 216–26.

MORAY, N. and ROTENBERG, I., 1989, Fault management in process control: eye movements and action. *Ergonomics* 32(11): 1319–42.

ROBERTSON, C.R. and EBERTS, R.E., 1987, Comparison of speech and pictorial displays in a cockpit environment. *Human Factors* 29: 31–44.

STANTON, N.A., 1994, *Human Factors in Alarm Design* (Taylor & Francis: London).

STANTON, N. A. and BABER, C., 1997, Comparing verbal media for alarm handling: speech versus textual displays. *Ergonomics* 40(11): 1240–54.

STANTON, N.A., BOOTH, R.T. and STAMMERS, R.B., 1992, Alarms in human supervisory control: a human factors perspective. *International Journal of Computer Integrated Manufacturing.* 5(2): 81–93.

STOKES, A., WICKENS, C.D. and KITE, K., 1990, *Display Technology: Human Factors Concepts* (Warrendale, PA: Society of Automotive Engineers).

SMITH, A., 1990, Noise, performance, efficiency and safety. *Int. Arch. Occup. Environ. Health* 62: 1–5.

WICKENS, C.D., 1984, *Engineering Psychology and Human Performance* (Columbus, OH: Merrill).

Structured Integration of Human Factors and Software Engineering Methods

K. Y. Lim

Nanyang Technological University, School of Mechanical and Production Engineering, Nanyang Avenue, Singapore 639798

1. INTRODUCTION

A perennially reported problem of human factors input to interactive systems development is that its contributions are often requested late. In particular, human factors inputs are frequently excluded from design specification and its involvement is confined mainly to late evaluation. Design inputs (human factors or otherwise) at this late stage of system development are difficult and costly to implement and thus are less likely to be realized in full. Here, the design of the system would have been more or less fully developed, with interlocking dependencies that would imply wide ranging design disruption if modifications were to be implemented. In addition, affective considerations may complicate matters. For instance, proposed design changes might violate feelings relating to design "ownership" (leading to undue defensiveness on the part of the system designer and avoidance of potential offense or "stepping on toes" on the part of the ergonomist). Further, design modifications proposed at this late stage tend to be subjugated by other more pressing concerns relating to project budget and schedule. The result is thus a resistance to design modifications, manifested as the "too-little-too-late" problem of human factors input (Lim and Long 1994).

This poor state of human factors input may have arisen paradoxically, from too rapid a recognition of its importance around the time of the personal computer (PC) revolution. At this time (1980s), the scope of human factors input to interactive systems development was poorly defined. Human factors became a victim of its own success, as the development of its methods could not keep pace with the rapid increase in demand for a wider scope of design contribution. Human factors methods then comprised largely of an incomplete collection of disparate techniques, traditionally focused on design evaluation as opposed to design specification (Lim 1996). The scope of its methods (such as task analysis) was thus found to be limited and incomplete in the coverage of the system design cycle. Further, the human factors design process was criticized as largely implicit, and consequently impossible to time or intersect with the design activities and contributions from other disciplines. Worse, since the scope and process of human factors design were incomplete and implicit, its project resource demands could not be anticipated and planned for. The result was a lack of management support and commitment, and the prolonged exclusion of human factors from the project budgeting and planning agenda. In the absence of an explicit allocation of resources, human factors considerations were observed to be accommodated only opportunistically and sacrificed first (when needed most) when

a project hitch arose. In these circumstances, even when project resources were allocated to human factors design, they were encroached upon as managers could not support their retention on the basis of its implicit design scope and process requirements. In an attempt to cope with these difficult constraints, the human factors inputs that could be proffered were found to be inadequately explicit, or worse no more than advice. Consequently, in the late 1980s, it became clear that existing human factors design processes and methods had to develop in earnest.

2. TWO PROMISING SOLUTION APPROACHES: RAPID PROTOTYPING AND STRUCTURED ANALYSIS AND DESIGN METHODS

To advance the methodological foundation of human factors design, and so counter the "too-little-too-late" problem of its input to interactive systems development, the design framework offered by two key system development approaches were examined; namely rapid prototyping and structured analysis and design methods. The objective was to assess their potential to support the location and specification of more explicit processes and methods for human factors design. These system development approaches will now be considered briefly in turn.

First, rapid prototyping that advocates the rapid construction of a design artifact was developed to facilitate the iterative cycle of design modification and test. Although the approach offered many advantages and solved a number of problems of human factors input, some key concerns remained unaddressed; in particular the following:

- A rapid prototyping approach usually generated little or no design documentation. This is due to its emphasis on the construction of a design artifact directly on a computer.
- The emphasis of rapid prototyping on the fast generation of an artifact is predisposed towards inadequate requirements analysis and design.
- A rapid prototyping approach may relegate human factors input to prototype evaluation as opposed to active involvement in design specification. Poor designs may thus be discovered late.
- Although rapid prototyping was originally equivalent to throwaway prototyping, it was not always the case in reality. Instead, there was considerable resistance to discard a prototype since considerable effort was usually invested in its construction. The practice may be attributed to the reluctance of designers to show crude prototypes to clients and end-users, due either to professional pride or the fear of losing credibility. The result may be premature commitment to a design. Late discovery of a poor design may again arise.

Structured analysis and design methods, in contrast, advocate comprehensive design analysis and specification before the undertaking of any design implementation. Examples of such methods include Structured Systems Analysis and Design Method (SSADM), Structured Analysis and Structured Design (SASD) Method, Yourdon and Constantine Method, and Jackson System Development (JSD) Method, to name a few. This class of methods was developed with the key objective on ensuring the design of a correct system (this may be contrasted with formal methods which exploit algebraic descriptions to ensure "correct"

development of a system). Structured analysis and design methods are promising as they provide the following:

- Complete and comprehensive life-cycle support for system development.
- Explicit methodological characteristics with well-defined scope, process and documentation/notation, for each stage of system development. Project planning and quality audit are, therefore, well supported.

The explicit methodological characteristics of structured analysis and design methods afforded the following advantages:

- A well-developed framework that facilitates the definition or identification of more specific intersections and timing of cross-disciplinary design activities and contributions, including those of human factors design.
- Explicit intermediate design descriptions that support design communication among designers, end users and managers.
- A set of comprehensive design notations which may be exploited by human factors to enhance the specificity of its design descriptions, and hence the potential of a common language across contributing disciplines.

Although structured analysis and design methods showed promise, they were targeted at addressing Software Engineering design concerns only. In particular, a survey undertaken in 1987 revealed that user requirements analysis and specification and user interface design, were excluded from their design scope. Thus, these human factors design processes had to be incorporated into the methods. Unfortunately, existing human factors design methods were inadequately explicit, and direct incorporation into existing structured analysis and design methods could not be achieved. Further, to specify appropriate intersections and timing of cross-disciplinary design activities and contributions, an explicit and complete human factors design method would be required. Since such a structured human factors method did not exist in the late 1980s, it would have to be developed. In this respect, the framework offered by software engineering structured analysis and design methods could be used to guide the development of a structured human factors method. The latter may then be integrated explicitly with structured software engineering methods to ensure more efficient system development. Work in this direction began in earnest at this time, as evidenced by the work of Walsh *et al.* (1988), Sutcliffe (1988), Damodaran *et al.* (1988) and Blyth and Hakiel (1989). A brief account of the resulting conceptions of cross-disciplinary methodological integration follows.

3. EXEMPLARS OF STRUCTURED INTEGRATION OF HUMAN FACTORS AND SOFTWARE ENGINEERING METHODS: AN OVERVIEW

Here, published work on the structured integration of human factors with leading software-engineering methods is briefly reviewed. In particular, conceptions of methodological integration of human factors with the JSD, SASD and SSADM methods are examined.

3.1. Integration of Human Factors with the JSD Method

The structured integration of human factors and JSD methods was the most actively pursued in the late 1980s, as JSD was the preferred method for real-time systems development. Besides Lim *et al.*'s (1990) work on integrating MUSE with JSD (named MUSE–JSD), a second conception of such an integration was also published by Sutcliffe and Wang (1991).

MUSE–JSD is the result of a project commissioned specifically to develop an explicitly integrated human factors and software engineering method. From the outset, JSD was targeted as the software engineering method for integration. As a similarly structured human factors method did not exist, the initial work of Walsh *et al.* (1988) was focused on the development of a structured human factors method. Several versions of such a method was developed, culminating in a generic structured human factors method named MUSE. As the scope, process, products, procedures and notation of each design stage are defined explicitly in MUSE, specific intersections and timing of human factors activities and contributions could be identified with respect to specific JSD design stages. Obligatory cross-disciplinary reference or contact points are thus specified to ensure a common design basis and information pool. These contact points help to ensure the development of a consistent and convergent system design. As MUSE recruits the structured diagram notation from the JSD method, the potential of a common notation is exploited to facilitate cross-disciplinary design communication. From the literature, it would appear that MUSE–JSD offers the most explicit and developed conception of methodological integration. For a detailed account of MUSE–JSD, the reader is referred to a related article included in this encyclopaedia, and to Lim and Long (1994).

Sutcliffe and Wang's (1991) work on methodological integration was directed at recruiting existing human factors techniques to extend the scope of the JSD method. Their main concern was to locate task analysis and user–computer dialogue specification against pertinent JSD design stages. In particular, they suggested that descriptions of user task activities may be linked to the JSD model by extending the scope of JSD interactive function specification. Further, as in MUSE–JSD, Sutcliffe and Wang (1991) also supported the use of JSD structured diagrams as a common notation for cross-disciplinary design specifications. Concerning user–computer dialogue specification, three intersections were proposed between human factors and JSD design specifications:

- Actions of JSD entities should be mapped onto actions of user interface objects. Their argument for this mapping was that the JSD model is essentially an object and event model. For instance, they suggested that user inputs and errors may be related to the system context defined by the JSD model.
- User–computer dialogue specification should be based on the functional supports required by the user's task.
- User–computer dialogue specifications should be linked to JSD input subsystem specifications. Sutcliffe and Wang proposed that dialogue control requirements for error recovery, and prompts and feedback messages, may be specified as simple JSD filter processes. To avoid segmentation of the user task description into incoherent units scattered across JSD filter process specifications, they reasoned that the latter specifications would have to be modified.

Although Sutcliffe and Wang proposed some interesting ideas for integrating human factors and JSD methods, their conception

did not include the development of a structured human factors method. Thus, it is unclear from the literature whether their proposed methodological integration could be achieved.

3.2. Integration of Human Factors with the Structured Analysis and Structured Design (SASD) Method

In response to in-house demands, Blyth and Hakiel (1989) took up the challenge of integrating human factors design with SASD, their in-house software engineering method. Their efforts involved incorporating task analysis as a basic activity within the SASD specification of an Essential System Model. Function allocation activities were also intersected with the SASD specification of a System Function Model. With these extensions, the existing machine-oriented perspective of the Essential System Model may be tempered with a user-oriented perspective. Further, to extend the scope of SASD to include the specification of human–computer interactions, Blyth and Hakiel conducted a review of existing human factors techniques. On the basis of the review, they constructed an eight-level conception of system specification as follows:

1. Goal and Task Levels: detailed analysis of goals, tasks and function allocation. A set of task models is thus derived from domain abstraction.
2. Conceptual Level: definition of conceptual objects to be represented at the user interface.
3. Semantic, Syntactic, Lexical, Alphabetic and Physical Levels: top-down specification of human–computer interactions.

These levels of system specification are supported by five sequential steps to advance human factors design concerns; namely analysis of extant systems; user characterization; task analysis; function allocation; conceptual model specification.

In summary, the work of Blyth and Hakiel (1989) addressed the much-needed extension of early stages of the SASD method, to include human factors design concerns and activities. However, the potential of a common notation to facilitate cross-disciplinary design specification, was not exploited.

3.3. Integration of Human Factors with the Structured System Analysis and Design Method (SSADM)

The integration of human factors with SSADM was undertaken by Damodaran *et al*. (1988) as part of a consultancy project. Their objective was to incorporate into SSADM four areas of human factors design; namely user analysis, job design, task allocation and prototyping. Strangely, user interface specification was excluded from their scope of integration. Since the scope of human factors design addressed was incomplete, it is unclear how high level descriptions of task allocation and job design

could be translated into system specifications. In particular, procedural details of their integrated method were obscure. It seems that Damodaran *et al*. were only concerned with locating modules of declarative human factors knowledge (e.g. checklists, design guidelines) against specific stages of SSADM. This limited focus could be attributed to a key requirement of their project; namely that the integrated method must support application by a designer without formal human factors training. The aim of the project is, therefore, rather ambitious. However, there is no evidence from the literature that this requirement has been met. Similarly, it is doubtful that methodological integration was achieved. From the evidence, it can only be concluded that their work achieved a loose or high level mapping of a limited set of existing human factors techniques and concerns, onto specific design stages of SSADM.

This account completes an overview of the integration of human factors with specific software engineering methods.

REFERENCES

BLYTH, R.C. and HAKIEL, S.R., 1989, A user interface design methodology and the implication for structured systems design methods. In *Proceedings of the IEE 3rd International Conference on Command, Control, Communications and Management Information Systems,* Bournemouth, UK, 2–4 May.

DAMODARAN, L., IP, K. and BECK, M., 1988, Integrating human factors principles into structured design methodology: a case-study in the UK civil service. In H.J. Bullinger *et al*. (eds), *Information Technology for Organisational Systems* (Amsterdam: Elsevier Science), 235–41.

LIM, K.Y., 1996, Structured task analysis: an instantiation in the MUSE method for usability engineering. *Interacting with Computers*, 8, 31–50.

LIM, K.Y. and LONG, J.B., 1994, *The MUSE Method for Usability Engineering* (Cambridge: Cambridge University Press).

LIM, K.Y., LONG, J.B. and SILCOCK, N., 1990, Integrating human factors with structured analysis and design methods: an enhanced conception of the extended Jackson System Development method. In D. Diaper *et al*. (eds), *Proceedings of the 3rd IFIP Conference on Human–Computer Interaction (INTERACT '90)* (Amsterdam: Elsevier Science), 225–30.

SUTCLIFFE, A., 1988, Some experiences in integrating specification of HCI within a structured system development method. In D.M. Jones and R. Winder (eds), *Proceedings of BCS HCI'88, People and Computers IV* (Cambridge: Cambridge University Press), 145–60.

SUTCLIFFE, A. and WANG, I., 1991, Integrating human–computer interaction with Jackson System Development. *Computer Journal*, 34, 132–42.

WALSH, P., LIM, K.Y., LONG, J.B. and CARVER, M.K., 1988, Integrating human factors with system development. In N. Heaton and M. Sinclair (eds), *Designing End-User Interfaces*. State of the Art Reports 15:8 (Oxford: Pergamon Infotech), 111–20.

Top Ten Mistakes in Web Design

J. Nielsen

Nielsen Norman Group, 2704 Fairbrook Drive, Mountain View, CA 94040, USA

1. INTRODUCTION

Design of home pages on the World-Wide Web is very challenging. Some of the common mistakes that designers make are discussed below and include:

- Using frames.
- Gratuitous use of bleeding-edge technology.
- Scrolling text, marquees, and constantly running animations.
- Complex URL.
- Orphan pages.
- Long scrolling pages.
- Lack of navigation support.
- Non-standard link colors.
- Outdated information.
- Overly long download times.

2. TOP TEN MISTAKES IN WEB DESIGN

2.1. Using Frames

Splitting a page into frames is very confusing for users since frames break the fundamental user model of the web page. All of a sudden, you cannot bookmark the current page and return to it (the bookmark points to another version of the frameset), URL stop working, and printouts become difficult. Even worse, the predictability of user actions goes out the door: who knows what information will appear where when you click on a link?

2.2. Gratuitous Use of Bleeding-edge Technology

Do not try to attract users to your site by bragging about use of the latest web technology. You may attract a few nerds, but mainstream users will care more about useful content and your ability to offer good customer service. Using the latest and greatest before it is even out of beta is a sure way to discourage users: if their system crashes while visiting your site, you can bet that many of them will not be back. Unless you are in the business of selling Internet products or services, it is better to wait until some experience has been gained with respect to the appropriate ways of using new techniques. When desktop publishing was young, people put twenty fonts in their documents: let's avoid similar design bloat on the Web.

As an example: use VRML if you actually have information that maps naturally onto a three-dimensional space (e.g. architectural design, shoot-them-up games, surgery planning). Do not use VRML if your data are N-dimensional since it is usually better to produce two-dimensional overviews that fit with the actual display and input hardware available to the user.

2.3. Scrolling Text, Marquees, and Constantly Running Animations

Never include page elements that move incessantly. Moving images have an overpowering effect on the human peripheral vision. A web page should not emulate Times Square in New York City in its constant attack on the human senses: give your user some peace and quiet actually to read the text. Of course, <BLINK> is simply evil. Enough said.

2.4. Complex URL

Even though machine-level addressing like the URL should never have been exposed in the user interface, it is there and it has been found that users actually try to decode the URL of pages to infer the structure of web sites. Users do this because of the horrifying lack of support for navigation and sense of location in current web browsers. Thus, a URL should contain human-readable directory and file names that reflect the nature of the information space.

Also, users sometimes need to type in a URL, so try to minimize the risk of typos by using short names with all lower-case characters and no special characters (many people do not know how to type a '~').

2.5. Orphan Pages

Make sure that all pages include a clear indication of what web site they belong to since users may access pages directly without coming in through your home page. For the same reason, every page should have a link up to your home page as well as some indication of where they fit within the structure of your information space.

2.6. Long Scrolling Pages

Only 10% of users scroll beyond the information that is visible on the screen when a page comes up. All critical content and navigation options should be on the top part of the page. More recent studies show that users are more willing to scroll now than they were in the early years of the Web. It is still recommend to minimize scrolling on navigation pages, but it is no longer an absolute ban.

2.7. Lack of Navigation Support

Do not assume that users know as much about your site as you do. They always have difficulty finding information, so they need support in the form of a strong sense of structure and place. Start your design with a good understanding of the structure of the information space and communicate this structure explicitly to the user. Provide a site map and let users know where they are and where they can go. Also, you will need a good search feature since even the best navigation support will never be enough.

2.8. Non-standard Link Colors

Links to pages that have not been seen by the user are blue; links to previously seen pages are purple or red. Do not mess with these colors since the ability to understand what links have been followed is one of the few navigational aides that is standard in most web browsers. Consistency is key to teaching users what the link colors mean.

2.9. Outdated Information

Budget to hire a web gardener as part of your team. You need somebody to root out the weeds and replant the flowers as the website changes but most people would rather spend their time creating new content than on maintenance. In practice, maintenance is a cheap way of enhancing the content on your

website since many old pages keep their relevance and should be linked into the new pages. Of course, some pages are better off being removed completely from the server after their expiration date.

2.10. Overly Long Download Times

This issue is placed last because most people already know about it; not because it is the least important. Traditional human factors guidelines indicate 10 s as the maximum response time before users lose interest. On the web, users have been trained to endure so much suffering that it may be acceptable to increase this limit to 15 s for a few pages.

Even websites with high-end users need to consider download times: it has been found that many customers access Sun's website from home computers in the evening because they are too busy to surf the web during working hours. Bandwidth is getting worse, not better, as the Internet adds users faster than the infrastructure can keep up.

3. TOP TEN MISTAKES REVISITED 3 YEARS LATER

My paper from May 1996, 'Top ten mistakes in Web design' is still surprisingly relevant today, 4 years later. The paper has become a minor Web classic with ~400 000 page views so far. It is still read 17 000 times per month. Even discounting the possibility that some people have read the article more than once, a readership of 400 000 means that the 'top ten mistakes' have been read by < 10% of the people responsible for the world's 5 million websites. So most of these mistakes are still being made and I still recommend that new Web designers read the older article.

4. ARE THE MISTAKES STILL WRONG?

4.1. Frames

Frames are no longer the disaster they were in 1995 and early 1996 due to some advances in browser technology: Netscape fixed the Back button with version 3, and since virtually nobody uses version 1 and 2 any more, this means that users can now navigate through frames with fewer problems. Version 4 reduced the problems printing frames (though users still often get a different printout than they expected), and Internet Explorer 5 has finally regained the ability to bookmark pages despite the use of frames. Frames still prevent users from e-mailing a recommended URL to other users and they also make the page clumsier to interact with.

4.2. Bleeding-edge Technology

If anything, users have less patience for bleeding-edge technology these days as the Web gets dominated by later adopters and the upgrade speeds for new browsers and plug-ins slow down. Users who encounter as much as a single JavaScript error usually leave a site immediately. It is just not worth the time to figure out how to make something work when there are 5 million other sites to go to.

4.3. Scrolling Text and Looping Animations

It is as hard as ever to read scrolling text, but aggressive use of distracting animation now causes even more problems than in 1996: users have started equating such designs with advertising which they routinely ignore. These days, it is extremely important for any content and navigation elements to look very different than prevailing advertising designs since users tune out anything that they do not think will be relevant to their task. Very severe

4.4. Complex URL

Users pay less attention to URL these days than they did in the early days of the Web. Since most sites now have navigation support, users are also relying less on the URL to tell them about their location on the site. But long URL still cause problems when users e-mail page recommendations to each other.

4.5. Orphan Pages

Less likely to make users stuck since most people have learned the trick to get to the home page of a site by 'hacking' the end off the URL. Still a disaster for novice users; still annoying for experienced users.

4.6. Scrolling Navigation Pages

Of users, 90% used not to scroll navigation pages but simply pick from the visible options. This has changed since most Web users now know that pages scroll and that important links sometimes are not visible 'above the fold.' Even so, the visible options still dominate and users sometimes overlook alternatives lower down the page. This is particularly bad if the visible part of the page seems to clearly communicate a certain purpose or a certain best approach: users may then happily conclude that they know what to do and not bother spending time on the rest of the page.

4.7. Lack of Navigation Support

Rarely seen, but a problem when it occurs. People are now getting used to certain canonical navigation elements such as a site logo in the upper left corner (linked to the home page) or a clear indication of what part of the site the current page belongs to (linked to the main page for that section). So if these elements are missing, users feel lost. Severe

4.8. Non-standard Link Colors

Continues to be a problem since users rely on the link colors to understand what parts of the site they have visited. I often see users bounce repeatedly among a small set of pages, not knowing that they are going back to the same page again and again. (Also, because non-standard link colors are unpleasantly frequent, users are now getting confused by any underlining of text that is not a link.)

4.9. Outdated Information

Worse now since so many other sites on the Web are continuously updated. Also, with the growth in e-commerce, trust is getting increasingly important, and outdated content is a sure way to lose credibility. (Note that archival information and information about old products are plusses and very different from outdated information.)

4.10. Slow Download Times

Contrary to many Internet pundits' pronouncements, the bandwidth problem has not been solved during the past 3 years; nor will it be solved during the next 3 years. Not until 2003 will high-end users have sufficient bandwidth for acceptable Web

response times. Low-end users have to wait until about 2008. I conclude that:

- All 10 mistakes from 1996 are still mistakes in 2000.
- Nine of the 10 mistakes still cause significant usability problems and should be avoided in modern websites

Scrolling navigation pages cause fewer usability problems these days and can be allowed if caution is taken in their design (any time you have overly long navigation pages, I would take it as a warning signal and as a requirement for usability testing).

Universal Design in Human–Computer Interaction

C. Stephanidis, D. Akoumianakis and A. Savidis

Institute of Computer Science, Foundation for Research and Technology-Hellas, Science and Technology Park of Crete, Heraklion, Crete, GR-71110, Greece

1. INTRODUCTION

In the past, accessibility in Human–Computer Interaction (HCI) was primarily concerned with the selection of suitable equipment to enable alternative computer access for people with disabilities. As a result, it was mainly considered as an afterthought and reflected a *reactive* approach, whereby Assistive Technology solutions addressed problems introduced by a previous generation of technology (Savidis and Stephanidis 1995, Stephanidis 1995, Akoumianakis and Stephanidis 1999). This reactive approach entails primarily *adaptations*. Adaptations facilitate access to the interface via suitable mechanisms, such as filtering, dedicated interaction techniques, such as scanning and specialized input/output devices (e.g. Braille displays, switches, eye-gaze systems). Typically, the result of adaptations includes the reconfiguration of the physical layer of interaction, and, when necessary, the translation of the visual interface manifestation to an alternative modality. For example, access to a graphical user interface (GUI) by a blind user requires "filtering" of the contents of the screen, using appropriate software (e.g. screen reader), so as to present them in an alternative modality (e.g. tactile, audio).

Despite the short-term benefits that such *a posteriori* adaptations may bring about, it is important to mention that there are serious shortcomings that render this approach inadequate in the long run. Some of these shortcomings have been identified in the relevant HCI literature. For the purposes of this chapter, it is important to add the following: first, adaptations introduce a programming-intensive approach towards accessibility, which increases the cost of implementing and maintaining accessible software. Second, technological progress may render adaptations harder to implement; there may be restrictions imposed either by the target application or by the operating system. Finally, it is increasingly recognized that, in the context of the emerging Information Society, accessibility should no longer be considered as mere translation of visual interface manifestations to alternative modalities (e.g. *a posteriori* adaptations), but as a requirement demanding a generic solution (Mueller *et al.* 1997, Stephanidis *et al.* 1998, 1999). As a result, there have been several efforts in the direction of advancing and articulating the principles of universal design to address a range of quality attributes, including accessibility, in the context of the emerging Information Society.

2. UNIVERSAL DESIGN

2.1. Definitions

Universal design or *design for all* (the two terms are used interchangeably) is frequently associated with different connotations. Some consider it as a new, politically correct term, referring to efforts to introduce "special features" for "special users" in the design of a product. To others, universal design is a deeply meaningful and rich topic that elevates what designers like to call "good user-based design" to a more encompassing concept of addressing the needs of all potential users.

Here (see also Stephanidis *et al.* 1998), the term is used to reflect a new concept or philosophy for design that recognizes, respects, values and attempts to accommodate the broadest possible range of human abilities, requirements and preferences in the design of all computer-based products and environments. Thus, it promotes a design perspective that eliminates the need for "special features" and fosters individualization and end-user acceptability. As already pointed out, the term is used interchangeably with the term *design for all* users. This does not imply a single design solution suitable for all users. Instead, it should be interpreted as an effort to design products and services in such a way so as to suit the broadest possible end-user population. In doing this, it is more than likely that there will be different solutions for different contexts of use.

Thus, universal design in the context of HCI is the conscious and systematic effort to proactively apply principles, methods and tools, in order to develop Information Technology and Telecommunications (IT&T) products and services, which are accessible and usable by all citizens, avoiding the need for *a posteriori* adaptations or specialized design. The rationale for universal design is grounded on the claim that designing for the average user, as the case has been with traditional design, leads to exclusionary designs and does not provide for the needs of the broadest possible population. As a result, the normative perspective of universal design is that there is no average user and consequently design should be targeted towards the broadest possible population.

In this context, universal accessibility implies the global requirement for access to information by individuals with different abilities, requirements and preferences, in a variety of contexts of use; the meaning of the term is intentionally broad to encompass accessibility challenges as posed by diversity in: (1) the target user population (including people with special needs) and their individual and cultural differences; (2) the scope and nature of tasks (especially as related to the shift from business tasks to communication and collaboration intensive computer-mediated human activities); as well as (3) the variety of technological platforms and associated devices through which information is accessed.

2.2. The State of the Art

Until recently, accessibility was technically supported through programming intensive efforts which relied on intuition and non-standardized techniques to allow assistive devices to obtain information possessed only by a particular platform (e.g. Windows), or a target application. By nature, therefore, these techniques are very specific to each application and each operating system. As a result, it is hard for them to handle upgrades, foreign versions, any applications they were not originally tested with, and it is also hard for them to migrate to new platforms such as Windows NT.

In the past few years, there have been a number of initiatives by mainstream actors (e.g. Microsoft, Sun, IBM, Apple) aiming to shift the focus of accessibility from the level of the application to the level of the overall interactive environment. Such efforts

seek to overcome the problems identified with contemporary practices by providing accessibility tools as integral components of mainstream interaction platforms and environments. Some of the promising alternatives, which have been developed, include the *Active Accessibility* initiative by Microsoft and the *Java Accessibility* by Sun. There have also been research projects aiming to provide architectural guidance and high-level tools for designing and implementing accessible user interface software. Indicative examples include the *FRIEND21* project (IPIE 1995) funded by Japan's MITI and the *Unified user interface development* platform (Stephanidis 1995, Stephanidis *et al.* 1997) developed in the context of the ACCESS project of the Commission of the European Union.

In addition to the above technological developments, and in line with a broadened notion of accessibility as postulated by universal design, there have been attempts to provide a critical mass of knowledge in the form of universal design recommendations. The existing corpus includes either general principles (i.e. Story 1998) or platform specific guidelines (i.e. HFES/ANSI 1997, W3C-WAI).

It should be noted that recent efforts towards universal design have met the wider appreciation of an increasing proportion of the research community (e.g. research consortia in the context of various Programmes of the European Commission), industry (e.g. the USA Telecommunications Policy Roundtable, Microsoft Active Accessibility, Java Accessibility), scientific and technical committees (e.g. USACM — the ACM public policy committee), legislative acts (e.g. Americans with Disabilities Act 1993; USA Telecommunications Act 1996 — sect. 255), as well as the United Nations General Assembly Standard Rules (1995).

In contrast to the above supporting initiatives and efforts, there have been claims debating the practicality and cost justification of universal design. In particular, there is a line of argumentation raising the concern that "many ideas that are supposed to be good for everybody aren't good for anybody" (Lewis and Rieman 1994). In spite of the truth behind this argument, universal design should not be conceived as an effort to advance a single solution for every body, but as an approach to provide environments that are designed in such a way that suit the broadest possible end-user needs, requirements and preferences. Another common argument is that universal design is too costly for the benefits it offers. Though the field lacks any real data and comparative assessments as to the costs of designing for the broadest possible population, it is felt that the cost of inaccessible systems is comparatively much higher and is likely to increase even more, given the current statistics classifying the demand for accessible products. It is important, however, to underline that any particular technology (in the broad sense of the term) towards universal design should satisfy much more than mere demonstration of technical feasibility, in order to be acceptable. Strictly speaking, even technical efficiency, which presupposes both technical feasibility and technical reliability, would still be an inefficient attribute. What is really needed is economic feasibility in the long run, leading to versatility and economic efficiency.

3. SKETCHING THE HCI CHALLENGES

To address universal accessibility in HCI in the context of the emerging Information Society, several obstacles need to be alleviated, including (1) the lack of a consolidated theory to guide and facilitate universal access, and (2) the intuitive and ad hoc character of the majority of recent development efforts in the field of Assistive Technology. These two shortcomings give rise to a wide range of issues, which for the purposes here, will be classified into design and development challenges.

3.1. Design Challenge

The lack of consolidated theories to guide and facilitate universal access is evident from the limited input and impact that prominent HCI design strands, namely *Human Factors evaluation* and *Cognitive Science*, have had on the study of (universal) accessibility.

In particular, Human Factors evaluation has delivered a wide range of general and platform specific guidelines for HCI design, as well as a rigorous scientific approach to system's evaluation. The available collections of guidelines cover a variety of topics and application domains. Some of them provide insights and recommendations towards accessible design of computer-based equipment (e.g. Thoren 1993, HFES/ANSI 1997), while others offer principles of good practice and examples of universal design (Story 1998). However, such accessibility guidelines, though useful to start with (when applied), suffer from well-known shortcomings, related both to the relevance of their actual human factors input to universal design, as well as to their practical use by designers (Akoumianakis and Stephanidis 1999).

Similarly, cognitive models, such as GOMS, have had minor impact, both on the design of systems accessible by people with disabilities, as well as the study of universal access. One study, by Horstman and Levine (1990), concentrated upon a word prediction task in an augmentative communication system. Their work, however, has come under fire by Newell and colleagues which has resulted in a lively exchange (Horstman and Levine 1992, Newell *et al.* 1992a, b), regarding the suitability of cognitive user modeling in augmentative and alternative communication.

In general, the models produced by cognitive science are designed to capture the generalities of a population rather than individual differences, or the requirements of small groups of specialized users. This is due to the need for collecting statistically valid data to justify the models. On the other hand, universal access stresses the importance of individual differences and seeks to develop models and tools for accommodating them throughout a product's life cycle.

3.2. Development Challenge

The above picture is further complicated by the lack of practical means to guide developments towards universal access systems. In the past, the issue of accessibility has been primarily addressed through programming-intensive efforts resulting in re-engineering or re-implementation of the user interface. In the field of HCI, part of the problem is attributed to the lack of tools to ease the task of creating accessible interactive software, which in turn, resurfaces one of the issues that has long been of concern to the HCI community, namely the user interface architecture (Mueller *et al.* 1997, Stephanidis *et al.* 1998).

In particular, existing architectural abstractions for user interface software, such as the Seeheim meta-model (Ten Hagen 1990), the Arch meta-model (UIMS Tool Developers Workshop 1992) or the PAC model (Coutaz 1990) fail to provide sufficient guidance towards building universally accessible applications.

First of all, they do not provide any guidance as to how to design and implement user interface adaptation, which is a central theme in universal access. More importantly, however, such models offer implementation-oriented views of user interface architectures, thus delimiting the role of design and not addressing how design knowledge can be propagated to development and implementation phases. As a result, they implicitly lead to re-implementations (reactive approach) rather than instantiation of an alternative design (proactive approach).

4. NEW RESEARCH AGENDA

Having pointed out the constraints that undermine the choice of a scientific and technological base for studying (universal) access, we now briefly outline a tentative research agenda for meeting the challenge. This agenda was developed in the first meeting of the International Scientific Forum (ISF) "Towards a Information Society for All" (Stephanidis *et al.* 1998) and it was subsequently refined specifically in the context of HCI (Stephanidis *et al.* 1999).

The agenda points out a broad range of required actions relevant to three main themes, namely technology and user-oriented issues, critical application domains and services, and support measures. Under the theme of technology and user-oriented issues, the agenda highlights the need for additional work covering the development of critical technologies, the advancement of suitable design frameworks and the evolution of powerful user interface architectures. Critical application domains and services include life-long learning, public information systems, terminals and information appliances, transactions services, social services and electronic commerce, as well as global issues, such as security, reliability, etc. Support measures that would facilitate a favorable environment towards an Information Society for *all* should cover the articulation of demand for universal design, support to industry, awareness raising and knowledge dissemination, and technology transfer.

In a subsequent effort, the above main themes were further elaborated to provide a roadmap for HCI research activities. Four main research clusters have been identified; the first three are proposed as Research and Technological Development (RTD) topics whereas the fourth is proposed as a support measure, or "horizontal" activity: (1) promote the development of environments of use; (2) support communities of users; (3) extend user-centered design to support new virtualities (and novel usage contexts); and (4) establish suitable accompanying measures. The four topics are interrelated. Thus, recommendations under one topic link with recommendations under a different topic. The type of actions envisaged, with the exception of the accompanying measures cluster, cover of all phases of technological development, ranging from feasibility studies, to basic and applied research, and demonstration. In what follows, we present a brief description of the meaning and rationale for each one of the four high-level recommendations, while below we elaborate on specific recommended RTD activities.

4.1. Promote the Development of Environments of Use

Environments of use imply integrated systems sharable by communities of users. They should, in contrast to the traditional notion of computers as productivity tools, allow for richer communications and signifying the progressive integration of the
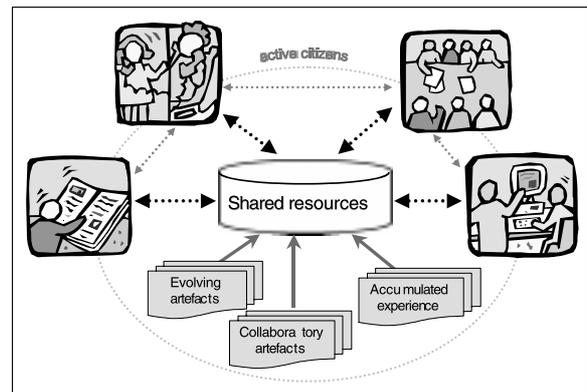

Figure 1. From productivity tools to environments of use.

computing environment with the physical environment (Figure 1). Moreover, in contrast to tools, which enhance the productivity of individuals, environments of use would promote the concept of loveable systems suitable for a broad range of communication and collaboration intensive activities among groups of people. Such environments should be characterized by sympathy and care for users and non-users and should be accessible by anyone, anytime, anywhere. (The term "non-user" is used to refer to members of a community who, though not interacting with the environment themselves at a particular point in time, are being affected by this environment, or its use by other active users.) Finally, they should provide unobtrusive means for supporting social activities.

As depicted in Figure 1, environments of use are likely to become integral components of daily activities among communities of users and facilitate the establishment of new forms of social endeavors. Consequently, they should be conceived and designed as community-centered, sharable, expandable, cooperative, collaborative and responsive media, catering through user and environment monitoring, for a broad range of human needs for both users and non-users. Additionally, they should offer voluntary and context-specific user support, and facilitate error tolerant behavior and preventive actions against unforeseen circumstances and/or misuse.

4.2. Support Communities of Users

Another critical trajectory en route to an Information Society is the one that progressively shifts the focus of attention from individual users to communities of users. The important element in this trajectory is the emphasis on social interaction in virtual spaces. To design interactions in such virtual worlds, it is pertinent to enhance the currently prevailing interaction paradigms (e.g. GUI and the World Wide Web) to support the broad range of group-centric and communication-intensive computer-mediated human activities (Figure 2).

Such a community-wide design perspective, requires that activities among members of communities of users become the primary unit of analysis, as opposed to an individual's keystrokes or performance measures. Moreover, the design focus should be on the cumulative experiences of the community's users with the shared resources, as well as on the way in which communities move from early formation to maturity. To this end, there is a compelling need to study and understand how such communities (e.g. the virtual city) are formed, evolve, grow and intra-/
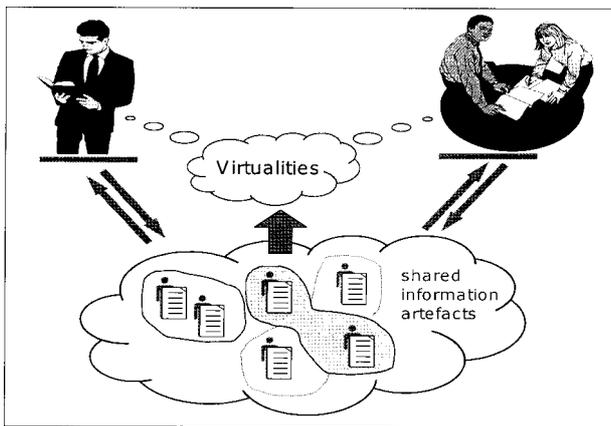
Figure 2. Supporting community-wide experiences.

Figure 3. New virtual spaces for individual and collective experience.

interoperate to synthesize methods that facilitate the design of suitable virtualities and computer-mediated activities for all potential community users.

4.3. Extend User–centered Design to Support New Virtualities

To facilitate the design of new virtualities likely to be encountered in the Information Age, the existing inventory of methods, techniques and tools for user-centered design should be suitably applied and enhanced (Figure 3). To this end, attention should be drawn upon the accumulated knowledge and results in the social sciences (e.g. human communication theories, language theories, action theories, etc), to promote and facilitate the use of more developmental approaches to the study of computer-mediated human activities. (The term *developmental approach* to studying computer-mediated human activities is used to refer to various established theoretical strands within the social sciences and/or psychology that take explicit account of and model development in human behavior and capability. Such developmental approaches have recently started to progressively find their way into HCI. Examples include *activity theory* (Bødker 1989, 1991, Nardi 1996), *situated action plans* (Suchman 1987), *distributed cognition* (Hutchins 1995), *language/action theory* (Winograd 1988), etc.) In all cases, the tight evaluation-feedback loop advocated by user-centered design should provide the primary channel for timely input into design processes, so as to ensure that deficiencies are corrected at an early stage, while updates are less costly to make.

4.4. Establish Suitable Accompanying Measures

Support measures cover a whole range of multidisciplinary and cross-sector actions needed to facilitate the development of an industrial environment favorable to an Information Society for the broadest possible end-user population. Actions are needed to promote and facilitate the adoption and diffusion of good practice in the areas of accessibility and usability to ensure quality in the use of products and services. To this end, it is important that accompanying measures are initiated to articulate demand (Kodama 1992) for universal design, support the industry in adopting novel methods and practices, raise awareness, promote knowledge dissemination, and transfer technology in the form of know-how and know-why.
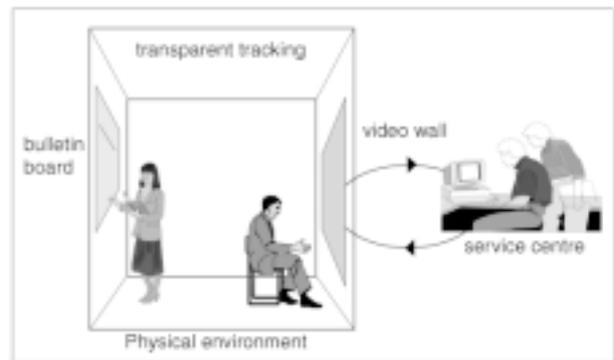
5. DISCUSSION AND CONCLUSION

The above are only some of the goals of international collaboration aiming to promote universal access in the Information Society. They demonstrate the benefits resulting from an international, multidisciplinary effort to advance the concepts and principles of universal design in the context of Information Society technologies, as well as the added value of undertaking this effort through a network of partners cooperating and collaborating towards a shared vision.

One possible option regarding the implementation of such an agenda is that national governments fund the required RTD policy mechanisms. However, many of the items and recommendations of the Research and Development (R&D) agenda have either an explicit international dimension, or can be more effectively addressed at an international level. As a result, what is needed is international collaboration and cooperation, which, in any case, would provide the necessary input to national and transnational RTD policy forums. In this direction, the ISF has initiated, and is planning to define activities aiming to consolidate current practice and experience in the area of universal design, and make it widely available as reference material, or provide contributions to on-going national and international standardization activities. Similarly, in the short- to medium-term, work can be directed towards the identification of key accessibility criteria or requirements to be met by products and services. These efforts can be consolidated into an appropriate form (e.g. accreditation scheme) that would guide subsequent efforts by both industry and academia, towards information products and services accessible and usable by the broadest possible end-user population. Furthermore, such activities can help industry to gain a renewed focus on the issue of universal design, and facilitate justification for the costs and benefits of alternative technologies. Additionally, it can stimulate new developments, and establish the ground whereby universal design informs and improves practice.

REFERENCES

AKOUMIANAKIS, D. and STEPHANIDIS, C., 1999, Propagating experience-based accessibility guidelines to user interface development. *Ergonomics*, 42.

BØDKER, S., 1989, A human-activity approach to user interfaces. *International Journal on Human–Computer Interaction*, 4, 171–215.

BØDKER, S., 1991, *Through the Interface: A Human Activity Approach to User Interface Design* (Hillsdale: Lawrence Erlbaum).

COUTAZ, J., 1990, Architecture models for interactive software: failures and trends. In G. Cockton (ed.) *Engineering for Human–Computer Interaction* (Amsterdam: North-Holland), 473–90.

HFES/ANSI 200, 1997, *Draft HFES/ANSI 200 Standard, Section 5: Accessibility* (Santa Monica: Human Factors and Ergonomics Society).

HORSTMAN, H. and LEVINE, S., 1990, Modeling of user performance with computer access and augmentative communication systems for handicapped people. *Augmentative and Alternative Communication*, 6, 231–41.

HORSTMAN, H. and LEVINE, S., 1992, Modeling AAC user performance: response to Newell, Arnott and Walter. *Augmentative and Alternative Communication*, 8, 231–41.

HUTCHINS, W., 1995, *Cognition in the Wild* (Cambridge, MA: MIT Press).

INSTITUTE FOR PERSONALISED INFORMATION ENVIRONMENT, 1995, *FRIEND21 Human Interface Architecture Guidelines* (Tokyo: IPIE).

KODAMA, F., 1992, Technological fusion and the new R&D. *IEEE Engineering Management Review*, 20, 6–11.

LEWIS, C. and RIEMAN, J., 1994, *Task-centered User Interface Design: A Practical Introduction* [http://www.syd.dit.csiro.au/hci/clewis/contents.html].

MULLER, M.J., WHARTON, C., MCIVER, W.J. and LAUX, L., 1997, Toward an HCI research and practice agenda based on human needs and social responsibility. In *Proceedings of ACM Conference on Human Factors in Computing Systems (CHI '97)* (New York: ACM Press), 155–61.

NARDI, B., 1996, *Context and consciousness: Activity Theory and Human Computer Interaction* (Cambridge, MA: MIT Press).

NEWELL, A.F., ARNOTT, J.L. and WALTER, A., 1992a, On the validity of user modeling in AAC: comments on Horstman and Levine (1990). *Augmentative and Alternative Communication*, 8, 89–92.

NEWELL, A.F., ARNOTT, J.L. and WALTER, A., 1992b, Further comment on the validity of user modeling in AAC. *Augmentative and Alternative Communication*, 8, 252–3.

RAHMAN, M. and SPRIGLE, S., 1997, Physical accessibility guidelines of consumer product controls. *Assistive Technology*, 9, 3–14.

SAVIDIS, A. and STEPHANIDIS, C., 1995, Developing dual user interfaces for integrating blind and sighted users: the HOMER UIMS. In *Proceedings of ACM Conference on Human Factors in Computing Systems (CHI '95)* (New York: ACM Press), 106–13.

STEPHANIDIS, C., 1995, Towards user interfaces for all: some critical issues. In *HCI International '95, Panel Session "User Interfaces for All — Everybody, Everywhere, and Anytime"*, 137–42.

STEPHANIDIS, C., SALVENDY, G., AKOUMIANAKIS, D., ARNOLD, A., BEVAN, N., DARDAILLER, D., EMILIANI, P.-L., IAKOVIDIS, I., JENKINS, P., KARSHMER, A., KORN, P., MARCUS, A., MURPHY, H., OPPERMANN, C., STARY, C., TAMURA, H., TSCHELIGI, M., UEDA, H., WEBER, G. and ZIEGLER, J., 1999, Toward an information society for all: HCI challenges and R&D recommendations. *International Journal of Human–Computer Interaction*, 11, 1–28.

STEPHANIDIS, C., SALVENDY, G., AKOUMIANAKIS, D., BEVAN, N., BREWER, J., EMILIANI, P.-L, GALETSAS, A., HAATAJA, S., IAKOVIDIS, I., JACKO, J., JENKINS, P., KARSHMER, A., KORN, P., MARCUS, A., MURPHY, H., STARY, C., VANDERHEIDEN, G., WEBER, G. and ZIEGLER, J., 1998, Toward an information society for all: an international R&D agenda. *International Journal of Human–Computer Interaction*, 10, 107–34.

STEPHANIDIS, C., SAVIDIS, A. and AKOUMIANAKIS, D., 1997, Tutorial on "Unified Interface Development: Tools for Constructing Accessible and Usable User interfaces." Tutorial no 13 in HCI International '97 [http://www.ics.forth.gr/proj/at-hci/html/tutorials.html].

STORY, M.F., 1998, Maximising usability: the principles of universal design. *Assistive Technology Journal*, 10, 4–12.

SUCHMAN, L.A., 1987, *Plans and Situated Actions: The Problem of Human Machine Communication* (Cambridge: Cambridge University Press).

TEN HAGEN, P.J.W., 1990, Critique of the Seeheim model. In D.A. Duce, M.R. Gomes, F.R.A. Hopgood and J.R. Lee (eds) *User Interface Management and Design* (Eurographics Seminars, Springer).

THOREN, C., 1993, *Nordic Guidelines for Computer Accessibility. The Nordic Committee on Disability* (Stockholm: Swedish Handicap Institute).

THE UIMS DEVELOPERS WORKSHOP, 1992, A metamodel for the run-time architecture of an interactive system. *SIGCHI Bulletin*, 24, 1.

W3C-WAI [http://www.w3.org/WAI/].

WINOGRAD, T., 1988, A language/action perspective on the design of cooperative work. *Human–Computer Interaction*, 3, 3–30.

Use of Modern Chinese Language

B. Han and L. Luo

Institute of Psychology, Chinese Academy of Sciences, Beijing 100101, China

ABSTRACT

Statistical research on modern Chinese has produced a great number of results in the form of articles and dictionaries, as well as offering a significant platform for psychological, ergonomic, and linguistic research on Chinese information processing, from both a theoretical and a technical aspect. This contribution is profound, given that many of the results have been adopted as international and domestic standards. Three statistical studies on character/word frequency of modern Chinese are briefly reviewed, together with an introduction to four representative dictionaries on Chinese features for presenting basic achievements of research on Chinese information processing. Furthermore, related international and domestic standards of Chinese information processing are introduced. For the sake of application, in particular for ergonomic research, it is of utmost importance to understand the fundamental issues in this field.

1. INTRODUCTION

Chinese has been regarded as the oldest language, used for more than 3,500 years, and currently with more than 1.3 billion users. It is one of the easiest languages to learn because there are no morphological changes in the sentences, and yet it is one of the most difficult in that so many characters which cannot be pronounced directly must be learned. Although Zhong Hua Zi Hai collected 85,568 characters, most of them were used in ancient times. Only a few thousand are used in modern Chinese literature.

The application of information technology in China always has to do with language-specific problems — a typical example is, how can a computer process Chinese information? This question involves technical issues such as how many characters (or words) are necessary for an information exchange in modern society, and how to transfer these characters into binary information in order for it to be processed by computer. It is also related with psychological or ergonomic issues such as how to design and improve the usability of the Chinese information processing system. Therefore, statistical study on the frequency of Chinese characters/words and related features, as well as ergonomic research, is needed.

More than two decades of research have produced many results, in the form of articles, dictionaries, databases, national standards, Chinese keyboard input systems, Chinese word-processing systems, desktop publishing systems, etc. However, human–computer interface research is relatively rare compared to statistical and technical research (Zhu and Xu 1990).
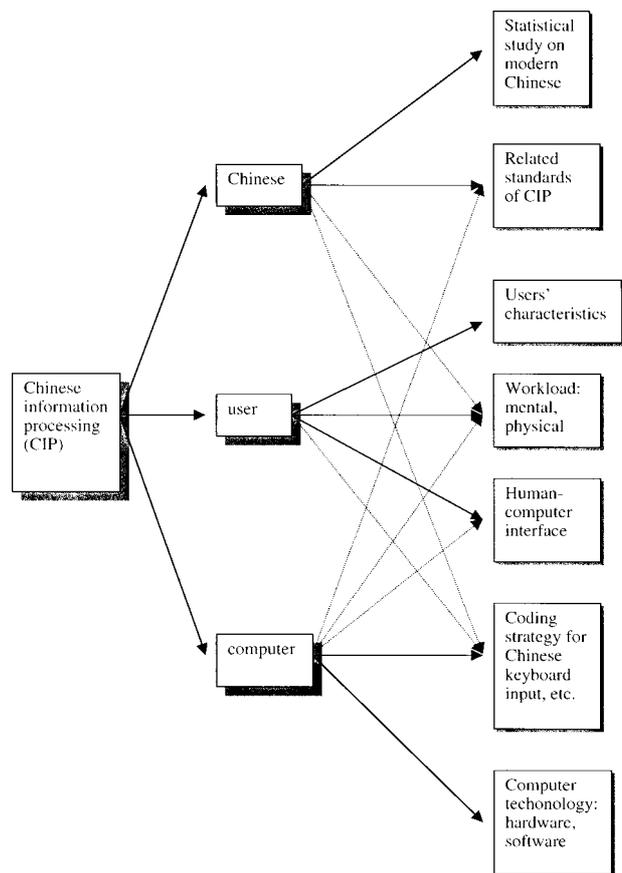

Figure 1. Research subjects related to Chinese information processing

Furthermore, most of these studies have been introduced in Chinese with only a few published in English, thus making them difficult to retrieve. This has had an adverse effect on Western colleagues' comprehension of the progress in related fields, as well as on communication between Chinese researchers and their Western colleagues. There is an increasing need for mutual communication in this net-wise world. Comparative studies on Chinese and Western languages, from information processing techniques to human cognition, will promote the process of global information from academic, cultural, and economic perspectives.

Figure 1 show the related study subjects for Chinese information processing (CIP). Please note the horizontal and vertical relationship between different subjects. This study reviews related statistical research on, as well as international and domestic standards of modern Chinese for, information interchange.

2. STATISTICAL RESEARCH ON, AND DICTIONARIES OF, CHINESE INFORMATION

Statistical research on modern Chinese has aimed at constructing databases for information exchange. The research can be divided into two stages chronologically, as is shown in the following three examples.

The first stage (1974–82) was characterized by three nationwide statistical studies on modern Chinese printed material. Some dictionaries were published and the first national standard of CIP (GB 2312-80) was issued, based on this research. Meanwhile, various Chinese keyboard input coding systems were

developed. The second stage (1983–98) was characterized by the publishing of more dictionaries, and the issuing of new standards, based on the basic research of the first stage. During this time, theoretical and technical developments had also made great progress, both quantitatively and qualitatively.

The second stage is also characterized with studies on micro-features of Chinese characters — for example, stroke order, components, components' combination, etc. Han (1994) studied the frequency of component combination and developed related databases. Two national standards, "Regular stoke order in modern Chinese" and "Chinese character component standard of GB 13000.1 character set for information processing" were also issued in 1997.

2.1. Statistical Study of Chinese Character Frequency (1974–79)

This is the first study to chart statistics on the utility frequency of modern Chinese — the study was known as the 748 project after its commencement date of August 1974 (it was finished in 1979). During this study some 21.65 million printed Chinese items were collected, showing the use of 5,991 characters and, for the first time in history, much first-hand data about the distribution of Chinese orthographic, semantic, and phonological features (Bei and Zhang 1988).

Based on these results, the first national standards of Chinese information processing were produced. *The Code of Chinese Graphic Character Set for Information Exchange: Primary Set* (GB 2312-80) was issued and *The Chinese Information Dictionary* (Li and Liu *et al.* 1988), which includes more than 30 entries of different features such as the radical, stroke number, component sequence, frequency, etc. of 11,254 characters, was also published. A database in disk format, called "The Database of Chinese Character Features" with the same indexes for 6,763 characters, was also open to public use (Shanghai Jiao Tong University 1985).

2.2. Frequency Statistics of Modern Chinese (1979–85)

From November 1979 until July 1985, 1.8 million samples of printed Chinese material were collected from different types of literature. In total, 31,159 words were found, together with 4,574 characters used. *The Modern Chinese Frequency Dictionary* (Wang *et al.* 1986) was published, based on these results. It includes eight tables listing characters or words according to frequency, usage, and distribution in different types of literature, as well as some analysis.

2.3. Word Frequency Statistical Studies on Modern Chinese (1981–86)

This is the largest study in this field to date, based on 25 million samples of printed Chinese, randomly and/or regularly collected from social science and natural science literature (about 0.3 billion printed characters), covering the years 1919 to 1981. The study began in 1981, and was concluded in 1986, during which a frequency statistical system as well as automatic recognition technology were developed. In total, 130,691 words, combined with 7,611 characters, were found.

As a result of this study, *The Frequency Dictionary of Common Words in Chinese* (Liu *et al.* 1990) was published. This dictionary lists 130,691 words, combined with 1-7 characters in phonetic order, according to the Chinese *pinyin* system. Each word was shown with its pronunciation, token frequency, percentage of frequency, accumulated frequency, etc.

2.4. International Standard Chinese Dictionary (1998)

The International Standard Chinese Dictionary (Fu *et al.* 1998) is the most recent of a number of dictionaries on Chinese information, published simultaneously in both book format and as a CD-ROM. It describes 20,902 characters (defined by ISO 10646) in a way that corresponds to all related standards already issued. Entries of each character include character form, stroke number and order, number and sequence of components, pronunciation (provided by a professional reader in the CD-ROM), definition, utility frequency, radicals, the 15 codes for international or domestic information exchange, and help functions. This dictionary represents the latest development of Chinese information processing.

3. INTERNATIONAL AND DOMESTIC STANDARDS OF CHINESE INFORMATION PROCESSING

Since the first national standards of Chinese information processing (GB 2312-80) were issued in 1982, many other related international and domestic standards have also been issued, as is showed in table 1. Most of the standards on Chinese information processing have been issued in Mainland China, while others were issued in Taiwan, Japan, or Korea. Some of them were adopted by the International Standards Organization as official international standards. Due to limited space, technical details of each standard cannot be described in this article. Please refer to the Internet resources listed in the references section, or to the original copy of a specific standard.

As was shown by earlier studies, several thousands of characters is enough for daily usage. Hence GB 2312-80 defined 6,763 characters. However, it is neither enough for special usage involving traditional Chinese documents (e.g. ancient literary or medical literatures), nor for international information interchange between China and Asian countries and areas, therefore ISO 10646 was issued after several years of preparation as a subset (Ideograph zone) of "The UCS Universal Multiple-Octet Coded Character Set". Table 2 shows the character sources in ISO 10646.

It is helpful to clarify the internal relationship between these standards, since some of them refer to the same thing, or cover similar character sets although they were issued in different countries or areas. Six standards issued in Mainland China were for 21,039 simplified characters and 21,142 corresponding traditional characters. These offer a basic platform for the information processing of Chinese documents. Table 3 shows the correspondence between the six standards.

Most of the standards describe only the form of single characters. This is because the single character is the basic element of Chinese words; statistical studies on characters are independent of cultural differences, while the single character is widely used in China, Japan, and Korea. However, more research is needed on components and stroke level in order to develop related standards. This will be useful to the evaluation of CKI methods, and to the Chinese information interchange at language level.

Table 1. International and Domestic Standards of Chinese Information Processing issued in recent years

Title of Standards	Issuer	Issued and time	Sample	Entries and Remarks
Scheme of the Chinese Phonetic Alphabet, Pinyin	SCRWS	1958		ISO adopted it as International Standard for spelling and marking Chinese in 1982.
GB2312-1980 Code of Chinese graphic character set for information interchange (CCCSII) – Primary set	SBS	1982.3.9	7445 graphic symbals included 6763 simplified characters	Binary code of QuWei and National Standard Code for Information interchange
Unitary Radical List of Chinese Characters (Draft)	SCRWS, PPA	1983	201	
GB/T 7589-1987 CCISII – The 2nd supplementary set	SBS	1987.3.27 1987.12.1	7237 simplified characters	Identical document: ISO 2022-86, REF
GB/T 7590-1987 CCISII – The 4th supplementary set	SBS	1987.3.27 1987.12.1	7039 simplified characters	Identical document: ISO 2022-86, REF
GB12345-1990 Code of Chinese ideogram set for information interchange (CCISII) – The 1st supplementary set	SBTS	1990.6.3	7583 graphic symbals included 6866 unsimplified characters	Correspondent with GB2312-80, included some punctuation used in upright typesetting, and 35 national phonetic alphabet. Some simplified characters match several unsimplifed characters
GB13131-1991 CCISII – The 3rd supplementary set	SBTS	1991.8.15	7237 unsimplified characters	Correspondent with GB7589-87
GB13132-1991 CCISII – The 5th supplementary set	SBTS	1991.8.15	7039 unsimplified characters	Correspondent with GB7590-87
ISO/IEC 10646-93	ISO	1992	20902	Subset of UCS universal multiple – octet coded character set
GB/T 13000.1-1993	SLWC, SBTS	1993.12.30	20902	Included characters proposed by China Mainland (17124), China Taiwan (17258), Japan (12157), Korea (7478). Identical document: ISO/IEC 10646-93, IDT
GF 3001-1997 Chinese Character Component Standard of GB13000.1 character Set for Information Processing	SLWC	1997.12.1; 1998.5.1	560 (disassemplied from 20902 characters)	
Regular Stroke Order in Modern Chinese	SLWC, SBNB	1997.4		

Note: Abbreviations used in the table stand for what shown as follows:
SCRWS = State Commission for Reformation of Writing System
SLWC = State Language Work Committee
SEC = State Education Commission, named as Ministry of Education— since 1998)
SBTS = State Bureau of Technological Supervision
ISO = International Standard Organisation
SBNB = State Bureau for News and Publishing
PPA = Press and Publication Administration of the PRC
MRFT = Ministry of Radio, Film and Television
SBS = State Bureau of Standards

Table 2. Character source of ISO 10646

Sources	Number of characters	References
Mainland China	17124	GB 2312-80 (GB 12345-90) GB/T 7589-87 (GB 13131-91) GB/T 7590-87 (GB 13132-91)
China, Taiwan	17258	CCCII CNS 11643-1992 Big Five
Japan	12157	JIS X 0208-1990 JIS X 0121-1990
Korea	7478	KS C 5601-1992 KS C 5657-1991

Note: Total number of characters in ISO (CJK Unified Chinese Coded Character Set) 10646 is 20902.

Table 3. Correspondent relationship between 6 standards issued in Mainland China

	Simplified characters	Traditional characters
	GB 2312-80 (0: 6763)	GB 12345-90 (1: 6866)
Standards (Set No.: Character number)	GB/T 7589-87 (2: 7237)	GB 13131-91 (3: 7237)
	GB/T 7590-87 (4: 7039)	GB 13132-91 (5: 7039)

Note: Set number and number of characters was shown in brackets. There are several situations in GB12345-90 where more than one old characters were simplified so correspondent to 1 character, hence total number of characters (6866) is more than 6763.

4. DISCUSSION

4.1. Some Basic Results of Modern Chinese Research

Statistical studies figured out the outline of Chinese character features which form the basis of designing CKI methods. For example, in 6,763 characters defined in GB 2312-80, the average number of stroke components are 10 and 3 respectively. In the ISO 10646 character set, 560 components are used. There are a total of 417 initial–final syllables used as phonological elements in Chinese. Of the characters commonly used in modern Chinese, 18.16% can be read with differing pronunciations.

4.2. Possible Application of Statistical Results of an Ergonomic/Psychological Study

Statistical research on modern Chinese offers basic data for constructing related standards, and has also been a prerequisite for the design, development, and application of the CKI method. These three aspects simultaneously complement each other, having developed on the increasing platform of computer technology. Still more are needed for research on ancient books, information retrieval, and indexes.

Such achievements have been very helpful to linguistic and psychological research, machine translations, artificial intelligence, language reform, character font design, as well as Chinese information processing. In particular, these achievements can be a good resource for ergonomists in such fields as human–computer interfaces and network communication (information exchanges via the Internet).

ACKNOWLEDGEMENT

REFERENCES

BEI, G.Q., and ZHANG, X.T., 1988, *Statistical Study on Chinese Character Frequency* (Beijing: Electronic Industry Press).

CHEN, Y.F., and HU, X.H., 1994, *Fundamentals of Technology and Theory for Chinese Character Inputting* (Beijing: Qinghua University Press and Nanning: Guangxi Science and Technology Press) (in Chinese).

FU, Y.H., *et al.,* 1998, *International Standard Chinese Dictionary*: CD-ROM version (Beijing: Peking University Press) (in Chinese).

HAN, B.X., 1994, Development of the database of Chinese constituents information: statistical analysis of the frequency of constituents and their combinations. *Acta Psychologica sinica*, 26(2), 147–52 (in Chinese with English abstract).

HAN, B.X., 1995, Component combination: a latent Chinese character structure level. *Journal of Chinese Information Processing*, 9(3), 27–32 (in Chinese with English abstract).

LI, G.Y., and LIU, R., *et al.*, 1988, *Chinese Information Dictionary* (Beijing: Science Press) (in Chinese).

LIU, Y., LIANG, N., WANG, D., *et al.*, 1990, *Frequency Dictionary Common Words in Modern Chinese—Section in Phonetic Order* (Beijing: Aviation Press) (in Chinese).

SHANGHAI JIAOTONG UNIVERSITY, 1985, *Database of Chinese character features* (Shanghai: Shanghai Scientific Literature Press).

WANG, H., CHANG, B.R., LI, Y.S., LIN, L.H., LIU, J., SUN, Y.L., WANG, Z.W., YU, Y.X., ZHANG, J.W., LI, D.P., 1986, *Modern Chinese Frequency Dictionary* (Beijing: Beijing Language Colleague Press) (in Chinese).

ZHU, Z.X., and XU, B.H., 1990, The development of ergonomics in China. *Applied Psychology: An International Review*, 39(3):265–73.

INTERNET SOURCES

http://www.cssn.net.cn/cgi-bin (China Standards Service Net) for detailed information about Chinese standards, standards publications, directory of organisation, etc. It offers both Chinese and English versions.

http://ftp.ora.com/pub/examples/nutshell/ujip/doc/cjk.inf for a brief description of international and domestic standards CJK (China, Japan, Korea) Chinese character set, and a detailed introduction about the technique of the CJK encoding system, together with valuable introduction of relative Internet resources.

http://www.kudpc.kyoto-u.ac.jp/~yasuoka/CJK.html for CJK character tables, Kanji (hanzi) variant tables, Unicode related tables, Domestic ISO646 character tables in gif format files.

User Requirements in Information Technology

I. Vasmatzidis

iNautix Technologies Inc., Jersey City, NJ, 07399, USA

1. INTRODUCTION

Building information technology (IT) systems require a user-centered design approach. The core of such an approach is studying, understanding, and meeting user needs. In addition, user interaction with IT systems is subject to characteristics and limitations imposed by the *human information processing* system. Consequently, designers of IT systems should have a good understanding and knowledge of the users:

- information processing system
- target audience
- jobs and tasks
- knowledge and experience
- physical characteristics and cultural characteristics
- physical environment

User requirements are defined by all these factors. Requirements reflecting the information processing system are relatively constant across individuals. Requirements reflecting the rest of the factors should be determined for each IT project separately. The goal of an IT system designer should be to develop systems which meet user requirements to ensure system usefulness and usability and an error-free and efficient user-performance.

2. THE HUMAN INFORMATION PROCESSING SYSTEM

A tremendous amount of research has been conducted on how humans' process information (see Ellis and Hunt 1993 for an overview). In this article, we summarize these aspects that are most important to IT systems designers. The human information processing system can be divided into three subsystems:
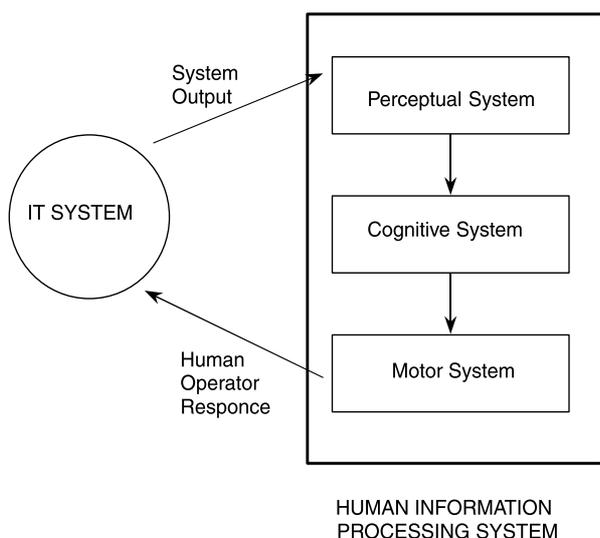
Figure 1. The three human information processing subsystems in a typical human operator–IT system interaction.

- the perceptual system
- the cognitive system
- the motor system

When interacting with IT systems, the system's output is sensed and recognized by the human operator's perceptual system. This information is then processed further by the cognitive system. As a result, a decision is being made, which is executed by the motor system (see Figure 1).

2.1 The Perceptual System

Information processing begins by the activation of the receptor cells. Receptor cells are specialized cells located in the eyes, ears, nose, tongue, and skin that sense the physical energy of the environment (light, sound, vibration, etc.), and initiate the processes of sight, hearing, smell, taste, and feeling. Energy received by the receptor cells is temporarily stored in the memory system known as the *sensory register* or the *sensory buffer*. The stored record is known as the *sensory trace*. The sensory buffer has a large capacity for storing information, but a very short duration. For example, visual sensory traces last for up to only 300 milliseconds. The role of the sensory buffer is to temporarily hold information so the processes of pattern recognition and memory can further process it.

Information in the sensory buffer is meaningless. It becomes meaningful only after pattern recognition has been activated. During this process, the sensory trace is contrasted against knowledge stored in the long-term memory. How exactly this contrasting takes place is not clear, and certainly is much more complex than a simple "pattern matching". In any case, it involves the long-term memory, which is the knowledge depository necessary for supplying meaning into the sensory trace. Furthermore, it appears that pattern recognition is based on an analysis of the features of the information retained in the sensory register. Finally, pattern recognition depends on the context in which information is received.

2.2 The Cognitive System

There are two memory structures within the human cognitive system:

- Short-term memory (STM)
- Long-term memory (LTM)

STM is where we temporarily store information that has been recognized. To store information in the STM, we need to pay attention to it by invoking the process of *selective attention*. STM has a limited capacity and a limited duration. We can retain a very small number of information items in STM (7 ± 2) for a period of few seconds only. In order to retain information in our LTM we need to rehearse; by rehearsing, information remains in the LTM longer and eventually is transferred to the LTM.

STM is not only a storage buffer, but also a cognitive processor. STM is where we perform conscious mental operations such as arithmetic calculations and thinking. Thus, STM is also known as *working memory*. The limited capacity of STM is the major bottleneck in the human information processing. To offset this limitation we use the technique known as *chunking*. By chunking information we devise rules to organize information into meaningful groups. For example, it is much easier to remember the phone number 123-1234 than the phone number 472-8175. The first number consists of two "chunks" combined in a

meaningful way: they are two groups of numbers generated by a simple rule.

Information rehearsed long enough in STM is eventually transferred to LTM. LTM is unlimited both in capacity and duration. Information in LTM is coded semantically (at the level of meaning), in contrast with coding in STM which is acoustic or phonetic. Despite the unlimited capacity of LTM, storing and retrieving information from it is unreliable and slow. For example, consider how often we failed to remember a familiar name when needed, only to recall it at a later time!

Registering information in our STM requires selective attention. The process of selective attention prevents STM from being overloaded. It enables us to selectively focus our attention on the stimuli relevant to the problem at hand. However, attention also occurs involuntarily. For example, while carrying out a conversation in a crowded room, we can detect our names spoken from a third person, a phenomenon known as the "cocktail party phenomenon". Such involuntary attention occurs at the semantic level. Attention can also be *divided*, or *time-shared*. Divided attention refers to paying attention to two or more tasks simultaneously. Divided attention has a detrimental effect on operator performance.

Another process of the cognitive system of great importance to IT system designers is learning. The following characteristics of learning are of practical significance:

- Learning is faster when the material to be learned is analogous to material already learned;
- Learning is facilitated when the material to be learned is well organized;
- Learning is more effective when information is presented in incremental units.

2.3 The Motor System

User decisions and thoughts are transformed into actions through a series of discrete micromovements. In rapid movements (for example, in skilled typing), the micromovements reflect motor instructions that are preprogrammed. These instructions are subject to modification, depending on the visual and tactile feedback. Execution of fine motor movements involves different muscles than execution of gross motor movements.

2.4 The Cognitive System: Design Implications

This section discusses implications for IT system design based on the knowledge of the human cognitive system.

Implications from the perceptual system:
- Present information in a clear way to facilitate its recognition and analysis;
- Match information with expectations from the LTM;
- Avoid overloading the sensory system.

Implications from the cognitive system:
- Do not overload the working memory with excessive information;
- Design so as information is recognized rather than remembered;
- Design tutorials in interactive, incremental ways;
- Minimize tasks that require divided attention. Implement design techniques (e.g. blinking fields) that enhance the mechanism of selective attention;
- Organize information in meaningful groups.

Implications from the motor system:
- Minimize gross motor movements;
- Require users to exert different fine motor movements to accomplish different operations (e.g. in the Windows operating system, require single clicking for selecting a document icon versus double-clicking for opening and viewing the contents of the document).

3. USER–CENTERED DESIGN: THE USER REQUIREMENTS PERSPECTIVE

Designing successful IT systems requires much more than a good understanding of the user cognitive system. It requires a knowledge and understanding of whom the users are (target audience) and of the users' jobs and tasks, knowledge and experience and physical environment. In other words, IT design should be conducted within the context of the user–work domain. This concept is similar to the concept of ecological design (Rasmussen and Vicente 1989), or use-centered design (Flach and Dominguez 1995).

3.1 Target Audience

The first step in developing IT systems is to determine the target audience. The target audience can be very specialized (for example, pilots or air-traffic controllers for aviation IT systems) or very generic (for example, the general public for recreational IT systems). Knowledge of the target audience will guide the subsequent steps of determining the user requirements as discussed below.

3.2 Users' Jobs and Tasks

Defining the users' jobs and tasks is very critical. IT systems are built for a purpose: to enable users perform their jobs in an efficient and safe way with a high degree of satisfaction. Therefore, a good understanding of their jobs and tasks is necessary for meeting these objectives. Job and task analyses are the techniques used to determine the users' jobs and tasks. A problem with these techniques lies in their very basics: how are jobs and tasks defined? Here, we provide the following definitions:

Job: The process through which the goal of one's work is accomplished.
Task: The *constituent elements* of a job.

Job analysis is conducted in order to determine:
- The goals of the users doing the job.
- The job workflow: how various people and subsystems are involved in the job process. What their roles and their responsibilities are.
- A list of the tasks involved in completing the job. The output of such an analysis should be:
  A *comprehensive list* of the tasks
  Frequency of task performance
  Criticality of task performance
  Time to complete each task
  Difficulty of task performance

There are various task analysis techniques (see Hackos and Redish 1998). Here, we discuss the procedural analysis technique. This technique can be used not only for developing new IT systems, but also in evaluating existing ones. Procedural analysis describes the actions taken and the decisions made during task
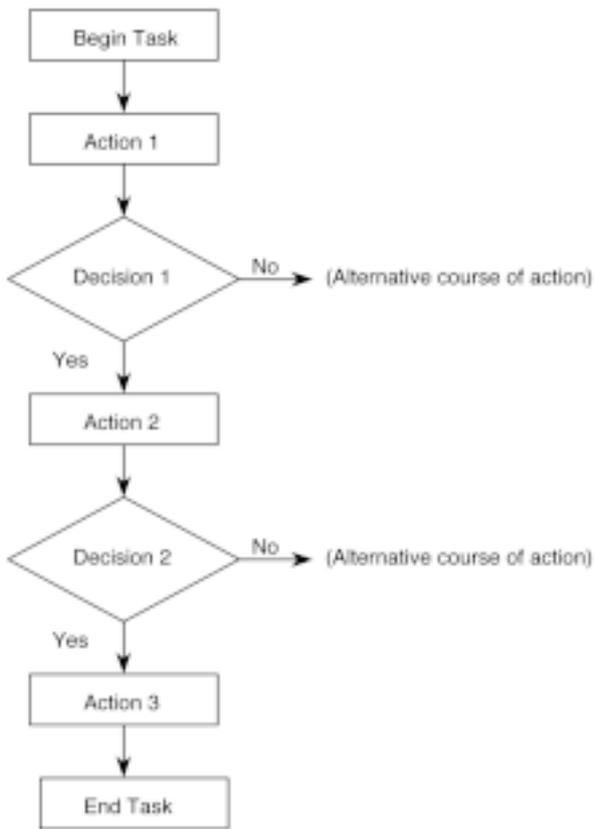
Figure 2. A procedural analysis diagram.

performance. A diagram illustrating procedural analysis is shown in Figure 2.

Procedural analysis requires direct observation of the users; it identifies the actions and decisions of the tasks currently performed by the users. As such, procedural analysis can be used for:

- Identifying user requirements in terms of actions and decisions for building new or improving existing IT systems;
- Developing more efficient ways of satisfying user requirements by deciding which of the actions should be left to the IT system, and which should be left to the user/human operator.

3.3 Users' Knowledge and Experience

User knowledge and experience can be *domain-specific (semantic)* or *system-specific (syntactic)*. Domain-specific experience refers to experience with the tasks and functions the IT system is designed to support. System-specific experience refers to experience with operating other IT systems. An understanding of the users' knowledge and experience assists in classifying users along the following dimensions:

- domain novice users
- domain intermediate users
- domain expert users
- system novice users
- system intermediate users
- system experienced users

When designing IT systems, the target audience should be analyzed in terms of domain and system experience. To satisfy the various types of users, various design approaches can be taken:

- Provide domain explanation and instructions for the IT system operation, and good error prevention and recovery procedures for users who are both domain and system novices;
- Minimize such explanation and instructions and provide efficient methods of system operation (for example, command languages) for users who are both domain and system experts;
- Provide domain explanation and efficient methods of system operation for users who are domain novices and system experts;
- Provide instructions for the IT system behavior, but very limited domain explanation for users who are system novices and domain experts.

Another aspect of user experience reflects the concept of *mental or conceptual models*. A mental model is the internal understanding users have of how the visible and invisible parts of a system interact in order to bring about the observed system behavior. Users develop mental models over time drawing upon their experiences with system interaction. Mental models help users to remember and predict system behavior, and may or may not be accurate. When learning new systems, users will attempt to transfer knowledge from existing mental models. Therefore, IT designers should:

- facilitate the development of an accurate mental model
- take advantage of existing user mental models

To meet these goals, the following guidelines are recommended:

- Provide consistency in IT system interface design: same user actions should result in similar system responses
- Make system processes visible to the users
- Provide descriptive and informative messages
- Use familiar metaphors in the system design

3.4 Users' Physical and Cultural Characteristics

The physical characteristics of the users play an important role in the design of IT systems. In general, the following physical characteristics should be studied:

- age
- gender
- handedness
- color deficiency
- physical handicaps

3.4.1 Age

In many countries, the elderly represent the fastest growing segment of the population. For example, it is estimated that in the US 50 million people will be over the age of 65 by the year 2010. It is known that cognitive and motor skills are diminished in elderly people. Designers of IT systems should make the appropriate design accommodations for elderly users (for example, increased font size).

3.4.2 Gender

Physical gender characteristics that are important to IT system designers are anthropometric in nature. The anthropometric characteristics of female populations are smaller than those of male populations. Such differences are critical when deciding on the dimensions of hardware devices.

3.4.3 Handedness

A relatively large percentage of the users are left-handed (11% in the

US). The design of IT system hardware (keyboards, computer mice, etc.) should be optimized for both right- and left-handed users.

3.4.4 Color deficiency

A significant percentage of the users (8% in the US) have some form of color deficiency. In addition, fewer women are color deficient than man. To accommodate color deficient populations, color should not be used as the only way for information coding. Other visual cues (such as underlining) should be used in a redundant manner to convey the same meaning as color coding.

3.4.5 Physical handicaps

Physically challenged users (impaired vision, impaired motor skills, deafness, etc.) may not be able to use existing IT systems. Specialized technologies are necessary to allow accessibility of IT systems by such users. Such technologies include text-to-speech screen readers for blind people, which convert text into voice output, and voice recognition input devices for users with impaired motor skills.

3.4.6 Cultural characteristics

With the globalization of information technology, the cultural characteristics of users gain more and more importance. This is especially true in the dissemination of information through the Internet where the target audience is literally global. When building IT systems for national, specific or international audiences the culture of the potential users should be well understood and accommodated in the design.

3.5 Users' Physical Environment

Poor environmental conditions will render IT systems useless no matter how well user requirements have been understood. Consequently, efficient and error-free use of IT systems should take place in environments optimized for user performance.

When considering the physical environment the following issues emerge:

- *Lighting.* Poor lighting may cause visual fatigue or glare on computer monitors.
- *Noise.* A noisy environment can mask auditory warnings and message provided by the IT system.
- *Work space.* Limited or poorly designed workspace will impede users' work while interacting with the IT system.
- *Temperature and humidity.* Temperature and humidity extremes have an adverse impact on user information-processing capabilities. Such extremes should be controlled to ensure error-free user performance.

REFERENCES

ELLIS, C. H. and HUNT R. R., 1993, *Fundamentals of Cognitive Psychology*, (5th ed.) (WCB Brown & Benchmark: Wisconsin).

FLACH, J.M. and DOMINGUEZ, C.O., 1995, Use-centered design. *Ergonomics in Design*, July, 19–24.

HACKOS, J.T. and REDISH, J.C., 1998, *User and Task Analysis for Interface Design.* (Wiley: New York).

RASMUSSEN, J. and VINCENTE, K.J., 1989, Coping with human error through system design: Implications for ecological interface design. *International Journal of Man-Machine Studies*, 31, 517–34.

User-centered Graphic Design

J. Frascara

Department of Art and Design, University of Alberta, Edmonton, Alberta T6G 2C9, Canada

1. INTRODUCTION

User-centered graphic design or human-centered graphic design refers to a design approach that considers the final user as the departure point, in contrast to other approaches that seek economic efficiency, aesthetic currency, personal expression of the designer or technological improvement as their main concern. "User"-centered, instead of "receiver"-centered graphic design refers to a position in communication theory that rejects the traditional terminology coined by Shannon and Weaver (1949), who defined the three elements of the communication chain as transmitter–message–receiver. These terms reduce people's behaviors to those of electronic devices, and do not recognize cognitive styles, cultural differences, expectations, feelings, intentions, value systems and levels of intelligence as dimensions that contribute to the construction of a communication situation. The messages produced are not received, but interpreted by people (Frascara 1997).

2. HISTORICAL REVIEW

A concern for the user of designed products has existed for a long time; different periods attended to different perceptions of what was seen as adapting the design of objects and information to the characteristics, needs and abilities of people. These concerns, however, were initially more connected to the design of objects than to the design of information. The industrial revolution generated different reactions from designers in the 19th century, the most famous of which was the Arts and Crafts Movement. It was opposed to mechanized manufacturing and proposed a revival of the handicrafts and of the value of aesthetics for daily use objects. These concerns were further developed in the 1920s around the Bauhaus School in Weimar (1919–33) and the European avant-garde. At that time, the main objective, rather than to oppose technology, was to integrate it with art to educate the designer and to design modern products. Artists and craftsmen were brought together to create the ideal 20th-century designer, exemplified by Herbert Bayer, Max Bill and other Bauhaus graduates. The designer was very much seen as a person who conceived functional and beautiful objects for the benefit of the public. The public, however, was not part of the conception and production processes, and from today's vantagepoint one can see that certain dimensions of what today is defined as user-centered design were still missing.

The Ulm design school (1955–68) added an interesting dimension to the knowledge of the user in the communications field by developing a theory of communication based on semiotics, pioneered by Tomás Maldonado. This approach emphasized the possibility to analyze messages according to semiological structures, adapting to visual communications the ideas proposed by Ferdinand de Saussure in the field of linguistics at the beginning of the 20th century. While semiology includes the problems of syntactics (the organization of stimuli), semantics (the organization of meaning) and pragmatics (the particularities of communicational situations), in practice the central concern of design theory in the 1950s and 1960s was on the first two dimensions.

Over the past 20 years it has become apparent that, in addition to the general structure of human cognition, individual differences are substantial enough between human groups to require that we adapt messages to different users. The awareness of differences between individuals within any human group has resulted in the development of the notion of "market mix", as a supplement to the notion of "market segment" that dominated marketing strategies in the 1950s and 1960s. This notion permits a better consideration of the users of information, since it considers the individual differences that are found in any human group.

3. PEOPLE, NOT GRAPHICS

The notion of user-centered graphic design becomes more complex when one acknowledges that design is not really concerned with objects, but with the impact that those objects have on people. The central problem of the designer is not the construction of graphics, products, services, systems or environments, but the creation of means for people to act, to interact, to realize their wishes and satisfy their needs. It is the needs and the wishes of people that design have to serve; the objects of design are, fundamentally, means. This requires a better understanding of people, of society and of the ecosystem.

When looking at design as centered on people, one should look at the operational impact, the cultural impact and the ecological impact objects have. Every design project has an operational objective: it is supposed to affect the knowledge, the attitudes or the behavior of people in a desired way. But any object deployed in the public space – be it communicational or physical – has a cultural and a physical impact or side-effect. The cultural impact affects the way people operate with other people and with things, and creates cultural consensus. More has to be done to understand this cultural impact so that designers can operate more responsibly in society. It is in connection with the study of both the operational impact and the cultural impact of objects and information where design becomes strongly connected to the social sciences. The physical impact of artificial objects affects the ecology of materials and energy (Papanek 1985), conditions our lives and may have positive or negative effects on our well-being.

4. INTERDISCIPLINE

Product development is now far from being conceived as the province of the individual designer, craftsman, manager or manufacturer: the conception, production, distribution and use of products have become complex parts of corporate strategies, and include not only extensive research based on marketing, but also rely on anthropology, psychology and sociology. The same could be said of the design of instructional and educational materials: this is not any longer the terrain of teachers specialized in specific content areas.

The evaluation of the effectiveness of any design product or idea requires the contribution of several fields in the social

sciences. Leading advertising agencies, manufacturers and institutions such as the International Standards Organization (ISO) have for over 20 years gathered together designers and social scientists around the challenges posed by the design of products and communications, with the purpose of ensuring that the driving criteria of any design decision are based on a knowledge of the capacities and conceptions of the users.

5. LEARNING ABOUT THE USER

The variety of objects and user group profiles that the discipline of design has to address today, makes it impossible to predict with certainty the reaction of a specific group of people to a specific object or communication. The effectiveness of a design product derives from a good fit in this relation user-product. This implies a change from the notion of the user as receiver to that of the user as partner, and looks at communication as an act of negotiation, where the position of the originator of the information and that of the interpreter enter in contact searching for a common terrain. This has placed more emphasis on the need to study every case, and has given birth to the concepts of user-testing and iterative design (MacKenzie 1993, Penman and Sless 1994). Iterative design is based on the production of working prototypes, their testing and evaluation, the incorporation of adjustments, and the development of new testing until the product appears to have been developed to a satisfactory level of performance. Norman (1990) has been an intense advocate of products that match users' needs, motivations and abilities, and his writings are rich in analyses of specific cases. Communication design that is not centered on the user is inefficient, and promotes a passivity in the public that in the long run weakens the possibilities to create a strong society, based on conscious choices and on a sense of sharing.

Several strategies assist the user-centered approach in its commitment to include the user in the design process. These include focus groups, interviews, surveys and user testing. Focus groups are semi-structured meetings with small groups of representatives of user populations aimed at collecting qualitative information about issues and perceptions surrounding a given design problem. Interviews are one-on-one conversations that lead also to qualitative information. Surveys intend to obtain quantitative information concerning the frequency of a given perception in a given population, or the percentage of people in a given population that coincide on a given issue. User testing refers to the production of prototypes of a given product and the systematic study of its actual use by a sample of the user population. The intention in this case is to observe the effectiveness of different aspects of a given product in order to correct its design before mass production and distribution take place.

The conception of user-centered design has resulted in what can be recognized as the "softening of design." Upon adopting the observation and analysis of the user as central elements of the design process, the design of workstations, for instance, has been changed to the design of work. It is not possible to invent and design the perfect chair on which a person could be sitting for 8 h everyday without becoming physically fatigued. It is necessary to see the design problem not only as the ergonomic design of the furniture and other elements operators have to use, but also as the design of the activities to be performed by a person.

The same applies to the evolution of design for education: it has changed from the design of teaching aids to the design of teaching/learning situations. The success of a learning experience cannot be trusted to the design of a teaching aid. The whole activity has to be planned so that the teaching aid contributes its best to the experience. Many details enter this terrain, but certainly the teacher's actions, the student's actions and the environment in which the intervention occurs, all contribute to the learning event, and must be seen as part of the design problem. In education, actually, the notion of paying attention to the receiving side goes back to the 18th century with J.J. Rousseau (1712–78) and continues with J.H. Pestalozzi (1746–1827) and M. Montessori (1870–1952), among others. They saw childhood as a distinct cognitive and emotional stage, rather than as an imperfect adulthood, and proposed the need to adapt the educational tasks to the intellectual and psychological characteristics of children. Their learning theories developed as attempts to look at education not as a process of transmission of information but as a process of development of capacities. This was done on the basis of taking cognizance of child development and of strategies believed to be most effective for the fostering of maturation.

6. USER–CENTERED DESIGN AS A CONSCIOUS NOTION

User-centered design began with the application of physical ergonomics to the design of objects. Intuitive at first, as in the medieval shoemaker that measured the clients' feet, and systematic later, as in the production of mass-made clothing, passenger vehicles or workstations. The conscious notion of user-centered design began with the ergonomic studies of Alphonse Chapanis during World War II, and then moved onto the cognitive, the emotive and the cultural, including today ecological and health issues.

It could be argued that commercial advertising has for the past 50 years been user-centered, in that, rather than being centered on descriptions of the virtues of the products (as it was at the beginning of the 20th century) it has been centered on desires and values of people. Unlike education and ergonomics, however, the main purpose of advertising is not to benefit the public, but to benefit the business that produces the advertised goods. User-centered design in all its implications not only intends to attend to the characteristics of the public to be addressed in order to reach it, but also it attends to the needs and the welfare of that public. This is the case in urban design that improves the quality of living, or in computer interface design that facilitates the tasks of the user (Pradeep 1998).

The aim of user-centered design that takes all dimensions into account is to benefit the user. This concern does not end with the user's wishes and capacities but it extends to the users' physical and psychological health and well-being. This is the point at which user-centered design touches on science and culture, becoming contextualized in knowledge and value systems. User-centered design is, therefore, culture and knowledge dependent, and conceptions of it will vary from culture to culture and from time to time, but the common element is that user-centered design attends to the needs and the capacities of the user, including physical and non-physical considerations of people as biological, social and cultural beings.

7. CONCLUSION

User-centered graphic design is based on an understanding of design as a problem-oriented, interdisciplinary, creative action, in which the nature of the problem – normally an active relation between a user and a product (information, object or environment) resulting in an effect – defines the set of disciplines that must contribute to the creation of a design response. These disciplines frequently include several branches of psychology (perception, cognition, development, behavior and emotion), some aspects of sociology (statistics, survey design, culture, demographics), and anthropology (ethnographic observation or cultural theory).

To summarize, user-centered graphic design places significant emphasis on users' needs, abilities and well-being; it does so aided by scientific research; it includes user-testing and applies an iterative design process; and it includes systematic evaluation of outcomes directly and indirectly connected to the design brief.

REFERENCES

FRASCARA, J., 1997, *User-centred Graphic Design: Mass Communication and Social Change* (London: Taylor & Francis).

NORMAN, D., 1990, *The Design of Everyday Things* (New York: Doubleday).

PAPANEK, V., 1985, *Design for the Real World* (Chicago: Academy Chicago).

MACKENZIE, M., 1993, Appearance and performance. *Communication News*, 6, 1–3.

PENMAN, R. and SLESS, D., 1994, *Designing Information for People* (Hackett: Communication Research Press).

PRADEEP, H., 1998, *User-centered Information Design for Improved Software Usability* (Norwood: Artech House).

SHANNON, C. E. AND WEAVER, W., 1949. *The Mathematical Theory of Communication* (Urbana: University of Illinois Press), 4–6.

Video Telephony

S.T. Väyrynen† and P.J. Pulli‡

†Laboratory of Work Science and ‡Department of Information Processing Science, University of Oulu, FIN-90401 Oulu, Finland

1. INTRODUCTION

Video telephony (VTP) is a method of telecommunication, which makes possible on-line interactive contact using both audio and video information links. The remotely sited parties can see and hear each other synchronously from different geographical locations, replicating face-to-face contact. VTP is typically one form of multimedia technology, which includes pictorial representations and written and spoken language. For humans, the eyes constitute a "information highway," and VTP is an extension to it.

VTP is based on hardware, software and netware. The nets utilized nowadays are most often wired, but more wireless mobile possibilities will be available in the future. In the past, VTP was generally expected replace the normal audio telephone in personal use. Later, professional applications, especially videoconferencing, have been used in business and education. Face-to-face communication is not the only form. It is possible to arrange group meetings and give education showing many things other than only the people who are lecturing and, vice versa. Via VTP, many actions, objects and places can be seen along with the participants.

VTP can be utilized in various remote expert and diagnostics applications. Many new approaches are possible. For example, a system with good image quality and motion reproduction enables sign language communication for deaf people. For them, VTP is an essential (assistive) technology for everyday life.

Also, paradoxically, VTP can benefit the blind. The EU project PROMISE emphasizes the use of the videophone as a remote eye for blind people living alone: letters and other documents can be shown to distant parties who can then read them aloud.

Today, VTP is developing towards a wider diversity of professional applications, e.g. desktop or portable videophones, partly usable as extensions of PCs. These scenarios are often equally well suited to home use, as they mostly utilize commercially available hardware. They are based on special software and the Internet. In 1999, the Internet does not routinely provide high-quality full-scale VTP.

VTP means more possibilities, but simultaneously also more demanding cognitive tasks. For instance, the transmission of both image and sound must be controlled during the communication. New ergonomic methods must to be developed to prevent shortcomings: constrained postures, discomfort in the neck–shoulder area, poor photometric and acoustic conditions, sustained visual load, rapid, demanding stimuli through both the eyes and the ears, and the mental load of being filmed all the time.

On the other hand, the VTP situation is natural; you can see and hear your partner remotely as in real life when you are together. Partly this still feels quite abstract: one must be able to handle virtual features, i.e. telepresence.

This article addresses the essential viewpoints of VTP ergonomics and future developments, covering both design and use perspectives.

2. BALANCED ERGONOMICS

Products cannot be used or designed in isolation. This article takes as a basis the model of a system consisting of five components: user, videophone, task, context, and environment (cf. Table 1)

2.1. Model Component: User

The users consist of very different individuals or groups ranging from an individual user of a customized unique videophone to all the potential users of ordinary videophones, i.e. consumers of high-tech and high-volume electronics. Hence, the whole range of principles, i.e. design for all or for one, such as a disabled person or a professional expert, may be applicable to VTP.

Providing a summary of the characteristics of the product's intended users should be one of the first steps in the design (Cushman and Rosenberg 1991). The user profile should include:

- age;
- gender;
- nationality;
- education;
- previous experiences with similar products;
- native language, including culture-dependent body language, e.g. the hands;
- reading skills in native and foreign language;
- mobility, sensory or cognitive impairments;
- occupation;
- skills; and
- motivation.

Figure 1. Typical (roll-about) VTP system is comprised of the following subsystems/components: (1) video input (camera), (2) video output (VDU), (3) auditory output (loudspeakers), (4) display of transmitted picture (PIP "viewfinder" feedback), (5) input device (mouse), (6) codec and connections to ISDN network in roll-about rack/housing, and (7) auditory input (microphone), input device for VDU (remote control) and another video input (document camera) on desk. System is in telemedical use by virtual team of three physicians. Photo: courtesy Videra Ltd.

Table 1. Guidelines (GI) by main headings. All the essential ergonomic interactions of VTP, called main headings, are presented as gray boxes. Each box indicates the place of one or more GI in the user-technology field. The last three GI are only described.

INTERACTION USER AND USING VS.	VIDEO TELEPHONY VTP					
	video input and output	auditory input and output	input devices i.e. controls	user-interface software	system configuration	
					terminal - specific	work station -specific
USER • eyes (seeing, showing) • ears (hearing, speaking) • head, neck and shoulders • upper limb • other body parts, body as whole • perception, memory • decision	13th GI 14th GI 15th GI 1st GI 2nd GI 3rd GI, 16th GI	17th GI 4th GI,	20th GI 6th GI 7th GI 5th GI 21st GI 22nd GI	18th GI 19th GI	11th GI 12th GI	8th GI 9th GI 10th GI
TASK AND CONTEXT	23rd GI:	Understand first deeply users' and organisations' interaction with and utilisation of VTP, utilise user-involvement, and look later for that VTP system will be systematically demonstrated, verified and validated				
ENVIRONMENT	24th GI: 25th GI:	Plan carefully and adjust properly as far as lighting conditions are concerned Remember that acoustic conditions and characteristics are very essential to VTP				

2.2. Model Component: Videophone

The generic key factors at the background of VTP systems are:

- networks (telephone, leased lines, local area networks, Internet, wireless, others); and
- connection to a network via a *co*der and *deco*der (*codec*) or corresponding software for digitizing and compressing image and voice signals, and vice versa.

Figure 1 shows a typical system configuration of commercial VTP in 1999 (cf. Table 1).

The major videophony systems are:

- Roll-about systems (Figure 1).
- Large wall panel systems (cf. information panels in public buildings).
- Desktop systems (often resembling multimedia PC).
- Portable systems (often resembling laptop PC).
- Handheld compact systems (features of mobile phones; see Chapter 5).

2.3. Model Component: Task

The tasks are two-fold: the tasks needed to operate the videophone and the tasks done by the videophone. The "substance tasks"

(VTP as a tool) range from a gentle, even "accidental," informal face-to-face discussions to urgent, complex tasks done between many geographical places. Such examples include telemedicine, engineering design, maintenance of machinery, etc., where remotely, transmitted, real-time, visual, high definition live and still information is the most significant value-added feature of VTP (cf. Kautto *et al.* 1997).

At the end of the 1990's, the most typical application is VTP for business meetings and distance learning. Travelling and travel costs can be decreased by multimedia on-line telecommunication, by the substitution of physical commuting with electronic commuting VTP may be feasible. Even between places not so remote, e.g. sections of a big building or different sites of a decentralized company.

One important type of communication is that between colleagues or peers. There may be two or more geographically distant designers who need to communicate in real time via an eye-to-eye or at least face-to-face video link with an audio link and a shared drawing surface, i.e. an electronic whiteboard equipped with a telepointer.

2.4. Model Component: Context

Private persons mainly use computer-based VTP models. It is expected that, in the future, there will be available a variety of handy mobile videophones for private use as well as for domestic, leisure and business purposes (see Chapter 5).

Most of the VTP end-users are professional employees in enterprises and the public sector. Various special groups of citizens, e.g. elderly or disabled people, are served remotely using home videophones. The reasons for using VTP for telework vary, including a desire to live in the country or a need to be near one's elderly parents or children. Virtual teams can be built: employees are connected via telematics interactively to the physical workplace, colleagues and the management.

2.5. Model Component: Environment

When used for videoconferencing, the videophone is typically placed in a normal meeting room, auditorium or small classroom. PC-based videophones are typically placed in an office, on management or technical help-desks, or on shop floors in industrial plants. Portable and future mobile VTP terminals can be moved from place to place in factories, offices or hospitals, or even be used outdoors and in vehicles.

3. INTERACTION–BASED ERGONOMIC CRITERIA: ELEMENTS FOR EVALUATION

3.1. User–Product Interface

VTP systems are analyzed in the light of human characteristics (Table 1). Interactions are characterized by a total of 25 heuristic Guidelines (Gl). More specific single points are also listed in the text, giving design rules and solutions developed for and applied to VTP by the authors and based generally or partly on the papers by Ivergård (1989), Shneiderman (1992), Dul and Weerdmeester (1993), Helander (1995) and Kroemer and Grandjean (1997).

3.2. Main Headings on Material Interaction

3.2.1. Video input and output versus head, neck and shoulders

1st Guideline: *Try to arrange VDU(s) and camera(s) so as to prevent unnecessary static load, fatigue and discomfort in the neck and shoulder*
- by preferring low enough placement: if the object of focus is close (up to 1 m) and the back posture is fairly erect, the optimal line-of-sight angle is ~30° below horizontal, and the maximal angle is 60° (Kroemer and Grandjean 1997);
- by preferring a horizontal or slightly lower line-of-sight for average-sized persons, which allows them to keep their trunk and head erect if the visual target is far away(>1 m); and
- by locating the built-in camera below rather than above the screen.

2nd Guideline: *Try to minimize the need to twist the head and trunk when looking at the VDU(s) and camera(s)*

3.2.2. Subsystems of VTP versus body as a whole

3rd Guideline: *Guarantee optimal body (-part) positions during the session*
- by ensuring that the joints are mostly in a neutral position;
- by locating all essential system components so as to minimize the need to twist the body;
- by preventing other constraints and/or static postures whether due to any subsystem; and
- by preventing poor hearing or vision that may cause poor posture, e.g. for elderly persons, with a limited range of clear vision, it is necessary to place the items in the system at a distance where they are clearly visible (cf. 14th Guideline).

4th Guideline: *Ensure that all subsystems are within easy and functional reach*
- by keeping the number of items that have to be touched minimal;
- by organizing well the VTP system, distinguishing between primary and secondary items, and giving the primary ones optimal accessibility; and
- by guaranteeing that the primary items are located within the reach of the 5th percentile.

5th Guideline: *Pay enough attention to the user profile*
- by taking into account specific user groups, e.g. the disabled and elderly;
- by supporting the diversity of human characteristics, e.g. left-handed and right-handed users; and
- by adapting to individual features affecting vision, e.g. eyeglasses and color blindness.

3.2.3. Input devices versus the upper limbs and the head, neck and shoulders

6th Guideline: *Optimize the functions of the input devices*
- by taking into account the importance, order and frequency of use of the different controls;
- by adjusting the C/R ratio optimally (Sanders and McCormick 1993);
- by avoiding cumulative repetitive movements of the mouse button or any key and awkward key combinations (the operation of several controls is only recommended to prevent unintentional touching);
- by employing the smallest, quickest and most accurate body

motions (response time is the sum of reaction time and motion time); and

- by considering remote controls which give the user more freedom.

7th Guideline: *Consider care of the anthropometrics and optimal postures when placing the input devices*

- by determining the distance between the controls anthropometrically. Cushman and Rosenberg (1991) recommend that the center to center spacing between any two buttons of a keypad or a touch-screen should be 19 mm, while the key size recommendation is 13 _ 13 mm);
- by allowing the wrist be mostly in a straight, neutral posture, especially if one has to move simultaneously or repetitively a hand or a finger;
- by remembering that wrist supports reduce the static load on the shoulders and arms, but the improper ones many contribute to some repetitive strain injuries; and
- by allowing the forearms to be supported on the chair or the work surface or in some other way.

3.2.4. Workstation versus whole body

8th Guideline: *Ensure care of proper dimensioning as far as different body sizes are concerned*

- by considering that the small person gives the measure for reach and the large for fit;
- by providing enough space for all the essential activities on surfaces;
- by reserving enough space for the thighs, knees, legs and feet;
- by allowing the feet to be firmly supported on the floor or a foot-rest; and
- by applying the recommendations for VDU workstations if feasible.

9th Guideline: *Give adjustability, support and variability a high priority when designing a workstation*

- by determining the adjustability range according to the user profile;
- by avoiding a work station without an adjustable keyboard or work surface height and adjustable height and distance of the screen meant for continuous use;
- by considering alternatives for the keyboard and mouse;
- by providing the chair with armrests when possible/desirable, and with a backrest with lumbar support up to 50 mm; and
- by increasing the degrees of freedom, e.g. by allowing a choice between different postures, by allowing easy adjustment, and by permitting natural movements;

10th Guideline: *Think carefully about all the essential joint angles of a sitting body*

- by remembering the pros and cons of the all proper sitting postures as far as VTP is concerned; and
- by realizing that the sitting posture makes it possible to focus one's attention better than the standing posture (Helander 1997): a lot of attention is needed in VTP, which favors classical and armchair sitting to sit–standing and standing.

3.2.5. Terminal versus whole body

11th Guideline: *Recognize essential system-specific key factors and take them into account ergonomically*

- by considering for weight, size and shape;

- by not accepting components that are unwieldy, unstable or difficult to grasp or handle (cf. center of gravity);
- by remembering the easy positioning of items if they are required to be held or manipulated at a distance from the trunk; and
- by benchmarking usability with successful known designs.

12th Guideline: *Remember safety, aesthetics and comfort characteristics*

- by analyzing kinetic, mechanical (sharp edges and protruding parts), chemical, electrical, thermal, pressure, radiation, EMC and software hazards;
- by noting the lifting, pushing or pulling characteristics/limits; and
- by giving comfort, appearance and physical degrees of freedom high priority.

3.3. Main Headings on Immaterial Interaction

3.3.1. Video input and output versus eyes

13th Guideline: *Make alphanumeric characters easy to read and understand*

- by paying attention to the demands set by screen and paper documents;
- by preferring a white-dominated screen with black letters if documents are frequently needed;
- by using text with familiar typeface and not consisting entirely of capitals;
- by using the common rule of thumb for character height h: h = 1/200th of viewing distance, which means a visual angle of 17 minutes of arc (capitals on a typical close VDU = 3 mm);
- by remembering that in a high-resolution display with good contrast, 9- or 10-minute characters may be legible enough;
- by recognizing that characters > 24–25 minutes of arc are inappropriate;
- by considering that the elderly, at least without glasses, can see precisely only so distant characters that are then at too small visual angle;
- by remembering the recommendations for "easy" line length, one of the most common being one-third of viewing distance; and
- by forcing the line spacing to be at least 1/30 of the line length.

14th Guideline: *Optimize watching conditions for vision, perception, fidelity and comfort*

- by using familiar symbols, pictograms and icons;
- by remembering that, for easy and quick eye movements, the regular viewing tasks should be within a 30° cone around the (principal, central) line of sight (cf. 1st Guideline); it means the following diagonals of the screen at different viewing distance:

Distance (cm)

	30	50	70		90	150	250	400
in inches	6.3	10.5	14.7(≈A4)		18.9	31.7	52.7	84.4
in cms	16	27	37		48	81	134	214

- by trying to minimize the number of items to be watched and focused on, e.g. the viewing distances to the screen, to the documents and to the keyboard should be as identical as possible because of productivity in general and for older users" eyes (Helander 1997);
- by weighting to allow for significant relationships between the photometric characteristics of the computer display and visual discomfort;
- by eliminating the effects of disturbances caused by mechanical shocks or vibrations affecting the camera; and
- by keeping in mind that, for special purposes, there are combinations of the screen and the camera that give the impression of being in direct eye-to-eye contact, and stereoscopic displays for enhanced fidelity.

15th Guideline: *Guarantee hi-fi quality of the transmitted image*
- by making the focusing of the camera easy during viewing;
- by ensuring optimal camera placement for unconstrained and realistic video communication, e.g. by utilizing voice-directed camera;
- by providing feedback helping to control the camera and to direct the behavior of the persons in front of the camera (e.g. PIP and possibly others corresponding viewfinder systems that guarantee properly placed and sized feedback of the transmitted image); and
- by providing clear status indication of whether the camera and transmission are switched on.

3.3.2. Video input and output versus perception and memory

16th Guideline: *Get acquainted with the basics of cognitive ergonomics related to display design*
- by considering the possibilities to strengthen signals due to the environment or the user's age (brightness, display time);
- by considering about information density and format on the video display; and
- by keeping in mind the 7 ± 2 rule, which gives the typical range of identifications people can make (Sanders and McCormick 1993).

3.3.3. Auditory input and output versus user in general

17th Guideline: *Try to find means, including noise control, to enhance the quality of audition in VTP*
- by noticing that auditory display is a challenge for VTP: provide directional accuracy and cues as to where to look on the VDU, and try to enhance situational awareness through the auditory sense when using VTP;
- by utilizing stereophony, which improves the listener's ability to localize the direction of sound;
- by applying echo cancellation;
- by utilizing a headset with a combined micro- and earphone;
- by attenuating the noise from the system, e.g. a noisy fan cooling the components;
- by preventing confusion which the synthesized speech or speech-based input the devices may generate (spoken messages); and
- by providing clear on/off indication of mute of auditory input or output.

3.3.4. User-interface software versus sensing

18th Guideline: *Emphasize the possibilities of software approaches when developing VTP further*
- by preferring software control that accounts for usability and such things as localization needs, elderly users and users with disabilities, i.e. a customized user interface;
- by consulting the standards before designing new pictographic symbols for the user interface software or other parts in the product; and
- by developing new symbols by selecting from several symbols the one that appears most effective in tests.

3.3.5. User-interface software versus perception, memory and decision

19th Guideline: *See that the general principles of human information processing and good practices are observed in the software*
- by preferring consistency in action sequences, terms, layouts, colors, etc.;
- by considering a direct-manipulation interaction style of the kind used in word processors, video games or CAD programs, which are transparent and concrete;
- by considering direct manipulation as a good way to master cameras, their sequence, focusing and zooming;
- by considering remote direct manipulation (e.g. using a keypad with arrow and function keys) as a useful possibility to be linked with VTP;
- by remembering that icons on the screen allow of recognition instead of recall and often provide spatial representations of the commands, which helps to make abstract concepts more concrete, especially for novice users (Majchrzak *et al.* 1987);
- by ensuring help and memory aid (user assistance);
- by designing the system so that operation involves a moderate rather than high short-term memory load;
- by designing the symbols and icons to be readily understandable, i.e. easy to decode by the user;
- by ensuring that operator errors can be easily corrected; and
- by utilizing standardization which refers to the common and familiar user-interface practices across the operation system.

3.3.6. Input devices versus eyes

20th Guideline: *Plan carefully in accordance with the established principles and standardized practices of cognitive ergonomics of controls*
- by keeping the use of colors limited because colors strongly affect attention;
- by designing the relative location of the cursor key on the keyboard so that it corresponds to the direction of movement (compatibility);
- by supporting blind orientation to the different keys on the numeric pad even by providing key 5 with a tactile cue;
- by making the minimum height of the lettering on the keys 2.5 mm;
- by using a standard layout for the keyboard, though a split and tilted version may be preferable; and
- by selecting one of the standard layouts for the numerical keypad, probably the one used in push-button telephones.

3.3.7. Input devices versus perception, memory and decision

21st Guideline: *Look for the recognized ergonomic design principles of input devices to enhance perception*

- by making operation easier by (1) anticipating the user's expectations concerning compatibility and standardization, (2) combining display and control, or (3) dialogue and anticipation between the user and the system;
- by promoting spatial compatibility, i.e. controls that follow the same layout as the system or displays; and
- by providing enough time between the execution of a response and the signal for the next response, whenever possible, let the user set the pace.

22nd Guideline: *Consider carefully the pros and cons of remote controls, touch-screens or speech control as well as other non-frequent possibilities*

- by designing the remote control so that the way is should be held is evident;
- by locating the important function keys of the remote controls consistently; and
- by remembering that as far as inexperienced users are concerned touch-screens have the advantage of providing

direct, concrete relationship between what the eyes see and what the hands do.

4. ERGONOMIC EVALUATION

A VTP system is subject to numerous and often contradictory requirements. Special attention should, therefore, be given to tradeoffs during the design or purchasing. In any case, the demands related to "the eyes and upper limbs up to the shoulders" are of great importance. These must be given enough weight when using multi-criteria decision aiding procedures for making tradeoffs (e.g. use value analysis; Figure 2). These two are though not only centers when designing VTP ergonomics. The key principle of effective design might be a balanced compromise in ergonomic specification.

5. EVOLUTION TOWARDS WIRELESS HAND–HELD MEDIAPHONE SYSTEMS

Developments in the wireless communication technology open up new possibilities in the implementation of the videophone as
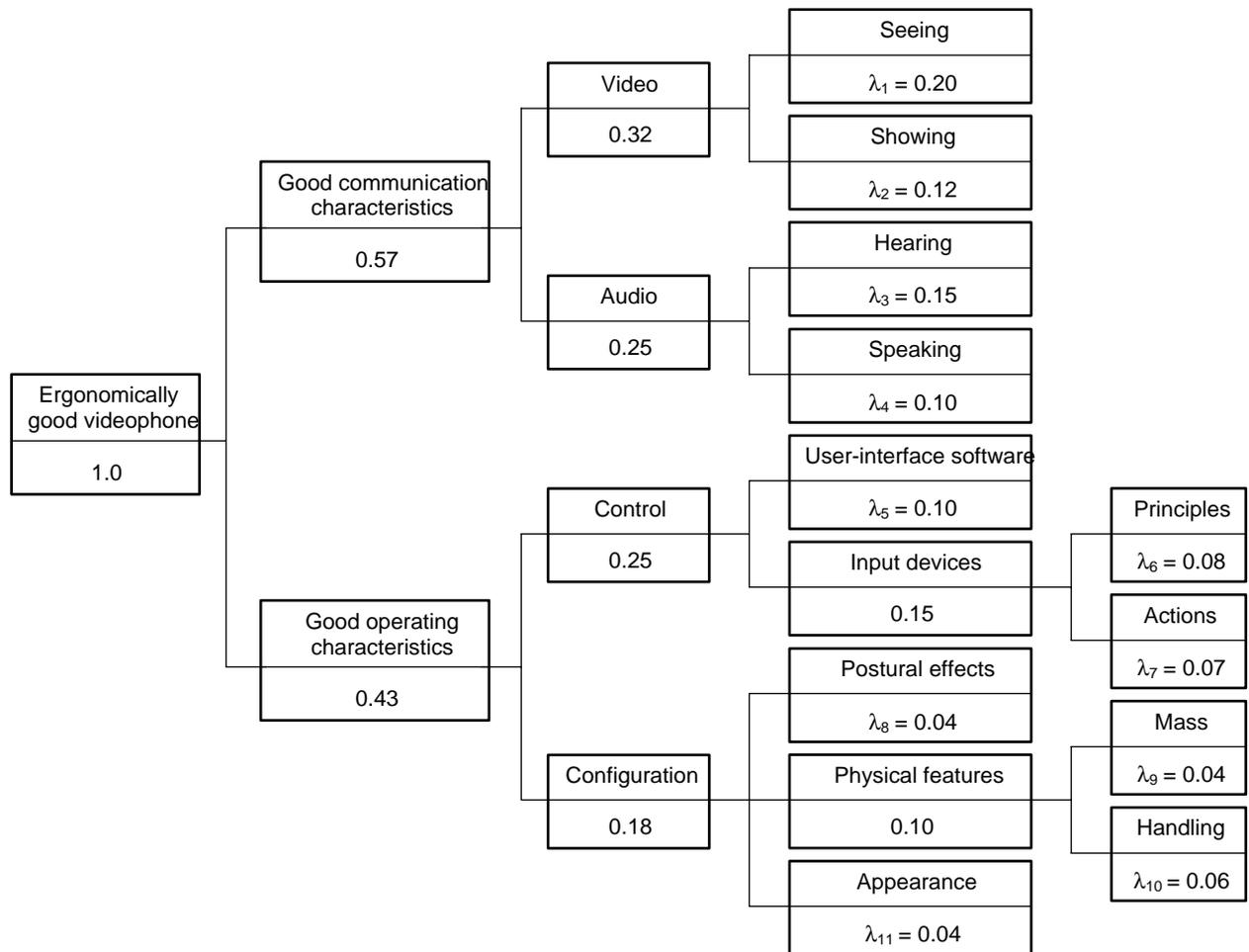
Figure 2. Example of one possible ergonomic objectives tree model for evaluating videophones. This model enables a multi-criteria tradeoff based on 11 final objectives, i.e. criteria ($j = 11$). Each criterion has a weight factor l_j corresponding to the relative importance the criterion, e.g. $l_1 = 0.20$ means that the proportional effect of good visual output characteristics on the total goodness is 20%. Optional i alternatives can be score-rated for each criterion (e_{ij}), e.g. on a five-point scale, from 0 = unsatisfactory to 4 = very good (Pahl and Beitz 1988). Total ergonomic goodness EG of each VTP alternative can be calculated as a weighted average ($EG_i = \hat{A} l_j e_{ij}$), expressed on the same scale. The cumulative sum of the hierarchic and subtractive shared weights is hence a total of one ($\hat{A} l_j = 1.00$).

an extension of the current cellular phone. Especially the broadband digital wireless networks that will emerge ~2001–05 hold promise for the widespread use of media phones, e.g. cellular phones with VTP characteristics. Mobility adds a completely new dimension to the basic videophone characteristics: telepresence and augmented reality. Telepresence is enabled by a "carry along" arrangement, where the videophone is stereoscopic and follows the movement of the users' head. Augmented reality is a further extension of telepresence, where synthetic, i.e. computer-generated, information is added to the telepresence image to convey some meaningful additional information belonging to the context, such as navigational information, personal identities, etc.

Telepresence and augmented reality bring in new requirements in addition to the videophone requirements: field of vision, registration accuracy and real-time responsiveness. Visual field involves the question of how many degrees of the normal viewing range of humans need to be covered? Increasing the visual field improves the telepresence effect, i.e. immersion, but at the same time detracts from resolution. Registration accuracy is defined for the accuracy of positioning annotations (graphical symbols, alphanumeric characters) on top of a stereoscopic video telephone image. Real-time responsiveness refers to the dynamic aspect of registration accuracy, i.e. how well the annotations are able to follow the respective objects in the image when the user moves.

Figure 3 presents the Cyphone Mediaphone, which was designed to study portable real-time stereoscopic video input and output enabling telepresence. The receiver is able to see the same view as the transmitter. The problem of different user sizes has been solved by a binocular-like mechanism. The technology of miniature displays is expected to reach a level sufficient for high quality and resolution of image viewing.

The user interface has been kept as simple as possible by eliminating the conventional numeric keypad. Instead, a numbering input device has been integrated in a touch-sensitive surface, which also serves as a touch-pad. Cyphone has a transparent and logically intuitive user interface, because the touch-pad and the screen have the same size. This gives a realistic feeling to the menus. The shape of the left side has been designed to bring the mini-displays near the eyes and also to support grip. The microphone has been placed at the edge of the touch-pad and the loudspeaker on the upper side of the display.

Cyphone has two cylindrical shapes going through it. There is no need for a physical connection between the cameras and the mini-displays, but a binocular impression has been wanted to create in order to give the users intuitive help on how to use the mediaphone. The shape also gives good support for grip. For some applications, such as personal navigation, a head-mounted display in the form of intelligent eyeglasses or sunglasses is foreseen, although the technology may only become mature somewhat later.

6. CONCLUSION

VTP in its different forms is a technology of today, which becomes more feasible every day. Ergonomics, e.g. usability, and various contextual factors, e.g. organization of jobs, are crucial as far its success is concerned. It remains to be seen what kind of physical systems are most effective and best preferred by the end-users.
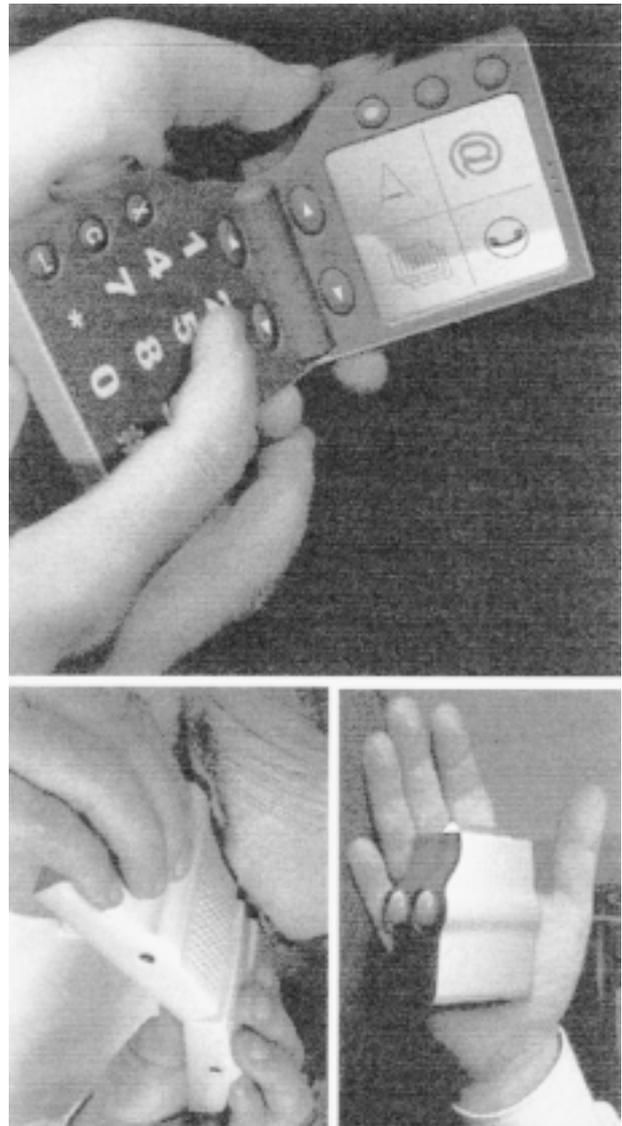

Figure 3. Cyphone, a mobile mediaphone developed at the University of Oulu. Telepresence and augmented reality applications are supported by the stereoscopic image input and output.

Various all-in-one systems will obviously become more common, among them hand-held compact systems, which have a high priority in the plans of telecommunication companies in 1999.

VTP normally causes an increased level of alertness, which further demands the user to have corresponding vigilance abilities. It also means that the user must have an adequate mental model of the devices and the whole system. The human stereotypes, if they are taken into account in a VTP system, can support in demanding using situations. The user interface software meets many challenges. So do for instance augmented reality and annotations.

It is evident that the users must put enough weight on the ergonomic characteristics of VTP. It is therefore possible to gain benefits both at utilitarian and subjective, i.e. hedonic, experiential and emotional, levels. To meet the expectations of the users of this demanding hybrid product, laborious design tradeoffs are needed by manufacturers. Most probably,

increasingly diverse VTP solutions will be needed in the future to satisfy the growing number of users and ergonomically challenging applications.

REFERENCES

CUSHMAN, W. and ROSENBERG, D., 1991, *Human Factors in Product Design: Advances in Human Factors/Ergonomics 14* (Amsterdam: Elsevier).

DUL, J. and WEERDMEESTER, B., 1993, *Ergonomics for Beginners: A Quick Reference Guide* (London: Taylor & Francis).

HELANDER, M., 1995, *A Guide to the Ergonomics of Manufacturing* (London: Taylor & Francis).

IVERGÅRD, T., 1989, *Handbook of Control Room Design and Ergonomics* (London: Taylor & Francis).

KAUTTO, J., VÄYRYNEN, S. and KIRVESOJA, H., 1997, Ergonomic requirements of a videotelephony-based telesupport system for demanding industrial maintenance. In Seppälä, Luopajärvi, Nygård and Mattila (eds) *From Experience to Innovation IEA '97 Proceedings* (Helsinki: Finnish Institute of Occupational Health), vol. 7, 72–4.

KROEMER, K. and GRANDJEAN, E., 1997, *Fitting the Task to the Human: A Textbook of Occupational Ergonomics* (5th ed.) (London: Taylor & Francis).

MAJCHRZAK, A., CHANG, T.-C., BARFIELD, W., EBERTS, R. and SALVENDY, G., 1987, *Human Aspects of Computer-Aided Design* (London: Taylor & Francis).

PAHL, G. and BEITZ, W., 1988, *Engineering Design — A Systematic Approach* (Berlin: Springer).

SANDERS, M. and McCORMICK, E., 1993, *Human Factors in Engineering and Design* (7th Ed.) (New York: McGraw-Hill).

SHNEIDERMAN, B., 1992, *Design the User Interface: Strategies for Effective Human–Computer Interaction* (2nd Ed.) (Reading: Addison-Wesley).

Virtual Environments

W. Barfield† and E. B. Nash

Virtual Environment Laboratory, Department of Industrial and Systems Engineering, Virginia Tech, Blacksburg, VA 24060, USA

1. INTRODUCTION

Virtual environments are interactive, multisensory, three-dimensional computer-synthesized environments that can be manipulated by virtual environment participants (Barfield *et al*. 1995b, Furness and Barfield 1995). By providing an egocentric frame-of-reference, virtual environments let the user be immersed within the computer simulation, resulting in a suspension of disbelief in which the participants feel like they are actually present in the virtual environment. This chapter discusses the visual, auditory, haptic and olfactory components of virtual environments, and the concept of presence in virtual environments.

2. COMPONENTS OF VIRTUAL ENVIRONMENTS

The sensory representations that characterize virtual environments are typically conveyed to participants using several different types of virtual environment display technology. Using virtual environment display technology, combinations of visual, auditory and/or haptic stimuli are displayed to the senses at sufficient fidelity such that they appear to originate from within the three-dimensional space surrounding the user. This chapter will briefly discuss the features of such displays and how they convey information to the visual, haptic and olfactory modalities.

2.1. Visual Display Technology

The visual component of virtual environments typically consists of a stereoscopic, computer-generated image viewed using an opaque head-mounted display (HMD) or a stereoscopic image viewed using time multiplexed shutter glasses (Figure 1).

Using a HMD, a virtual environment is created using two separate displays (e.g. nominally 1 inch^2) which provide two slightly offset views of the same scene. The magnitude of the lateral disparity provided by the two images is dependent on the interocular distance between the viewer's eyes. Through the use of optical elements that magnify, collimate and project the images via a mirror combiner, the computer-generated image appears as a large picture or full-scale three-dimensional scene suspended in the world. Furthermore, a scene viewed using a head-tracked HMD will provide the user an egocentric frame of reference, where objects are viewed from the perspective of the user's position in the three-dimensional space provided by the virtual environment.

Interactivity within the virtual environment is achieved when the psychomotor (and in some cases, physiological) behavior of the user can be monitored by the system and used as input to manipulate the virtual display. Physical movements of the user can be integrated with visual display information when a position sensor system is used. Such a system will typically track the user's head and hand positions. With these types of direct physical mappings, the user can control the virtual environment within natural visual and motor skills, providing a display that is "visually coupled" to the user (Kocian and Task 1995). It is important to note that coupling is an important concern when interactivity is desired, regardless of the type of display.

Figure 1. A virtual environment viewed through an opaque head-mounted display (HMD). Also shown is a glove with a 6° of freedom position sensor allowing the user to manipulate virtual objects.

2.2. Auditory Display Technology

For virtual environments, the goal of auditory display technology is to manipulate sound sources so that they assume localized positions within the virtual environment, mapping the source channel into the three-dimensional space of the user (Cohen and Wenzel 1995). The most conventional approach to spatial sound generation employs a hardware- or software-based convolution engine that convolves a monaural input signal with pairs of digital audio filters to produce output signals for presentation over stereo loudspeakers or headphones.

2.3. Haptic Display Technology

Haptic displays can be classified into two groups: tactile displays, which provide sensations of shape and texture on the surface of the skin, and kinesthetic displays, which provide the sensations of position and force in muscles and joints. The goal of haptic displays when used with virtual environments is to persuade the user to attribute haptic sensations to the parameters of the virtual objects rather than to the immediate location of the sensation (i.e. the haptic display's contact with the skin). This psychological phenomenon is referred to as distal attribution (Loomis 1992).

2.3.1. Tactile displays

The three primary tactile display technologies associated with virtual environments are static mechanical tactile stimulation, vibrotactile stimulation and electrotactile stimulation. Although the former would readily provide the means necessary to produce the sensation of skin deformation at the exact moment of touch with a virtual object, these types of displays are not regularly used due to high power consumption and the rapid adaptation to static stimuli (Kaczmarek and Bach-y-Rita 1995).

2.3.2. Kinesthetic displays

Kinesthetic displays used in virtual environments operate to produce sensations of mechanical energy flow. Specifically, electric motors provide force feedback against the limbs of the body to mimic the forces of physical interaction experienced while interacting with real world objects. Coupled with the actions of the virtual environment participant, kinesthetic displays can model energy flow relationships bi-directionally. This entails both sensing the parameters of the virtual objects and their actions and reproducing these as physical forces to the user, and sensing forces input by the user and reproducing these as actions within the virtual environment.

2.4. Olfactory Displays

Other senses are not exempt from application within virtual display technology. Though olfaction has received relatively little attention among researchers (and gustatory none at all), the olfactory modality provides a potentially rich source of information to users in virtual environments. Potential applications include training for the recognition of hazardous materials, room identification by odor and the demarcation of exit pathways (Cater 1992). In practice, virtual olfaction displays can be created for use with virtual environments using a system of hardware, software and chemicals to deliver the olfactory stimulus. However, before they can be of use, psychophysical issues such as detection thresholds, accuracy of odor recognition, odor localization, odor discrimination and channel capacity must be fully understood (Barfield and Danas 1996).

3. SELECTED HUMAN FACTORS ISSUES IN VIRTUAL ENVIRONMENT TECHNOLOGY

3.1. Perceptual Capabilities and Limitations of Humans

It is understood that human senses have certain capabilities and limitations. Thus, the capabilities and limitations of the senses can be used to provide guidelines for the design of virtual environment display technology (Barfield *et al*. 1995a). For example, visual, haptic or olfactory displays need not be designed with a resolution that exceeds the capabilities of the human's senses to perceive the presented information. In the context of vision, because the retina contains fewer cones towards the periphery compared with the fovea, it is relevant to ask whether visual display resolution should also decrease toward the periphery of the display and how realistic color should be represented in the periphery.

At present, the full integration of human sensing and manipulation capabilities with current display technologies is impeded by either the lack of knowledge of human capabilities or by limitations in technology. With haptic displays, this means that there are human movements that cannot be adequately sensed, or, more importantly, human senses that are under-utilized. The present disparity between rate of task performance in teleoperated or virtual environments, compared with real ones, is the net result of such disparity. If we can completely match the specifications of human sensing with virtual environment display characteristics, then it may reduce this disparity to zero. This will not likely happen in the near future.

3.2. Relevance of the Task

The optimal coupling of display fidelity with human's sensory capacities will also be task-dependent. That is, not all tasks will require the largest field of view, the most realistic tactile feedback, or the most realistic spatialized sound technically possible. In this regard, Barfield et al. (1995a, b) have discussed the tradeoff between field of view and display resolution in the context of the required application and task. If the task is to navigate through a virtual environment, then a HMD with a wide field of view may be necessary. However, if the task is to control a slave manipulator at a remote site, then a HMD providing a smaller field of view with higher resolution may suffice.

3.3. Prolonged Exposure to Virtual Environments

Before the use of virtual environments will gain widespread prevalence, researchers must address the question of the effect of prolonged exposure to virtual environments on the human's ability to re-adapt spatially and cognitively to the real world. This consideration becomes especially important when the stabilization of the virtual environment is poor. In such cases, exposure may lead to symptoms of simulator sickness, manifested as general discomfort, apathy, drowsiness, headache disorientation, fatigue, pallor, sweating, salivation, stomach awareness, nausea and vomiting after or during a session with a virtual environment. Isolation of these causal factors must occur before prevention techniques can exist.

4. PRESENCE IN VIRTUAL ENVIRONMENTS

Because virtual environment display technology has the capacity to surround the user with three-dimensional stimuli from multiple sensory modalities, under ideal conditions these components may interact to provide the viewer with a sense of presence within the virtual world (Barfield and Hendrix 1995). In our view, presence is not a physical but rather it is a cognitive state that emerges from the information the human sensory organs collect and convey to the brain. Presence is a desirable effect of a participant's interaction with a virtual environment. In fact, most virtual world applications are designed with the objective of creating an environment that will engage and lure the participant into "feeling" present in the environment.

Hendrix and Barfield (1996) have shown that presence in virtual environments is dependent on the type of input device, display characteristics (such as stereopsis and field of view [FOV]), and the presence or absence of head tracking. Among input devices, these include the nature of the input device itself, the ease of input control and how well the input device allows the participant to move within the virtual space of the virtual environment. Among display characteristics, these include the level of engagement of the displayed environment, geometric field of view and stereoscopic cues (Hendrix and Barfield 1996). In addition, researchers have postulated that an increase in display fidelity will influence the level of presence in a virtual environment (Hendrix and Barfield 1995) and that other display variables, such as scene update rate, will influence the fidelity of the scene and, hence, the sense of presence (Sheridan 1992).

5. CONCLUSIONS

In this chapter is presented the display components and interface concepts related to virtual environments. Once the gamut of available display technologies become more fully developed, these must be integrated into the virtual environment such that optimal combinations are provided to fit the capabilities and limitations of humans and the tasks they perform. In the future this strategy may serve to provide users with minimal disparity between sensations experienced during tasks performed in virtual environments and those experienced during tasks in the real world.

REFERENCES

BARFIELD, W. and DANAS, E., 1996, Comments on the use of olfactory displays for virtual environments. *Presence: Teleoperators and Virtual Environments*, 5, 109–21.

BARFIELD, W. and HENDRIX, C., 1995, The effect of update rate on the sense of presence within virtual environments. *Virtual Reality: Research, Development, and Application*, 1, 3–15.

BARFIELD, W., HENDRIX, C., BJORNESETH, O., KACZMAREK, K.A. and LOTENS, W., 1995a, Comparison of human sensory capabilities with technical specifications of virtual environment equipment. *Presence: Teleoperators and Virtual Environments*, 4, 329–56.

BARFIELD, W., ZELTZER, D., SHERIDAN, T. and SLATER, M., 1995b, Presence and performance within virtual environments. In W. Barfield and T.A. Furness III (eds) *Virtual Environments and Advanced Interface Design* (New York: Oxford University Press), 473–513.

CATER, J.P., 1992, The nose have it! *Presence: Teleoperators and Virtual Environments*, 1, 493–4.

COHEN, M. and WENZEL, E.M., 1995, The design of multidimensional sound interfaces. In W. Barfield and T.A. Furness III (eds) *Virtual Environments and Advanced Interface Design* (New York: Oxford University Press), 291–348.

FURNESS III, T.A. and BARFIELD, W., 1995, Introduction to virtual environments and advanced interface design. In W. Barfield and T.A. Furness III (eds) *Virtual Environments and Advanced Interface Design* (New York: Oxford University Press), 3–15.

HENDRIX, C. and BARFIELD, W., 1995, Presence within virtual environments as a function of visual and auditory cues. *IEEE Virtual Reality Annual International Symposium*, 74–82.

HENDRIX, C. and BARFIELD, W., 1996, Presence within virtual environments as a function of visual display parameters. *Presence: Teleoperators and Virtual Environments*, 5, 274–89.

KACZMAREK, K.A. and BACH-Y-RITA, P., 1995, Tactile displays. In W. Barfield and T.A, Furness III (eds) *Virtual Environments and Advanced Interface Design* (New York: Oxford University Press), 349–414.

KOCIAN, D.F. and TASK, H.L., 1995, Visually coupled systems hardware and the human interface. In W. Barfield and T.A. Furness III (eds) *Virtual Environments and Advanced Interface Design* (New York: Oxford University Press), 175–257.

LOOMIS, J.M., 1992, Distal attribution and presence. *Presence: Teleoperators and Virtual Environments*, 1, 113–19.

SHERIDAN, T.B., 1992, Musings on telepresence and virtual presence. *Presence: Teleoperators and Virtual Environments*, 1, 120–5.

Virtual Reality

R. C. Williges

Department of Industrial and Systems Engineering, Virginia Polytechnic Institute and State University, Blacksburg VA 24061, USA

1. INTRODUCTION

Virtual reality (VR) is a term that describes a variety of computer-mediated techniques that are used to represent various aspects of the physical world to the user. Although most of the current techniques deal with representations of the visual environment, consideration has also been given to representing the entire human sensory environment including visual, auditory, touch, and motion domains. Consequently, virtual environment (VE) is often used as the more descriptive term of this family of computer-augmented techniques.

1.1. Degree of Virtuality

A more comprehensive view of the VE is to consider a reality-virtuality (R-V) continuum. For example, Milgram and Drascic (1997) describe an R-V continuum from the real world to a VE created totally with computer modeling. In between these two extremes various forms of mixed reality can exist. If the environment is primarily the real world with aspects of the VE superimposed on it, then one has an augmented reality environment. On the other hand, if the environment is primarily a VE with aspects of the real world included within it, then one has an augmented virtuality representation.

Representing the VE along this R-V continuum allows the designer to consider a variety of mixed reality alternatives for design application rather than just a pure real world or completely modeled VE. For example, augmented reality can be used to enhance or highlight aspects of the real world such as guiding industrial quality control inspection through computer augmentation. Alternatively, augmented virtuality can be used to enhance the VE by depicting a picture of the user's real hand pointing at or grasping an object in the VE so that the user has a higher degree of perceived presence and sensory immersion in the VE.

1.2. Telepresence

One of the most important considerations in the design and use of VR is to facilitate the user's perceived presence of actually being in the VE. Draper, Kaber, and Usher (1998) provide a comprehensive review of various parameters involved in the perception of presence in a physically remote or computer-simulated site as characterized by VR systems, telecommunications, and teleoperations. They describe simple, cybernetic, and experiential telepresence. Simple telepresence is the user's ability to operate the computer-augmented environment. Cybernetic telepresence is a quality index of the human–machine interface itself based on the compatibility of human performance with the computer-mediated environment. Experiential telepresence refers to the mental state of the user who feels physically present in the synthetic environment.

Draper, Kaber, and Usher (1998) review both technology-based and psychologically-based explanations of telepresence. Technology-based explanations relate factors such as sensory feedback, fidelity of sensory information, control dexterity, display quality and consistency, and degree of sensory immersion to telepresence. Psychological explanations of telepresence, on the other hand, involve characteristics such as task flow, self-attribution of presence in the VE, situation awareness, and cognitive factors related to information processing and decision making that affect perceived presence in the VE. They proposed a structured attention resource model as a means of describing and predicting the degree of telepresence.

2. DESIGN ISSUES

The human factors specialist must consider a variety of design issues when using VR. Stanney, Mourant, and Kennedy (1998) provide a comprehensive review of over 150 scientific articles related to human factors issues in the VE. They organized their review around three major issues including human performance efficiency, health and safety issues, and the social impact of technology. Human performance efficiency included issues pertaining to task characteristics, user characteristics, human sensory and motor physiology, multi-mode interaction, and design metaphors. Health and safety issues included cybersickness and deleterious physiological after-effects. Social issues deal with the negative impact of user's misuse of VE technology such as excessive exposure to violence in VR gaming applications. Based on this review and current research in VEs, one must consider several critical issues when designing applications of VR.
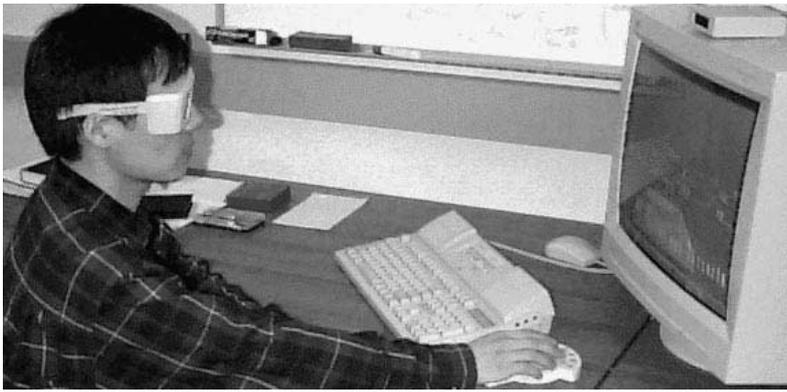
2.1. Display Alternatives

Visual displays need to be considered carefully when developing VR applications especially in augmented reality displays that combine real and virtual objects in the same field of view. Localization of these objects, superimposing images, and aligning virtual and real three-dimensional (3D) objects are critical for user acceptance and performance.

Current VR display technology allows several alternatives for presenting the computer-mediated visual scene. Figure 1 depicts three of the most common classes of visual display of the VE. Desktop VR can be used to present either perspective 3D scenes on the display screen or true stereoscopic 3D images by viewing the display screen through shutter glasses as shown in Figure 1a. A head-mounted display (HMD) can be used to provide a more immersive visual environment with a limited field of view (FOV). The HMD can be an augmented reality display as shown in Figure 1b in which the user can see through the HMD to superimpose the virtual image on the real world. The visual display can also incorporate head tracking shown in Figure 1c to adjust the visual scene depicted on the HMD according to head movements of the user. The most immersive VE visual scene uses full wall projections providing a large FOV shown on one to six screens in a completely enclosed VR room as shown in Figure 1d.

2.2. Control Alternatives

Various control alternatives are available for users of VEs. Four commonly used controllers used in VR are a 3D ball, 3D puck, wand, and dataglove as depicted from left to right in figure 2.

Figure 1. Display alternatives for virtual environments; a. Desktop display; b. Augmented reality display; c. Head tracking display; d. Room display.

Figure 2. Four control alternatives for virtual environments.

Most often these controllers are used to control visual orientation and locomotion within the visual scene in the VE. A certain amount of training and practice is required to achieve proficiency in the use of these controllers. Care must be taken to chose the appropriate controller for the specific application so as not to interfere with task performance.

2.3. Perceived Presence

Instilling a feeling of presence in the VE as compared to simply viewing the VE from an outside real-world perspective is often a design goal in VR applications. The feeling of presence is an evaluation made by the user and measured by subjective opinions, surveys, and rating scales. However, perceived presence does not have a perfect positive correlation with performance in the VE, and it depends upon a variety of design parameters. For example, Snow and Williges (1998) isolated seven parameters that significantly influenced perceived presence. They found that FOV,

sound, and head tracking, had almost three times more influence on presence than visual display resolution, texture mapping, stereopsis, and scene update rate.

2.4. Cybersickness

Negative side effects ranging from mild discomfort, to headaches, to nausea can occur in VR. This type of discomfort is somewhat similar to motion sickness and is generally referred to as cybersickness. Individuals differ in terms of their susceptibility and severity of exhibited symptoms, but usually their discomfort is short lived after exiting the VE. Several parameters of the visual scene including limited FOV, visual vection or flow in peripheral vision, sensory cue conflicts, lags, etc. need to be considered by the human factors engineer as ways of minimizing potential health and safety risks related to cybersickness.

2.5. Orientation and Navigation

Navigation throughout the VE can be problematic for new users and can hamper their performance. Human factors issues include both features that affect orientation and aids to improve navigation. Many factors related to the design and representation of features such as the number of edges used graphically to depict an object in the VE affect user orientation and navigation. Understanding the influence of these factors is a critical to fostering effective use of the VE. Navigational aids such as visual movement or momentum, maps, visual portals or doorways, and 3D representations can be used to improve navigation in the VE.

2.6. Sensory Modality

Although the primary focus has been on the visual display of the VE, other sensory modalities have been considered. Three other prominent sensory channels include auditory, motion, and touch modalities. Alternative devices are also available for special applications such as speakers for 3D spatialized sound, force sensitive controllers for providing haptic feedback, and treadmills to allow locomotion. Both the primary effects of each sensory modality and the interaction of various multiple sensory inputs need to be considered in designing the VE.

2.7. Individual Differences

Individual differences among users also influence the design and use of VR. Two important classes of users that require special consideration are children and the elderly. Most VR equipment is designed to be used by young adults, but there are age differences in the use of this equipment. As children begin to use VR more often in educational and leisure activities, the equipment needs to be designed to fit their physical capabilities, to minimize their discomfort and cybersickness, and to foster their spatial orientation in the VE. Additionally, the reduction in physical as well as sensory capabilities of vision and hearing need to be considered by the human factors engineer when using VR devices with the elderly.

3. APPLICATION AREAS

The use of VR technology has been expanding rapidly. Different application areas often raise unique human factors and ergonomic issues that need to be considered in order to facilitate appropriate use of this emerging technology.

3.1. Entertainment

Probably the first use of VR was computer games in which the player is immersed into the VE. The popularity of these games is growing and has a large economic impact in our society. Additionally, VR has been used extensively in the motion picture industry as a means of providing realistic animation and representation of environments, characters, and objects that do not exist in the real world but are critical components of the fantasy world depicted on film. Given the commercial market potential, entertainment will continue to be a leading application for VR. Little human factors attention has been given to determining the characteristics of the virtual representation that is engaging and provides the greatest entertainment for the user.

3.2. Design

Applications of VR have been useful in design applications. Specifically, in architectural design, 3D representations and virtual walkthroughs of structures during various phases of design have proven useful in facilitating design decisions, isolating design problems, and promoting user acceptance. The exact nature of the role of visualization and the level of visualization required in various stages of architectural design is still an unresolved human factors engineering issue.

Engineering design applications of VR have proven to be a valuable aid to workers in manufacturing systems. In particular, augmented reality is useful as an aid to workers in the assembly of complex components in aircraft systems and in the reduction of visual inspection time while maintaining inspection accuracy as compared to purely manual inspection. By superimposing a virtual projection of instructions, measurement points, etc. in the physical manufacturing process, the augmented reality display provides valuable guidance to the worker. Likewise, VR applications can be made to the design of workplace layout and ergonomic analysis of workstations.

3.2. Conferencing

An emerging application area is the use of VR to facilitate group and team work through computer conferencing using a VE. For example, conferences of distributed team members could be accomplished through networked, fully immersive VEs. The degree of virtuality on the R-V continuum needed to facilitate team communications, techniques to coordinate task performance, and improved decision making procedures are examples of macroergonomic issues that need to be considered.

3.4. Education

Educational use of VR is rapidly increasing. For example, scientific visualization of complex mathematical and chemical structure using VR technology can lead to improved understanding and new insights. Likewise, many areas of medical science related medical diagnosis and surgical treatment could be demonstrated and practiced in a VE without danger to live patients. The necessary and sufficient level of telepresence in the VE visualization required to promote learning still needs to be specified.

3.5. Simulators

The area of application of VR that provides the most important set of human factors issues is the design and use of VEs as a

synthetic training devices. The visual display and multisensory feedback provided by VR technology is ideally suited for use in training simulators of space, air, land, and naval transportation vehicles. Simulation training can improve operator safety and reduce operational costs as compared to real-world training in these vehicles. The VE can be used both in part-task and whole-task training environments. However, issues related to the effect of perceived presence on transfer of training and the minimization of cybersickness need to be addressed by human factors engineers when using VR simulators.

4. AREAS OF DEVELOPMENT

Although applications of VR have increased tremendously, several areas of development are being addressed by current research in order to facilitate further improvements. Most of these developments are methodological.

4.1. Metrics

Standard evaluation metrics are used in order to facilitate comparison of various alternatives of VR. Rating scales of discomfort and perceived presence in VE have been developed. One example of specifically defined metrics in VE is the use of free modulus magnitude estimation as a convenient psychophysical technique for assessing a user's perceived presence in VE.

4.2. Integrated Database

The results of empirical studies in VR need to be integrated into a database that includes the results of several studies. Both a common set of tasks and a procedure for representing and interrogating the integrated VR database need to be developed before integrated databases can be built.

Lampton, Kneer, Goldberg, Bliss, Moshell, and Blau (1994) developed a battery of generic VE tasks to support training applications of VR. Their battery includes five task categories and several subtasks within each category. The visual task includes subtasks related to acuity, color, object recognition, size estimation, distance estimation, and search. The locomotion task has both walking and flying components. The walking components include straightaway, backup, turns, figure eight, and doorway subtasks; whereas, the flying category includes window and elevator subtasks. The manipulation task includes slide, dial, and bins subtasks that involve various ways of grasping objects in the VE. The tracking task includes both head and device control in regard to stationary and moving targets. Their fifth task is a reaction time task including simple and choice reaction times. Developers of a VE application can choose the appropriate set

of tasks and subtasks from this battery for their application and, thereby, have a common baseline for comparing alternative designs of the VE application.

Sequential experimentation can be used as a mechanism for combining data across several experiments into an integrated database using a common subset of this VE task battery. Fractional-factorial designs and central-composite designs along with a common data point and data bridging can be used to collect data in an efficient manner on a large set of factors affecting VE performance. Empirical models based on polynomial regression can subsequently be generated from the resulting integrated database. Snow and Williges (1998) demonstrated the use of these techniques as a means of predicting the effect of several system factors and their interaction on perceived presence in a VE.

4.3. Design Guidelines

Little guidance exists for the appropriate user-centered design and application of VR. As more research is conducted, as case studies of applications increase and as integrated databases are developed, empirically-based human factors and ergonomic design guidelines and standards will develop for the use of VEs.

5. CONCLUSIONS

The technology involved in generating VR systems is advancing rapidly, and applications of this technology are escalating. Because VR is still an emerging technology, this area offers exciting and challenging issues for human factors engineers to address in order to harness the potential of this technology for human use.

REFERENCES

DRAPER, J.V., KABER, D.B. and USHER, J.M., 1998, Telepresence. *Human Factors*, 40(3), 354–75.

LAMPTON, D.R., KNERR, B.W., GOLDBERG, S.L., BLISS, J.P., MOSHELL, J.M. and BLAU, B.S., 1994, The virtual environment performance assessment battery (VEPAB): Development and evaluation. *Presence*, 3(2), 145–57.

MILGRAM, P. and DRASCIC, D., 1997, Perceptual effects in aligning virtual and real objects in augmented reality displays. *Proceedings of the Human Factors and Ergonomics Society 41st Annual Meeting* (Santa Monica, CA: Human Factors and Ergonomics Society).

SNOW, M.P. and WILLIGES, R.C., 1998, Empirical models of perceived presence in virtual environments based on free-modulus magnitude estimation. *Human Factors*, 40(3), 386–402.

STANNEY, K.M., MOURANT, R.R. and KENNEDY, R.S., 1998, Human factors issues in virtual environments: a review of the literature. *Presence*, 7(4), 327–51.

Virtual Reality: Virtual and Synthetic Environments – Technologies and Applications

R.J. Stone

Virtual Presence Ltd, Chester House, 79 Dane Road, Sale M33 7BP, UK

1. INTRODUCTION AND DEFINITION

With most new and exciting technologies, there is often an early tendency for the pioneers and "evangelists" to become preoccupied with novel hardware and software and divorced from the requirements of commercial applications and user needs. Typically the result is a painfully slow and often costly uptake process on the part of industry and commerce. "Technology push" outweighs both "market pull" and sensible implementation based on sound ergonomics principles. So it was in the past with speech recognition, robotics, artificial intelligence and neural networks. So it was with Virtual Reality (VR). That is, until quite recently. Since 1997, and after an incubation period lasting some 6–7 years, VR has experienced something of a massive revival. A revival that has taken the form of a number of important developments that have helped VR to become an *accessible*, usable and justifiable technology for many industrial, commercial, educational and public sectors.

To appreciate the reasons for this revival, one has to understand what VR actually is: "Virtual Reality refers to a suite of technologies which permit intuitive, real-time interaction with three-dimensional databases …" (Stone 1996a). Absent from this definition are what many still believe are the two most important features of VR (prerequisites, almost): computer graphics and *"immersion"*. The term immersion (or "inclusion") describes the situation where users experience a strong sense of presence in a virtual world, typically having donned a head-mounted, stereoscopic display and special forms of clothing, such as instrumented gloves, suits, even mechanical exoskeletons. It will be seen later that what is referred to today as "wearable computing" is but one item in an ever-increasing catalogue of data input and output technologies for VR. Returning to the definition, the four key issues are:

- *Suite of technologies*. This recognizes the fact that VR is a convergence of technologies from a much wider range of disciplines — telerobotics, multimedia, computer-aided design (CAD), process simulation, computer-generated imagery (CGI), animation and, of course, human factors and ergonomics. Nowhere has this been demonstrated to better effect than in the exemplary historical and ongoing efforts of the human factors teams at NASA's Ames Research Center in California. Their early Virtual Environment Workstation (VIEW) project, described below, paid full attention to the human factors issues associated with VR and associated interface technologies, including the integration of auditory cueing, multi-sensory data representation and speech recognition. Today, similar attention is paid to other safety-

Figure 1. NASA air traffic tower simulator. Courtesy National Aeronautics and Space Administration.

critical applications of virtual environment technology, such as NASA's Air Traffic Control Simulator (Figure 1).

- *Intuitive interaction*. The term "intuitive" reflects an ambition of VR to provide users with a *transparent* interface which makes navigation through, and interaction with virtual environments as "natural" as possible, thereby avoiding lengthy and costly training in specific computing skills. In the recent past, it was assumed that intuitive interaction could only be provided by immersing the VR user within a virtual environment. Today, there is a variety of highly usable multi-function display and control devices on the market. There are also many other products that lay claim to an ergonomics quality but leave a lot to be desired when they are used.

- *Real-time*. This important feature of VR relates to the importance of allowing users the freedom to explore and interact with virtual objects and environments without incurring disruptive time delays between their actions (e.g. head or hand movement) and the result of those actions within the virtual environment. Often one comes across a tendency for VR to be compared with many of the CGI and computer-animated films one can experience at the cinema. While these productions capture the attention of their audiences with their advanced graphics, they are nothing more than cartoons — highly detailed graphics which are drawn, or "rendered," frame-by-frame over very lengthy periods and, where necessary, superimposed on live action. In other words, they are *scripted* computerized sequences which, while visually impressive, cannot be interacted with in real time. Unlike VR, users are not free to choose where they can move next, or which object they will grasp and manipulate next.

- *Databases*. The significance of VR lies not in the nature of the sensory experience *per se* (with vision typically providing the dominant channel), but in the structure and behaviour of the data underpinning the sensory experience. The software qualities and "intelligence" endowed to objects in a virtual environment, together with their geometric and behavioural relationship with other objects, permit an

approximation to the nature of their real-life counterparts. Alternatively, in cases where such objects or processes do not exist in reality or are not visible to the naked eye (e.g. financial records, molecular structures and fluid dynamics), such features of VR make those objects and their behaviours more accessible and understandable to the human user. Some examples of these features will be illustrated below.

2. A BRIEF HISTORY

Without doubt, the USA played a significant and formative rôle in the establishment of a global VR market and community (Moody 1999). Many researchers will acknowledge the 1950s' contributions of such pioneers as the late Morton Heilig with his *Sensorama* entertainment system — the "forerunner" of the single-person virtual arcade system (http://www.telepresence.com/org/ sensorama/index.html). However, the first concrete example of a concept involving an immersive VR system is attributed to the 1960s' work of Ivan Sutherland. His experimental system, called *Sketchpad*, was a computerized design tool that allowed users to design and manipulate graphical objects on a screen using a light pen. Sketchpad ran on a TX-2, the most powerful computer available at the time. It was not only a tool for visualization, it was an operating environment in its own right. Sutherland's later work, described in 'The ultimate display' (Sutherland 1965), was based on research for the Advanced Research Projects Agency. Using a ceiling-mounted, cantilevered stereoscopic display headset boasting a 40° field of view — the *Sword of Damocles* — primitive 3D wire frame graphics with hidden line removal could be inspected from numerous direction by the user.

In the 1960s, an alternative form of VR had been the subject of much research by Myron Krueger, who had been investigating the human–computer interaction aspects of what he called "artificial realities" (Krueger 1991). In 1969 Krueger became involved in a project known as *Glowflow*, creating illusions of perspective and slope by fitting out normal, darkened rooms with tubular lighting at various orientations. Through other, more interactive projects, Krueger's research led to the development of the *Videoplace* or *Videodesk* System. By backlighting a translucent screen behind the participant (or participants, in the case of a networked system), body form could be distinguished from the background by video digitizers. A special executive computer processor translated the digitized "behavior" of the participant into silhouette form which could interact with the graphical world and its constituent objects appropriately (including some early forms of "artificial life").

However, the VR technologies being marketed today have only been under development since the early 1980s, through pioneering ergonomics research into the natural human control of space robots ("telepresence") at NASA Ames, and in the development of the *SuperCockpit* concept at the Wright Patterson Air Force Base in Dayton, Ohio. The goal of the *SuperCockpit* work was to develop advanced avionics and cockpit management systems to enhance information transfer and to protect pilots' eyes from the dazzle threat from laser weapons and nuclear airbursts. Instead, virtual representations of the environment external to the cockpit were generated from airborne sensors and reconnaissance data for presentation using large screens or sophisticated helmet-mounted displays (HMD), such as the *VCASS* System (Visually-Coupled Airborne Systems Simulator;

Harvey 1987). In stark contrast, NASA's VIEW System, mentioned above (http://telepresence.com/org/nasa/index.html), combined a low-cost HMD with an interactive glove equipped with magnetic spatial hand tracker and fiberoptic finger flexion/extension transducers (the *DataGlove*, one of the products from the late VPL, Inc.; Fisher *et al.* 1988).

In the UK, the other country accredited with major innovations in the field, VR first came to the notice of the British public late in the late 1980s. Commercial research teams who had been involved in developing the technology presented their work for the first time at the 1990 London Computer Graphics Conference. Even before then VR projects had been under way, the first notable instance being in the early 1980s. The *Spatial Workstation*, a mobile television, displaying simple 3D wire frame images to the wearer of shuttered glasses, was developed by Jonathan Waldern, who went on to establish the VR games company W Industries (later Virtuality Plc). In 1989, a VR and telepresence research program was started at the recently opened National Advanced Robotics Research Centre in the North of England. Following a period of some 4 years (during which time the now-familiar corporate names of Division and Superscape had emerged from embryonic teams of hardware and software specialists), the center launched the first collaborative industrial VR initiative called *VRS* (Virtual Reality & Simulation; Stone 1994). *VRS* brought together organizations from such market sectors as defence, medical, retail, utilities, aerospace and petrochemical. The legacy of *VRS* can still be seen today in the existence of some quite prestigious corporate VR facilities and projects throughout the country.

3. VR TECHNOLOGIES

Over the 3 years that *VRS* was striving to introduce companies to the concept of VR, one witnessed a considerable upturn in the development of new human–system interface technologies, particularly from the USA. Some of these developments were stimulated by a marketing drive that required many defence-related companies to become more open and competitive with their products. To do this, they had to target new markets and "down-grade" their display and control technologies from the restrictive and costly requirements of military standards. Predictably, however, most of the developments were stimulated by an almost obsessive need for all manner of companies across the globe to become involved in the VR market in one way or another. New peripherals, many of which were being launched even before the first prototype had been completed, were featuring in catalogues with implicit claims that each would solve the outstanding problems associated with immersing a human being into a computer-generated world, or with controlling sophisticated robots at a distance. However, these claims proved to be "flights of fancy," with potential purchasers of new display and control devices experiencing first hand the problem of placing orders for "vapourware." For a time, the VR hardware and software supply community began to lose face. Another factor contributing to the problems of a technology desperately striving to acquire credibility was the absence of any serious applications for the technology — games and entertainment systems were still seen to dominate the market. Of those products that did eventually reach the customer, many were considered to be no more than expensive and none-too-reliable toys, with no evidence of an

Figure 2. Example of an immersive virtual reality. Courtesy Virtual Presence.

Figure 3. Example of a CAVE virtual environment system. Courtesy Silicon Graphics, Inc. (UK).

ergonomics input to their design and little prospect of being used in anger in real industrial applications. Fortunately, the situation has changed for the better. Robustness, reliability and value for money are the norm in today's VR hardware and software rather than the exception.

As mentioned above, in some cases, the exploration of, and interaction with virtual environments or CGI requires the use of special peripheral devices to create a sense of presence, or immersion, within the virtual environment. In one form of immersive VR (Figure 2), users don a stereoscopic head-mounted audio-visual display (HMD) and can interact with the environment in real time using such devices as an instrumented glove (e.g. Virtual Technologies' *CyberGlove*), hand controller ("wand") or even complete body suits. The HMD and other peripherals are tracked in real time, for example using a proprietary electromagnetic tracking system (Stone 1996b). Movements and control actuations of the users in real space are transformed by the computer to update their position and orientation within the virtual space. The computer then renders the 3D virtual scene for display to the user, adding other appropriately transformed sensory data as necessary, such as 3D sound, touch and even smell.

Immersive VR was first commercially demonstrated on 7 June 1989 when the American company VPL, Inc. launched their RB2 system (Reality Built for 2). RB2 featured such revered product names as the *EyePhone* HMD, based on liquid crystal displays (cannibalized pocket televisions) with a special optical assembly and the VPL *DataGlove*, as described above.

Yet, total immersion in VR, using head-mounted displays and other bodily-worn equipment, still remains VR's "Holy Grail. The actual need for immersive VR accounts for no more than 15–
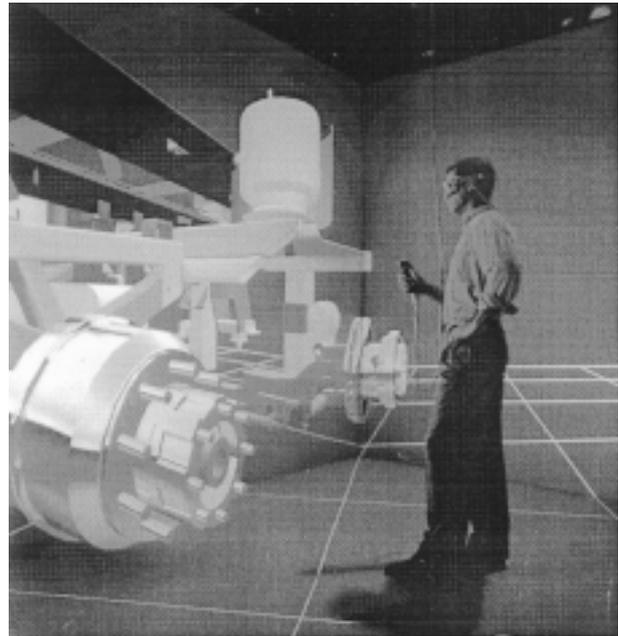
20% of the real applications evident throughout industry. Nevertheless, the quest for lighter weight and lower cost devices, such as the Sony Glasstron and Olympus personal information displays, continues unabated, fuelled with hopes of a lucrative domestic or mobile business market. Proponents of total immersion have many technical hurdles to overcome before their users can freely explore and interact with a computer-generated world they inherently feel part of, using the skills they were born with — walking, looking, grasping, listening, speaking, smelling. There are other equally important variants of VR, most of which tend to be based around the method by which information is presented visually to the user(s).

For example, an impressive, although expensive alternative to the HMD approach is the CAVE (Cave Automatic Virtual Environment — the trade name applied by the University of Illinois; Figure 3).

A CAVE is, in effect, a small room within which a small number of users are surrounded by whole-wall displays onto which the virtual images are back-projected using high quality video projectors. CAVE users may be provided with special liquid crystal "shutter" glasses, synchronized with the projectors, so that each alternate scan line of the display triggers one of the shutters, presenting left-eye or right-eye images only, thus creating a 3D effect ("active stereo"). A variation on the CAVE concept is the immersive projection sphere, one example of which has been developed by VR Systems in Southampton. This system consists of an acrylic sphere, 3.5m in diameter, into which the user stands (Figure 4). The sphere is supported by bearings mounted within a large horizontal ring. Within the area of the ring is a smaller sphere surrounded by rotational sensors. By walking within the main sphere, the user can explore virtual environments, displayed on its translucent surface by means of a series of high power projectors.
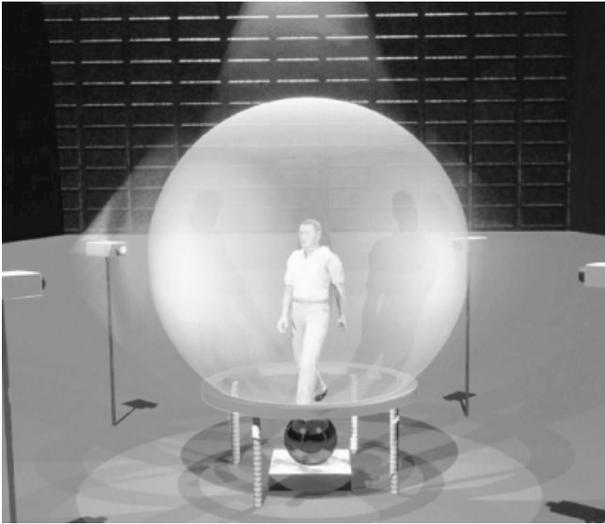
Figure 4. VR Systems' immersive projection sphere. Courtesy VR Systems Ltd.

"Partial" immersion can be achieved by means of what is popularly known as a BOOM mechanism. BOOM is actually a trade name of a device developed by Fake Space Labs in the USA, and stands for Binocular Omni-Orientable Monitor. BOOM users do not have to "wear" the display housing, as with HMD. Rather, they simply pull a counterbalanced stereoscopic display module towards their faces on demand and interact with the virtual environment by means of small buttons located at the side of the optical enclosure (or other means).

Desktop VR is a term applied to standard computer terminals, such as PCs or more sophisticated Microsoft Windows-based workstations. The virtual environment is presented to the user via a conventional monitor (sometimes modified to display stereoscopic images using optical assemblies or LCD shutter glasses, as described above). Movement through, and interaction with the environment can be achieved using a range of devices, from the conventional mouse and multi-axis isometric joysticks, such as the Spacetec IMC Spaceball or LogiCAD3D's Spacemouse/Magellan, to novel finger-mounted systems such as the Spectrum Ring Mouse.

Projection VR (Figure 5) is a display technique whereby individuals or groups of viewers sit in front of a single wall-mounted screen, multiple "wrap-around" screens or look down onto a special desk-or bench-mounted display. The virtual imagery is projected onto these screens using conventional three-tube video projectors. Again, LCD shutter glasses can be used to achieve a stereoscopic effect, although pairs of projectors with polarizing filters may also be used, so that viewers (wearing low-cost spectacles with complementary polarizing filters) can experience a "passive stereo" effect.

On a smaller scale, 3D virtual images can be created for single users, or small groups of users, by projecting the output of two liquid crystal or cathode ray tube displays onto screens of various sizes. One method of achieving such projection involves the use of (for example) holographic optical elements, with a resulting stereoscopic effect that avoids the need for users to wear cumbersome headsets or special glasses (autostereopsis). Another, more advanced form of projection — the Volumetric Display — offers 3D images that literally "float" in space (apparently occupying a real volume) and is viewable from many different angles by many observers.

A version of VR which has found popularity in the medical and surgical arena is referred to as Augmented Reality (AR). AR makes use of special headgear, or modified optical instruments (such as microscopes), and superimposes virtual images onto the user's view of the real world. AR requires careful attention to the accuracy of head tracking, to avoid disruptive visual problems caused by poor registration of the virtual and real images.

Other, more speculative systems, such as direct neural interfaces for widespread human use in VR, despite their "popularization" on TV programs, remain a science fiction dream, although this could easily change within the next 10 years. Less intrusive technologies exploiting human physiological and sensori-motor systems (e.g. eye tracking, laser retinal scanning, electromyography) may well develop over the next few years to the point when the need to don cumbersome and expensive VR headsets gloves or suits will become unnecessary. Sound generation for virtual environments is quite a mature subject and,

Figure 5. Example of a large-scale projection display. Courtesy CADCentre.

Figure 6. *PHANToM* haptic interface. Courtesy Sensable Technologies.

while the cost–benefit of adopting full 3D sound has yet to be proved (as opposed to straightforward mono or stereo), there is no doubt that the incorporation of sound in even the simplest virtual world adds immensely to the user's experience. Other forms of feedback to the human VR user — haptic (touch/force) and olfactory (smell) — are still the subject of international R&D programs (one impressive olfactory project being undertaken in the US, funded by the Advanced Research Projects Agency, ARPA). Some of these developments have yielded impressive results. Some are even mainstream products, such as Sensable Technologies' *PHANToM* haptic feedback systems (Figure 6).

A final and very important point to bear in mind is that VR is no longer the sole province of the so-called and expensive graphics "supercomputer." There is a relentless quest for faster, higher quality and more interactive virtual worlds to be deliverable on industry-standard Windows NT/95 workstations (e.g. Intergraph, Silicon Graphics), even multimedia PCs using high-performance graphics accelerators (witness the pervasion of low-cost Open GL/Direct 3D boards). Software packages for building and distributing 3D and digital panoramic virtual worlds on CDs (e.g. Superscape's VRT Visualiser/Viscape, Apple's QuickTime VR, Black Diamond's Surround Video, Live Picture's Reality Studio, Infinite Picture's SmoothMove, etc.) are readily available. In fact, access to quite sophisticated demonstrations, model libraries and low-cost, even free modeling and run-time software over the Internet for non-commercial use (e.g. AC3D modeling package: http://www.comp.lancs.ac.uk/computing/users/andy/ac3d.html; Swedish Institute of Computer Science DIVE — Distributed Interactive Virtual Environment: http://www.sics.se/dive; Manchester University, UK; Maverik GNU License: http://hegel.cs.man.ac.uk/systems/Maverik) are also contributing to a more widespread uptake of VR than was the case 2 years ago.

4. APPLICATIONS

Since 1997 it has been recognized that the key to the future of VR lies in the uptake of the technology by industrial and commercial organizations. Space does not permit a complete review of where the successful applications can be found. However, those interested in pursuing the subject can find a wealth of information on a Web site dedicated to improving VR awareness in the UK (UK VR Forum: http://www.ukvrforum.org.uk). Also available from this site is a 1998 report prepared for the UK VR Forum (Cydata 1998) which looks at the future of the VR market and presents trends for uptake within certain key sectors. Right at the outset, the report concludes: "the research results provide clear evidence that VR technology and implementation expertise have advanced sufficiently for a wide range of commercial applications to be possible, some of which seem now to be showing the prospect of substantial business benefits." There are other encouraging messages. For example, one well-known international telecommunications company has been exploiting VR for some time now. The results, insofar as their business is concerned, are well worth mentioning:

- 1168 days have been saved per year over three projects in just 6 months.
- VR models were derived from original engineering CAD data.
- Models were not just geometric in nature, but included version control, associated text and picking lists.

- VR is today hosted on an organization-wide Intranet.
- VR has stimulated a 97% reduction in requests for hard-copy material.
- Technical authors derived pictorial information from the VR model, reducing the need for camera crews and video by four sessions per year.
- Electronic access to the VR models generated savings on overseas visits by 25%.

Other applications areas are summarized below.

4.1. Large-scale Engineering

The ability to convert otherwise computer memory-intensive CAD data into visually acceptable, real-time interactive models is a feature of engineering VR that is now well established. Endowing those models with behavioral capabilities culled from associated engineering databases is also becoming an accepted and proven process (as will be seen in the sections dealing with scientific simulation and defence). The applications covered by the term "large-scale engineering" are many and varied, from petrochemical plant design or construction sites in general, to the ergonomic and aesthetic evaluation of automobile and civil aircraft interiors; from the verification of maintenance strategies for aircraft engines (Figure 7) to the prototyping of nuclear control rooms (UK VR Forum: http://www.ukvrforum.org.uk). Ergonomics is playing an ever-increasing rôle in engineering VR and simulation. For example, there has been a marked rise of recent years in the number of commercial human mannequin packages, themselves endowed with anthropometric and biomechanic data, not to mention physiological libraries encompassing energy expenditure, clothing effects, psychomotor parameters and so on. Not only do these packages provide the VR and human factors practitioner with a cost-effective assessment technique for engineering assessments, they also enable new industrial users of VR to do away with antiquated (and highly inaccurate) forms of "ergonomic" assessment based on toy figures, such as "GI Joe" or "Action Man!"

4.2. Physical/Scientific Simulation and Micro-engineering

This category is one of great promise for VR in the mid-to-far term future. VR has always held the potential to unlock the
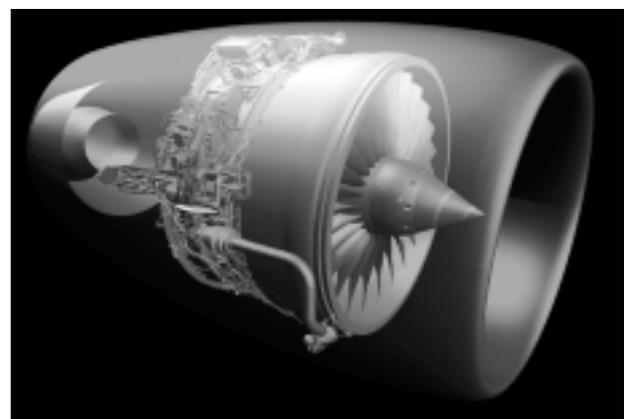

Figure 7. Virtual Rolls-Royce Trent 800 aero engine. Courtesy Virtual Presence.

Figure 8. Virtual computational fluid dynamics process, as displayed using VR techniques. Courtesy Virtual Presence and UK Health & Safety Laboratory.

"snapshot" and rather static world of the optical, electronic or digital microscope, to bring to the human eye minute features of living and inanimate material which would otherwise be invisible — and certainly inaccessible. Similarly, VR can help scientists to visualize all manner of complex artifacts and processes, be they new protein chains, viral compositions, chemical drug concepts or gaseous particles in motion (i.e. computational fluid dynamics, or CFD).

As an example of VR used for CFD visualization, NASA Ames Research Center's Numerical Aerospace Simulation Facility has developed a highly impressive facility based on VR technologies to allow users to explore numerically generated, three-dimensional unsteady flow fields. The project allows users, equipped with a BOOM stereoscopic display and *DataGlove* (described above), to view and inject smoke-like virtual tracers in and around aerospace vehicles. One of NASA's sister sites (Langley) has, for international research and education purposes, now ported the concept of the Virtual Wind Tunnel on to the World Wide Web. Another example of using low-cost (PC) computing technologies to deliver fully interactive VR representations of CFD processes was supported by the UK's Health & Safety Executive. This project assessed the design of chemistry laboratory fume cupboard handles and their effect on the propagation of potentially harmful particles back into the laboratory and towards the working scientists (Figure 8).

As with large-scale engineering today, ongoing developments in micro-, even nano-engineering will establish VR as a *de facto* standard in the provision of advanced interfaces for scientists and engineers, delivering a new form of telepresence for the manufacture, inspection, assembly and operation of miniature systems for such pioneering applications as microsurgery and biotechnology.

4.3. Defence

Defence establishments and military forces across the globe — already home to many of the world's human factors specialists — have long been exploiters of VR technology (or "synthetic environments," to use military phraseology), primarily in large-scale simulators designed for such activities as operations planning, war gaming, command–control–communications and intelligence (C^3I) and, of course, tri-service pilot, navigator and driver training. However, this exploitation has, of recent years, extended to part-task or "off-mission" activities, such as those military trainers which endow basic CAD or VR models of military platform subsystems with realistic behaviors, thereby enhancing the training of such procedures as maintenance, fault-finding and refit. VR has been developed to create realistic military environments for such tasks as helicopter machine gun training, parachuting experience, explosive ordnance disposal, naval helicopter deck landing, submarine and surface ship blind piloting, officer of the watch training, etc. Also, as weapons platforms become more advanced, the inevitable reduction in real systems available for training means that computer-based lessons, many featuring VR, will become an essential tool of the military classroom. Nowhere is this problem more acute than is the case with future submarines. Here, smaller fleets will spend much more time at sea and less time in dock at the disposal of the inexperienced naval rating. The need for VR classroom trainers to familiarize the incoming ratings with submarine systems and layouts is incontestable. Furthermore, the success of networked simulators, courtesy of pioneering initiatives such as distributed interactive simulation (DIS; Smart 1998), brings a major new dimension to military VR trainers. For example, networked VR allows geographically remote infantry personnel to train together in a shared virtual theatre of campaign, acting as a single, coordinated battalion, complete with fighting vehicle back-up and other hardware support. It also provides the classroom educator with an important practical tool for assigning tasks and responsibilities to individual or collaborating trainees, not to mention providing a more quantifiable method of assessing the extent of trainees' learning.

Figure 9. Minimally invasive surgical trainer (MIST) in use. Courtesy Virtual Presence.

Figure 10. VR for retail space planning and consumer behavior assessment. Courtesy Virtual Presence and Nestlé.

4.4. Medical and Surgical Applications

The recent major advances in computing technology have delivered the means to construct sophisticated and comprehensive anatomical and physiological simulations of the human body. From the digital reconstruction of microtomed bodies of executed convicts (e.g. the Visible Human Project: http://www.nlm.nih.gov/pubs/factsheets/ visible_human.html) to speculative deformable models of various organs and vascular systems, the quest to deliver comprehensive "virtual humans" using dynamic visual, tactile, auditory and even olfactory data goes on and on. Yet, with one or two exceptions, the uptake of these simulations by surgical research and teaching organizations has been poor. One cannot attribute this failure to a lack of technological appreciation or foresight on the part of individual specialists or administrators within the target user organizations. The poor uptake actually stems from an equally poor understanding — sometimes on the part of simulation developers — of the medical needs and ergonomic requirements of the surgical users and trainees. Furthermore, it is all to often easy to forget that most medical organizations simply cannot justify the excessive initial costs of so-called graphics "supercomputers" — not to mention crippling annual maintenance charges, depreciation and, in today's rapidly changing IT world, rapid technological redundancy.

Figure 11. Sunrise over a virtual Stonehenge. Courtesy Virtual Presence and English Heritage.

The prohibitive costs and technological difficulties of implementing a surgical trainer based on comprehensive virtual humans using dynamic visual, tactile, auditory and even olfactory data have prompted a number of VR proponents to carry out a radical rethink of their methodological approaches. One example of such a rethink, MIST (Minimally Invasive Surgical Trainer), evolved from a comprehensive in-theatre task analysis. MIST is a British PC-based "keyhole" surgical trainer that uses commercial VR and database software to foster and document trainees' acquisition of minimally invasive surgery skills (Gallagher *et al.* 1998, Taffinder *et al.* 1998). In stark contrast to those expensive simulators that attempt to generate high-fidelity anatomical images, MIST (Figure 9) presents the trainees with simple geometric tasks that extrapolate to those surgical procedures evident in theatre (e.g. clamping, diathermy, tissue sectioning, etc.).

At the time of writing, numerous medical and academic institutions are revisiting the controversial topic of surgical training and assessment. It will not be long before the VR and human factors communities witness an influx of training and evaluation systems, each claiming to be able to deliver objective measures of surgeons' dexterity performance and skill levels. Each claim will need to be assessed on its own merits and care will need to be taken in defining exactly *what* is being measured. The issue of transfer of training (from the virtual to the real), as discussed below, is one of the major research opportunities for human factors and VR specialists alike, especially in the demanding arena of medicine and surgery. Human factors specialists interested in researching these opportunities are recommended to read the excellent collections of papers in Westwood *et al.* (1998, 1999).

4.5. Retail

In the retail arena, it is now accepted that VR is totally capable of eliminating the need to build full- or part-scale models for ergonomic, "visual impact" and usability assessment, be they models of individual products, concept shelf and storage designs, or even complete stores (Figure 10). By avoiding the wasteful use of materials in the iterative construction, modification and reconstruction of product and packaging examples, VR has, unintentionally, become an environmentally friendly technology — one that some market research organizations and many leading retail and "brand name" companies are keen to exploit. For example, studies have already shown the cost-effective nature of VR in assessing consumer behavior in immersive supermarket and domestic settings. Unlike conventional assessment practice, whereby supermarket areas have to be cordoned off (a very costly practice), or cumbersome eye tracking equipment has to be deployed in-store, VR trials can be set up almost anywhere.

4.6. Heritage

There has been a strong international growth of interest in the prospect of using VR to recreate historic sites and events for such purposes as education, special project commissions and showcase features at national and World Heritage visitor centers (Figure 12; Virtual Heritage: http://www.geocities.com/Athens/Acropolis/5014/vh-6.html). To use just some of the words put forward in the early 1990s by two Nynex researchers — Rory Stuart and John Thomas — VR can give the general public access to places

and things not normally accessible, to explore objects and experience events that could not normally be explored without "alterations of scale or time," to interact with remote communities and to interact with virtual (historical) actors. In the context of heritage, VR goes much further, however, in that it offers a means of protecting the fragile state of some sites and can help educate visitors not so much about their history, but in how to explore, interpret and *respect* them. For example, VR can display the potential damage caused by the ravages of human intervention and pollution by accelerating the simulated destructive effects to monuments, even large areas of land and sea, over short periods.

5. THE FUTURE FOR VR AND ERGONOMICS

It should be appreciated that VR is not just a suite of tools capable of delivering radically new design and evaluation techniques to the engineering, design or human factors practitioners involved in projects such as those presented above. It is also a technology with its own genre of ergonomics *needs*. At the outset the need for sensible implementation of VR facilities based on sound ergonomics principles was stressed. However, therein lies a problem.

Apart from one specific public domain summary publication by Wickens *et al.* (1989), with some reference to 3D in the likes of Boff and Lincoln (1988) and Farrell and Booth (1984), there have been no serious attempts to produce practical guidelines — suitable for application in real industrial settings — for the design and assessment of VR or interactive 3D visualization systems. Most of the available books on VR are of little practical use and date very quickly. A substantial chapter (Chapter 52) exists in the well-known Salvendy *Handbook* (1997), although again one has to question the practical use of its contents. International efforts to develop a Handbook of Virtual Environments are, at the time of writing, being coordinated by the University of Central Florida, which will include quite an extensive coverage of usability issues.

As VR matures, so its accelerated uptake by industry and commerce will depend on the provision of hard (financially supported) case study material and the availability of practical guidelines and techniques for assessing user performance and application relevance. As far as the former issue (case study material) is concerned, there are now examples which can be used to help overcome the initial barriers to adopting VR by companies large and small. As far as the latter issue is concerned, in my view, there are three key areas (research opportunities, even) that are currently lacking in the pursuit of human factors excellence for the industrial and commercial application of VR.

5.1. Measures/Metrics of Situational Awareness

Situational Awareness (sometimes referred to as *spatial* awareness) is a general term that refers to several aspects of human performance in a real or synthetic environment. The key components contributing to a VR participant's awareness of a situation include:

- Judgement/perception of one's own position relative to the world and objects therein ("egocentricity").
- Overall comprehension of the current situation.
- Projection of future status.

As part of a methodological approach to assess operators'

performance and the transfer of training from the virtual to the real environment, situational awareness (SA) measures can be developed, based around reasonably established US and UK developments, such as the Situational Awareness Global Assessment Technique, SAGAT (Endsley 1995a, b), and the Situational Awareness Rating Technique, SART (Taylor 1990).

The SA measure development and test procedure should involve a classification of the tasks expected of the operators (e.g. based around Wehrend's task decomposition approach, as contained in Keller and Keller 1993) leading to a specification and application of:

- the most appropriate pretraining indices of trainees' propensities to adapt to VR, including cognitive style (e.g. Group Embedded Figures), mental rotation qualities (e.g. Shepard and Metzler 1971) and experiential factors. Note that Parker and Harm (1992) proposed that an individual's mental rotation capabilities are important for the reduction of motion sickness and for the fostering of competent performance (navigation and interaction) in and with virtual environments;
- within-training indices (cognitive mapping, spatial awareness, object-relation/path recall); and
- post-training validation of SA and adaptation techniques, correlated with standard acceptance test measures and mental workload (e.g. using the NASA Task Load Index, TLX).

5.2. Transfer of Training

"Attention comes first, learning after attention is focused. And learning is primarily action …." (Dewey *et al.*, in Bricken 1991)

"The most important principle of classroom activity design is that the students' actions determine what will be learned…." (Walker, in Bricken 1991)

The most important three words in these quotes are attention, action and learning. By its very nature, VR is an attention-grabbing medium. A VR stand at an exhibition, be it focused on engineering, heritage or education, attracts the interest of individuals of all ages and from all walks of life. The real-time nature of VR and its capability to represent aspects of real and conceptual worlds are appealing to many people, especially those who would not normally have the opportunity to experience such worlds. Again by its very nature, *VR demands interaction*. It is not a passive medium, as one might find with conventional learning and presentational media — "chalk-and-talk," slides, overheads, videos, etc. The user is, within certain constraints, free to explore and interact with objects in the virtual environment. This freedom is believed by many to help enhance trainees' *intrinsic motivation*, thereby improving their learning (retention) and subsequent real-world performance (see also the comments under situational awareness).

A classic example of the efficacy of VR in training future factory workers on manufacturing lines which use human operators in the supervision of industrial robots is a study carried out by Motorola, in conjunction with Adams Consulting Group in the USA. The production line in question, which was recreated in VR for both desktop and immersive presentation to test subjects, was Motorola's Pager Robotic Assembly Line. When

comparing the performance of those trainees who had exposure to the VR set-up with those who trained using real equipment in the laboratory, the results of the trials (while not subjected to any significance testing, due to the small sample sizes; seven in each group) suggest that VR trainees learn faster and make fewer errors.

Projects addressing the advantages of virtual environments in the human acquisition of spatial information are also a topic of study within the VR laboratories of the Massachusetts Institute of Technology. MIT's *SPLOC* (Spatial Location) program was designed to address the transfer of spatial knowledge from virtual to real spaces and to provide guidelines on stimulus simplification routines (to take account of real-time limitations of a variety of computer platforms) without incurring performance penalties in spatial navigation performance. In addition, *SPLOC* aims to advise on the augmentation of simplistic virtual worlds using other technologies (e.g. multimedia, basic audio-visual image inserts). One of the key applications generating information of use to the *SPLOC* program is a submarine Officer of the Deck (OOD) simulator which enables trainees to guide a submarine into dock using a range of display formats, from actual bridge views to plan position displays and virtual "binoculars."

It may be of interest to note here that specific experimental design guidelines for training effectiveness and training transfer have been attempted by Meister (1985, 1986, AIAA 1993).

5.3. Health and Safety Issues

Since the beginning of the spread in the uptake of VR technologies in the early 1990s, there have been anecdotal reports and limited academic research results which suggest the existence of serious side effects associated with prolonged immersion in a virtual environment. This, coupled with fierce debate on longer-term consequences, has captured the attention of potential VR users, not to mention the press. In a number of cases VR or "cyber-" sickness, as it has been called, has been listed as *the* major barrier to adoption for many commercial organizations. While the situation is still less than satisfactory — very few unambiguous or constructive guidelines have resulted from R&D programs — the quality of investigation has improved steadily since 1995.

One key issue is that there are significant individual differences in the nature of the symptom(s) reported, the severity of the report and the time of onset. The important work by Kolasinski (1995) summarizes many of these differences. For example, it is claimed that any symptom that is likely to occur will do so within 10 min from the start of immersion, with the severity increasing thereafter. It has sometimes been stated that individuals (e.g. pilots) with real-world experience of the tasks used in the design of simulators (for training, design, prototyping, etc.) are also those most susceptible to sickness when performing in the virtual "reconstruction." Some VR users who experience repeated exposure to virtual environments (an initial 15–30 min with a second exposure occurring within 2–3 days) adapt well to those environments, with an accompanying reduction in the incidence of sickness reports. Postural instability ("ataxia"), measured before and after an immersive VR experience is considered by some researchers to be one of the more important indicators of disorientation and possible malaise.

Human factors researchers intending to conduct experiments into VR-induced malaise should be aware that any observed or recorded effect can be caused by a combination of factors as well as those attributed to individual differences. These include how well the VR system has been set up, the content of the target virtual world, the degree of interaction expected of the subjects and the nature of the task being undertaken. Another frustrating problem is that *specific* conclusions drawn using a VR system procured today may well be of little value to industrial users. The results may well be invalidated with 2–3 months, typically as a result of the launch of new head-mounted displays (with higher resolution and wider-field optics), faster graphics cards supporting more photorealistic environments, higher-quality interactive control devices, etc.

The VR and human factors communities are still attempting to implement and validate both subjective *and* objective methods of recording VR users' performance, susceptibility to sickness and general well-being. Among the *objective* methods currently being researched are eye/head motion tracking and playback, brain activity (electroencephalography, EEG), muscle activity (electromyography, EMG) and gastric activity (using electrogastrograms, EGG). While many of these techniques are beginning to yield interesting results, they are not yet available in a form that supports their effective application in industrial settings. Among the more commonly referenced subjective measures available are pen-and-paper questionnaires, such as the Simulator Sickness Questionnaire (SSQ; Kennedy *et al.* 1993), the Motion Sickness Questionnaire (MSQ; Reason and Brand 1975), General/Specific "Self Efficacy" (G/SSE) Questionnaires, NASA's Task Load Index (TLX) measure of mental workload, and so on. Although a good number of these measures can be applied pre-, during and post-immersion in their "raw" form, it has to be recommended that they are reviewed before use, with the aim of tailoring their contents to the target VR application.

Until more in-depth and longitudinal studies of within or post-immersion effects are announced, one can only recommend caution (and, to a certain extent, common sense) when setting up an in-house VR facility, be it for academic or industrial use. Practical guidelines are available, such as those published by the author (Stone 1998) and by the naval Training Systems Center in Orlando (NAVTRASYSCEN 1988; Stanney *et al.* 1997). These go some way to helping VR-equipped organizations satisfy their in-house health and safety inspectors that the subject is being taken seriously.

There are those in the industrial and VR market sectors who may well be critical of research projects attempting to quantify the extent of VR-induced malaise. Nevertheless, academic researchers should take heart by the fact that, even though their objective and subjective measures of performance may not correlate well with symptoms or expressions of sickness, they are nonetheless contributing to an ever-growing library of good, basic applied psychology data and documented human factors experiences.

6. CONCLUSIONS

The successful adoption of VR technologies into organizations, be it for commercial gain, for streamlining operational procedures, or even basic education is not just a case of trying to impress potential users with the capabilities of an exciting technology. Understanding the needs and characteristics of the individual user and his or her organization is essential to the future development of VR as a stable form of information technology. It

is all-too-easy (even for qualified human factors specialists) to fall into the trap of striving for visual excellence at the expense of usability and content, not to mention losing sight of the needs of the user organization. For example:

many developers (especially in the human factors field) believe that one effect brought about by the existence of advanced [VR] hardware and software technologies has been a reduction in the application of scientific rigour to the design of human–system interfaces. Suddenly, reasonably user-friendly software tools have become readily available which have, in some cases, permitted the designers of information displays to "go to town" in their design approach. The result? "3D works of art" — visually impressive interface formats — but of questionable usability. The drive for visual impact appears to have over-shadowed the crucial issue of concentrating on the underpinning human factors issues surrounding the need for sophisticated 3D display formats …. (Stone 1997)

It should always be borne in mind that VR is, first and foremost, a suite of technologies which provides the ergonomics and human factors community with a "toolkit" for optimizing the design of the human–system interface for numerous applications. Ergonomics, sometimes (ignorantly) underrated as a technological field of endeavor, has a significant contribution to make to the development of VR into the next Millennium. Not just as a means of alerting VR users to negative and sometimes scare-mongering issues, such as the potential side effects of immersion, but in the development of methodologies to measure and report the positive effects of applying this exciting human-centered technology throughout industry. Human factors, or ergonomics, is often defined as the study of the relationship between the human and his or her working environment. It should make no difference whatsoever if that working environment is real or virtual.

REFERENCES

AIAA, 1993, *Guide to Human Performance Measurements*. American National Standard (ANSI approved). ANSI/AIAA G-035-1992.

BOFF, K.R. and LINCOLN, J.E., 1988, *Engineering Data Compendium: Human Perception and Performance*. Integrated Perceptual Information for Designers Program; 3 vols plus User's Guide (H.G. Armstrong Aerospace Medical Research Laboratory, Wright-Patterson AFB).

BRICKEN, W., 1991, Training in VR. In T. Feldman (ed.), *Virtual Reality '91: Impacts and Applications. Proceedings of the First Annual Conference on Virtual Reality*, London, June 1991 (Meckler).

BURDEA, G.C. and COIFFET, P., 1994, *Virtual Reality Technology* (Chichester: John Wiley).

BURDEA, G.C., 1996, *Force and Touch Feedback for Virtual Reality* (Chichester: John Wiley).

COBB, S., NICHOLS, S. and WILSON, J.R., 1995, Health and safety implications of virtual reality: in search of an experimental methodology. In *Proceedings of the FIVE Conference*, December.

COBB, S., NICHOLS, S., RAMSEY, A. *et al.*, 1996, *Health and Safety Implications of Virtual Reality*. VIRART Report No. VIRART/96/131. Summary Report of Program of Work (UK Health & Safety Executive Grant No. 3181/R53.133).

CYDATA, 1998, *UK Business Potential for Virtual Reality*. Executive Summary of Market Research Study conducted on behalf of the UK Government's Department of Trade & Industry.

EARNSHAW, R., GIGANTE, M. and JONES, H. (eds), 1993, *Virtual Reality Systems* (Academic Press).

EARNSHAW, R., VINCE, J. and JONES, H. (eds), 1993, *Virtual Reality Applications* (Academic Press).

ENDSLEY, M.R., 1995a, Measurement of situation awareness in dynamic systems. *Human Factors*, 37, 65–84.

ENDSLEY, M.R., 1995b, Toward a theory of situation awareness. *Human Factors*, 37, 32–64.

FARRELL, R.J. and BOOTH, J.M., 1984, *Handbook for Imagery Interpretation Equipment*. Report No. D 180-19063-1 (Seattle: Boeing Aerospace Co.).

FISHER, S., WENZEL, E.M., COLER, C. and McGREEVY, M., 1988, Virtual interface environment workstations. In *Proceedings of the Human Factors Society 32nd Annual Meeting*.

GALLAGHER, A.G., McCLURE, N., McGUIGAN, J., CROTHERS, I. and BROWNING, J., 1998, Virtual reality training in laparoscopic surgery: a preliminary assessment of Minimally Invasive Surgical Trainer, Virtual Reality (MISTVR). Paper presented at the Medicine Meets Virtual Reality 6, San Diego, CA; 28–31 January.

HARVEY, D., 1987, VCASS: A second look at the Super Cockpit. *Rotor and Wing International*, 21, 32–33, 63.

KALAWSKY, R., 1993, *The Science of Virtual Reality and Virtual Environments* (Addison-Wesley).

KELLER, P. and KELLER, M., 1993, *Visual Cues — Practical Data Visualization*, 2nd edn (London: IEEE Computer Society Press).

KENNEDY, R.S., LANE, N.E. *et al.*, 1993, Simulator sickness questionnaire: an enhanced method for quantifying simulator sickness. *International Journal of Aviation Psychology*, 3, 203–20.

KOLASINSKI, E.M., 1995, *Simulator Sickness in Virtual Environments*. US Army Research Institute Technical Report No. 1027 (Project No. 2O262785A791).

KRUEGER, M. W., 1991, *Artificial Reality II* (Addison-Wesley).

MEISTER, D., 1985, *Behavior Analysis and Measurement Methods* (New York: Wiley).

MEISTER, D., 1986, *Human Factors Testing and Evaluation* (New York: Elsevier).

MOODY, F., 1999, *The Visionary Position* (New York: Times Business).

NAVTRASYSCEN, 1988, *Simulator Sickness*. Field Manual MOD3, supported by Program Element 62233N (Office of Naval Technology).

PARKER, D.E. and HARM, D.L., 1992, Mental rotation: a key to mitigation of motion sickness in the virtual environment? *Presence*, 1, 329–33.

REASON, J.T. and BRAND, J.J., 1975, *Motion Sickness* (London: Academic Press).

SALVENDY, G. (ed.), 1997, *Handbook of Human Factors and Ergonomics*, 2nd edn (Chichester: John Wiley).

SHEPARD, R.N. and METZLER, J., 1971, Mental rotation of three-dimensional objects. *Science*, 171, 701–03.

SMART, E., 1998, [http://www.ist.ucf.edu/labsproj/projects/dis.htm].

STANNEY, K.M., KENNEDY, R.S. and DREXLER, J.M., 1997, Cybersickness is not simulator sickness. In *Proceedings of the Human Factors and Ergonomics Society 41st Annual Meeting*, 1138–42.

STONE, R.J., 1992, Virtual reality and telepresence. *Robotica*, 10, 461–7.

STONE, R.J., 1994, A year in the life of British virtual reality: will the UK's answer to Al Gore *please* stand up? *Virtual Reality World*, 2.

STONE, R.J., 1996a, A study of the virtual reality market. Report prepared for the UK Department of Trade & Industry (Communications & Information Industries Directorate).

STONE, R.J., 1996b, Position and orientation sensing in virtual environments. *Sensor Review*, 16, 42–48.

STONE, R.J., 1997, Report on 3D representations for tactical displays. Report prepared for UK Defence Research Agency Portsdown's Centre for Human Sciences. Project No. CHS0/5002S, performed under subcontract to Gregory Harland Ltd (Windsor, UK).

STONE, R.J. 1998, Health & Safety guidelines for the operational usage of immersive virtual environments. *Virtual Presence Report*, 5.

SUTHERLAND, I.E., 1965, The ultimate display. *Proceedings of IFIP Congress 1965*, 2, 506–8.

TAFFINDER, N., McMANUS, I., JANSEN, J., RUSSELL, R. and DARZI, A., 1998, An objective assessment of surgeons' psychomotor skills: validation of the MISTVR laparoscopic simulator. *British Journal of Surgery*, 85 (suppl. 1), 75.

TAYLOR, R.M., 1990, Situational awareness: aircrew constructs for subject estimation. Report IAM-R-670 (Farnborough: Institute of Aviation Medicine).

WESTWOOD, J.D., HOFFMAN, H.M., STREDNEY, D. and WEGHORST, S.J. (eds), 1998, *Medicine Meets Virtual Reality 6*. Studies in Health Technology and Informatics, 50 (IOS Press).

WESTWOOD, J.D., HOFFMAN, ROBB, R.A. and STREDNEY, D. (eds), 1999, *Medicine Meets Virtual Reality 7*. Studies in Health Technology and Informatics, 62 (IOS Press).

WILSON, J.R., NICHOLS, S. and RAMSEY, A., 1995, Virtual reality health & safety: facts, speculation and myths. *VR News*, 4, 20–4.

WILSON, J.R., 1996, Effects of participating in virtual environments: a review of current knowledge. *Safety Science*, 23, 39–51.

VR Technology for Ergonomics Design

R.H.Y. So

Human Performance and Virtual Reality Laboratory, Department of
Industrial Engineering and Engineering Management, Hong Kong
University of Science and Technology, Kowloon, Hong Kong

1. INTRODUCTION

Here, a specific technology, namely virtual reality (VR), will be introduced with an emphasis on its applications in ergonomics design.

The word "ergonomics" has its roots in the Greek language and it can be defined as the "natural laws of work." For human beings the most natural way to select field-of-view is by turning our heads. When viewing conventional monitor displays, users cannot fully utilize this "natural" capability. When the displays are used to present two-dimensional (2D) materials such as text, graphs and charts, the loss of "head-pointing" capability does not affect the performance and the satisfaction levels of the viewers. However, when presenting three-dimensional (3D) graphics simulation, the lack of "head-pointing" can degrade the sense of realism and, hence, reduce the satisfaction levels of the viewers (e.g. in a driving simulator). Examples of applications with 3D graphics include computer-aided designs (CAD), interactive digital tours and training simulations. With VR display systems, the ability to select the field-of-view using head movements has been retained. A typical VR display system consists of a head-mounted display (HMD) and a head position tracking system. As the viewer turns his/her head, the orientation of the head is measured so that images presented on the HMD can be updated according to the head-pointing angle. In other words, the viewer can select the field-of-view by turning his/her head. This VR display enables the viewer to view and "immerse" into a visual environment generated by a computer. Not surprisingly, in the early days of VR displays, they have been referred to as "visually-coupled" displays. A user wearing a commercially available HMD is shown in Figure 1. With an HMD, stereoscopic images can also be presented. One of the earliest VR displays was developed in 1968 by I.E. Sunderlands in the USA.

Figure 1. A user wearing a head-mounted display.

Since then, the technology has advanced rapidly. Even as this chapter is being written, break-through in high-resolution displays, haptics interfaces and 3D audio interfaces are taking place. This chapter would be of limited use if the author focused only on the capability of the current VR technology. Therefore, the focus will be on the fundamental promises of VR technology and their contributions towards better tools for designers, ergonomists and engineers to design more user-friendly and safer products, equipment, workplaces and living environments.

Below will be explained the fundamentals of VR systems (Section 1.2) and ergonomics designs (Section 1.3) so that readers will have a clear picture about the needs and issues concerning ergonomics design and the capability of VR technology. A summary of typical procedures for conducting ergonomics design is then presented. This is then followed by a discussion on how VR technology can complement existing computer-aided tools and offer added values. Examples of current VR tools are sited and this chapter concludes with a suggested approach towards developing a comprehensive VR tool for ergonomics design.

2. FUNDAMENTALS OF VR SYSTEMS

2.1. Brief History of VR Technology

HMD technology was originally developed to improve interactions between US Air Force pilots and their flight control and weapon delivery systems. One of the earlier HMD prototypes, the Visually Coupled Air-borne Simulation System (VCASS), can be found in the Research Laboratory at the Wright-Patterson Air Force Base at Dayton, Ohio. This HMD used cathode ray tube (CRT) technology and is quite bulky. With the use of liquid crystal display (LCD) technology, the size and the weight of a HMD have been significantly reduced. Researchers at the NASA Ames Research Center in California were among the first to apply LCD technology to the construction of HMD. As the HMD technology becomes mature, VR displays find their way into the non-military market (e.g. the game and entertainment industry, and the engineering design industry).

2.2. Types of VR Systems

In general, VR systems can be classified into four categories: (1) immersive VR; (2) desktop VR; (3) augmented VR; and (4) quick-time VR systems. Although there are overlapping areas among these categories, each does have its own unique characteristics. Most VR flight simulators developed for the military belong to the category of immersive VR. This type of VR system requires some sort of surround displays such as a HMD or a CAVE™ (a surround projection system developed at the University of Illinois). Users can view computer-generated images appropriate to their head positions and orientations. In other words, head positions are tracked in immersive VR applications. Besides a head-tracking system, special input devices translate natural body movement into control commands. One example is the use of CyberGlove™ to measure and use the position and gesture of the hands to control the direction of navigation. Other examples include the use of a modified treadmill and hand-held joysticks with position sensors attached. Examples of immersive VR applications are found in many entertainment theme parks (e.g. Disney World). At the Institute of Simulation and Training, University of Central Florida, a multi-user immersive VR system with two HMD has been developed. Two users can "enter" the
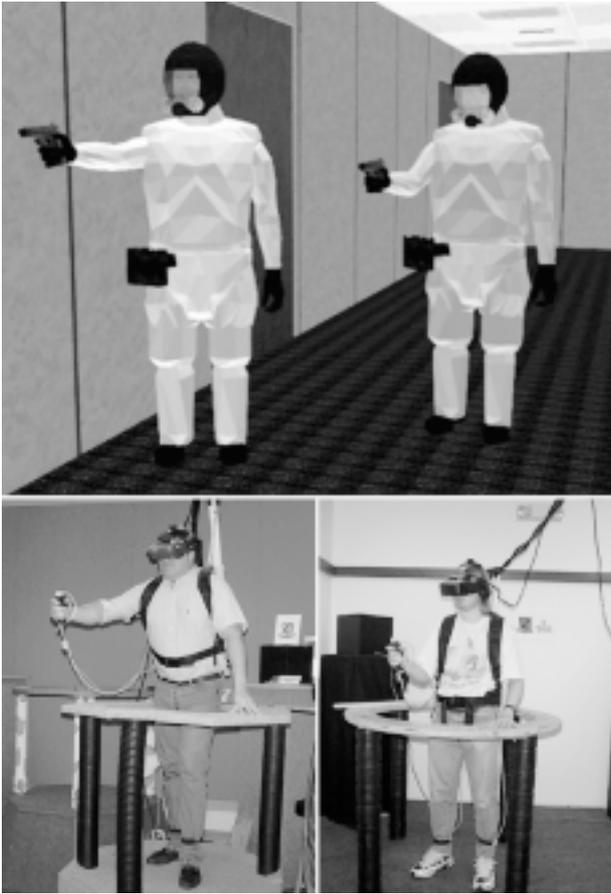
Figure 2. Two physically separated trainees practicing team mission procedures in a shared VE using the Fully Immersive Team Training (FITT) research system. FITT was developed by the US Army Research Institute for the Behavioral and Social Sciences and the University of Central Florida's Institute for Simulation and Training (courtesy of the IST, UCF).

same computer-generated environment and interact with each other as well as with animated manikins. Magnetic position sensors are put on the users' bodies and limbs so that the computer can track the body movements and recreate them and their effects within the computer-generated virtual environment (VE). Also, by stepping on the same spot, users can instruct the computer-generated virtual bodies to walk. This system is used for training purposes and the Army Research Institute has been the sponsor of this project (Figure 2). In Stuttgart, Germany, a CAVE™ display presents actual sized computer models of motor proto-types so that designers can visualize their designs and make appropriate changes.

In desktop VR systems, graphics with 3D perspective cue are presented on a monitor display. Navigation is done through conventional computer input devices (e.g. a PC mouse). Examples of desktop VR applications written in Virtual Reality Mark-up Language (VRML) can be readily found on the Internet. With desktop VR systems, no head-tracking is involved and viewers cannot select their field-of-views with head movements. Figure 3 illustrates two viewers watching a desktop VR system in stereo through two pairs of shutter-glasses™. Graphic images for the

Figure 3. Two users viewing a desktop VR system in stereo, each wearing a pair of shutter-glasses™.

left and the right eyes are presented alternatively on the monitor and the shuttering filters on the shutter-glass™ are synchronized with the presentation so that appropriate images are shown to each eye alternately.

In immersive VR systems with HMD, HMD are usually non-see-through. In other words, users cannot view the outside world when they are watching images generated by the computers. In augmented VR applications, HMD with see-through capability are used. The working principle is similar to that of a head-up display in which computer-generated images are displayed on top of the outside world. This is a very specific application of VR technology and is mostly used in manufacturing industry where computer-generated markers are spatially over-laid on top of some real objects. Head-tracking is usually required to maintain the correct spatial positions of the computer-generated markers. Augmented VR has long been used in the military for target acquisition (Barfield and Furness 1995).

Quick-time VR, originally developed by the Apple Computer Ltd, refers to the use of digitized photographs to generate a VE in which users can select their field-of-views. Currently, head-tracking is not used in most quick-time VR systems. Navigation with the quick-time VR environment is normally performed using a PC mouse. Details of this technology are on the World Wide Web using quick-time VR as a search word. A series of pictures is taken from a set of predetermined viewpoints and a special browser allows users to navigate through the pictures as if they are watching a movie. Special technologies to compensate for the camera distortion and mismatches between consecutive photographs are applied so that the "movie" will appear smoothly. The quick-time VR technology is mostly used in 3D cataloging and Web-based advertisements.

In the following sections the author explores how immersive and desktop VR systems can help in ergonomics design.

2.3. Uniqueness of VR Systems

To explore how VR technology can help in ergonomics design, it would be helpful to understand the uniqueness of VR technology. As described above, one of the most powerful features of a VR system is its ability to surround users with a computer-generated 3D visual environment (VE) in which users can interact with images in real-time. This visual environment can be further

enhanced by the presentation of 3D audio information carrying the correct spatial characteristics. An example would be a virtual construction site in which users experience the visual environment of working at height and also relying on their hearing to detect any dangers around them. Besides visual and audio interfaces, the third useful interface in VR systems is the haptics and thermal display. Appropriate force, vibration and thermal energy can be applied to the users' bodies and limbs. Although the current technology for haptics interfaces has not matured yet, future VR systems will certainly have these capabilities. In recent years, research on computer-controlled olfactory interfaces has also been emerging.

3. FUNDAMENTALS OF ERGONOMICS DESIGN

3.1. Types of Ergonomics Design

Readers should find ample information concerning ergonomics design in the encyclopedia or in *Ergonomics Design for People at Work* (Eastman Kodak 1986). The author does not intend to duplicate the effort here. Rather, this section will classify the types of ergonomics design from a practitioner's point-of-view so that readers can appreciate the challenge of ergonomics design and how VR technology can help.

The aim of ergonomics design is to design products, equipment, workspaces, environments and tasks so that they will fit the users. In general, the focuses of ergonomics design can be divided into three main categories: (1) workspace and environment design; (2) task flow and equipment design; and (3) workload and work patterns design.

3.2. Conventional and Computer-aided Methods for Ergonomics Design

In the first category, workspace and environmental design, the objectives are to keep (1) the users in comfortable body postures and (2) their sensory channels (e.g. vision and hearing) free from unnecessary and harmful disturbances. The conventional methods to achieve these objectives include the measurement of users' body dimensions (i.e. anthropometric data), application of workstation design guidelines, measurement of the environmental stressors (e.g. vibration, temperature, noise), and reduction of stressors if they exceed their relevant standards and limits. With the advances in computing technology, CAD packages for workspace design are available. Besides the general facility and product design tools like AutoCAD, 3D Studio and ArchiDesign, some software has built-in manikins that are sized according to an anthropometric database. Examples of such software include JACK, ManneQuin_Pro, SAFEWORK and MASoft. In designing a user-friendly product, the physical properties of the materials can be of importance. CAD packages that will simulate the physical properties of materials (e.g. Working Model 3D) and perform finite element analyses (e.g. ANSYS) are available. Concerning the reduction of environmental stressors, programs to compare the measured data with relevant standards and threshold limits are also readily available. At the Army Research Laboratory in the USA, a research group under Dr G.R. Price has developed a computer model of the ear to predict hazards from intense sound. With this program, animation of how a cochlea responds to a sound impulse can also be seen. At the Ruhr-Universitaet, Bochum, Germany, Professor J. Blauert has developed computer programs to simulate the acoustics properties of concert halls. Listeners can select the seating position and hear the simulated acoustic effect of an orchestra performing on the stage. In the Computer Aided Systems Human Engineering Performance Visualization System (CASHE) developed by the Crew System Ergonomics Information Analysis Center, Dayton, Ohio, a series of on-line test benches is available for users to see and hear simulated effects of display vibration, image flickering, speech degradation and hearing loss.

In the second category, task flow and equipment design, the objective is to optimize the logistics flow of information and materials so that a task can be completed efficiently and comfortably. Ergonomists have been tackling this problem by matching mental models of the operators with the actual workflow and patterns of equipment usage. In the field of work flow and work process optimization, the simple method and time study remains a popular technique. Computer simulation software such as ARENA, Pro-model, SimWork and Micro-Saints can identify the critical elements within a workflow that need attention. Currently, with the exception of Micro-Saints, neither ARENA nor Pro-model includes a database of predetermined human capability. To fill this gap, MODAPTS PLUS can be used to determine a reasonable cycle time for a given work sequence before conducting the work flow simulation.

In the third category, work load and work pattern design, the aim is to design the task requirement and schedule so that the persons involved can fulfill the task without hurting themselves. Specific objectives will depend on the exact nature of the tasks. For example, if the task is of a repetitive nature, one of the objectives must be to minimize the risk of cumulative trauma disorders (CTD) to the workers. If the task involves the handling of heavy materials, then the objective would be to minimize the risk of muscle and spine injury. Likewise, other objectives would include proper shift scheduling to maintain workers' vigilance and proper training so that the task requirement will be lower than the workers' capability. Conventional methods to conduct ergonomics assessment rely heavily on the expertise of the ergonomist and the published data concerning the physical and psychological limits of humans. Although computer programs are available for individual design procedures such as task analysis (e.g. Micro-Saint) and heart rate analysis (e.g. Polar software), the author has yet to find a comprehensive software that can assess multiple ergonomics risks involved in a given task. Currently, software is available to analyze and predict the loading on the spine as functions of weights and lifting postures (e.g. ManneQuin_Pro). Also, an extension of MODAPTS PLUS software can analyze the repetitive nature of a work sequence and offer predictions concerning potential muscle-skeletal injury.

4. PROMISES AND USES OF VR TOOLS IN ERGONOMICS DESIGN

4.1. Summary of the Current Process of Conducting Ergonomics Design

To assess and appreciate the benefits of VR technology for ergonomics design, the inadequacy of current computer-aided tools are explored. The current process of ergonomics design comprises field measurements, proto-typing, usability testing, expert opinion solicitation, guidelines and standards solicitation, computer simulation, and customer survey. An illustration of this
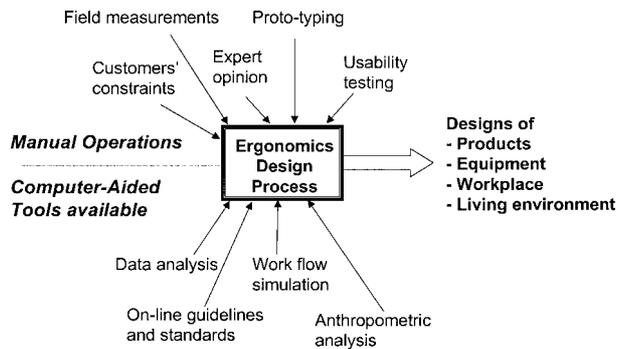
Figure 4. Information flow during a typical ergonomics design process.

process is shown in Figure 4. This generalized process applies to the design of products, equipment, workspaces and living environments. Currently, individual computer-aided tools can conduct simulation, on-line guidelines and standards search, and analyses of field data. Although some computer-aided tools have multiple functions (e.g. CASHE: both simulation and guideline search), comprehensive computer-aided tools for conducting ergonomics design are rare. The author believes that VR technology can help the integration of individual computer-aided ergonomics design tools. For a more academic examination of this subject, see Wilson (1997).

4.2. Benefits and Limitations of VR Technology

VR technology offers a natural way for designers, end-users and future maintenance workers to interact with designs of products, equipment and workplaces. However, the author would like to stress that VR alone could not and would not provide a comprehensive solution for ergonomics design. A combination of VR interfaces, existing knowledge databases, simulation algorithms and data analysis software is needed for the formulation of a comprehensive solution for ergonomists and designers. When this is achieved, computers would be able to perform design procedures that were previously labor intensive (e.g. the conventional procedure of physical proto-typing can be replaced by digital proto-typing). This is not to say that humans can be taken out of the design loop, but that humans can have more time to think creatively while the VR tools handle the more routine design procedures.

An inspection of Figure 4 shows that conventional use of computers in the design process has been biased towards mathematical computations (e.g. data analyses and simulation) and data storage and retrieval (e.g. guidelines and standards). Neither of those requires much real-time interactivity between humans and computers. This phenomenon is understandable since, traditionally, computers have always been used as number-crunching machines (e.g. mainframes and Cray supercomputers). Only in recent years has graphics simulation become important (e.g. Silicon Graphics workstation). As explained above, immersive VR technology offers a natural way for users to interact with computers through their visual, audio and tactile interfaces. As a result, design processes like proto-typing, usability testing and simulation of field data (e.g. noise) can now be implemented as part of an immersive VR simulation. When these previously non-computer-aided processes are featured in the same computer

application with conventional features like on-line ergonomics guidelines and workflow simulation, a comprehensive VR tool for ergonomics design is born. As one could easily imagine, the integration of all these features in a user friendly way is itself a huge ergonomics design challenge! However, positive steps have already been taken at many VR and ergonomics research laboratories throughout the world. In the next section, the author will browse through some examples of existing VR tools for ergonomics design.

4.3. Examples of VR Tools for Ergonomics Design

In the USA, a VR design tool for digital prototyping called "Working Model 3D" (from MCE software) has been developed on the Window 95/NT platform. This tool can simulate the physical properties of mechanical parts so that design engineers can test and visualize the mechanical properties of their prototypes (www.mcesoft.com). Within this VR simulation environment, computer-generated objects can move, slide and collide with each other. At the Department of Mechanical Engineering, University of Wisconsin-Madison, a desktop VR tool called "Virtual Design Studio, VDS" has been developed for the design of mechanical and electromechanical components (icarve.me.wisc.edu). This VDS uses shutter-glasses™ to present stereo images and can take voice commands from the designers. At the Bosch research laboratory in North America, a 3D graphical design tool called MASoft 3D has been developed for workspace design (www.boschautomation.com). This MASoft 3D is similar to the ManneQuin_Pro (www.nexgenergo.com) where scalable manikins can be inserted into a workspace design for anthropometric analyses. CadPLUS" Products Co. in the USA has a range of software products for facility layout, and mechanical system design. To benefit from desktop VR technology, a VR module called CadPLUS" VR Flyer™ has been developed. This module allows the viewers to navigate through the CAD models in real-time (www.cadplus.com). At the University of Alberta, Canada, a VR authoring tool called Minimal Reality (MR) is available at no cost to educational establishments so as to promote research and development in VR applications. A group of researchers at the Human Interface Laboratory, University of Washington, utilize augmented VR technology to assist patients of Parkinson's disease to walk normally (ww.hitl.washington.edu).

In Europe, the Engineering Link (formerly part of British Rail) at Derby, UK, has used Superscape (a PC-based VR authoring tool) to develop a desktop VR tool to design the interior carriages of a train (www.superscape.com). This tool enables engineers and designers to check the access for wheelchairs, and catering trolleys. At the VIRART center, University of Nottingham, UK, a VR design tool has been developed for the Nottinghamshire Social Services Department to assess houses and flats for their suitability for wheelchair access (www.nottingham.ac.uk/meom). Professor Robert Stone, based in Manchester, UK, has used immersive VR technology to help a company called Airline Services to visualize and market interior design of airplane cabins. As can be seen in ManneQuin_Pro, SAFEWORK and MASoft, the availability of scalable manikins enhances the usefulness of VR ergonomics tools. At the School of Psychology, Cardiff University, UK, research efforts have been focused on the development and use of virtual
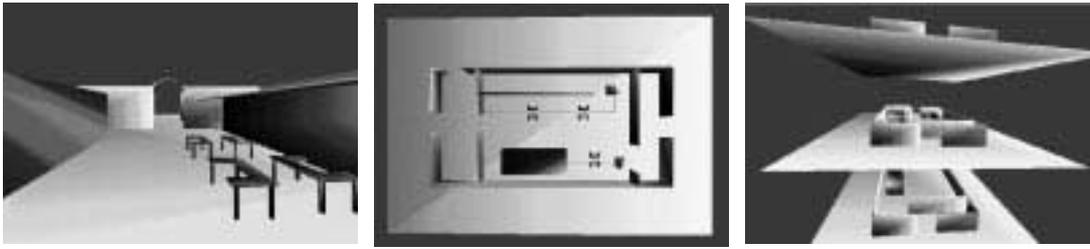
Figure 5. VR computer-aided design tool for multi-floor factory design. The tool was developed at the Hong Kong University of Science and Technology.

manikins in VR design environments (www.cf.ac.uk). A series of case studies concerning how VR technology has helped the industry in the UK to design products and facilities can be found at www.ukvrforum.org.uk. Examples include the design of a nuclear plant control room and an offshore oil platform. In view of the needs for a more comprehensive VR tool for ergonomics design, the center for VE at the University of Salford, UK, has developed an Interactive Product Simulation Environment for Assessing Assembly and Maintainability Tasks (IPSAM). Details can be found at www.salford.ac.uk/cve. At the Department of Civil Engineering, University of Strathclyde, UK, VR technology has been applied to evaluate benefits of different construction equipment (www.strath.ac.uk). As mentioned above, motor designers in Stuttgart, Germany, are using immersive VR systems to visualize and evaluate their designs. In Sweden, researchers at the Division of Ergonomics and Aerosol Technology, University of Lund utilize VR technology to enhance the cooperation and communication among designers. Their progress can be viewed at www.cw.lu.se

At the Hong Kong University of Science and Technology, a VR CAD package based on the RENDERWARE library (a PC-based VR authoring tool) has been developed to enable designers to layout multi-floor factories (Figure 5). A VR training environment for operating computer numeric control (CNC) milling machines has also been developed to optimize the training procedures of workers (www.ieem.ust.hk). This application is developed using World-Tool-Kit software (a VR authoring tool) and runs on a Silicon Graphics workstation.

5. FINAL REMARKS: TOWARDS A COMPREHENSIVE VR TOOL FOR ERGONOMICS DESIGN

As seen from the examples described above, the scope and capabilities of individual VR tools tend to be limited and specific to one or two applications only. This may have been due to the complexity involved in simulating the real world properties of individual products and environments. For example, in VR tools for mechanical design, complex algorithms are needed to simulate the effects of physical interactions. In interior design applications, special attention is needed to simulate high quality graphics and effects of lighting. In addition, acoustics properties are also important for certain designs such as concert halls and noisy workplaces. In other words, it is difficult to develop a general purpose VR tool that enables users simultaneously to examine multiple ergonomics issues. An alternative and more realistic approach is to develop a base program on which additional testing modules can be added. This base simulation program should perform desktop VR simulation of objects and environments with engineering precision and with simulated Newtonian properties such as gravitational force, reaction force and friction. Examples of add-on modules include immersive VR module for usability testing, acoustics test modules, finite element analysis modules, aesthetic modules (with fine graphics and lighting effects), anthropometry modules and workflow modules. With advances in computational speed and VR technology, we can look forward to a VR tool that can significantly reduce the time taken to make prototypes and conduct usability tests.

REFERENCES

BARFIELD, W. and FURNESS III, T.A., 1995, *Virtual Environments and Advanced Interface Design* (New York: Oxford University Press).

EASTMAN KODAK, 1986, *Ergonomics Design for People at Work* (New York: van Nostrand Reinhold).

WILSON, J., 1997, Virtual environments and ergonomics: needs and opportunities. *Ergonomics*, 40, 1057–77.

Part 5

Display and Control Design

Auditory Warnings and Displays: Issues Relating to Design and Selection

G. S. Robinson, S. E. Lee and J. G. Casali

Department of Industrial and Systems Engineering, Virginia Polytectonic Institute and State University, Blacksburg, VA 24061, USA

1. INTRODUCTION

Auditory displays, alarms and warnings are ubiquitous in our technologically advanced society. Buzzers in cars tell us that the seatbelt is not fastened or that the headlights are on, gas pumps beep to remind us to pay for our gasoline purchase, and computers prompt us whenever we make a mistake. Although auditory displays permeate nearly every aspect of our lives, few people, including those responsible for the design and selection of such displays, are aware of or consider the many factors that affect the utility of auditory displays and ultimately affect the usability of the systems in which they are implemented. This article will discuss some of the more important factors affecting the design and selection of auditory displays, alarms and warnings. The purpose is strictly introduction; topics will be introduced, and the importance of these issues within the context of system design will be briefly discussed. The discussion is not intended to represent a comprehensive review of the field or serve as a design guide. For more detailed discussions of these issues, see the references cited herein.

2. APPROPRIATENESS OF AUDITORY DISPLAYS

The first question that must be answered when contemplating the use of an auditory display is: "Is an auditory display appropriate for this application?" (as opposed to a visual display or perhaps a speech display). The answer can be complex and depends on several factors, including the message content and complexity, environmental conditions, and the expected workload of the message recipient(s). For example, in situations where lighting conditions are poor or unpredictable or where the operator's visual channel is already fully occupied, an auditory warning would probably be appropriate (Deatherage 1972). In general, an auditory display should be considered when: the message is short, simple and will not be referred to later; when the operator's visual channel is already (over)loaded; when immediate action is required; when the operator's location or the orientation of his/her head is unknown (auditory signals are omnidirectional and omnipresent) or when environmental conditions (darkness or glare) preclude the use of a visual display.

3. MASKING

Perhaps the most important issue related to the design and selection of auditory displays is masking. Even a well-designed auditory alarm, warning or display is useless if it cannot be heard. Masking is defined as an increase in the absolute threshold of one sound, the *masked* sound, caused by another, *masking*, sound (Gales 1979). In industrial situations, the masked sound might be an auditory alarm or warning, speech or a sound produced by a machine (e.g. the sound of a slipping drive belt or chipped gear teeth). Typical masking sounds include engine or motor noise, ventilation or fan noise, conveyor noise, and machine noise. It is also possible for one intentional signal to mask another intentional signal if both are active at the same time. Masking becomes a problem when an intentional or incidental sound conveying useful or important information is rendered inaudible by another sound. Several methods have been proposed for calculating masked thresholds in noise. Two of these methods are briefly described below.

3.1. Critical Band Method

The masked threshold for a tonal signal can be approximated using critical band theory (Fletcher 1940) if the spectrum of the masking noise is sufficiently flat. In many cases, this is a valid assumption and the masked threshold (L_T) can be expressed as:

$$L_T = L_S + 10 \log_{10}(BW),$$

where L_S is the spectrum level of the masking noise in the vicinity of the signal component being considered and BW is the width of the auditory filter centered around the signal component in question.

It is important to note that the spectrum level of the noise as presented in the equation above *is not* the same as the octave band or one-third octave band level measured using a spectrum analyzer. Spectrum level refers to the level per Hz or the level that would be measured if the noise were analyzed using a 1-Hz wide filter. If we assume that the noise is flat within the bandwidth of a one-third octave band filter, for example, then the spectrum level can be estimated using the following equation:

$$L_S = 10 \log \left(\frac{10^{BL/10}}{BW_{1/3}} \right)$$

where L_S is the spectrum level of the noise within the one-third octave band filter, $BW_{1/3}$ is the bandwidth of the one-third octave band filter, calculated by multiplying the center frequency (f_c) of the filter by 0.232 (Beranek 1988), and BL is the noise level measured in the one-third octave band in question.

The bandwidth of the auditory filter can be approximated simply by multiplying the frequency of the masked signal/tone by 0.15 (Patterson 1982, Sorkin 1987). Again, this procedure assumes that the signal being listened for is tonal or has strong tonal components. (If the calculations presented above are to be based on octave band analyses, $BW_{1/1}$ can be found by multiplying the center frequency of the octave band by 0.707.)

3.2. ISO 7731–1986(E)

International Standard 7731-1986(E) (1986) presents methods for calculating the masked threshold for a signal in noise based on broadband as well as octave-band and one-third octave band noise measurements. The general procedure for the full and one-third octave band measurements follow.

- Step 1: Starting at the lowest full or one-third octave band level available, the masked threshold (L_{T1}) for a signal in that band is:

$$L_{T1} = L_{N1},$$

where L_{N1} is the noise level measured in the full or one-third

octave band in question, referred to as *band level* (*BL*) in the preceding example.

- Step 2: For each successive full or one-third octave band filter *n*, the masked threshold ($L_{\mathrm{T}n}$) is the noise level in that band or the masked threshold in the preceding band, less a constant; whichever is *greater*:

$$L_{\mathrm{T}n} = \max (L_{\mathrm{N}n}; L_{\mathrm{T}n-1} - C),$$

where *C* = 7.5 dB for octave band data or 2.5 dB for one-third octave band data.

This procedure, unlike the earlier procedure, takes upward spread of masking into account by comparing the masked threshold in the band in question to the masked threshold in the preceding band. (Upward spread of masking is the phenomena wherein a tonal or band-limited noise will mask sounds at frequencies above the center frequency of the noise in addition to those sounds occurring at the center frequency of the noise.) It is also more adaptable to a wide variety of signals, both tonal and broadband, whereas the previous method assumes a tonal signal.

3.3. Ensuring Audibility

It is important to remember that these calculations represent threshold levels, not the level at which the signal can be reliably heard. It is generally agreed that a signal should be presented 6–15 dB above its masked threshold to ensure audibility (Wilkins and Martin 1982, Sorkin 1987). However, to avoid a startle response, the signal should not exceed its masked threshold by > 25 dB. Furthermore, when alarms are to be presented in high-noise environments, caution should be exercised to ensure that the signal level does not exceed 115 dB to minimize the potential for noise-induced hearing loss (OSHA 1989).

In general, the louder the background noise is relative to a signal, the more difficult it will be to hear the signal. Conversely, if the level of the background noise is reduced or the level of the signal is increased, the masked signal will be easier to hear. Whenever it is possible to do so, it is preferable to reduce the background noise level to reduce masking problems. If reducing the background noise is infeasible or impractical, it may be possible to increase the level of the signals, but close attention must be paid to the resulting sound pressure levels to avoid hearing loss for the listener. Although most off-the-shelf alarms and warning devices have a preset output level, it is possible to increase the effective level of such devices by placing multiple alarms or warning devices throughout the workplace instead of relying on a single centrally located device.

4. INATTENTION

Most laboratory-based masked-threshold experiments are conducted in a manner such that the experimental participants not only expect a specific signal, but are also listening intently for it. While such procedures are useful for determining true detection thresholds, they are not representative of the real world. Workers do not normally sit around waiting for an alarm; instead, they are doing their jobs, concentrating on the task at hand. It is intuitively obvious how someone engrossed in his/her job might fail to notice an auditory alarm that would otherwise be audible. Unfortunately, attention is an extremely difficult variable to

quantify and control in an empirical study. Typically, experimental participants are simply asked to perform some secondary task (e.g. playing a video game or driving a vehicle simulator) while simultaneously performing the listening/detection task. Few such studies have been performed, and the results are contradictory. Some researchers find no difference in detection thresholds (Wilkins and Martin 1982, 1985) while others find significant differences (Fidell 1978).

Some of the factors that can adversely affect the audibility of an auditory display (both in the laboratory and the real world) include the motivation of the listeners, the consequences of a false alarm, the benefits or rewards for a correct response, and fatigue. In laboratory experiments, it is very easy inadvertently to affect a test subject's criteria for deciding when to respond to a signal, thus biasing the results of the experiment. Such experimental artifacts are at least partially responsible for the contradictory results reported in the literature. On-site tests are even more difficult due to the researchers' nearly complete lack of control over the experimental conditions (noise and signal level, workload, etc.). Lacking any truly reliable, quantitative data, the best and most conservative course of action is to assume that workers may not notice an alarm or warning if they are engrossed in their jobs. The usual solution to this problem is to increase the sound pressure level at which the alarm is presented, staying within the guidelines presented earlier.

5. CONTRAST AND COMPLEXITY

In addition to manipulating the levels of the auditory displays, alarms, warnings, and background noise, it is also possible to increase the likelihood of detection of an auditory display or alarm by manipulating its spectrum so that it contrasts with the background noise and other common workplace sounds (Wilkins and Martin 1982). For example, in an environment characterized by high-frequency noise such as sawing or planing operations in a lumber mill, it might be best to select a warning signal with strong low-frequency components, perhaps in the 700–800 Hz range. On the other hand, for low-frequency noise such as might be encountered in the vicinity of large-capacity ventilation fans, the better choice would be an alarm with strong high-frequency components in the 2000–3000 Hz range. Unfortunately, there are no quantifiable guidelines on signal contrast to aid designers in this pursuit.

A concept intimately related to contrast is the spectral and temporal complexity of the signal itself. Simply put, complex signals are more easily detected in noisy environments than are simple tonal signals. There are several reasons why this is so. First, a complex signal with multiple fundamental frequency components and strong harmonic and inharmonic elements is more likely to have at least one component that exceeds the masked threshold in nearly any noise. Furthermore, a complex signal (i.e. a signal possessing multiple frequency components, a periodic or aperiodic temporal pattern, and with modulating frequency and/or amplitude envelopes) contains much more information than is contained in a simple pure tone. The brain can use this additional information in processing the auditory stimulus and will thus be better able to distinguish the desired signal from the background noise. Manipulation of the various acoustic characteristics of a complex auditory signal also allows

for the manipulation of the signal's perceived urgency (discussed below), its contrast with other signals and the background noise, and its attention-getting ability.

6. RECOGNITION AND IDENTIFICATION

Regardless of how well an auditory alarm or warning is designed or how well it contrasts with the background noise and other signals, humans have a limited ability to distinguish one item from a group of N similar items. A basic principle of ergonomics places this number between five and nine, depending on the stimulus (Miller 1956). Patterson (1982) found that individuals could quickly learn the first seven of a set of 10 auditory alarms and could learn the entire set with only 1 h of training. However, just 1 week later, the same individuals recognized, on average, only seven of the 10 alarms. Although recall improved after additional training, Patterson suggested that the number of auditory alarms be limited to seven or eight to avoid confusion and eliminate the need for constant retraining.

7. LOCALIZATION AND SPEED JUDGMENT

In addition to masking an auditory display, alarm or warning, elevated noise levels such as those encountered in typical industrial environments can have a detrimental impact on an individual's ability to localize sound. The brain localizes sounds in space by utilizing subtle phase and intensity differences between the sounds received in each ear (to judge azimuth) as well as the balance between the direct and reflected sound (to judge distance). Speed and movement are judged by changes in the sounds' position over time, the rate of change, and perceived pitch changes as the sound source approaches or recedes. As background noise levels increase, sound sources are judged to be closer to the observer than they actually are. This is because the noise masks the reflected component of the sound reaching the observers' ears (McMurtry and Mershon 1985). Localizing the sound's direction is less affected by the presence of noise. However, if the noise source and signal are co-located or the signal level is not sufficiently above its masked threshold (by ~10–15 dB), then problems in localizing high-frequency signals can occur (Small and Gales 1991).

Why is this important when considering auditory alarms and warnings? Consider the conditions encountered by someone working in a warehouse or factory. For safety reasons, that person must be able to hear an approaching forklift or automatic guided vehicle (AGV). While many such vehicles do have auditory warnings that function when the vehicles are in motion, they can also create considerable noise (an example of a moving alarm co-located with a noise source). The problem is exacerbated if additional background noise sources (e.g. machinery or HVAC noise) are present. Similar situations exist on loading docks. In both scenarios, workers must not only be able to hear the auditory warnings, but also must be able to pinpoint their location with a high degree of accuracy to avoid a potentially catastrophic accident.

8. PERCEIVED URGENCY

Another important factor in the design and selection of auditory warnings is their *perceived urgency*. Problems often arise when the reaction elicited by an alarm does not match the severity of the circumstances related to the alarm. Disruptive alarms can not only draw an operator's attention away from the task at hand, but can also divert attention from an emergency while the operator attends to the alarm. Conversely, disaster could result if an alarm warning of a critical situation were perceived to be routine or unimportant. To alleviate this problem, Patterson (1982) recommends that auditory alarms be differentiated by type: advisory alarms or "attensons," intended to draw the operators' attention to additional information explaining the situation, and immediate action alarms. Furthermore, he suggests that an alarm should be carefully designed or selected so that its perceived urgency matches the urgency of the situation responsible for the alarm.

The problem remains, however, as to which acoustic parameters can be manipulated and how they should be manipulated to achieve a desired level of perceived urgency. Edworthy *et al.* (1991) found that the perceived urgency of a warning signal could be varied in a predictable manner by manipulating signal characteristics such as fundamental frequency, amplitude envelope, harmonic delay, rhythm, speed, pitch range and pitch contour. Haas and Casali (1995) found that perceived urgency increased and reaction time decreased as pulse level increased and inter-pulse interval decreased.

Unfortunately, it is not yet possible to assign a quantitative value of urgency to a given situation and then specify signal parameters so that an appropriate alarm can be designed or an off-the-shelf unit selected (although with continued research, this scenario might someday become a reality). The important thing to remember is that it is not appropriate to use the loudest, most obnoxious alarm available for every situation. Each situation must be carefully considered and the alarm that will elicit the appropriate response selected; good judgment, common sense, and a thorough understanding of the situation should guide this process until quantitative guidelines are available.

9. DISPLAY SETS

In complex systems in which multiple alarms are implemented, the alarms should not be designed independently of one another. Instead, the alarms should designed *as an integrated set*. By doing so, the system designer can optimize, to the extent possible, the acoustic parameters which affect the audibility of the alarms, their perceived urgency, and the contrast of the alarms with one another and any background noise. In addition to optimizing the alarms to the environmental conditions, this approach, if implemented successfully, will have the added benefit of maximizing operator performance by reducing response times and minimizing the possibility of confusion between alarms.

10. AUDITORY ICONS – A NEW CLASS OF AUDITORY DISPLAY

Auditory icons are representational sounds in that they have specific, stereotypical meanings associated with their sound. While traditional auditory displays and alarms are typically defined by their acoustic parameters (e.g. a 1000 Hz tone with a 1 s period and a 50% duty cycle), auditory icons are defined by the objects or actions that created the sounds. (An example application is using the sound of a metal trash can being emptied to convey the action of discarding the items in the "Trash" folder

on a computer.) While auditory icons have been demonstrated to be superior to conventional signals in terms of recognition, there has been little success in implementing them in complex systems. This lack of success is often attributed to a mismatch between the operator's perception of the warning signal and the event the signal is intended to represent. Recent research (Belz *et al.* 1997) has attempted not only to formalize the auditory icon selection process, but also to determine if the use of auditory icons can actually reduce operator response time in critical situations. For example, in simulated driving tasks, auditory icons elicited significantly shorter response times and improved recognition compared to conventional auditory warnings for both front-to-rear and side collision scenarios.

Although auditory icons are promising, they are still a long way from being deployed as displays in real systems. Some of the questions still needing to be answered include: (1) how can the perceived urgency of such sounds be quantified and manipulated?; (2) how will the use of such sounds in "artificial" displays affect a listener's response in situations where the "real" sound occurs? (for example, if a tire skid is used as a collision-avoidance warning in a vehicle, will it have a long-term effect on a driver's reaction, in terms of response time and action taken, when a real tire skid is heard?); (3) will habituation occur and how will this affect a listener's response?; and (4) How will the occurrence of false alarms affect a listener's response and acceptance of the system in which the auditory icon(s) are implemented? These and other questions must be answered before auditory icons can be selected with confidence for use in complex, real-world systems.

11.   SUMMARY

The above discussion was intended to introduce the reader to the complex process of designing and selecting auditory displays. There is more to the process than simply picking a buzzer or horn from an industrial supply catalog. Briefly to recap, when contemplating or implementing auditory displays, warnings or alarms into a workplace or system, caution should be exercised to ensure that:

- an auditory display is appropriate for the application;
- the displays, alarms or warnings will be audible;
- the displays, alarms or warnings are presented at levels sufficiently above their masked threshold to command attention without provoking a startle response;
- the displays, alarms or warnings contrast sufficiently with the background noise and one another;
- the number of different auditory displays, alarms or warnings is appropriate;
- the perceived urgency of the display, alarm or warning is appropriate for the situation; and
- multiple alarms (when used) are designed as an integrated set rather than individually.

Finally, many industrial organizations and consortia have developed standards and/or design guidelines regarding implementation of auditory displays, alarms, and warnings in specific situations or for specific purposes. The American National Standards Institute (ANSI) and the International Standards Organization (ISO) publish many such standards. Whenever relevant standards or guidelines exist, they should be followed.

REFERENCES

BELZ, S.M., WINTERS, J.J., ROBINSON, G.S. and CASALI, J.G., 1998, Auditory icons: a new class of auditory warning signals for use in intelligent transportation subsystems. *SAE 1997 Transactions — Journal of Commercial Vehicles/Section 2,* 106, 431–9.63

BERANEK, L.L., 1988, *Acoustical Measurements* (Cambridge, MA: American Institute of Physics).

DEATHERAGE, B.H., 1972, Auditory and other sensory forms of information presentation. In H.P. van Cott and R.G. Kincade (eds), *Human Engineering Guide to Equipment Design* (New York: Wiley), 123–60.

EDWORTHY, J., LOXLEY, S. and DENNIS, I., 1991, Improving auditory warning design: Relationship between warning sound parameters and perceived urgency. *Human Factors*, 33, 205–31.

FIDELL, S., 1978, Effectiveness of audible warning signals for emergency vehicles. *Human Factors*, **20**, 19–26.

FLETCHER, H., 1940, Auditory patterns. *Reviews of Modern Physics*, 12, 47–65.

GALES, R.S., 1979, Hearing characteristics. In C.M. Harris (ed.), *Handbook of Noise Control* (2nd ed.) (New York: McGraw-Hill), 8-1–18.

HAAS, E.C. and CASALI, J.G., 1995, Perceived urgency of and response time to multi-tone and frequency-modulated warning signals in broadband noise. *Ergonomics*, 38, 2313–26.

ISO, 1986, *Danger Signals for Work Places – Auditory Danger Signals*. ISO 7731–1986(E) (Geneva: IOS).

McMURTRY, P.L. and MERSHON, D.H., 1985, Auditory distance judgments in noise, with and without hearing protection. In *Proceedings of the Human Factors Society 29th Annual Meeting* (Santa Monica: Human Factors Society), 811–13.

MILLER, G.A., 1956, The magical number seven, plus or minus two: some limits on our capacity for processing information. *Psychological Review*, **63**, 81–97.

OSHA, 1989, *Occupational Noise Exposure Limits*. Code of Federal Regulations, 29 Parts 1900–1910 (OHSA), 179–95.

PATTERSON, R.D., 1982, *Guidelines for Auditory Warning Systems on Civil Aircraft*. Paper 82017 (Cheltenham: Civil Aviation Authority, Airworthiness Division).

SMALL, A.M. and GALES, R.S., 1991, Hearing characteristics. In C.M. Harris (ed.), *Handbook of Acoustical Measurement and Noise Control* (3rd ed.) (New York: McGraw-Hill), 17.1–25.

SORKIN, R.D., 1987, Design of auditory and tactile displays. In G. Salvendy (ed.), *Handbook of Human Factors* (New York: Wiley), 549–76.

WILKINS, P.A. and MARTIN, A.M., 1982, The effects of hearing protection on the perception of warning sounds. In P.W. Alberti (ed.), *Personal Hearing Protection in Industry* (New York: Raven), 339–69.

WILKINS, P.A. and MARTIN, A.M., 1985, The role of acoustical characteristics in the perception of warning sounds and the effects of wearing hearing protection. *Journal of Sound and Vibration*, 100, 181–90.

Calibration and Characterization of Color Displays

C. S. Calhoun† and D. L. Post‡

†Logicon Technical Services, Inc., Dayton OH, USA

‡Air Force Research Laboratory, Wright-Patterson Air Force Base OH, USA

1. INTRODUCTION

Visual information designers and researchers must sometimes specify, predict, or control the luminance and chromaticity of self-luminous displays. (Luminance is a measure of radiant energy, weighted according to its visibility for humans. The preferred units are candelas per m^2 (cd/m^2), as specified in the Standard International (SI) notational system, but foot-Lamberts (fL), from the English system, are still used sometimes; chromaticity is used to refer to those psychophysical measures of color that are independent of luminance.) Accurate image representation can be very important because the human visual system responds differently to different levels of light intensity, color, size, duration and other factors. Visual information designers attempt to exploit visual system characteristics through appropriate manipulation of light. For an information display design to produce a desired response consistent with population stereotypes, for example, or to avoid an undesired response such as a perceptual artifact, a display image's luminance and chromaticity may have to meet certain specifications.

For a color display, calibration consists mainly of setting the display's white point to specific chromaticity coordinates, although peak luminance, convergence, purity, and focus adjustments may also be necessary. (A display's white point is the color produced when all three RGB channels are driven equally.) Characterization is a measurement process yielding data that describe the input-voltage to luminance transfer functions of the display's red, green and blue (RGB) channels. Several good methods are available for using characterization data to predict a display's colorimetric behavior; the best choice for a given application depends on the desired prediction accuracy, one's tolerance for implementation difficulty and the available measurement equipment.

1.1. Color Perception

Color perception is a complicated process, affected by physical, psychophysical, and psychological variables. The Commission Internationale de l'Eclairage (CIE) developed a system in the early 1900s for converting measurements of light into psychophysical measures, e.g. luminance and chromaticity coordinates; however, the CIE system does not account for all phenomena that affect color perception. Walraven (1992), for example, details several interesting perceptual artifacts that can result from factors such as the observer's visual adaptation state, the image's spatial and temporal characteristics, surrounding lights and environmental influences such as ambient light. Therefore, although an image's luminance and chromaticity coordinates can be determined precisely using appropriate instruments, this information does not specify the resulting color perception, and if the perception is important it must be checked in the environment in which the display will be used.

1.2. Display System Fundamentals

A typical digital display system includes a computer, video-generating device, and display. Display output results when image-generating software on the computer sends appropriate digital values to the video-generating device, which converts them to analog voltages using a digital-to-analog converter (DAC) and sends the voltages to the display. The display generates images through spatial or temporal combination of the RGB channels. Most DAC parse their output voltage range into (2^8 =) 256 levels. This process is easily seen with computer software that permits user-adjustable colors in text or graphics. Generally, the user has access to 256 (0–255) adjustments per channel. Thus, (256^3 =) 16.7 million RGB combinations are possible. The goal in display system characterization is to select and measure a few well-chosen DAC that permit accurate predictions for any RGB DAC combination.

2. DISPLAY CALIBRATION

Display calibration is needed to ensure proper color rendition and has direct consequences on the subsequent characterization. Calibration parameters include contrast, brightness, individual channel gains, and white point. The contrast or cutoff control sets the maximum achievable luminance, whereas the brightness or gain control sets the background level. It is usually best to maximize the display's peak luminance while maintaining the ability to display a reasonable black. Many new monitors now include front-panel, electronic channel-gain controls. If available, these controls are adjusted such that the desired white point is achieved; otherwise, internal controls must be used. The display's white point is usually set equal to a standard white, such as CIE standard illuminant C or D$_{65}$, and confirmed using a colorimeter or spectroradiometer (see Section 3.4).

The brightness and contrast calibration procedure is (1) Make sure the display is located in its intended environment; (2) reduce brightness to its "off" state; (3) reduce contrast to a low level while maintaining a viewable image; (4) increase brightness until the background raster can be seen, then reduce brightness until the background just disappears; and (5) increase contrast until the image can be viewed comfortably. The process can be repeated if necessary, but generally yields a satisfactory result on the first iteration.

3. CHARACTERIZATION

Characterization is needed only when predictive equations are going to be used to either predict the luminances and chromaticity coordinates that will be produced by specified DAC, or to compute the DAC needed to produce desired luminances and chromaticity coordinates. Characterization consists usually of measuring the luminance of the RGB channels separately at several (usually equally spaced) DAC. The resulting data are then used to fit the predictive equations. An alternative characterization method, developed by Berns *et al.* (1993b), is to measure the display system's white point at several DAC triplets, and measure the chromaticity coordinates of the RGB channels. The chromaticity coordinates are then used to decompose each white measurement into its constituent RGB luminances, thus yielding three RGB luminances for each white measurement and reducing the required number of measurements.

3.1. Display Gamma

A display's voltage to luminance transfer function is often characterized by a power function, the exponent for which is called its gamma. One difference between CRT monitors and flat-panel displays is that CRT typically have a gamma ranging from 2–3, whereas flat panel displays have a gamma of ~1; this means that flat-panel display luminance changes fairly linearly with drive voltage. Therefore, to characterize a CRT adequately, it is especially important to sample each channel's transfer function at several DAC.

3.2. Display System Characteristics

CRT often exhibit a lack of independence among the RGB channels: The whole does not equal the sum of its parts. Channels interact due to circuit interference and power supply limitations. It is also possible to have interactions among different areas of the display, due either to electronic limitations or internal reflections off the display's faceplate, which cause light from one area on the display to exit at another. Therefore, when performing characterization measurements, it is best to display an image that is representative of those for which colorimetric predictions are desired. Non-uniformities in the display's temporal and spatial characteristics can also affect characterization. For example, Berns *et al.* (1993a) found that a CRT can take as long as three hours to warm up, and even then as much as 90 s may be needed before the image's luminance and chromaticity stabilize. Longer-term instabilities occur over a month or more. One must also be aware that luminance and chromaticity may vary across the display screen (for CRT, luminance is greatest at the center of the screen and decreases toward the sides and corners by as much as 25%), so a characterization performed at one location may not yield accurate predictions for others. Finally, viewing angle should be considered. Viewing angle does not usually have much effect for CRT, but can impact both luminance and chromaticity for flat panel displays—liquid-crystal displays, in particular.

3.3. Ambient Illumination

Ambient illumination affects image legibility and readability. It does this by adding luminance to the image, which reduces contrast and desaturates the colors. It can also create specular glare off the faceplate. If the display will be viewed under ambient illumination, characterization should be performed with the illumination present.

3.4. Photometric and Colorimetric Instruments

Three types of devices are used to characterize displays: photometers, colorimeters, and spectroradiometers. Photometers measure luminance. Colorimeters measure chromaticity coordinates and either luminance (if they have optics to gather light from a target) or illuminance (if they have an integrating diffuser instead). Properly calibrated devices are reliable to about ± 2% in luminance or illuminance and to within ± 0.005 in chromaticity coordinates. Spectroradiometers provide a color's spectral power distribution (SPD), which is the equivalent of a fingerprint for light. An SPD graphs radiant power as a function of wavelength. Given an SPD, the associated luminance and chromaticity coordinates can be computed using the appropriate CIE equations.

4. METHODS FOR ACHIEVING DESIRED COLORS ON A DISPLAY

There are essentially two ways to get predetermined luminances and chromaticity coordinates to appear on a display. The best choice depends mainly on the number of colors that are needed. If only a few colors are needed — for example during prototype evaluation — iterative manual search is best. If many colors are needed, display characterization and the use of predictive equations are usually more efficient in the long run.

4.1. Iterative Manual Search

Iterative manual search requires little knowledge about the display system. The operator simply adjusts DAC until the desired luminance and chromaticity coordinates are achieved. Before attempting this, it is a good idea to measure the peak luminances and chromaticity coordinates of the RGB channels and ensure that the colors one seeks lie within the display's color gamut. (See Post and Lloyd (1994) for a tutorial on color gamuts.) It is also helpful if the channels are monotonic, i.e. each increase in DAC produces an increase in luminance.

Iterative manual search involves: (1) presenting a color on the display; (2) measuring the color's luminance and chromaticity coordinates; (3) determining whether the measurement meets the desired specifications; (4) adjusting the color's RGB DAC if necessary; and (5) performing another measurement to determine whether the adjustments produced the desired color. This process repeats until the desired color is achieved. When performing the adjustments, it is useful to know that the R DAC affects mainly the CIE *x*-chromaticity coordinate, G affects mainly the *y*-coordinate, and B affects mainly the *z*-coordinate, which is to say it affects both *x* and *y*. All three DAC affect luminance, of course, although G usually has the greatest effect because the G channel typically has the greatest maximum luminance.

4.2. Predictive Methods

Predictive methods are used to identify either the DAC needed to produce a desired luminance and chromaticity coordinates, or conversely, the luminance and chromaticity coordinates resulting from a particular set of DAC. There are basically two approaches to predicting a display's colorimetric behavior: theory-based modeling and empirical modeling. Both approaches require fitting functions to the characterization data set.

To predict the luminance and chromaticity coordinates that will be produced by specified DAC, the first step is usually to use predictive equations to determine the RGB channel luminances that the DAC will produce. These luminances are converted to CIE XYZ-tristimulus and the tristimulus are then converted to luminance and chromaticity coordinates. The conversion from channel luminances to tristimulus is:

$$
\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} \dfrac{x_R}{y_R} & \dfrac{x_G}{y_G} & \dfrac{x_B}{y_B} \\ 1 & 1 & 1 \\ \dfrac{z_R}{y_R} & \dfrac{z_G}{y_G} & \dfrac{z_B}{y_B} \end{bmatrix} \begin{bmatrix} L_R \\ L_G \\ L_B \end{bmatrix}, \tag{1}
$$

where X, Y, Z are the desired CIE 1931 XYZ-tristimulus; x_R/y_R, x_G/y_G, etc. are the CIE 1931 x,y-chromaticity coordinates for the RGB channels and will be referred to here as the

transformation matrix; and L_R, L_G, L_B are the RGB channel luminances necessary to produce the desired tristimulus. Typically, the transformation matrix is determined at each channel's maximum DAC to maximize measurement signal-to-noise ratio. Colorimetric predictions need not use only a single transformation matrix, though. A separate matrix can be generated for each display parameter that affects predictive accuracy, such as location on the display, image size, number of images presented simultaneously and different ambient lighting conditions.

Sometimes, an instrument capable of measuring channel chromaticity is not available, so channel chromaticity must be estimated. The display's manual may contain values; if not, they can usually be obtained from the manufacturer. For CRT, the channels' chromaticity coordinates can also be guessed using published standards for phosphors.

To predict the DAC needed to produce a desired luminance and chromaticity coordinates, the inverse of equation (1) is used to compute the required RGB channel luminances. Predictive equations are then used to determine the RGB DAC that will produce those luminances.

4.2.1. Theory-based modeling

Berns *et al.* (1993b) presented a method for modeling the relationship between RGB DAC and RGB channel luminances that is based on theory of how CRT behave. (Actually, they modeled the CRT's output in units they called "monitor tristimulus values," which are more generic than luminance; however, their model is easier to understand and explain in terms of luminance.) The basic form of the equation they used is:

$$L_i = L_{i,max}[k_{g,i}(DAC_i/DAC_{i,max}) + k_{o,i}]^{i}, \qquad (2)$$

where L_i is channel i (R, G or B) luminance, $L_{i,max}$ is channel luminance at maximum DAC ($DAC_{i,max}$), $k_{g,i}$ is display system gain for channel i, DAC_i is DAC for channel i, $k_{o,i}$ is display system's or offset for channel i, and γ_i is the display system's gamma for channel i. Equation (2) can be inverted to compute DAC for a given luminance. The gain, offset and gamma terms are estimated for each channel using nonlinear regression, i.e. search techniques. The resulting power function provides accurate predictions and has been endorsed by the CIE (1996); however, special software is required to perform the nonlinear regressions, human intervention is required to ensure the solutions converge, and each solution requires multiple searches using different starting values for the parameters to assure optimality.

4.2.2. Empirical modeling

Empirical modeling seeks a function that minimizes prediction error without regard to whether the function corresponds with theory. Post and Calhoun (1989) found that the predictive accuracy of several empirical polynomial and logarithmic functions, assessed in terms of error in luminance and chromaticity coordinates, was generally poor. Post and Calhoun (1989) presented a better approach, termed piecewise linear interpolation assuming constant chromaticity coordinates (PLCC), that provides good predictions with minimal implementation

effort and no more than 16 characterization measurements per channel. Lucassen and Walraven (1990) found that a logarithmic version of PLCC also provides good predictive accuracy. PLCC can be used with the white-point decomposition method introduced by Berns *et al.* (1993b) to reduce the required number of characterization measurements.

The aforementioned methods provide good predictive accuracy, however improvements can be made. Post and Calhoun (1989) described a method they called piecewise linear interpolation assuming variable chromaticity coordinates (PLVC), which provides improved accuracy. PLVC incorporates measured changes in channel chromaticity at each measured DAC. The advantage is that variations in channel chromaticity are accounted for in the predictions. The disadvantage is increased implementation complexity.

4.3. Correcting for Faceplate Internal Reflections

Del Barco *et al.* (1995) showed the improvement in predictive accuracy that can be achieved by incorporating a correction for internal reflections at the faceplate. After performing a characterization, the display is measured at zero DAC. Next, the zero-DAC measurement is subtracted from each measurement in the characterization. Predictive equations are then fitted to the corrected characterization data set. To predict the luminance and chromaticity coordinates produced by given DAC, the calculation is made in the usual manner and the zero-DAC measurement is then added to the prediction. To predict the DAC needed to produce a given color, the zero-DAC measurement is first subtracted from the color's tristimulus and then the calculation is made in the usual manner.

REFERENCES

BERNS, R.S., GORZYNSKI, M.E. and MOTTA, R.J., 1993a, CRT colorimetry, Part II: Metrology. *Color Research and Application*, 18, 315–25.

BERNS, R.S., MOTTA, R.J. and GORZYNSKI, M.E., 1993b, CRT colorimetry, Part I: Theory and practice. *Color Research and Application*, 18, 299–314.

CIE, 1996, *The Relationship between Digital and Colorimetric Data for Computer-controlled CRT Displays*. CIE publication 122 (Vienna: CIE Central Bureau).

DEL BARCO, L.J., DIAZ, J.A., JIMENEZ, J.R. and RUBINO, M., 1995, Considerations on the calibration of color displays assuming constant channel chromaticity. *Color Research and Application*, 20, 377–87.

LUCASSEN, M.P. and WALRAVEN, J., 1990, Evaluation of a simple method for color monitor recalibration. *Color Research and Application*, 15, 321–6.

POST, D.L. and CALHOUN, C.S., 1989, An evaluation of methods for producing desired colors on CRT monitors. *Color Research and Application*, 14, 172–86.

POST, D.L. and LLOYD, C.J.C., 1994, Color display gamuts and ambient illumination. *Displays*, 15, 39–43.

WALRAVEN, J., 1992, Color basics for display designers. In Widdel and D.L. Post (eds) *Color in Electronic Displays* (New York: Plenum), 3–38.

Chinese Characters and Computer Input

B. HAN and J. LI

Institute of Psychology, Chinese Academy of Sciences, Beijing 100101, China

1. INTRODUCTION

The past three generations of Chinese have had a tough but successful experience of processing their native language by computer. After the development of the English word-processing system of the 1960s, Japanese as well as Chinese scientists began designing the Chinese word-processing system (Ong and Shahnavaz 1987). Since the first Chinese keyboard input system (CKI) was produced in 1978 (Chen and Hu 1994), dozens of Chinese word-processing systems (CWPS) were developed. Some of them, e.g. CCDOS, UCDOS, CWPS, Xiao Jun 2.13 and WPS, became popular in the 1980s. In the 1999s, almost all the important international business programs have a Chinese equivalent, e.g. dBASE, FoxPro, SPSS, SAS, not to mention MS Office (Word, Power-Point, Excel).

CKI has been a bottleneck for Chinese information processing. Although the speed of CPUs and printers has increased, users must keep their typing at a relatively slow speed (50–300 characters/words per min for a normal/professional typist). Given the fact that the efficiency of CKI is already higher than that of English keyboard input (Chen and Hu 1994), and the successful design and marketing of the Chinese information processing system, one of the remaining problems is that each must include several CKI methods to obtain as many users as possible. On the one hand the user can choose the method he/she prefers, but on the other the system designer must offer several options, since no method can be regarded as superior over another. This is confusing for the beginner. Which method is the best and which is the most suitable for him/her? Although Chinese is not as difficult to input as may be imagined, the complexity of CKI can cause some initial fear or anxiety. In fact, even a skilled Chinese typist has a greater workload than the user who is typing English. This chapter will describe the causes of the workload of the CKI process, as well as the efforts to lighten this workload.

2. PHYSICAL/MENTAL DEMANDS OF CKI

2.1. General Workload

No matter what kind of features (orthographic, phonological, combination) a CKI method will take, all of them aimed to transfer the text to an internal binary code which can be processed by computer, according to specially designed rules, as shown by Figure 1, which depicts the flow chart of CKI. Note that compared with English keyboard input, stages 1, 2 and 4 are unique in CKI. Stage 5 is much more complex than the letter-internal code used for transforming the input of English, thus increasing the demands on both user and computer. As is shown, the types of workload differ between the initial stage of learning and the executive process, a fact which nearly all CKI users must deal with.

- Remembering the coding and mapping rules — each CKI method has at least two sets of rules (Han and Sun 1999). One is for encoding characters/words with a selected number of phonological or orthographic features as elements. The other is a set of mapping rules, whereby each element is allocated to one key on the board. It is the primary source of mental workload for the beginner of CKI. However, it is not always difficult since the coding method to some extent uses a rational code, which means that it is similar to a user's background knowledge. For example, the phonological coding method uses the Pinyin system (initials and finals) which most Chinese are familiar with from their early schooling. This makes it easier for users to learn, although they still have the coincident characters (those which share the same code) to deal with. Orthographic coding methods are difficult to remember but have far fewer coincident characters. For mapping rules, the phonological element (21 initials, 35 finals, four tones) is easy to map on the 26 keys of the keyboard, and is fairly simple to learn. However, orthographic elements (normally > 100 elements except for the stroke feature) are more difficult to remember and map. There are always some specific rules to balance the distribution of elements on keys, based on orthographic or phonological similarity between elements and keys. The user must simply learn these by heart.
- Encoding the characters/words — users encode the text manually, or mentally, according to the coding method used. This can be done one by one at stroke, component, character, word, phrase, or even sentence level. This will result in an elemental sequence or string for the basis of the next operation.
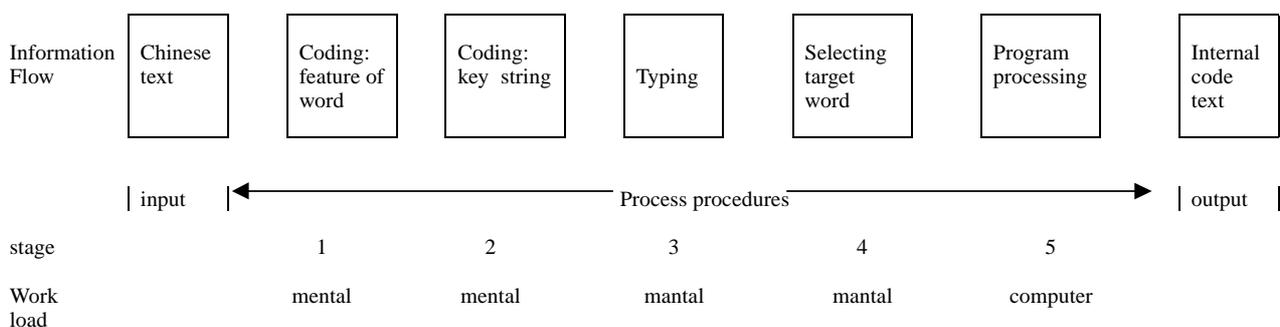
Information Flow	Chinese text	Coding: feature of word	Coding: key string	Typing	Selecting target word	Program processing	Internal code text
	input	← Process procedures →					output
stage		1	2	3	4	5	
Work load		mental	mental	mantal	mantal	computer	

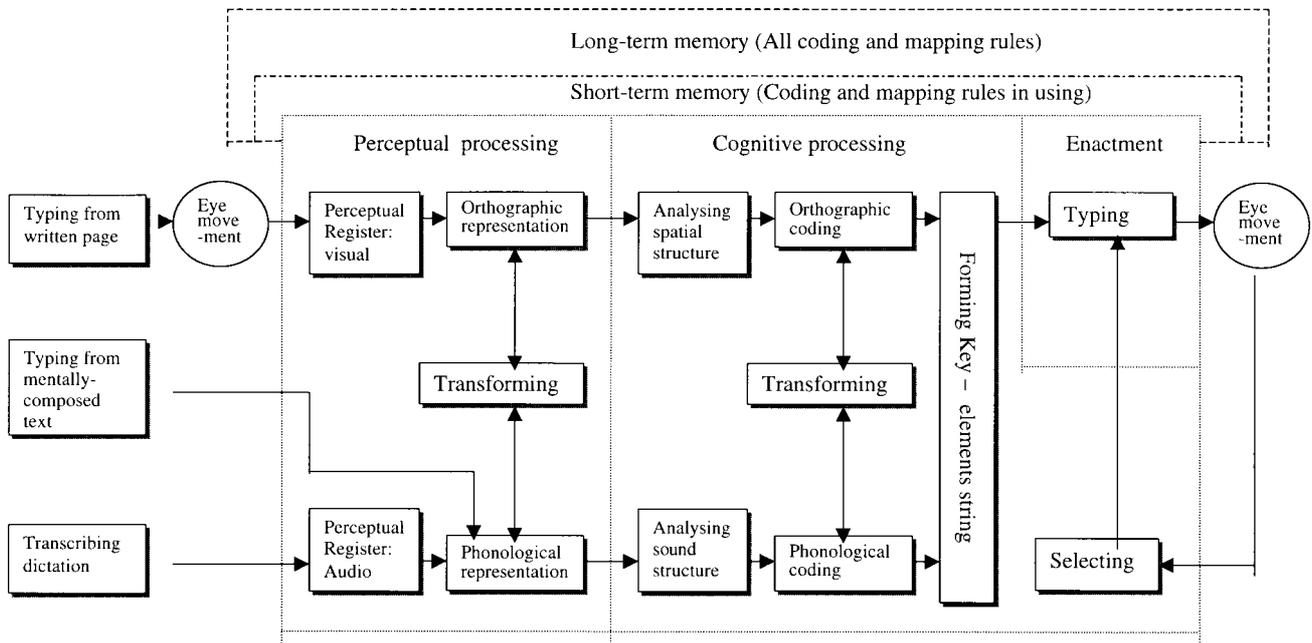**Figure 1. Flowchart, processing procedure, and workload of Chinese keyboard inputting.**

Figure 2. Cognitive model of Chinese keyboard input (adapted from Zhang and Chen 1991).

- Input of text by punching keys — the operator punches the keys along with the element sequence which is formed through encoding. Physical workload is the main characteristic of this step. There is also a certain degree of physical workload in connection with typing. For example, in typing from a written page, the operator must move his/her eyes back and forth between the screen and the text while typing. This causes tension in the ocular and neck muscles. However, the mental workload in this situation is relatively low if the typist uses an orthographic coding method, and need simply transcribe the text to an internal code. (The orthographic processing level in this situation infers a lower mental workload). Figure 2 shows the cognitive process along with different typing situations.

- Selecting the main character — one of the difficulties of CKI is that the user must select a main character from a number of coincident characters/words, especially regarding the phonological coding method (Han and Sun 1999). Occasionally it requires both time and effort since the user must turn over the small pages showing the coincident characters/words in the limited windows of the screen. The uncertainty of the result after keypunching produces some mental stress, which can only be eliminated by selecting the main character. It could even be disappointing if no matching character is produced after punching.

2.2. Specific Aspects of CKI Workload

The workload is also related to specific aspects, e.g. typing situations, coding methods, interface characteristics, and the users themselves.

- Input situations — CKI coding methods can be used daily for transcribing dictation, or for input from a written page, as well as mentally composed texts. Different situations involve different modalities. Whether the feature used in the coding method matches the modality involved decides how great the mental workload will be. For example, the auditory modality is involved in processing phonological information, so the phonological coding system is suitable for typing from dictation or mentally composed texts. If using the orthographic coding method in these two situations, the user must process the information by phonological modality, then encode it in a visual system, then transfer to auditory modality again in order to finish the input. This kind of transferring greatly increases the mental workload, not to mention the special rules that have to be referred to in some cases. In addition, the deep processing level (phonological and semantic process versus orthographic process; Craik and Lockhart 1972, Han and Lin 1998) involves more cognitive resources, hence the added mental workload.

- Coding methods — the degree of mental workload varies with the different coding methods, since the coding rules

do not always correspond with a user's background knowledge. As stated, the phonological code based on the Pinyin system is therefore easiest to remember, while the orthographic code normally uses components which the user is familiar with, but which nevertheless require some thought before using. As mentioned above, the greater the number of coincident characters, the greater the amount of eye and head movement, hence a greater physical workload. Although a combination of phonological and orthographic codes are used in most coding methods in order to obtain a balance of advantages/disadvantages, it is necessary to choose one of them as a main code, thus avoiding any confusion. Another aspect of the physical workload is the difference in code lengths, i.e. the difference in key punching time. For example, the character "jiang" by Pinyin has five letters, with only one initial (j) and one final (iang). This means that two punches are theoretically enough for inputting by initial (represented as j) and final (say, if represented by h), but at least five letters are needed if using the whole letter code (jiang). This is the reason for differences in movements and the amount of fatigue in muscles.

- Interface design — there has been very little research done at system level for interface designs; an important exception are the studies on Chinese character display (Zhu and Xu 1990). These are significant given the fact that most of the educated Chinese will have to use the CKI system sooner or later. This area contains much for ergonomic researchers to explore. Considering that there are several popular Chinese platforms to date, the availability of different CKI methods is not equal to users. Some of these methods are included as OEM programs (e.g. Biao-Xing-Ma, Wu-Bi-Zi-Xin), some often have to be installed by users themselves separately.

- The operator — the operator's characteristics, e.g., computer anxiety, background knowledge (composing, coding), personality, and movement/activity co-ordination, can affect the CKI efficiency. However, no research has been done on CKI users with respect to the above-mentioned variables. Psychologists have a great deal to explore in this area.

2.3. Other Related Problems

- Software design — Chinese version of many popular English information processing systems have been produced. Given the technical success of marketing the Chinese versions of software such as MS Office, SPSS, FoxPro, etc., more attention should be paid to the specific characteristics of Chinese information. Issues such as page layout, format, and document production/transferring need more research instead of just copying everything from the original foreign software. Nevertheless, several CIP systems (e.g. WPS, CCED, UCDOS, and Chinese Star) have a considerable number of users.

- Speed and memory of computer and printer — each Chinese character is saved and processed as a picture. The greater its discriminating rate, the greater the demand on the processing speed or memory of CPU and printer. Mismatching between input, processing, and printing often induces extra waiting time.

3. RELATED GROUPS FOR DECREASING THE WORKLOAD OF CKI

It has been a 20-year struggle for designers and operators to try to lower the demands of CKI. Achievements have been significant, in that more and more people use word processing daily, and many technical improvements have been made. However, there is still a long way to go. The participation of ergonomic researchers and psychologists will be helpful to both designers and users; these related groups can approach the problem from different angles.

3.1. The Designer

The designer always chooses a favorite feature, either orthographic or phonological, as a coding element. However, it is important to consider how to facilitate the use of the coding method. The first step, often taken instinctively, is that of making the coding element as rational as possible in order to match the user's background knowledge. The second step is to make the rules — coding as well as mapping — as simple as possible. However, the simplicity of these rules is often cancelled by the coincident code or the biased distribution of features among the keys on the board. Therefore, a certain number of special rules must be used as a control. This is an expedient measure for the programmer, although it results in a greater mental workload for the user. In short, it is not always possible for the programmer to make everything easy.

3.2. Users

Users must choose the best CKI method for his/her daily word-processing work, although this is not always a clear-cut choice to make. It is important to strike a balance between usability, ease of recall, and the individual's own language habits and profession.

Users in northern China who can speak Mandarin well often prefer phonological coding methods, while users in southern China often choose orthographic coding methods, and so avoid having to deal with the pronunciation of characters/words. Professional typists often prefer orthographic methods for higher speed and fewer coincident codes during inputting, while researchers and writers will most often choose phonological methods for the consistent auditory modality involved in thinking and encoding, and also for the ease of learning and remembering.

3.3. Ergonomic Researchers/Psychologists

The role of ergonomic researchers/psychologists in lowering the workload of CKI can be significant, with respect to the remaining human/computer interface issues. These issues have seldom been investigated; improved communication between researchers in different fields and between different countries is needed.

4. DISCUSSION
4.1. Workload and Evaluation of CKI

When we discuss the workload of CKI, we often refer to keyboard input of alphabetical languages, in particular that of English. However, no reasonable comparison on keyboard input between alphabetical languages and ideographic languages has been made to date, nor a cognitive process analysis or any experimental studies.

The key to decreasing the workload of CKI is to find suitable methods for evaluation and measurement, as well as effective methods for lowering the workload. This has, however, proven difficult because firstly, although earlier studies on evaluation realize the difficulties in copying variables related to users, further study is still needed, and secondly, more research is needed on ergonomic issues of CKI.

Han and Li (1998) have proposed a framework of CKI system evaluations, based on the cognitive process analysis and user-centred design strategy. It is argued that the "best" principle should be replaced by a "satisfying" principle in the design of a CKI method. Accordingly, the computer program strategy should be changed from "Chinese-centered" or "computer-centered" to "user-centered." A Chinese-centered strategy focuses on language attributes, while a computer-centered strategy focuses on the application of coding methods. The user-centered strategy focuses on the cognitive characteristics of different inputting situations as well as human/computer interface harmony.

4.2. Significance of Research on the Workload of CKI

It is, no doubt, very important to lower the workload of the CKI system. There are 1.3 billion people whose native language is Chinese, in addition to the increasing number of foreigners using the Chinese language to communicate with Chinese colleagues for business or academic reasons, plus Internet visitors of innumerable Chinese Web sites. This implies an increase in the use of CKI systems. Much time and effort could be saved through a decrease of workload in at least one aspect of CKI. Related topics, such as the limitation of Chinese attributes and coding designs, as well as user-training programs, need further study.

In addition, this is a promising field in which many areas remain unstudied and worthy of exploration by international ergonomic researchers, not only for the sake of technology itself, but also from the viewpoint of cross-cultural comparison.

REFERENCES

CHEN, Y.F. and HU, X.H., 1994, *Fundamentals of Technology and Theory for Chinese Character Inputting* (Beijing: Qinghua University Press and Nanning: Guangxi Science and Technology Press). [in Chinese]

CRAIK, F.I.M. and LOCKHART, R.S., 1972, Levels of processing: a framework for memory research. *Journal of Verbal Learning and Verbal Behavior*, 11, 671–84.

EDITORIAL DEPARTMENT, 1996/97, *People's Republic of China Year Book*, vol. 16 (Beijing: PRC Year Book Ltd, with Hong Kong: NCN Ltd).

HAN, B.X., 1997, On the mental, coding, and statistic component sets. *Computer World*, 11. [in Chinese]

HAN, B.X. and LI, J., 1998, On the "user-centered" design strategy of Chinese coding input system. Paper reported at annual conference of Chinese Ergonomics Society, October, Beijing, China. [in Chinese]

HAN, B. and LIN, Z., 1998, Coding process of orthographic, phonetic, and semantic features in Chinese word recognition. In Xinhui Yang (ed.) *Exploration of Psychology*. Serial no. 1, 267–74, (Nanjing, China: Nanjing University Press). [in Chinese]

HAN, B.X. and SUN, P., 1999, Fundamental issues and coding strategy of Chinese keyboard input. *International Encyclopedia of Ergonomics and Human Factor*.

KANG, A.B.H. and BOON L.H., 1987, Issues in software ergonomics: user interface design and Chinese computing input methodology. *Ergonomics*, 30, 1645–53.

ONG, C.N. and SHAHNAVAZ, H., 1987, Ergonomics in China: a historical perspective and some recent development. *Ergonomics*, 30, 1631–44.

WANG, H., CHANG, B.R., LI, Y.S., LIN, L.H., LIU, J., SUN, Y.L., WANG, Z.W., YU, WANG, Z.M., 1990, Recent developments in ergonomics in China. *Ergonomics*, 33, 853–65.

ZHU, Z.X. and XU, B.H., 1990, The development of ergonomics in China. *Applied Psychology: An International Review*, 39, 265–73.

Computer Mouse Use

A. Aarås and O. Ro*

Alcatel STK ASA, PO Box 60, Økern, 0508 Oslo, Norway and
*Premed AS, PO Box 275, Økern, 0511 Oslo, Norway

1. INTRODUCTION

In the last decade visual display units (VDUs) have become widespread for routine use in almost all workplaces in offices and production units. Development of computer networks and software programs like Windows have given the operator the opportunity to obtain almost all information at the computer. Thus, the time per day spent using the computer and particularly the time spent using the mouse as an input device has been increasing. Johnson *et al.* (1993) found that mouse use can account for up to two-thirds of the computer operation time, depending on the task performed and the software used. However, mouse use is nearly always combined with keyboard work, therefore consideration should be given to optimize all aspects of VDU activities. Several studies have documented a relationship between VDU work, particularly mouse use, and musculoskeletal illness. Karlqvist *et al.* (1996) reported an odds ratio of 2.2 with prolonged mouse use for CAD operators regarding shoulder/upper arm, elbow, wrist, and hand/finger symptoms compared with telecommunication laboratory workers. Several studies have documented increasing risk of upper extremity symptoms in terms of an odds ratio of 2 comparing VDU work more than 4 hours per day with an operational time of less than 4 hours per day (Punnett and Berqvist, 1997).

An exposure-response relationship between VDU use of more than 2 hours versus less than 2 hours for hand/wrist problems was found by Oxenburgh (1987). Punnett and Berqvist (1997), in their review of epidemiological studies of VDU work, found that VDU work indicated higher risk of neck, shoulder, arm, wrist and hand musculoskeletal illness compared with non-VDU work. Bergqvist *et al.* (1995a) compared VDU work with non-VDU work. They found a cumulative incidence of 2.8 for discomfort in the forearm and hand compared with non-VDU workers. Furthermore, a dose-response relation between VDU work and hand/wrist problems was indicated (Bergqvist *et al.* 1992). Many factors are associated with musculoskeletal discomfort in the neck and shoulder: rapid work pace, stereotyped repetitive keyboard and mouse use. Other factors are stressful work posture resulting from poor workstation and input device design as well as insufficient recovery time in terms of limited opportunity for rest breaks (Punnett and Bergqvist 1997). Limitation of spontaneous rest breaks was found to be associated with neck and shoulder discomfort (Bergqvist *et al.* 1995b).

In addition to musculoskeletal illness, eye discomfort is the main problem reported by VDU operators. Aarås *et al.* (1998) found a relationship between visual discomfort and pain in the neck and shoulder for VDU workers. The correlation coefficient was between 0.3 and 0.4. In the same study, a significant reduction of visual discomfort was reported by the VDU workers after having increased the illuminance level from 300 lx to 600 lx with luminaires giving high luminance of the ceiling and walls (approx. 80 cd m^{-2}),

Optometric corrections for presbyopic operators have to consider the viewing distances, which for VDU work are normally longer than reading distances. Laboratory studies by Horgen *et al.* (1989, 1995) showed that single-vision lenses gave less muscle load in the neck and shoulder compared with multifocal lenses. The importance of having optimal corrections was shown in a study where two intervention groups received optometric correction if needed; these groups reported significant reduction of visual discomfort but the control group, who did not receive corrections, showed no significant changes (Aarås *et al.* 1998). The importance of using single-vision lenses, if possible, is also supported by Bergqvist *et al.* (1995b), who found that VDU operators wearing monofocal glasses had lower risk for tension neck syndrome compared with those wearing bifocals or progressive glasses.

2. WORK POSTURE

There is little knowledge about an optimal postural load which gives a training effect of the muscles and prevents musculoskeletal discomfort. However, there is a consensus that the static muscle load should be kept at a minimum while the dynamic work pattern should be increased (Kilbom *et al.* 1986). Based on this knowledge, the VDU workplace should enable the operator to adopt a reference work posture giving minimum static load, along with easy variation in work posture, thereby increasing the dynamic muscle work.

2.1 Postural Load of the Neck

The muscle activity of the neck, shoulder and back is influenced by the trunk posture along with the head, neck, and arm position. In addition, any external load in the hands will increase load in the musculoskeletal system. Using electromyography (EMG), Schüldt *et al.* (1986) measured the lowest static load in the neck and shoulder muscles when the trunk was leant backwards 10-15°. The muscle load is very low, even for an extremely flexed position of the cervical spina (Harms-Ringdahl 1986). These results are supported by the biomechanical analysis of Dul *et al.* (1982), who found that in 40-45° flexion of the cervical spina, the neck muscles are stretched more than 30% of their length, i.e., the passive muscle and ligament forces are almost sufficient to counterbalance the forward turning movement of the head. Aarås *et al.* (1997) compared directions of gaze below the horizontal of 15° and 30° to the centre of the screen for data entry work. No significant difference was found between the two positions regarding static trapezius load or the static load of the erector spina lumbalis (L3 level).

Epidemiological evidence supports the idea that great flexion of the head and neck does not increase the incidence of musculoskeletal illness in the neck and shoulder. Aarås (1994) found that in workers who flexed their head between 39° and 58° for the predominant work position, the development of sick leave was low for musculoskeletal illness in the neck and shoulder. These data indicate that even great flexion of the head and neck in the sagittal plane gives a small risk of development of musculoskeletal illness in the neck and shoulder areas. This knowledge seems important for reducing visual fatigue. Investigation of the resting states of the oculomotor systems in different directions of gaze suggests that work with downwards gaze could reduce visual fatigue. Lie *et al.* (1997) have reported that visual discomfort and musculoskele-

Figure 1. The physiometer records EMG on four channels; six channels are available for postural angles measurements. Inclinometers are attached to the upper arm, head and back

tal pain and stiffness were reduced in a one-year follow-up study when the position of the VDU screen was lowered from 15° to 30° downward gaze inclination to the centre of the screen.

2.2. Postural Load of the Shoulder

Flexion and abduction of the upper arm in the shoulder joint are much more important risk indicators for musculoskeletal illness in the neck and shoulder than flexion of the neck (Sigholm *et al.*

1984, Kadefors 1994, Aarås 1994). Sigholm concludes that the shoulder muscle load depends on the degree of flexion and abduction of the upper arm in the shoulder joint (glenohumeral joint) and the load in the hands. This means that flexion and abduction of the upper arm in the glenohumeral joint should be kept as low as possible. This is particularly important if the operators have no possibility of supporting the forearms on the tabletop in front of them (Aarås *et al.* 1997). A laboratory study was carried out to measure EMG (figure 1) from the upper part of the musculus trapezius and from the lumber part of the musculus erector spina (L3 level) during data entry work. The muscle load was significantly less in sitting with supported forearms compared to sitting and standing without forearm support (Aarås *et al.* 1997).

Similar results were also found when using a mouse as input device (figure 2). In this test, a computer version of the card game solitaire was chosen because of short cognitive decision making, leading to very frequent movement of the mouse.

The median trapezius load was reduced significantly from 3.2% maximum voluntary contraction (MVC) to 0.4% MVC. Corresponding values for the erector spina were 1.7% MVC and 1.0% MVC. The results from this laboratory study were confirmed in a real work situation. Two groups were given intervention in terms of new lighting, opportunity to support the forearms, and optometric corrections if needed. A third group acted as a control. According to Aarås *et al.* (1998), Shoulder pain was reported with no significant differences between the three groups before intervention ($p = 0.66$). The average intensity of shoulder pain during the last six months showed a significant reduction in one of the intervention group ($p = 0.02$) and a clear tendency to reduction in the other group ($p = 0.08$), while no significant changes were found in the control group ($p = 0.92$), when comparing before and after interventions. The intervention groups reported significantly lower intensity of shoulder pain compared with the control group after the interventions ($p = 0.02$).

Both the keyboard height and the VDU placement are associated with pain in the musculoskeletal system (Punnet and Berqvist 1997). Thus, the workstation should be adjustable to accommodate the anthropometric dimensions of the operators. Training in adjusting the table and chair is important in reducing
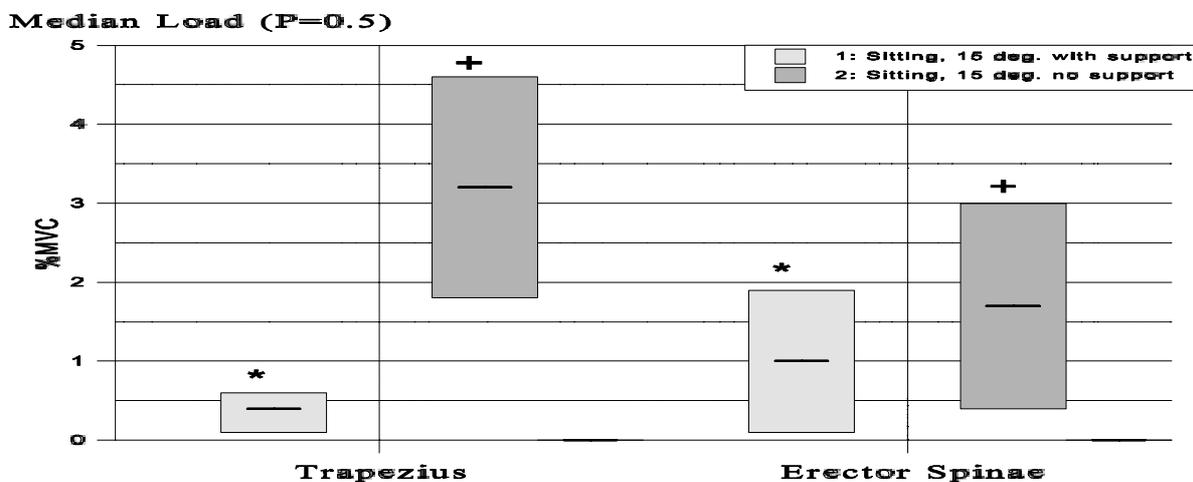

Figure 2. Group median values with 95% confidence interval for the two work postures; 20 subjets participated in the study

neck, shoulder, and forearm discomfort during VDU work (Lim and Carayon 1994).

2.3   Postural load of the forearms

Supporting the forearm on the tabletop did not affect the pain level in the forearm (Aarås *et al.* 1998). However, a relationship was found between the pain level in the forearm and the total time when the operators used the mouse. In a laboratory study, a mouse which gave the operator a more neutral forearm position was compared with a traditional mouse design. The muscle load of the forearm in terms of extensor digitorum communis and extensor carpi ulnaris was significantly less when using the mouse with the forearm in a more neutral position compared with the greater pronation of the forearm required for a mouse of traditional design (Aarås and Ro 1999).

The importance of the position of the upper arm and forearm is documented by Karlqvist *et al.* (1994). They found that mouse operators had a more outward rotated position of the upper arm in the shoulder joint compared with nonmouse users. For mouse operators the work posture and movements of the arm and hand were in the interval 15-30° ulnar deviation of the wrist as much as 34% of the working time. The corresponding percentage for those workers not using a mouse was 2%. Karlqvist *et al.* (1996) found an association of neck and upper extremity symptoms with hours per day of mouse use. The odds ratio with prolonged mouse use was 2.2 or higher for shoulder, upper arm, elbow, wrist, and hand/finger symptoms. Hünting *et al.* (1981) found that data entry work with more than 20° of ulnar abduction of the right wrist had an odds ration of 3 regarding pain in the right arm compared with those with less than 20°. They recommend that the ulnar deviation of the hand in the wrist should be less than 20°. Preferably, the hand should be held in the most neutral position possible at the wrist joint.

The importance of working with a neutral position of the wrist and hand is evident from several studies. A neutral position of the wrist may be important to prevent carpal tunnel syndrome (CTS) by giving low intracarpal tunnel pressure (ICP). The position of the wrist that creates the lowest ICP when the wrist has a flexion of 2.0-3.5° and ulnar deviation of 3-5° and a small pronation (Rempel and Horie 1994). Epidemiological studies support the importance of operating the keyboard and mouse with a more neutral position of the hand and wrist. Titteranonda *et al.* (1998) compared a keyboard of alternative design (splitting the keyboard in half and increasing the angle of the two halves from lateral to medial end) with a traditional keyboard. After six months the overall pain symptom severity of musculoskeletal illness (tendinitis) and functional status of the hand significantly improved while the operator using the traditional keyboard reported worsening of pain and discomfort.

It has also been suggested that changing the operation of the mouse between the right and left hand should be beneficial. Hoffmann *et al.* (1997) showed that VDU operators easily learn to change the mouse operation between their preferred hand and their nonpreferred hand, thereby reducing the period of activity in one hand. In a prospective epidemiological study, Aarås and Ro (1999) have documented the importance of operating the mouse with a neutral position of the wrist and hand compared with a more pronate forearm when using the traditional mouse design. Some 66 subjects suffering pain in the shoulder and fore-

arm with an intensity of at least 25 mm on a 100 mm Visual Analog Scale (VAS) were included in the study. The group was randomly divided into one intervention group and one control group. The intervention group got a new mouse, giving them the opportunity to work with the forearm in a neutral position; the control group continued with a traditional mouse design. After six months a significant reduction was reported in the intervention group regarding the pain intensity in the wrist/hand and forearm as well as the shoulder region and neck; no significant changes were observed in the control group. These results were valid for the average pain intensity as well as the frequency of pain.

3. CONCLUSION

The amount of knowledge already available should be applied to reduce the high prevalence of work-related musculoskeletal illness and visual discomfort for VDU workers. Primary prevention is of utmost importance, because when musculoskeletal illness has reached a chronic state, it seems very difficult to cure the disease, even if the most important risk factors are reduced to a minimum (Berg and Torell 1988). However, there is a clear indication that if the risk factors are abolished in the early stages of the musculoskeletal illness, the outcome will be promising (Jonsson and Persson 1988).

REFERENCES

AARÅS, A., 1994, The impact of ergonomic intervention on individual health and corporate prosperity in a telecommunication environment. *Ergonomics*, 37(10), 1679–96.

AARÅS A., FOSTERVOLD. K.I., RO, O., THORESEN, M., and LARSEN, S., 1997, Postural load during VDU work: a comparison between various work postures. *Ergonomics*, 40(11), 1255–68.

AARÅS, A., HORGEN, G., BJØRSET, H.-H., RO, O., and THORESEN, M., 1998, Musculoskeletal, visual and psychosocial stress in VDU operators before and after multidisciplinary ergonomic interventions. *Applied Ergonomics*, 29(5), 335–54.

AARÅS, A. and RO, O., 1999, Can a more neutral position of the forearm when operating a computer mouse reduce the pain level for visual display unit operators? A prospective epidemiological intervention study. *International Journal of Human–Computer Interaction*, 11(2), 79–94.

BERG, M. and TORELL, G., 1988, Persistence of musculoskeletal symptoms: a longitudinal study. *Ergonomics*, 31, 1281–85.

BERGQVIST, U., KNAVE, B., VOSS, M., and WIBOM, R., 1992, A longitudinal study of VDT work and health. *International Journal of Human Computer Interactaction*, 4, 197–219.

BERGQVIST, U., WOLGAST, E., NILSSON, B., and VOSS, M., 1995a, The influence of VDT work on musculoskeletal disorders. *Ergonomics*, 38, 754–62.

BERGQVIST, U., WOLGAST, E., NILSSON, B., and VOSS, M., 1995b, Musculoskeletal disorders among visual display terminal workers; individual, ergonomic and work organizational factors. *Ergonomics*, 38, 763–76.

DUL, J., SNIJDERS, C.J., and TIMMERMAN, P., 1982, Bewegungen und Krafte im oberen Koptgelenk beim Vorbeugen der Halswirbelsaule. *Manuelle Medizin*, 20, 51–58.

HARMS-RINGDAHL, K., EKHOLM, J., SCHÜLDT, K., NEMETH, G., and ARBORELIUS U.P., 1986, Load moments and myoelectrical

activity when the cervical spine is held in full flexion and extension. *Ergonomics*, 29, 1539–52.

HOFFMANN, E.R., CHANG, W.Y., and YIM, K.Y., 1997, Computer mouse operation: is the left-handed user disadvantaged? *Applied Ergonomics*, 28(4), 245–48.

HORGEN, G., AARÅS, A., FAGERTHUN, H.E., and LARSEN, S., 1989, The work posture and the postural load of the neck/shoulder muscle when correcting presbyopia with different types of multifocal lenses on VDU-workers. In: Smith and Salvendy (eds) *Work with Computers: Organizational, Management, Stress and Health Aspects* (Elsevier: Amsterdam), pp. 338–47.

HORGEN, G., AARÅS, A., FAGERTHUN, H., and LARSEN, S., 1995, Is there a reduction in postural load when wearing progressive lenses during VDT over a three-month period. *Applied Ergonomics*, 3, 165–71.

HÜNTING, W., LAUBLI, T., and GRANDJEAN, E., 1981, Postural and visual loads at VDT workplaces. I. Constrained postures. *Ergonomics*, 24, 917–31.

JOHNSON, P., DROPKIN, J., HEWES, J., and REMPEL, D., 1993, Office ergonomics motion analysis of computer mouse usage. In *Proceedings of the American Industrial Hygiene Conference and Exposition*, pp. 12–13. Fairfax VA: American Industrial Hygiene Association.

JONSSON, B.G. and PERSSON, J., 1988, Disorders of the cervicobrachial region among female workers in the electronics industry: a two-year follow-up. *International Journal of Industrial Ergonomics*, 3, 1–12.

KADEFORS, R., 1994, Welding and musculoskeletal disease: a review. *Proceedings of the 12th Triennial Congress of the International Ergonomics Association*, 2, 72–74.

KARLQVIST, L., HAGBERG, M., and SELIN, K., 1994, Variation in upper limb posture and movement during word processing with and without mouse use. *Ergonomics*, 37, 1261–67.

KARLQVIST, L.K., HAGBERG, M., KOSTER, M., WENEMARK, M., and ANELL, R., 1996, Musculoskeletal symptoms among computer-assisted design (CAD) operators and evaluation of a self-assessment questionnaire. *International Journal of Occupational Environments*, 2, 185–92.

KILBOM, Å., PERSSON, J., and JONSSON, B., 1986, Disorders of cervi-cobrachial region among female workers in the electronics industry. *International Journal of Industrial Ergonomics*, 1, 37–47.

LIE, I., FOSTERVOLD. K.I., AARÅS, A., and LARSEN, S., 1997, Gaze inclination and health in VDU-operators. *Proceedings of the 13th Triennial Congress of the International Ergonomics Association*, 5, 50–52

LIM, S.-Y. and CARAYON, P., 1994, Relationship between physical and psychosocial work factors and upper extremity symptoms in a group of occide workers. *Proceedings of the 12th Triennial Congress of the International Ergonomics Association*, 6, 132–34.

OXENBURG, M.S., 1987, Musculoskeletal injuries occurring in work processor operators. In: Steenson (ed.) *Readings in RSI: the ergonomic approach to repetition strain injuries* (New South Wales University Press), pp. 91–95.

PUNNETT, L. and BERGQVIST, U., 1997, Visual Display Unit Work and Upper Extremity Musculoskeletal Disorders. A Review of Epidemiological Findings. National Institute of Working Life, 17184 Solna SwedenæErgonomic Expert Committee Document No 1. ISBN 91–7045–436–1, ISSN 0346–7821.

REMPEL, D. and HORIE, S., 1994, Effect of wrist posture during typing on carpal tunnel pressure. *Proceedings of Working With Display Units: 4th International Scientific Conference, Milan*, p. C27.

SCHÜLDT, K., EKHOLM, J., HARMS-RINGDAHL, K., NEMETH, G., and ARBORELIUS, P., 1986, Effects of changes in sitting work posture on static neck and shoulder muscle activity. *Ergonomics*, 29(12), 1525–37.

SIGHOLM, G., HERBERTS, P., ALMSTRØM, C., and KADEFORS, R., 1984, Electromyographic analysis of shoulder muscle load. *Journal of Orthopedic Research*, I, 379–86.

TITTERANONDA, P., BURASTERO, S., ARMSTRONG, T., and REMPEL, D., 1998, Effects of alternative geometry keyboards on VDT users with work-related musculoskeletal disorders: clinical outcomes and keyboarding styles. *Proceedings of Premus-ISEOH'98*, *Finland*, p. 121.

Design Issues: Action Research In Control Room Operations

N. A. Stanton, M. Ashleigh
Department of Psychology, University of Southampton, SO17 1BJ, UK
Tony Gale
University of Portsmouth, Portsmouth, UK

1. CHANGES IN CONTROL ROOM OPERATION

There is a tradition of research being conducted into control room operation to enhance our understanding of the role of the human operator and learn about new ways of supporting those activities. According to Kragt (1994) technological developments in process control have led to dramatic changes in the nature of work practices and behaviors. This development has gone through four generations of control systems: local manual control, local automatic control, centralized panel displays and SCADA (system control and data acquisition) systems. The first revolution was to automate parts of the process so that workers were able to supervise larger areas of plant. The second revolution was to centralize the controls and displays into a single control room, again enabling workers to supervise larger areas of plant. The third revolution was to put all the information at the workers' fingertips via information technology, further reducing the personnel requirements.

A review of research into human supervisory control reveals three distinct phases over the past three decades. Research in the 1970s may be characterized by interest in cognitive control. Interest in the individual shifted to interest in team structure and performance in the 1980s. More recently, in the 1990s, researchers have been focusing on human behavior in context. Zwaga and Hoonhout (1994) argue that all of the technological developments in supervisory control have been based upon the conception of the control room engineers' task of "operation-by-exception": control room engineers only intervening in the process when called to do so by the alarm system (Dallimonti 1972). However, Zwaga and Hoonhout (1994) argue that this conception is fundamentally flawed. Rather, they propose, control room engineers' behavior is better characterized by a "management-through-awareness" strategy: control room engineers are actively extracting information from their environment rather than passively reacting to alarms. The dichotomy of "active extraction" and "passive reception" was noted by Stanton and Baber (1995) in an analysis of alarm handling activities.

In a review of the impact of control room technology on the behavior of the operators it was apparent that three research approaches were dominant: the use of surveys and questionnaires, small-scale observational studies, and simulation of control room environments. Many of the questionnaire studies involved selected samples of employees, often at a time of introducing new technology. This meant that they were being surveyed under conditions of task restructuring, when they were ill-trained for the technology in question, fearful of redundancy, and motivated to a particular set of response biases. Observational studies, which appear on the face of it to offer ecological validity, are often of a very short duration so that they are likely to suffer from the Hawthorne effect. Social psychologists have drawn a distinction, in the context of observation, between acts and actions, the former relating to behaviors observed by an outsider, the latter to the layers of meaning which surround an act, from the viewpoint of the actual participant or indigenous person. Unless the outsider is au fait with the group's history, purposes, beliefs, and values, acts remain as acts. Laboratory studies, which offered more control, often involved relatively brief samples of behavior generated by unrepresentative samples of volunteer participants with little sense of continuity of employment or of the personal significance of the tasks they were required to perform. We know that workers, after extensive operations in a working context, develop subtle adaptations in their interaction with work interfaces, establishing idiosyncratic patterns and habits which simply cannot be captured in brief simulations. In the search for an alternative approach which yields veridical data, ecological validity and experimental control seem to be at odds with each other; yet without appropriate controls it is hard to plot true causal paths. This proposal offers to test a compromise research approach which seeks to combine ecological validity with control in a dynamic and developmental way.

2. RELATIVE BENEFITS AND PITFALLS OF EXISTING APPROACHES

Gale (1984) was quick to realize that the benefits of traditional approaches should be preserved, whilst the disadvantages should be overcome if at all possible. The recognized benefits of laboratory studies include: control over independent variables, limitation on the number of independent variables, control over environmental variables, control over participants' behavior, event sampling at choice, construction of complete experimental designs, choice of representative or random samples, systematic manipulation of variables, systematic development of a series of studies eliminating specified sources of variability, simplification of data to a level manageable by existing theoretical power, partitioning of subject effects, systematization of error, capacity for repeated longitudinal measurement over time, planning in advance and replication. However, there are several shortcomings: failure of a differential partitioning of *in vivo* influences, the generation of statistically significant effects rather than practically significant recommendations, an interest in effects greater than chance rather than effects which apply to large populations, a focus on contemporary theoretical ideas to the exclusion of potential or actual factors impacting on a working environment, theory-driven rather than problem-driven orientation, the use of inexperienced participants (often from an inappropriate background), and sampling which is far too brief to enable participants to develop their own coping strategies. Field studies, whilst being more realistic, typically suffer from: multivariate influences on participants, uncontrolled environmental contexts, uncontrolled and unpredictable disruptive events, restrictions on access for the duration of the study, restrictions on experimental power and status, the use of incomplete designs with incomplete cells, low cooperation and high suspicion by participants, biased samples, incomplete designs which restrain inference, and an inability to replicate.

3. THE CAFE OF EVE METHODOLOGY

In 1984, Gale presented an internal report to the Human Factors Technology Centre at ITT Europe, in which he proposed a new

research strategy for assessing the impact of new technology and for guiding design. In 1987, Gale and Christie set out a detailed blueprint for the approach. The project was called the CAFE OF EVE — a Controlled Adaptive and Flexible Experimental Office of the Future in an Ecologically Valid Environment. Whilst originally conceived as an approach for investigation of human behavior with office technology, the CAFE OF EVE approach may be equally successfully applied to investigations in control rooms. This requires the researcher to reconceptualize the research paradigm by applying an action research approach to the investigation of human activity.

The aim of the CAFE OF EVE project is to combine the advantages of both laboratory and field studies while minimizing the disadvantages. The proposal involves taking over a control room within a company in a way which allows for the day-to-day operational function, combined with a parallel set of research studies. Staff operating within the selected control room would be included in the research function. For the duration of the work, their job descriptions would alter officially to include reflection on their working situation and upon task demands. Their roles as experts in this regard would be recognized. In so doing, they would be invited into a partnership of equals with the researchers, thereby diffusing some of the suspicion which attaches to clever outsiders. In turn, the researchers would share some of the control room functions with the aim of understanding the meaning of events and activities for participants. At the same time, through daily exchanges with the permanent employees, the barriers between participants and experimenters would break down. As a conscious act the researchers would gently move conversation and interaction beyond the working context, to issues of family and other non-work concerns, thereby cementing affiliate relationships.

By addressing issues beyond the work barrier, the researcher who adopts a willingness to self-disclosure lowers within-work barriers to free communication. Thus the researcher take on the role of participant observer as developed in anthropology (Vetere and Gale 1987), living and working within the human system in question but also recording daily events. Researchers and participants share a social world. As researchers and participants share a social world the barriers between researcher and participant become more permeable and participants feel freer to express their opinions and reactions about their working environment. In daily debriefing sessions, participants interact with researchers, with the goal of identifying problems from the participants' perspectives. End-of-day briefings are part of the extended job description of the worker, and the employer formalizes the use of the last half hour of the day for the worker to reflect on system operation during the day, his or her reactions, and any other thoughts which come to mind. Because of the relationship which has been built up with the researcher, the end-of-day briefing is more akin to an after-work report to the worker's spouse or partner than to an alien researcher. Within this context, therefore, the expectation is that worker reactions will be more unabridged than would be the case in more typical research interviews. Thus, the research questions which are generated are not dictated by existing theories but by the actual perceived experience of control room engineers.

So far as possible, video observation and analysis, diary keeping, interactive recording of subjective responses would be carried out in the control room and integrated with everyday task functions. The aim of the CAFE OF EVE approach is to use a longitudinal and developmental technique to capture real experience and to shape new technological developments. The research questions are not imposed by prior conceptions but emerge from the working context and the views and analyses of participants. Thus it involves a partnership in exploration in which researcher and participant have equal status. It is argued that objectivity is retained because the researcher is still apart, but ecological validity is ensured by drawing on the participants' day-to-day experiences.

The key features of the approach are as follows:
1. Ecological validity is approximated;
2. All psychologically significant variables are likely to be identified;
3. Participants and researchers develop a partnership of exploration;
4. Sampling is flexible and by mutual consent;
5. Participants reveal and are able to reflect upon coping strategies developed over time;
6. Research studies emerge naturally;
7. Participants themselves will suggest studies or identify salient variables;
8. Participant loss will be minimal thereby allowing longitudinal studies;
9. Cooperation will affect broader organizational structures, such as new ways of working.

The CAFE OF EVE methodology offers an action research approach. There is obviously an element of risk associated with investing in such a long-term project. The researcher is likely to encounter unperfected events and difficulties, but this likely to be outweighed by the quality of the data and the insights gained through research of this nature.

4. CONCLUSIONS

The CAFE OF EVE approach seeks to draw together a normal working context and a controlled laboratory to create a special human factors environment, capitalizing on the benefits of ecological validity and experimental control, while seeking to avoid the disadvantages of the two contrasting approaches. In so doing, the research benefits should surpass the benefits typically yielded by either approach taken separately or sequentially. What we are proposing, and its emergent properties, could constitute a minor revolution in human factors research.

REFERENCES

DALLIMONTI, R., 1972, Future operator consoles for improved decision making and safety. *Instrumentation Technology* 19(8): 23–8.

GALE, A., 1984, *Prelude to the Cafe of Eve: Report to the Human Factors Technology Centre* (Harlow: ITT Europe).

KRAGT, H., 1994, *Enhancing Industrial Performance* (London: Taylor & Francis).

STANTON, N.A. and BABER, C., 1995, Alarm initiated activities: an analysis of alarm handling by operators using text-based alarm systems in supervisory control systems. *Ergonomics* 38(11): 2414–31.

VETERE, A. and GALE, A., 1987, *Ecological Studies of Family Life* (Chichester: Wiley).

ZWAGA, H.J.G. and HOONHOUT, H.C.M., 1994, Supervisory control behavior and the implementation of alarms in process control, in N.A. Stanton (ed.) *Human Factors in Alarm Design* (London: Taylor & Francis).

Design and Use of Displays

S.J.P. McDougall and D.J. Oborne
Department of Psychology, University of Wales Swansea, UK

1. INTRODUCTION

Technological advances, particularly increases in computational power, mean that the ways in which displays can be used to support human–machine systems have multiplied. At the core of display design is the assumption that the information provided for the user should be appropriate to both user and system requirements at any given time. As Rolfe and Alnutt (1967) put it, "The display translates what is at first imperceptible to us into perceptible terms". At its best, providing appropriate information support in a way that is meaningful and easily interpreted will improve the overall performance of the human–machine system (Bennett, Nagy, and Flach 1997).

2. USES OF DISPLAYS

The diversity of ways in which displays are now employed means that it would be impossible to provide a comprehensive listing of their uses. At their simplest, displays may provide simple warnings or quantitative information. Other displays may provide information primarily about the quality of the machine state (e.g. in photocopiers). More complex displays often have multi-purpose goals. These include displays which are designed to track or control complex processes (e.g. in nuclear power and chemical plants) and should enable the operator to predict or anticipate the state of the system and provide warnings of system anomalies or danger. Other complex systems include tactical cockpit displays, air traffic control, and other command and control center displays. In these situations the amount of potential information which could be displayed is growing all the time and the trade-off between user support and display clutter constantly needs to be juggled in design. Furthermore, information requirements in complex systems may change quickly and dramatically and displays must be able to change and adapt to those needs.

Given the demands made of displays, it would be ideal if there were a "royal road" to optimal display design. There is not. What follows, therefore, are a series of considerations, heuristics, and procedures which can help the designer to create and evaluate displays. At the end of the day, however, it is often the imaginativeness and creativity with which these are employed by the designer that determines the effectiveness of a display.

3. DISPLAY DESIGN

Factors which need to be taken into account in display design will be considered in relation to (a) the design of symbols or icons for displays, (b) the use of words and alphanumerics, (c) the organization of displays, and (d) the process of creating and evaluating designs.

3.1. Symbols and Icons in Displays

Symbols and icons are commonly used in displays to indicate system functions and to provide information from the system. What follows is a general consideration of symbol characteristics that need to be considered in design and how these are likely to affect users' interaction with the display.

3.1.1 Pictorialness

Pictorial, or concrete, symbols depict objects, places, and people with which we are already familiar in the real world. Other symbols have less obvious connections with the real world and information is represented using graphical features such as combinations of shapes and arrows (figure 1 shows examples of pictorial and abstract symbols). Pictorial symbols are useful when someone is learning to use a display because real world knowledge can be used to guess what they mean. Users, particularly novices, prefer them and they give the feeling that the display is easy to use. Often lack of space and the complexity of the information to be presented makes frequent use of pictorial symbols impracticable. This does not matter too much since experienced users usually find more abstract symbols just as easy to use.

3.1.2 Visual complexity

This is the amount of detail or intricacy in a symbol (figure 1(i) shows a complex symbol, the others are simple). In general, it is important to follow a "simplicity principle" when designing symbols. Using many complicated symbols will mean that users will take more time to find the appropriate information in a display. Experienced users may be able to deal better with clutter on displays, but it will still slow them down and will lead to errors being made.

3.1.3 Color

Color can be used very effectively to support the organization and ease of use of displays (see Davidoff 1987 for a review). Color is particularly useful when displays are dense or complex because it can be used to direct attention to a particular symbols, to groups of symbols, or to particular detail within a symbol. As a result, it can reduce the time taken to identify appropriate information and help to create conceptually related symbol families. Color can be used to rank items in order along a scale (by using increasingly powerful shades of the same color) or to provide hazard warnings (red is commonly accepted in America and Europe). Unrestrained use of color should be avoided since it creates visual clutter and increases search times.

3.1.4 Size

The precise size of symbols in a display is usually determined by three features: typical viewing distances, display quality, and viewing conditions. Size requirements at different viewing
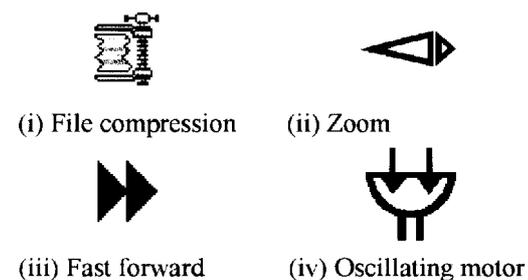

Figure 1. Examples of symbols used in displays.

distances are usually specified in detail by internationally agreed standards, individual company style, or by government. Display quality may vary in accordance with resolution, contrast, focus, and glare. Viewing conditions depend upon environmental factors such as noise, smoke or dust; they also include physiological and psychological factors such as fatigue, eye strain, workload, stress, and anxiety.

3.1.5 Shape

Shape can be used to improve discriminability between symbols in displays. If effective contrasts in shape are used they can reduce the time it takes for users to identify appropriate information. Symbol shape can also be used to help organize displays. Shape can highlight similarities between symbols which represent similar information and can be used. to create symbol families. Shapes can also be used to convey conventional meanings (e.g. figure 1(ii)).

3.2. Use of Words and Alphanumerics

Despite the increasing trend to move away from the use of words, and especially lines of text in displays, symbols in displays are very frequently accompanied by text. The most obvious reason for this is to reduce the ambiguity inherent in the use of symbols. Symbols often rely on visual associations and the context in which they appear for their meaning whereas words are rarely ambiguous. Words can also convey more complex meanings and ideas in a way that would be virtually impossible with symbols. Similarly, use of numbers is often the most effective way of conveying quantitative information, particularly where display space is limited. Despite the obvious differences between symbols and text, many of the design principles applied for symbols can be also applied in this domain.

3.2.1 Concreteness and meaningfulness

Where users may not always be experienced or have appropriate training, concrete words which have clear meanings should be used wherever possible. When words are abstract (e.g. "idea", "method", "justice') the meaning becomes more diffuse and harder to comprehend. For experienced users, however, the use of abstract or jargon terms which are specific to the use of the system can sometimes provide a useful shortcut to more complex meaning. Where words are being used as textual labels for symbols, there should be a clear match between symbol and label. Labels should be avoided where they are likely to be long (figure 1(iv)) or add little meaning.

3.2.2 Size, simplicity, and shape

As with symbols, the size of letters is largely determined by viewing distance, display quality, and viewing conditions. Font size is almost always determined via agreed international standards. Letters are usually kept as simple as possible in order to enhance legibility and decrease clutter on the display (e.g. avoiding the use of long or heavy serifs and hairline strokes). A feeling of clutter can also be avoided by ensuring that letters and words are adequately spaced. Providing they are not over-used, letters can provide a useful visual contrast to symbols on a display. This is because skilled readers can instantly identify text amongst other shape cues.

As Wright (1999) has noted, users tend to adopt the principle of "least reading effort'. Users' searches for relevant information may often be limited initially to viewing symbols; only where disambiguating or further information is required will text be consulted. Labeling or other textual information should be kept as legible and brief as possible to allow users to "skim" this information and direct their attention to what is relevant.

3.3. Display Organization

Good display organization is the most important determinant of whether or not users can direct their attention to the relevant information. It also plays an important part in determining how easy that information is to understand and respond to.

3.3.1 Configurality

Configurality refers to the way in which elements within displays are arranged in order to convey information effectively. Careful attention is paid to the nature of the relationship between elements in displays in order to allow easy interpretation. The different types of relationships are as follows:

(a) Separable relationships: There is no interaction between visual elements in the display (see figure 2(a)). These displays are better if users need to attend selectively to each display component (i.e. they need access to low-level data).

(b) Configural relationships: There is an intermediate level of interaction between visual elements. Figure 2(b) shows how providing a shape cue can aid interpretation of the graph in figure 2(a).

(c) Integral relationships: There is a very strong relationship between elements to the extent that unique perceptual identities are lost. Figure 2(c) shows an object display used to represent operation in a nuclear plant. The spokes of the polygon have been designed so that a regular polygon represents normal conditions and any distortions indicate an abnormality. This display integrates information from more than 100 individual sensors. What is clear is that this figure conveys meaning in a condensed way that would be impossible with words. Integral displays are better if users need to combine information to make a response or need to be aware of higher level constraints among several pieces of information.

Use of configurality is likely to be most useful where relationships between visual elements can be exploited to create new emergent features which aid interpretation, where information must be extracted from a complex screen, and where user workload is high.

3.3.2 Display complexity and layout

In addition to the use of simple symbols and text, the simplicity of the display as a whole needs to be considered. The need for display simplicity has become an increasingly important consideration as the amount of information has multiplied. Tullis (1983) suggests that four types of complexity need to be considered as follows:

(a) Overall density: this refers to the percentage of the total number of possible characters or symbols which could occur in the display space. High information density causes perceptual overload, increased errors and difficulties in finding appropriate information.

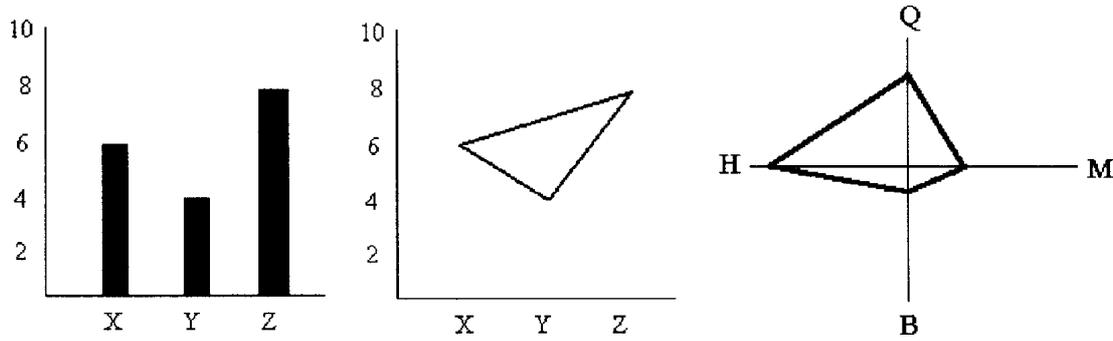(b) Local density: this is the amount of space which is filled

Figure 2. Use of configural relationships in displays.

around a target area or symbol. This is important because blank spaces can be used to provide structure in a display and areas of high local density alert designers to areas where user performance may be affected.

(c) Grouping: this differs from local density because it considers how close items are and how likely they are to be perceived as conceptual groups.

(d) Layout: this builds on consideration of grouping and considers the irregularity, or layout complexity, of functional groupings in a display.

(e) Consideration of display complexity and layout should lead to a well organized display which is easy to interpret. The use of appropriate conceptual groupings allows the user to develop strategies to reduce mental workload.

3.3.3 Creating contrasts between groupings

Creating contrasts between different parts of the display layout can help users to direct their attention quickly to appropriate parts of the display. It can also help users gain a quicker conceptual understanding of the information being presented. Discrimination between symbol and text clusters, or families, can also be achieved by the use of elementary features in the displays. These features include:

(a) color
(b) size
(c) shape
(d) orientation
(e) increasing the size of critical symbol features

Other contrasts include the use of simple symbols in an otherwise complex display. Similarly, pictorial symbols are distinctive when most other symbols in the display are abstract. Overuse of contrastive features should be avoided since this will simply reduce distinctiveness across the display.

3.3.4 Conceptual complexity and "cognitive fit"

The assumption underlying concern with the visual aspects of a display is that good visual elements in a display will facilitate the users' interpretation and integration of information. Consideration needs to be given to the complexity of the information in layout groupings and the amount of world knowledge activation and processing required when this information is presented. The display should fit in as well as possible with users' current knowledge and expectations.

4. CREATING AND EVALUATING DESIGNS

Given the complexity and diversity of displays the process of creating and evaluating designs is likely to differ from one to another. Some of the methods typically used are considered briefly below.

4.1. Creating Display Designs

There a number of steps which designers typically follow when designing displays. The earliest phase of design usually consists of formulating a clear statement of requirements for the display. This will include not only details about what should appear in the display but information about the system operating environment and likely user groups. This is followed by, or carried out in conjunction with, a task analysis which will consider the nature and order of the tasks to be carried out by the system and the user. This will include consideration of context of use and workload. Once this has been completed, prototype display designs will be created. Ideas for displays may be elicited from focus and user groups. The precise nature of the display, however, will be determined to a greater or less extent by a number of other considerations. These include the tradition and philosophy of the company for whom the display is being created, precedents created by displays of a similar nature, customer expectations, recommendations from international standards along with government requirements, and the likely costs and time-scales for development.

4.2. Evaluating Display Designs

Once a prototype has been created, the designer may choose one or more of a variety of evaluation techniques depending on the nature of the display being created. These are as follows:

(a) Comprehensibility testing: This may be carried out by asking potential user groups to name, match, or rate the meaningfulness of symbols which will appear on the display.

(b) Discriminability testing: User awareness of information groupings in the display can be determined by asking potential users to sort information provided in the displays into groups. This is often done using simple card sorting tasks using cards with each piece of information noted on the cards.

(c) Surveys, questionnaires, and interviews: These are typically used to troubleshoot and provide information about areas of difficulty in the displays.

(d) User preferences: Indications of the kinds of displays that users will like using can be obtained by presenting users with a series of display possibilities and asking for rankings, ratings or "most preferred" statements. User preference does *not* always predict user performance.

(e) Prototypes and simulation: These are often created after the initial evaluation process. Simulations are particularly important in estimating likely user performance where systems are complex and workload is high.

5. CONCLUSION

At the heart of display design is the need to provide easily interpreted information which supports decision making and system control. There is no simple formula to allow the designer to do this. It is the combination of knowledge, experience, and creativity which most often leads to good design.

REFERENCES

BENNETT, K.B., NAGY, A.L., AND FLACH, J.M., 1997, Visual displays. In G. Salvendy (ed.) *Handbook of Human Factors and Ergonomics* (New York: Wiley), pp. 659–94.

DAVIDOFF, J., 1987, The role of colour in visual displays. In D.J. Oborne (ed.) *International Reviews of Ergonomics: Current trends in human factors research and practice* (London: Taylor & Francis), pp. 21–42.

ROLFE, J.M. and ALLNUTT, M.F., 1967, Putting the man in the picture. *New Scientist*, 16, 401–6.

TULLIS, T.S., 1983, The formatting of alphanumeric displays: A review and analysis. *Human Factors*, 25, 657–82.

WRIGHT, P., 1999, Printed instructions: Can research make a difference? In H.J.G. Zwaga, T. Boersema, and H.C.M. Hoonhout (eds) *Visual Information for Everyday Use: Design and Research Perspectives* (London: Taylor & Francis), pp. 45–66.

Handwheels

H.–J. Bullinger and M. Braun

Fraunhofer-Institute for Industrial Engineering (IAO), Stuttgart, Germany

1. INTRODUCTION

Generally, controls are the interface elements in a man — machine system through which energy or information is transferred from the user to the technical system. The most common way to transmit movement by mechanical controls is rotation. There are numerous different types of rotary controls which are normally standardized. In order to operate with rotary controls a torque is required. According to their shape different types of rotary controls can be differentiated: (hand)wheels, rotary disks, rotary switches, knobs and cranks. Rotary controls can be equiped with pointers or scales.

Design, selection and arrangement of handwheels — as one important type of rotary controls — must be considered especially according to the criteria of human factors and ergonomics. The dimension and the position of a handwheel affect the strain to which the user is subjected as well as the effectiveness and safety of the system. Design dimensions of a handwheel must be compatible with the anatomical, anthropometric, and physiological marginal human conditions. It is the intention of this contribution to provide some guidance in the selection, arrangement and design of handwheels.

2. OPERATIONAL CHARACTERISTICS OF HANDWHEELS

Wheels as torque transmitting elements are assigned to the rotary controls and can be shaped as handwheels (to be used with just one or with both hands) or as fingerwheels. They can be differentiated between disk handwheels with a closed stay and spoked handwheels, which consist of a rim to induce power, several spokes and the stock. The induction of power can be improved by a crank that is arranged with the rim. Handwheels are usually produced by casting or pressing material (metal or synthetic). Characteristic dimensions of handwheels, as well as permissible actuating forces and torques, are given in Figure 1.

Handwheels are recommended especially for the following control functions: continuous adjustment, precise adjustment, and large force applications. Quick adjustment and accidental actuation are acceptable with functional application. Handwheels

Dimensions		Moment M = f(D)	
Diameter	D = 100-800 mm	D = 100-200 mm:	M = 2-40 Nm
Rim Diameter	d = 30-40 mm	D = 200-250 mm:	M = 4-60 Nm

Figure 1. Characteristic dimensions of handwheels

are not suitable for control tasks requiring tactile and visible feedback. The main operational characteristics of handwheels refer to the control task and the coupling conditions, which are discussed below.

Force, speed and accuracy are regarded as the most important parameters of control tasks. These performance parameters are particularly influenced by the dimensioning and positioning of the handwheel. The maximum isometric muscle forces of humans in the hand–arm motion range constitute the basis for the transmission of forces to handwheels. The most important influential variables regarding the maximum isometric actuating moments, besides human muscular capacity, are technical parameters such as the force application point, the direction of force, and the type of coupling. The actuation of a handwheel should take place within the maximum force output range. With handwheels, the actuating forces or moments can be increased if the direction of force transmission is chosen in such a way that the body can find a support (backrest) or that, if possible, weight can be translated into actuating forces. The maximum isometric actuating moments for spoked handwheels are shown in Figure 2 by an example. As can be seen in Figure 2, body weight can be used if the handwheel is arranged at low positions, resulting in hight actuating moments. However, the effective translation of body weight is only possible if the handwheels have a larger diameter.

In the case of a frequent actuation of the handwheel, the actuation force, on account of the continuous performance limit of man, should not exceed 15% of the maximum force. If the actuation times are very short (< 5 sec) the resistance can account for approximately 50% of the maximum force. When designing safety equipment it must be ensured that, regarding the force to be developed, even the weakest person (fifth percentile) will still be able to actuate the handwheel.

The actuating speed depends primarily on the resistance of the used handwheel. The highest actuating speed can be obtained by a large motion area using a crank, because no regrasping of the handwheel will be necessary due to the static coupling.

The actuating accuracy of handwheels depends on the following parameters: dimensions and arrangement of the handwheel (front-, side-, height-position), motion range, type of coupling, forcing function, environmental influences, etc. Actuating movements can either be continuous or discrete — that is, in steps or stages. To set the intermediate positions, provision will have to be made for mechanical ratchets, so that the position can be safely identified.

The type of coupling can be determined according to whether coarse-motor or sensory-motor control tasks are to be performed. Clasping by the hand should be applied for coarse-motor control tasks, whereas gripping by the fingers should be applied for sensory-motor control tasks. The movements are further reduced in the event of a two-hand coupling on the handwheel.

The transmission of energy to the handwheel is usually frictional. In frictional coupling, the actuating force is transmitted through the friction forces that exist between the handwheel and the hand, with the necessary normal forces produced by the coupling forces (compare Section 3). In some cases a positive coupling will appear if actuating forces are induced by spokes
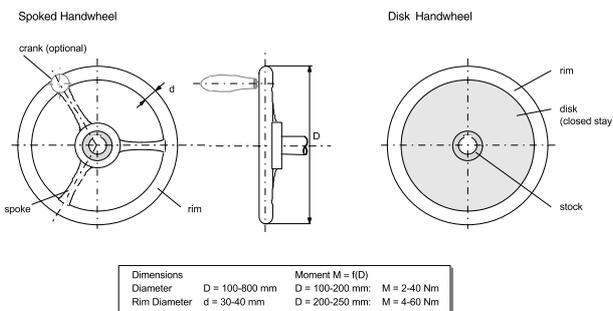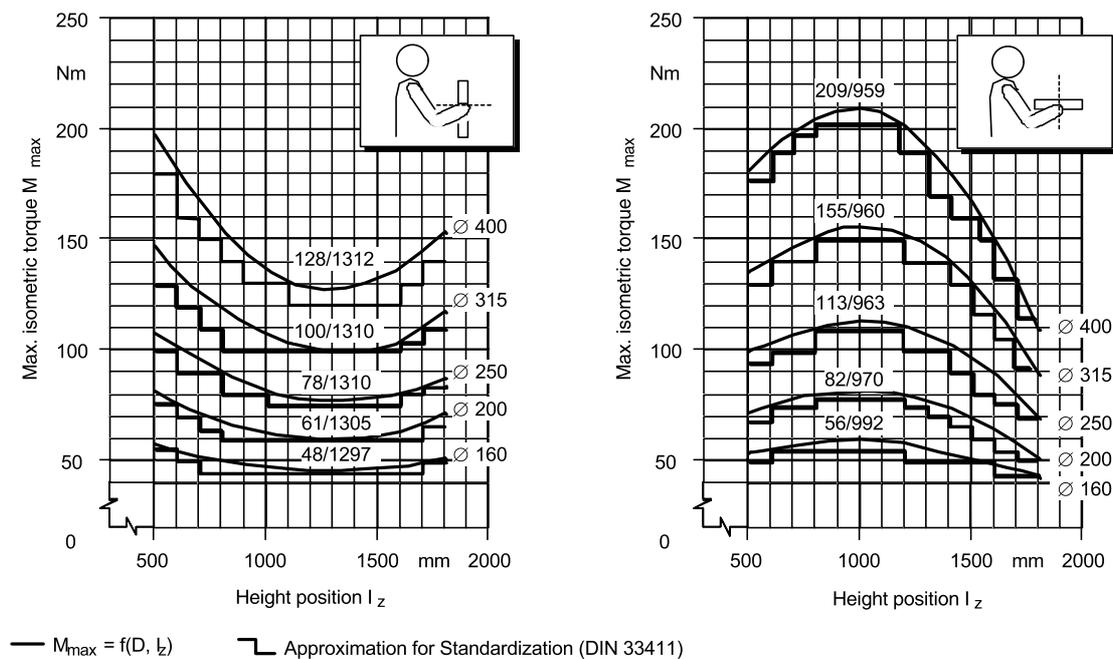
Figure 2. Isometric torques of handwheels (T = f {control diameter, height, direction}) (Mainzer 1982).

into the handwheel or if the handwheel is equipped with a crank. A positive coupling permits greater actuating force to be applied, but at the same time it also confines the possible movements of the hand–arm system.

Frictional coupling is favorable for dynamic coupling on the handwheel (regrasping). If the scope of movement of the hand–arm system is not sufficient for the required actuating distance, static coupling changes to dynamic coupling. In dynamic coupling, regrasping is necessary, so that the required movement can be performed. If the possibilities of movement are exhausted in static coupling, a larger scope of movement can be achieved only by movements of the trunk, resulting in additional stress and strain for the user (Kern, Muntzinger, and Solf 1984).

3. ERGONOMIC DESIGN PARAMETERS

The design dimensions of the handwheel — such as shape, size, material, and surface — are important factors influencing the operating effectiveness, with the characteristics performance, stress and strain of the user, and safety criteria. For defining the design variables, the control task, coupling conditions, and the user's capabilities and anatomy are of importance.

Designing the shape of one-hand- and two-hand-operated disk handwheels is regarded as being problematic. For safety aspects, disk handwheels must have a closed stay. The relationship between the maximum isometric actuating moment and the rim diameter has been approximated on the basis of experimental tests by a regression function for different handwheel arrangements (Kern *et al.* 1984). The regression function permits a rim diameter of approximately 45 mm to be derived. For reasons of economy and weight, the calculated rim diameter cannot be realized. Considerable savings of material and weight can be achieved by taking suitable design measures in which the stay of the handwheel can be used as coupling area. In the design of handwheels it will also have to be taken into account that the shape must satisfy criteria of ergonomics and safety for all relevant

positions and arrangements. As shown in Figure 2, the shape of the handwheel satisfies ergonomic criteria for both two-hand operation in the sagittal–horizontal axis and one-hand operation in the horizontal–frontal axis; that is, a large coupling area and thus minimum strain on the hand is guaranteed in all control positions.

In the case of frictional coupling of the hand with the handwheel, the power is transmitted by friction forces, where the normal force will be developed by the coupling force. The friction coefficient is determined by anatomical–physiological parameters such as the size of the coupling area, surface structure, and the skin's degree of moisture, on the one hand, and by the material's properties such as surface roughness and profiles, on the other hand. In frictional coupling the material must also be selected with respect to the frictional behavior that exists between hand/fingers and the handwheel. Unsuitable materials and surface structures lead very quickly to heavy strain and destruction of the upper skin layers. For detailed information on material friction properties, see Bullinger *et al.* (1997).

If the surface of the handwheel is profiled, the size and form of the profiles as well as the profile spacing will be important. Moreover, the direction of the profiles relative to the hand or the fingers is also relevant for power transmission. Because the strain on, and the danger of injury to, the skin becomes higher with increasing profile size, only fine profiles (profile spacing < 3 mm) are permissible for control element surfaces. In this case the profiles must be orientated vertically to the direction of force. The effectiveness of the profiles on smooth surfaces depends heavily on the normal force. Under smaller loads per unit area (normal forces), profiled surfaces show smaller coefficients of friction than smooth surfaces. The effective coupling area is considerably reduced by the profiles. Under higher loads per unit area, greater coefficients of friction result in the case of profiled surfaces, owed to a certain interlocking between skin and profiles.

If relative movements occur between handwheel and hand during operation, the thermal conductivity of the handwheel material will also have an influence on the operating effectiveness. The thermal conductivity of relevant control materials may vary between 0.15 W/Km (PVC rigid) and 70 W/Km (steel). At high actuating speed, the arising frictional heat can be dissipated only via the hand, which may result in an inadmissibly high temperature rise in the hand. With static coupling of the hand/fingers on the handwheel, materials with low thermal conductivity are required so that the hand´s temperature will not be imparted to the handwheel too rapidly, resulting in hypothermia of the hand.

If gloves have to be worn for performing control tasks, this must be accounted for in the dimensioning of the handwheels. Provision will have to be made for a plus or minus allowance of approximately 10% for inner and outer dimensions. With respect to the degree of fulfilling the control task, a reduced tactile sensitivity and lower mobility of the hand must be expected when wearing gloves.

4. ARRANGEMENT OF HANDWHEELS

The arrangement of handwheels is governed, on the one hand, by human capabilities such as anatomy, anthropometry, and physiology and, on the other hand, by the characteristics of the technical system to be manipulated. The most important criteria which have to be taken into account concerning the arrangement of handwheels are:

> Movement–physiological marginal conditions (as range of the hand–arm system, requirements made on actuating force, speed, accuracy)
> Coupling conditions
> Sequence of the process to be controlled (sequence and frequency of activities)
> Safeguards against inadvertent operation
> Operation while sitting or standing

According to the first aspect, it must be considered whether the handwheels will have to be actuated by women or men, or by both, and what percentile range will be relevant to the collective. For the transmission of greater forces and torques, provision will have to be made for operation while standing. On the other hand, operation while sitting largely eliminates the need for static posture energy. Depending on the type of handwheel used, its preferred vertical position will be between the elbow level and the shoulder level. Handwheels actuated by both hands should be positioned in the median plane.

To permit error-free operation of the handwheels, without inadvertently actuating any neighboring control element, certain distances must be observed. The minimum distance between two neighboring handwheels is about 20 mm, with the optimum distance at 50 mm, where the fingers are regarded as a crucial spatial condition (Grandjean 1991). If the operator wears gloves, provision will have to be made for a corresponding allowance. If the rear side has to be used as coupling area in one-hand- and two-hand-operated handwheels, a freespace between 20 and 35 mm will have to be provided between the rear side of the handwheel and the technical system.

5. COMPATIBILITY OF CONTROL–DISPLAY–SYSTEM

To ensure high effectiveness and safety of the system and a reduction of the response time as well as the learning phase, a high degree of compatibility is required between any control, technical system and display, taking into account the stereotypes of humans. Compatibility exists if the rotational movement direction of the handwheel coincides with the direction of movement of the technical system or the observable system variables. Thus, by clockwise turning of a handwheel, a response in a right direction or an extending or increasing function of the technical system is expected. Consequently, by counterclockwise turning a left direction, a retract or recrease response is expected.

The operation of a valve is an exception to these movement–effect stereotypes. To open the valve, it must be turned counterclockwise; to close it, it must be turned clockwise. These specifications may differ in different national standards and regulations.

6. SAFETY REQUIREMENTS

Improper actuation of a handwheel may be caused by the operator, unauthorized persons, or by surrounding influences. Inadvertent actuation of a handwheel by the operator may be caused by inadequate distance between one control and the next, getting one's clothing caught, or supporting oneself by holding onto a handwheel. Inadvertent actuation of a handwheel may also result from wrong operation by the operator, for example, due to unfavorable coding. The design principles discussed above should therefore also be seen in the light of safety aspects.

The marginal conditions of safety must be taken into consideration as early as in the definition of the design parameters. Shape and surface of the handwheel must be designed so that slipping off is prevented in order to avoid injury to the worker and improper actuation. If handwheels are mounted on rotating shafts, provision will have to be made for clutches so that the power transmission can be interrupted. If this cannot be realized, handwheels will have to be used in which clothing cannot get caught — that is, disk handwheels with a closed stay must be used instead of cranks and spoked handwheels. In the case of handwheels in which an inadvertent actuation would endanger persons and the system, provision will have to be made for
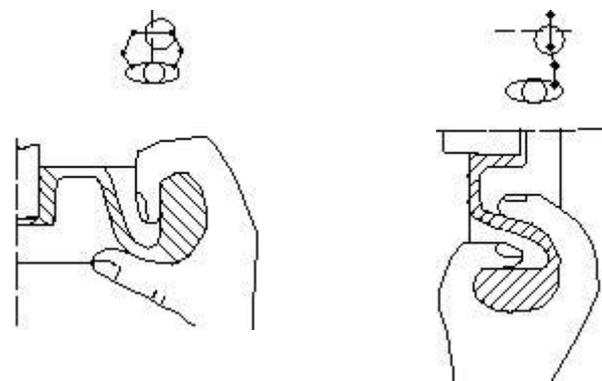

Figure 3. Shape for handwheels (disk type) and coupling conditions for different handwheel arrangements (From: Kern, Muntzinger, and Solf 1984).

additional safeguards. Such undesirable actuation can be prevented or minimized by making provision for large tripping forces, awkward handling by moving directions other than the preferred anatomical directions, shielding of the handwheel and using detachable or lockable handwheels.

The design measures mentioned above will have to be supplemented by notices or signs. It is obvious that in part safety measures do not satisfy the requirements of ergonomics. Thus it may be necessary to compromise, with due consideration to the priorities.

REFERENCES

BULLINGER, H.-J. and SOLF, J. J., 1979, *Ergonomische Arbeitsmittelgestaltung I-III* (Bremerhaven: Wirtschaftsverlag NW).

BULLINGER, H.-J., KERN, P. and BRAUN, M., 1997, Controls. In Salvendy, (ed.) *Handbook of Human Factors and Ergonomics* (2nd ed.) (New York: Wiley).

GRANDJEAN, E., 1991, *Physiologische Arbeitsgestaltung* (4th ed.) (Landsberg: Ecomed).

KERN, P., MUNTZINGER, W. F., and SOLF, J. J., 1984, *Entwicklung von normungsfähigen, ergonomisch richtig gestalteten Bedienteilen*, Report BMFT-HdA.

MAINZER, J., 1982, *Ermittlung und Normung von Körperkräften — dargestellt am Beispiel der statischen Betätigung von Handrädern*, Fortschritt-Berichte, Reihe 17, No. 12 (Düsseldorf: VDI-Verlag).

MUNTZINGER, W. F., 1986, *Ergonomische Gestaltung von Rotationsstellteilen für grob- und sensomotorische Tätigkeiten*, Dissertation (Berlin: Springer).

Input Devices and Controls: Manual Controls

S. K. Adams

Iowa State University, Ames, IA 50011, USA

1. INTRODUCTION

1.1. Human Control of Machines

Control by a person over his or her immediate environment is necessary for survival and to fulfill all physical needs and desires. The advancement of civilization and standards of living accompanying it have been strongly affected by the type and quality of tools and machines employed to enable necessary work activities to be accomplished. Throughout history until the past two centuries, machines were driven by water, wind, human and animal power. Control by the human was of the most basic types: either one in which the human supplied all of the force, energy and direction of movement as in using hand tools, or one in which the human directed forces as in plowing with animal power or steering a sailing vessel. All feedback was through human perception of the controlled activity and environmental conditions. Machines powered by external and internal combustion engines and by electricity extended the concept of control to include the production of mechanical energy and the operation of a great variety of machines and processes. Feedback from the process could now occur by direct observation as in sawing a board with a power saw or through displays through which the machine provides representative information (quantitative, qualitative, digital, analog or graphic) about what is taking place or has taken place in the process.

Later, the introduction of automation enabled internal control of basic machine processes using feedback generated by sensors within the machine (e.g. thermostat) replacing operator perception and skill with electrical and mechanical functions. Human control now assumed a higher level with more emphasis on planning, interpretation and cognition than on neuromuscular dexterity. Later, process control by digital computer created yet another level of control in which two computers are involved: one that monitors and directs machine or process functions based on feedback from the operation being performed according to predetermined parameters, and one that instructs the machine controlling computer in machine language based on input supplied by the human operator as well as feedback on the status of the operation from the machine controlling computer. At this level of control, the task of the human operator as a controller is largely cognitive. Control of the machine or process is exercised as keyboard or touchscreen input to a digital computer based on information displayed on the monitor. In some cases human control is eliminated almost entirely in the normal operating mode with the human operator only monitoring intermittently as in the use of an aircraft autopilot. Finally, when a machine or process is completely automated, the human operator is an observer only and exercises no control other than to start or stop the machine or process as in an emergency, for tool changes or maintenance.

There are obviously many levels of variation and degree of human versus automated control in the higher levels of control. Sheridan (1992: 1–97) discusses the many cognitive aspects and human–computer interactive features that are possible through modern technology, including telerobotics and supervisory control of complex and remote operations. Decisions made in this realm of system design determine the types of computer workstations and other design features needed for effective human–machine interaction.

Meister (1971: 12–13) lists the following questions related to the role of the human operator in the design of complex human–machine systems that should be discussed in dialogue between the overall system designer(s) and human factors specialists:

1. What inputs and outputs must be provided to satisfy system goals?
2. What operations are required to produce system outputs? (What must the man–machine system (MMS) do?)
3. What should be the assignment of system responsibilities (i.e. what functions should the man perform within the system)?
4. What kinds of men are required to perform these functions?
5. What tasks should they perform? Can they respond to system inputs and produce the necessary outputs?
6. What equipment interfaces does the man need to perform his tasks (e.g. controls/displays, test points, diagnostic information, procedures)?
7. What is the effect of the machine on the man; the man on the machine? (Since there is an interaction between the two, the machine may demand more of the man than he can produce; likewise, the man may hinder the functioning of the machine.)

Carefully defined answers to these questions should provide a framework upon which specific issues relating to human control of the system being developed can be structured and related, with necessary design tradeoffs to assure optimum utilization of human and machine capabilities, including computer utilization, software and the cognitive aspects of human control. Questions relating to control system hardware in the control station should be addressed after the basic foregoing system design questions have been answered since they depend upon the latter. Hardware questions include those related to the type, location, priority, functional grouping, sequential use and physical design features of controls. This section will concentrate primarily on the physical design features of controls.

1.2. Human Error and Control

Whenever a human operator must perceive stimuli, process information, make a decision and take action, there exists the possibility for human error. Error can occur in any of the activities listed above and it may occur consciously as in misjudging clearance for a part or vehicle or unconsciously as in the accidental tripping of a switch. Human error may involve the omission of an act as in neglecting to activate a lockout control to prevent accidental operation of a machine while it is being services or the commission of an act such as opening a valve that should be closed or performing operations on the wrong sequence. The severity of the consequences of human error tends to increase with the size and complexity of the system being controlled and the amount of energy being controlled. System integrated human factors in designing control systems are of major importance for

aircraft, ships, nuclear power plants, railroad traffic control centers, public utility systems, oil refineries, chemical plants and other facilities where human error can lead to a disaster.

Poor ergonomic design can result in design induced errors. These are control errors likely to occur even with well-trained and dedicated operators because they violate basic principles of design in human factors engineering and are incompatible with operator expectations or capabilities. Consider, for example, the likelihood of an operator correctly performing a control function that requires the first and last pair of switches in a row of five identical switches to be placed in an upward position while the middle switch must be placed in the downward position while the operator is visually engaged in another task. In such cases even what would be considered to be a minor error from a human factors perspective (inadvertent activation or non-activation of a switch) could have major consequences. In contrast, a well-designed system could have interlocking controls and display feedback that would prevent any incorrect sequence or omission even if it was performed intentionally.

Human errors can be classified in a number of ways. Swain and Guttmann (1980: 2–8) classify human error in the following categories:

1. Errors of omission.
2. Errors of commission.
3. Sequential errors.
4. Timing errors.
5. Extraneous acts.

The fifth category includes actions that are not part of assigned tasks and could cause the other four types to occur. In terms of the human–hardware interface that exists in a human-controlled system, the following categories are convenient from an ergonomic design perspective:

1. Perceptive errors — failure to detect or interpret a signal or condition.
2. Cognitive errors — failure to understand information from a display or a control action.
3. Decision — failure to select the appropriate action or to act in the appropriate time frame.
4. Response — performing an incorrect adjustment, failing to execute a control function, or performing functions in the wrong sequence.
5. Behavioral — use of an incorrect procedure or disregard of a procedure.

Controls and their associated displays should be designed and arranged in ways that prevent or greatly reduce the likelihood of these types of error.

Fitts and Jones (1947) as presented in Sinaiko (1961: 322–58) studied 460 pilot errors in flying US Army transport aircraft during and following World War II. While this study is old, it is considered classic in revealing errors that have been common in operating many types of industrial, military and consumer hardware through many years. Errors were classified as:

1. Errors in control selection or operation resulting from a lack of uniformity or consistency (standardization) in location and mode of operation.
2. Substitution errors resulting from variable arrangement patterns, inadequate separation and a lack of shape-coding or warning lights
3. Adjustment errors that can be reduced by automatic features,

simplified one-step operations, continuous (infinite settings) adjustment and control locks.

4. Forgetting errors that can be reduced by lock-outs for incorrect sequences, uniform "off" positions, check lists and warning systems
5. Reversal errors that can be reduced through improved control display compatibility and logical geometric relationships between the task or function and the control movement.
6. Unintentional activation of controls preventable through the use of uniform, user tested design and by adequate separation of controls
7. Inability to reach controls reduced through proper use and acquisition (including measurement) of anthropometric data and the use of functional anthropometry (measurements done under real operating conditions)

The use of ergonomic principles related to control and control system design coupled with mockups and user testing of layouts and controls during the early phases of system or machine design will eliminate or minimize these ergonomic design problems.

1.3. Modes of Control

There are many interactive modes that can be used to link a voluntary human response to a desired machine action. For ordinary human–machine systems, for large, slow controls requiring high levels of force the arms and legs are used with force and position feed back occurring through the hands and feet as occurs in the use of levers and footpedals. For smaller controls hand rotation and finger actions permit quick, accurate and coordinated inputs as with toggle and rocker switches, rotary dials, thumbwheels, small cranks, buttons, keyboards, keypads and touch panels.

Other more sophisticated controls are used in special applications. In some fighter aircraft, force on a joystick with little or no movement is used as primary input in controlling the direction, acceleration and rate of turn of the aircraft. The inertial reactions of head movements (accelerations and decelerations) have been used to control powered wheelchairs. Quadriplegics have used tongue-depressed buttons to guide wheelchairs or to activate robotic arms. Voice commands have been used in some applications, primarily to execute emergency stops or perform special functions in critical situations. Corneal reflection of an infrared beam has been used as a control input in advanced military systems but the unstable nature of eye position and movement patterns makes this unsuitable for most other applications. Electromyographic signals from limb muscles or electro-oculographic signals from eye muscles are also feasible only in advanced research or clinical applications. Glove-mounted optic bundles, potentiometers and strain gauges can also generate control signals from finger movements. These biofeedback methods often present challenges in reliability and accuracy. They can be of use in basic research on reactions during perception and response in dynamic human–machine task interfaces. McMillan *et al.* (1997) provide an in-depth discussion of non-conventional controls.

1.4. Control Resistance

The physical resistance of a control to being moved provides information regarding its position and movement in the form of force feedback. There are four basic types of control resistance:

1. Elastic resistance.
2. Friction resistance.
3. Viscous resistance.
4. Inertial resistance.

1.4.1. Elastic resistance

In the use of elastic resistance, the reactive force is proportional to the control displacement. Force equals a constant multiplied by displacement. The control returns to its neutral position when the force applied to it is released. The control can be held in any desired position by applying and maintaining the appropriate force for that position. It permits quick changes and is unlikely to be activated unintentionally or by outside forces. Common examples of elastic resistance controls include throttles, accelerator pedals and clutch pedals.

1.4.2. Friction resistance

For friction resistance, there are only two levels of reactive force: one for static friction to initiate control movement and one for dynamic friction during movement. Reactive force equals one constant value or the other. Since dynamic friction between solids is always less than static friction, overshooting the intended control position is likely. Precise adjustments are very difficult. Such controls are less likely to be inadvertently moved by the operator or an outside force but they do not permit accurate adjustments using single movements. They are suitable for on–off mechanisms and for multiposition displacement if detents are provided (as with a lever) to define each position.

1.4.3. Viscous resistance

For viscous resistance or viscous damping, force is proportional to velocity. Force equals a constant multiplied by the control velocity in any given direction. Two directions (forward and backward) are commonly used. Viscous resistance reduces the likelihood of accidental control movement and prevents making large control movements rapidly. It facilitates smooth, precise control movements while also permitting rapid changes in direction. Small adjustments in position are easily made.

1.4.4. Inertial resistance

For a control providing inertial resistance, force is directly proportional to acceleration, paralleling Newton's Second Law. In this case, starting and stopping control movement are difficult thereby hindering precise movements or incremental adjustments. This type of control is unlikely to be accidentally activated or changed in its position. Since it requires large forces, frequent use should not be required. It also presents the problem of overshooting. Such controls appear in the form of large valve control wheels or cranks used to control the flow of water or other fluids or to operate heavy doors or hatches.

2. HUMAN FACTORS IN MANUAL CONTROL

2.1. Anthropometric Factors

Anthropometric factors to be considered in designing manual control systems include those related to clothed body dimensions for the desired population percentile range (typically 5th to 95th) for both genders considered in combination with reach capabilities, working positions for seated and standing operators as applicable, and mobility and access requirements. Mockups and simulators provide the best methods of evaluating proposed workstation and panel designs for the intended population. Simulators, although more expensive, also permit studies of human–machine interface dynamics. Strength, speed and accuracy of hand, arm and leg movements vary greatly with limb flexion, extension, orientation and direction of applied force. Hand, finger and foot dimensions must also be considered in providing adequate grip surfaces, contact areas and control separation. Avoiding accidental activation of controls for major functions is very important, especially when safety could be compromised. Some controls may have to be covered or placed in a mechanically locked position when not being used. In any application, dynamic functional anthropometry of representative users in their normal work clothing should be used in determining work space, layout and control selection.

2.2. Biomechanical Factors

Biomechanical factors include strength in terms of specific force exertion capability, type of control motions required to operate controls, speed and precision of control motion required, reach capability for given control operation and effects of body acceleration (as on a vehicle) upon performance. Good anthropometric design facilitates good biomechanical design since limb strength, speed and precision of movement are capabilities that depend upon body configuration. Very few control operations require a large fraction of an operator's maximum strength but excessive control force can cause fatigue and possibly neuromuscular disorders from repetitive usage over an extended period of time. Force requirements for controls should be within ranges provided in control design force recommendations for a given type of control. Control resistive force should be sufficient to prevent unintentional control activation or movement. Emergency controls and the forces needed to operate them should be determined on an individual case basis, considering the capabilities of the intended user population. Chaffin (1991) provides an in-depth analysis of many whole body and limb segment related strength capabilities.

2.3. Psychological Factors

Psychological factors designing manual controls include those related to control location, arrangement, spacing, kinesthetic and tactile feedback generated, and logic as affected by interaction with the operation of other controls. In addition, there are factors related to the compatibility between a control and its associated visual display. Many design principles, guidelines and recommendations have been derived for a multitude of applications and operational settings. A number of general design principles have been discovered and developed over the years. These apply to a broad spectrum of applications and can, in some cases, be said to resemble laws in science. Boff and Lincoln (1988) compiled human factors engineering data compendiums covering many aspects of manual control, visual, auditory and tactile displays plus many other cognitive aspects of control. Psychological factors related directly to controls, their operation and feedback include:

- Control spacing and arrangement.
- Control operational logic (control–display compatibility, sequencing, directional compatibility.
- Control feedback: position, resistive force, surface tactile

features, shape, appearance, transillumination, acoustic signal, detent resistance, labeling.

- Population stereotypes (preferences of the user population derived from operational experience).
- Sanders and McCormick (1993: 301–33) list and illustrate rules for spatial and movement compatibility between controls and their associated displays.

2.4. Operational Factors

Operational factors relate to the means by which human control is conveyed to the machine or process. These include choices between discrete and continuously variable positions for controls, the use of linear versus rotary controls, the control–display movement ratio (C/R or control–response ration, formerly called the C/D or control–display ratio), the simultaneous or coordinated control of multiple units or functions, and predetermined control logic (preventing unacceptable or unsafe combinations of control settings, lockin and lockout procedures, proper control activation or deactivation sequences, time delays, and other programmable features).

3. DESIGN CRITERIA FOR SPECIFIC MECHANICAL CONTROLS

3.1. Classification of Controls

Manual controls can be classified according to any number of functional categories depending on the technical aspects of the machine or process being controlled. Specific and highly detailed classifications and spatial arrangements have been developed for some applications such as aircraft and nuclear power plants. In developing control devices for the operator in these and other high investment applications, strong interaction occurs directly through experimentation or indirectly through standardization, training of operators and licensing of designed systems among human factors engineers, licensed professional operating personnel and cockpit/control panel design engineers. The use of detailed mockups and simulators is typical in evaluating proposed designs and selecting the one judged to be the best in terms of operational reliability and effectiveness.

From an operational viewpoint, controls are classified by the nature of their machine function, limb usage in operating them and by physical structure and appearance. A given control may operate according to discrete or continuous or unlimited settings. For example, a small tractor could operate at various speeds by discrete selection of the appropriate gear while engaging a clutch and then releasing the clutch when in the desired gear. A hydrostatic transmission permits speed selection over a continuous range of speeds and requires only one forward motion on a control lever to set the forward speed. Backing the tractor is accomplished by setting the control lever behind the central neutral position at a position corresponding to the desired reverse speed. Any manual control capable of multiple settings to control a single variable over a defined range can be made to operate in continuous or discrete settings.

Controls will be classified in this section according to the following categories:

- LARGE LINEAR CONTROLS
- Pedals
- Levers
- Cranks
- Wheels (valve wheels)
- Steering wheels
- Yokes
- SMALL LINEAR CONTROLS
- Bars
- Slide controls
- Push–pull knobs and handles
- Push buttons
- SWITCHES
- Toggle
- Rotary
- Rocker
- LARGE ROTARY CONTROLS
- Cranks
- Wheels (valve)
- Steering wheels
- Yokes
- SMALL ROTARY CONTROLS
- Small cranks
- Thumb wheels
- Rotary knobs and dials
- L-handles
- KEYBOARDS
- Two-handed
- One-handed
- MOUSE
- TOUCH DEVICES
- Touch panels
- Touch screens and displays
- Membrane keyboards
- Light pens
- Pointers
- JOYSTICKS
- TRACKER BALL

3.2. Selection of Controls

Controls are summarized according to their use in Table 1 for the ease of discrete control settings. Examples of recommended functional uses for general classes of discrete controls are given in the right hand column. It is important in selecting controls to follow population stereotypes (preferences within a given application that has acquired a history of successful and favored control designs and arrangements). Population stereotypes often result in strong user habits that are likely to cause errors if control design is altered radically for a given function.

A guide for selecting controls whose settings can be made on a continuum permitting an unlimited number of positions to be selected over the range of control settings (hence the term "continuous controls") is given in Table 2 along with examples of recommended use. Both linear and rotary controls are included.

Controls should be selected or designed for functionality with aesthetic features or styling added later if desired, provided that these do not reduce operator performance. For example, a steering wheel for an automobile should be designed with proper consideration to turning force, grip strength and steering motions required in terms of its diameter and rim thickness before deciding upon the use of a wooden wheel or leather covering as classic material. Many functional criteria can be considered, depending on the application. A representative set of commonly used criteria

Table 1. Selecting discrete controls

Type	Users
Linear	
Pushbutton	Where a control or an array of controls is needed for momentary contact or for activating a locking circuit
Legend	Where an integral legend is required for pushbutton application
Slide	Where two or more positions are required or are arranged in matrix to allow easy recognition of relative switch settings (e.g., auditory levels across channels)
Toggle	Where two positions are required or space limitations are severe Three-position toggles used only as spring-loaded center-off type or where rotary or legend controls are not feasible
Rocker	In place of toggles where toggles may cause snagging problems or where scarcity of panel space precludes separate labeling of switch positions Three-position rockers used only as spring-loaded center-off type or where rotary or legend controls are not feasible
Push-Pull	Where two positions are required and such configuration is expected (e.g., auto headlights, etc) or where panel space is scarce and related functions can be combined (e.g., ON-OFF/volume control) Three-position push-pulls used only where inadvertent positioning is not critical
Rotary	
Selector	Where three or more positions are required In two-postion applications where swift visual identification is more important than positioning speed
Key Operated	In two-position applications to prevent unauthorized operation
Thumbwheel	Where a compact digital control-input device with readout is required

Table 2. Selecting continous controls

Type	Uses
Linear	
Lever	Where large amounts of force or displacement are involved or when multi-dimensional control movements are required
Isotonic (Displacement) Joystick	Where precise or continouse control in two or more relate dimensions is required
	Where positioning accuracy is more important than positioning speed
	Data pickoff from CRT or free-drawn graphics
Trackball	Data pickoff from CRT
	Where there may be cumulative travel in a given direction
	Zero-order control only
Mouse	Data pick-off or entry of coordinate values on a CRT
	Zero-order control only
Light Pen	Track-oriented readout device
	Data pick-off, data entry on CRT
Rotary	
Continous	Where low forces and precise adjustments of a continous variable are required
Ganged	Used in limited applications where scarce panel space precludes the use of single continous rotary controls
Thumbwheel	Used as an alternative to continous rotary controls where a compact control device is required

is presented in Table 3 along with a representative set of controls which may or may not meet a given criterion. It is important, therefore, that criteria be ranked or prioritized in a given application and that specific control selection is based on the prioritized criteria.

When many functions and controls are involved in control panel design, controls should be grouped according to function and arranged in sequential order of their use, when used sequentially, from left to right and from top to bottom. Importance and frequency of use should be used to decide which functional group should be located within the central visual and forearm reach areas and which ones should be located peripherally.

Table 3 lists some selected characteristics of controls, based on control design features discussed previously in addition to other control identification features along with ratings of commonly used discrete and continuous controls. It is important to note that every control has at least one fair or poor rating based on one or more particular characteristics listed. All controls listed also have at least one good rating based on a given characteristic. Many other characteristics and controls could be included in this type of table. The reasoning process involved in setting up such a table for a given application should be of value

in specifying or designing controls that will be effective functionally as well as ergonomically. Involving users in the design and selection of controls as through the use of mockups when final decisions need to be made can enhance operator compatibility even further. System users and operators can then be trained to operate a new system or become totally familiar with its controls before the system is introduced in an operational setting. This approach will eliminate or minimize design retrofit changes and complaints from system operators that "No one ever contacted us or asked our opinion. The designer has never had any operating experience."

Referring to Table 3, it is also important to note that the control itself provides operating feedback to the user through its design features, control setting and location. In this design aspect, the control serves as a tactile display indicating the status of operator commands to the system. Information from a control can be the result of several design features including its location, shape, size, mode of operation, labeling and color. Each of these forms of information or coding has advantages and disadvantages. These are summarized in Table 4. As with the comparison of controls in Table 3, no method of coding is without disadvantages or advantages. Depending upon the operating environment (e.g. use of gloves, poor lighting) tradeoffs can be made among the options available. In any case, simplification and standardization of controls and their operation are always important if errors are to be minimized.

Table 3. Comparison of common controls

| | Control Type | | | | | | |
| | Discrete | | | | Continous | | |
Characteristics	Rotary Selector	Thumb-wheel	Push-button	Toggle Switch	Continous Rotary	Thumb-wheel	Lever
Space requirement (location and operation)	medium	small	small	small	small to medium	small	medium to large
Likelihood of Accidental Activation	low	low	medium	medium	medium	high	high
Effectiveness of coding	good	poor	fair to good	fair	good	poor	good
Ease of visual identification of Control Position	fair to good	good	poor[a]	fair to good	fair[b] to good	poor	fair to good
Ease of Non-Visual Identification of Control Position	fair to good	poor	fair	good	poor to good	poor	poor to fair
Ease of Check Reading in Array of Controls	good	good	poor[a]	good	good[b]	poor	good
Ease of Operation in Array of controls	poor	good	good	good	poor	good	good
Effectiveness as Part of a Combined Control	fair	fair	good	good	good[c]	good	good

[a] Exception is when control is back-lighted and light comes on when control is activated

[b] Application only when control makes less than one rotation; round knobs must also have a pointer

[c] Effective primarily when mounted concentrically on one axis with other controls

Table 4. Advantages and disadvantages of various types of coding

| | Type of Coding | | | | | |
Advantages	Location	Shape	Size	Mode of Operation	Labeling	Color
Improves visual identification	x	x	x		x	x
Improves nonvisual identification (tactual and kinesthetic)	x	x	x	x		
Helps standardization	x	x	x	x	x	x
Aids identification under low levels of illumination and colored lighting	x	x	x	x	(when trans-illuminated)	(When trans-illuminated
May aid in identifying control position (settings)		x		x	x	
Requires little (if any) training: is not subject to forgetting					x	
Disadvantages						
May require extra space	x	x	x	x	x	
Affects manipulation of the control (ease of use)	x	x	x	x		
Limited in number of available coding categories	x	x	x	x		x
May be less effective if operator wears gloves		x	x	x		
Controls must be viewed (i.e., must be within visual areas and with adequate illumination present)					x	x

3.3. Recommended Design Features for Controls

3.3.1. Large controls

3.3.1.1. Foot pedals and switches

Recommended design criteria for foot pedals and switches are presented in Figure 1. Separation, as well as size and displacement, is important in avoiding accidental selection. Alternate shapes for foot switches are primarily for visual identification and familiarization. In terms of feedback, location provides the only mode of providing information in selecting a pedal or foot switch. In considering control resistance for any large control, it is important to limit control force to no more than 5% of maximum strength for continuous operation or 20%

when rest periods are at least ten times as long as holding time. Elastic resistance should be used in most applications with the pedal or foot switch returning to its neutral position when released. Depending on seat height the maximum pedal force should not exceed 89 N for high position seat resulting in a primarily downward stroke or 178 N for a mid-position seat resulting in a primarily forward (leg extension) stroke. If the seat is low enough to permit maximum leg extension force against a horizontally resisting pedal, the maximum allowable force based on strength should not exceed 800 N for male operators or 622 N for female operators.

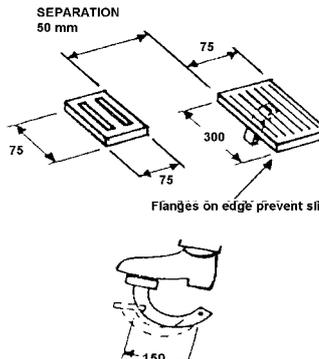For pedals requiring angular depression (pronation) of the foot, such as accelerator pedals, the resistive force at the ball

TYPICAL AND ALTERNATIVE SHAPES RECOMMENDED DIMENSIONS (mm)	FORCE (kg) MIN MAX	DISPLACEMENT (mm) MIN MAX	NOTES

FOOT PEDALS
LEG ACTUATED (BRAKE) TYPE

1-5   9

NB approx 30% reduction
in force 400mm below seat

Sitting operation for
foot pedals

Standing permissible if pedal
at floor level or
infrequent, short
duration operation.
Provide support for
weight of leg
(return spring)

ANKLE ACTUATED (ACCELERATOR) TYPE
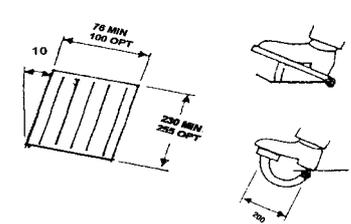
1-5   5-4
(1-5 kg is spring
return pressure)

Max force with knee at 160^0

Leg forward of hips
Pivot at heel or in front of toe

Max accuracy with knee
$95\text{-}135^0$ and ankle $85\text{-}110^0$

Pivot between toe and arch

Higher accuracy and
operating speed for
close spacing

FOOT SWITCHES
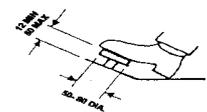
Return spring 8 kg
for standing
operation

Standing operation
permissible

Treads avoid slip

ALTERNATIVE SHAPES
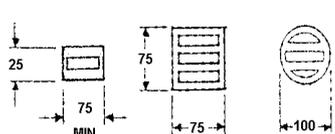
Figure 1. Foot pedals and foot switches. Distances in mm, forces in Kp (1 Kp = 9.8N)

position on the foot should not exceed 44 N. The recommended angle between a foot rotated pedal and the floor varies with seat height. It is typically 20° or less for higher seats as on highway trucks. It should not exceed 35° for a seat height of 43 cm or 45° for a very low seat height of 15 cm or lower.

Recommended foot pedal stroke length varies with the type of vehicle or hardware being controlled. For pedals that require displacement of the foot by leg extension, a displacement range between 5 and 18 cm could be preferred. Larger displacements (10–18 cm) are desirable for brake pedals to provide kinesthetic feedback through leg extension since pressure feedback from the foot can be largely masked by footwear and proper displacement is essential for safe braking especially under slippery road conditions. Automatic brake adjustment systems on power assisted brakes provide an alternative to providing this amount of feedback to the operator.

For an operator who is normally wearing heavy work shoes, boots or snow galoshes, additional pedal stroke length (1.0–2.5 cm) should be provided to compensate for the additional sole thickness. Foot pedals should be approximately the same width as the sole of the shoe, typically 8–10 cm. Pedal separation should be at least 5 cm edge to edge. Pedal shape is not a significant factor except in providing initial visual orientation regarding location, provided that the area is sufficient for rapid location, stable contact and reliable operation. Pedals used occasionally or for short working intervals should have a length of at least 7.6 cm. Pedals used continuously or for long periods should be between 28 and 30 cm in length.

Pedals operated by leg extension should have an elastic resistive force within an overall range between 35 and 270 N. For pedals operated by foot pronation (forward rotation) with the heel fixed, elastic resistive force should be within a range from 16 to 23 N. Foot switches should have a displacement between 2.5 and 6.5 cm for ankle rotation only and between 2.5
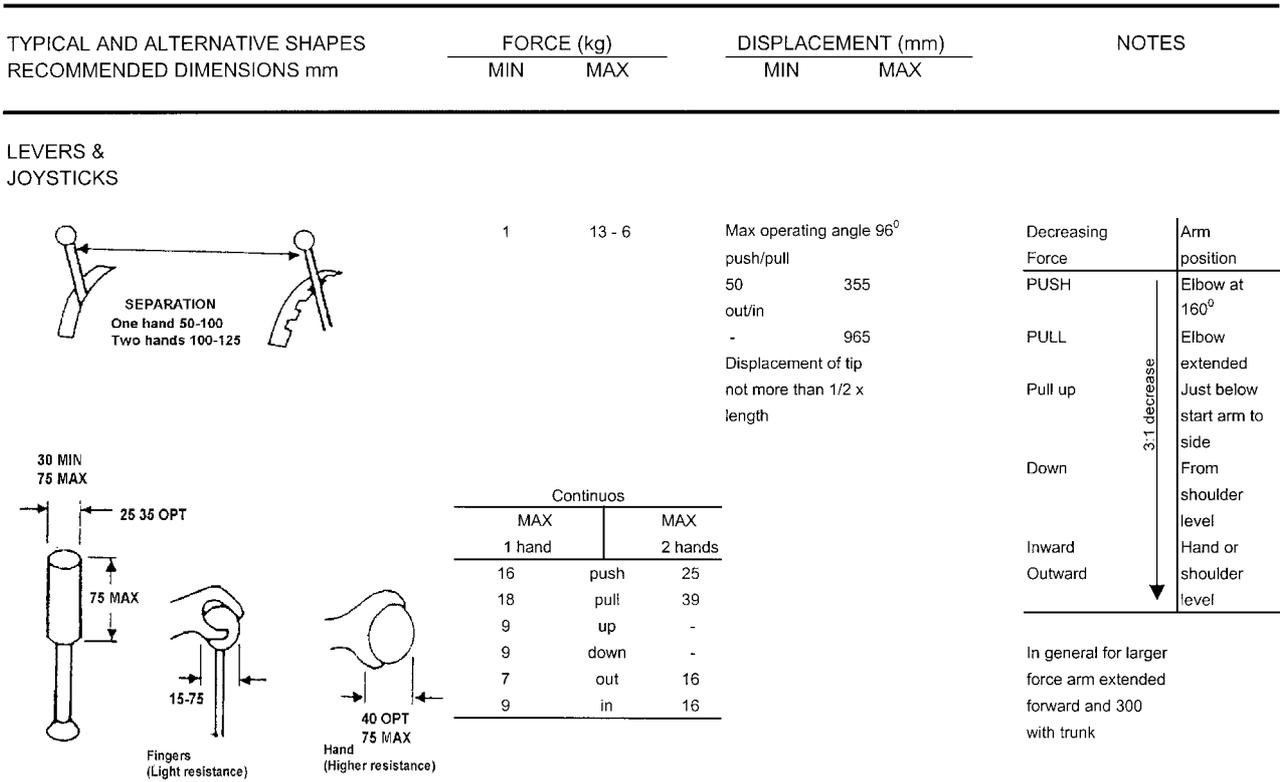
TYPICAL AND ALTERNATIVE SHAPES RECOMMENDED DIMENSIONS mm	FORCE (kg)		DISPLACEMENT (mm)		NOTES
	MIN	MAX	MIN	MAX	

LEVERS & JOYSTICKS

	FORCE (kg) MIN	FORCE (kg) MAX	DISPLACEMENT (mm)	NOTES
	1	13 - 6	Max operating angle 96^0 push/pull	
			50 355 out/in	
			- 965 Displacement of tip not more than 1/2 x length	

Continuos

MAX 1 hand		MAX 2 hands
16	push	25
18	pull	39
9	up	-
9	down	-
7	out	16
9	in	16

Decreasing Force	Arm position
PUSH	Elbow at 160^0
PULL	Elbow extended
Pull up	Just below start arm to side
Down	From shoulder level
Inward	Hand or
Outward	shoulder level

3:1 decrease

In general for larger force arm extended forward and 300 with trunk

Fingers (Light resistance)   Hand (Higher resistance)

Figure 2a. Levers and joysticks. Distances in mm, forces in Kp (1 Kp = 9.8 N)

HORIZONTAL

or

ALTERNATIVE SHAPES

For high speed and accuracy hand close to and in front of body

Handgrips Straight grips for displacement < 30^0

Ball grip for > 30^0

Overlap of finger and thumb for stronger grip
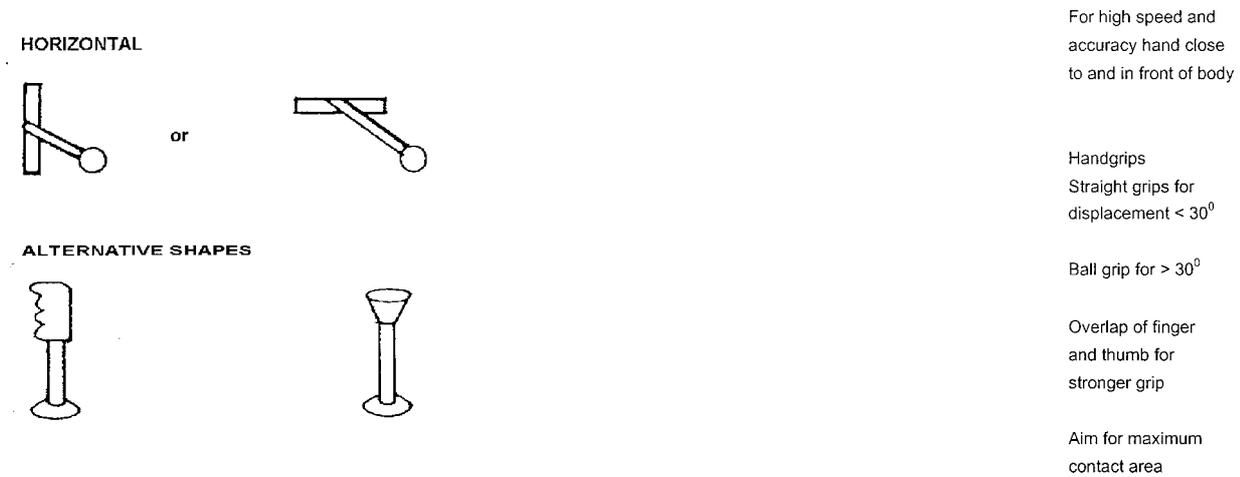
Aim for maximum contact area

Figure 2b.

and 10 cm when leg extension occurs in their operation. Their elastic resistance should not exceed 90 N and should be at least 18 N when the foot is not normally on the control or 45 N when the foot normally rests on the control.

3.3.1.2. Levers

Lever action controls occur in many forms. They are used for shifting gears, operating all types of large construction equipment, cranes, forklifts, farm machinery, aircraft and many types of industrial machines. The knob at the end of a lever can be shape-coded or labeled in many cases to facilitate identification with a particular function. A knob in the shape of an end rounded cylinder or ball is preferable when a firm or prolonged grip is required. Recommended ends diameters are 3.2 cm for full grip or 1.3 cm for a fingertip grip.

Levers can operate by discrete displacement settings as when detens are used to define control position as with a gearshift or by continuous positioning over a defined range such as a throttle lever.

A joystick is a special type of lever operable over a continuous range including combinations of front–back and side-to-side motions as if it were attached to a ball. Joysticks can also respond to force only and remain stationary. A joystick may have a large full-grip handle on which can be mounted other controls such as
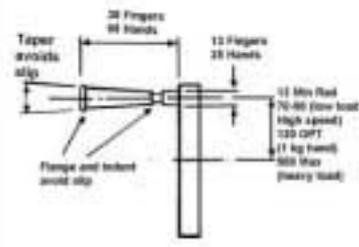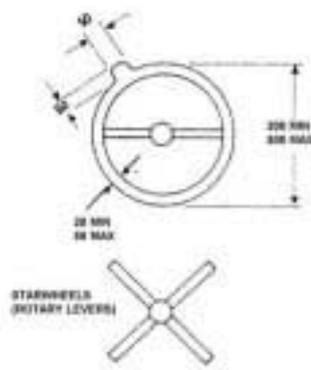
TYPICAL AND ALTERNATIVE SHAPES RECOMMENDED DIMENSIONS (mm)	FORCE (kg)		DISPLACEMENT (mm)		NOTES
	MIN	MAX	MIN	MAX	

CRANKS

	FORCE (kg)		DISPLACEMENT (mm)		NOTES
	Peripheral forces Rapid setting		SIZE (rad)	SPEED rpm	180 mm radius endured longer than smaller crank.
	0.9	2.5	70-90 (rapid)	120	
	2	4	500	200	
	Precise setting				Whole arm action more powerful than wrist/elbow
	-	3.5			
	Friction (Continuos)				Speed higher for vertical than horizontal
	0.9	2.2			
	(infrequent)				Accuracy higher for large radius, small displacement
		4.5			

HANDWHEELS

	FORCE (kg)		DISPLACEMENT (mm)		NOTES
					Good grip needed
	One Hand		Provide support (return spring) for weight of arm		
		2	13		Smaller speed with
	Two Hands	25			vertical wheel
	Largest force: horizontal, short vertical shaft, small displacement		Displacement not greater than +/- 600 from normal position unless hands can be removed in turning		
	Smaller force: shaft facing operators large displacement		Up to 3 revs without knob.		Less visual obstruction with 2 spokes
		20-23	5 revs with knob.		

Figure 3. Cranks and handwheels. Distances in mm, forces in Kp (1Kp = 9.8N)

triggers and pushbuttons. Joysticks are used widely in helicopters and in military aircraft. Human factors relating to joysticks are similar to those for other levers with regard to control of joystick shaft position. Recommended design criteria for levers and joysticks are shown in Figure 2. Maximum recommended forces in Kp (1 Kp = 9.8 N) are shown as related to the direction of application. The type of lever selected should also be related to the type of grasp that will be used. For a finger grasp only, the handle (or knob) diameter should be above 1.3 cm and below 7.5 cm. For a hand grasp, the handle diameter should be between 3.8 and 7.5 cm. Optimum recommended diameters are between 2.5 and 3.5 cm. The type of resistance force depends upon the application and the type of feedback and speed of operation desired.

As a general guideline, for a finger grasp, resistance should be between 3.3 and 9.0 N and for a hand grasp, between 9 and 90 N. An all purpose upper limit is difficult to specify since a large lever could be operated using two hands with push and pull forces up to 450 N. The maximum lateral (right–left) resistance for a one-handed lever operation should be no more than 90 N and for a two-handed lever operation no more than 130 N.

3.3.1.3. Cranks and handwheels

Cranks are effective as controls when continuous adjustments requiring multiple rotations must be made. Attainable turning rate increases with decreasing crank radius (arm length) and decreasing resistive torque. Cranks can also be attached to large rotary knobs or handwheels to facilitate large rapid changes in control setting. Cranks can be coded for identification by their location and color and also by labeling. While cranks can be designed for either one- or two-handed operation, handwheels are designed for two-handed operation. Handwheels permit greater torque to be exerted than cranks but are much slower for multirevolution movements. Recommended design criteria for cranks and handwheels are summarized in Figure 3. For small cranks (< 9 cm in radius) requiring rapid movement, the resistive force to be overcome in turning the crank or initiating turning should be at least 9 N but no more than 22 N. For large cranks (13–20 cm radius) requiring rapid turning at a relatively constant speed, the resistive force at the crank handle should be between 13 and 25 N. A crank should turn freely on its shaft. The handle surface should provide a high coefficient of friction for the hand or glove. The handle diameter should facilitate a firm grip.

Handwheels should facilitate two-hand control and permit rapid movements as well as accurate movements when needed. If rotations > 60° are required on a handwheel, the hands must be repositioned. Wheel displacement should be determined by considering the desired control/response or C/R ratio. For example, when the handwheel must be rotated on a large arc increasing the handwheel diameter can increase the C/R ratio. The thickness diameter of the handwheel diameter should not be less than 1.9 cm or more than 3.2 cm.

3.3.2. Small controls

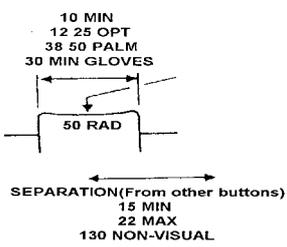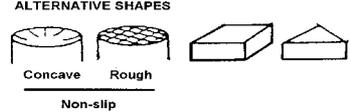Small controls may operate by either linear or rotary motion.

TYPICAL AND ALTERNATIVE SHAPES RECOMMENDED DIMENSIONS (mm)	FORCE (kg)		DISPLACEMENT (mm)		NOTES
	MIN	MAX	MIN	MAX	

PUSH BUTTONS

| | 0 - 25 (Frequent 0 - 8) | 1 - 2 | 3 | 16 40 (Ball hand or gloves) | Size, shape and spacing depend on the consequence of error and operator skill. |

10 MIN
12 25 OPT
38 50 PALM
30 MIN GLOVES

50 RAD

SEPARATION (From other buttons)
15 MIN
22 MAX
130 NON-VISUAL

Start with low resistance, build up, and then drop with click

Keyboards
0 - 15   0 - 3

For a sloping plane vertical buttons are preferred to buttons at 90^0 to the panel.

May include built-in illumination.

May require colour coding. See BS 4099(1986) Part I.

ALTERNATIVE SHAPES

Concave   Rough
Non-slip

Barriers or recesses can be used for separation

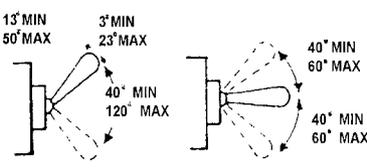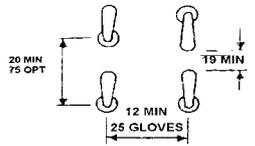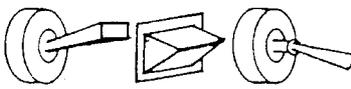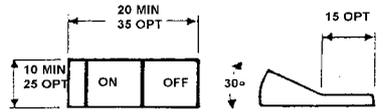Figure 4a. Small controls. Distances in mm, forces in Kp (1 Kp = 9.8N)

TOGGLE SWITCHES

13° MIN   3° MIN
50° MAX   23° MAX

40° MIN   40° MIN
120° MAX   60° MAX

40° MIN
60° MAX

0 - 25   10 40^0 between settings

Spring loaded near the central position. Resistance to build up, then fall off. No resting between settings.

2 SETTINGS
ASSEMBLIES

20 MIN
75 OPT

19 MIN

12 MIN
25 GLOVES

3 SETTINGS
UK CONVENTION

OFF
OFF ⊙ ON
ON

ALTERNATIVE SHAPES

ROCKER SWITCHES

20 MIN
35 OPT   15 OPT

10 MIN
25 OPT   ON   OFF   30°

Can be used instead of toggles. A legend is possible

Figure 4b. Toggle switches and rocker switches

Linear controls include push buttons, push–pull knobs and slide controls. Rotary controls include thumb wheels, rotary knobs, L- and T-handles, and rotary switches. Controls that rotate partially in making discrete settings include toggle switches and rocker switches. Examples of recommended design criteria for small controls are given in Figure 4a. There are many ways of coding small controls including shape, color, surface knurling, size, location and labeling. Knurling or fluting on thumbwheels also helps to prevent hand slippage when rotating a control. A dot or line can be used to indicate the "on-off" position. Rotary
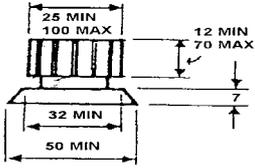
825

TYPICAL AND ALTERNATIVE SHAPES RECOMMENDED DIMENSIONS (mm)	FORCE (kg)		DISPLACEMENT (mm)		NOTES
	MIN	MAX	MIN	MAX	

ROTARY SELECTOR
(KNOB OR ROTARY BAR)

KNOB

A skirt gives a more visible scale
and prevents damage to scale.

Bar knobs are preferable for
checking settings in an array

30^0 detents can be
identified by feel

ROTARY BAR
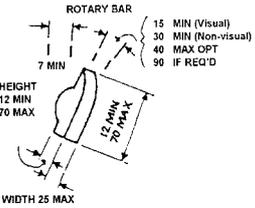
ALTERNATIVE SHAPES
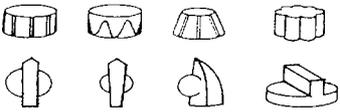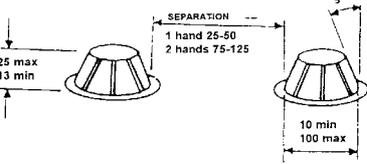
| 0.44 Nm (torque) | Unlimited (depends on operating time) | Unsuitable for rapid adjustment unless folding crack handle. |

Higher torque if
knurled or larger size.
Serrated edges for
grip.

Figure 4c.

STACKED

HIGH TORQUE (5 Finger grab)

FOLDING CRANK HANDLE
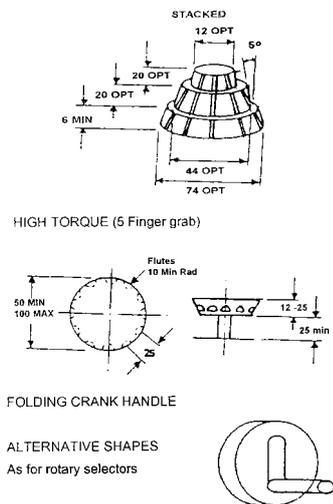
ALTERNATIVE SHAPES
As for rotary selectors

Figure 4d.

controls that produce a similar tactile sensation but correspond to different functions should not be grouped together. Toggle and rocker switches should be clearly labeled.

Push buttons controlling multiple functions or producing different operating conditions for the same machine or process can be coded by logical combination to prevent undesirable sequences or to lock out the system for safety or security. Many patterns of operating logic can be used. Examples include: momentary row lockout, clearing or neutralizing all functions, mutually exclusive actuation of functions, and singular fixed sequential operation.

All small controls should have a resistance force or torque sufficient to prevent accidental actuation or change in setting. For rotary selector switches and knobs, for example, a resistive force between 3.3 and 13 N at the outer edge or rim of the control is desirable. Control separation should also permit operation without accidental actuation of an adjacent control. For most applications, edge-to-edge control separation should be at least 5 cm for random selection or 2.5 cm for sequential selection. Allowance for the use of gloves will increase these values. Resistance resulting from friction or control inertia should be minimized.

When controls have discrete settings, elastic resistance should increase to a maximum and then cease as the control snaps into position. The recommendation applies to rotary selector switches, rocker switches, toggle switches and rotary L- and T-handles. If ambient noise is present, an audible click is recommended to indicate that the setting has been made. Click count also provides secondary feedback to the tactile sense of control position.

A series of toggle switches should be arranged horizontally rather than vertically. If toggle switches must be arranged vertically, larger spacing between switches is necessary to minimize accidental actuation.

3.3.2.1. *Direction of control movement*

To prevent control reversal errors, it is important for controls to operate in directions that are compatible with associated display or vehicle movement and operator expectations based on experience e with controls on and off the job. Factors to be taken into account include:

1. Location and orientation of the control relative to the operator.
2. Position of the display in relation to the control and the direction or system status change shown in the response indicated on the display.
3. Orientation of the operator relative to the controlled system or vehicle.
4. Type of action being caused by activating the control.

Many recommended control movements now have the status of being universal standards for nearly all applications. A representative set of common control directional recommendations is given in Table 5.

When it is necessary to locate controls and their associated displays in different orientations, right-to-left should correspond positionally with top-to-bottom. If several rows of controls are located beneath several rows of displays on a panel, the rows should be arranged similarly. Thus, displays 1-2-3-4 on a top row and 5-6-7-8 on the second row underneath should relate to respective controls 1-2-3-4 directly under the second display row and 5-6-7-8 located under the first control row. Another effective design, except for visual interference of some displays by the hand while making control adjustments, would locate each of the eight controls directly under its associated display. If controls are stacked vertically as concentric rotary dials (maximum of three with the largest on the bottom and smallest on top), their associated displays should correspond from left to right with the dials top to bottom. Rotating pointers and knobs with fixed surrounding numerical scales are preferable to fixed surface pointers and numerals on the rotating knob.

When several planes of motion are required for control panels, the relative positions, orientations and directions of control movement should be the same. That is to say that if the operator were to face a panel identical to one in a desktop position in front of him/her, located in a different place the positions, orientations and movements would be the same.

3.3.3. Keyboards and keypads

3.3.3.1. *Keyboards*

Alphanumeric keyboards serve a great variety of needs in providing data input as well as command and control of machines and processes. The standard QWERTY keyboard has dominated for many years as the chosen option for letter arrangement of keys despite the fact that this arrangement did not have its origin in ergonomic experiments or user surveys. Because of its extremely wide usage and standardization as a population stereotype throughout the western hemisphere and all countries where English is the standard language for commerce, finance and industry, the QWERTY arrangement will probably dominate for many years to come. Other arrangements, for example those using chord or combinatorial keying, could become popular as small computers become more powerful and versatile. Coupling the QWERTY keyboard with the computer has expanded the original typewriter keyboard with a top row of special function keys, cursor controls (keys or trackball) and a numeric keypad located on the right side. Function keys are sometimes placed on the left side.

The QWERTY keyboard arrangement had its origins in the 19th century when it was used to record telegraph messages and jamming of the mechanical typewriter (caused by collision and binding together of letter printing levers carrying the letter imprint pads) could be very disruptive as the levers were being manually

Qwerty Keyboard

Dvorak Simplified Keyboard
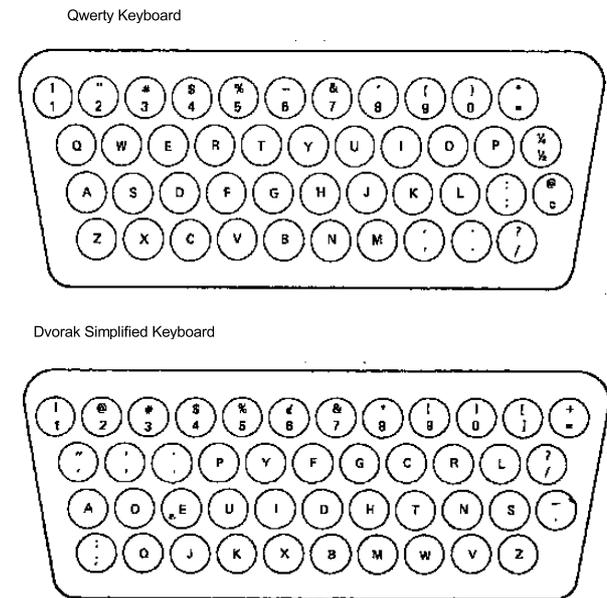
Figure 5. Current QWERTY and simplified keyboard arrangements. The simplified keyboard is specified in American National Standards Institute standard ANSI X4.22-1983 and was derived from the Dvorak keyboard. Source: Sanders, MS and McCormick, EJ, 1993, *Human Factors in Engineering and Design* (7th ed.) (New York: McGraw-Hill Inc.) p. 361, reproduced with permission of the McGraw-Hill Companies.

NOTE: The original Dvorak keyboard had numeric keys across the top row in the following order, from left to right: 7-5-3-1-9-0-2-4-6-8.

Table 5. Recommended control direction movements. Adapted form Woodson, WE, Tillman, B and Tillman, P, 1992, *Human Factors Design Handbook* (2nd ed.) (New York: McGraw-Hill Inc.) p. 433, reproduced with permission of the McGraw-Hill Companies

Action Desired	Control Movement
Turn on	Up, right, forward, press inward, clockwise
Turn off	Down, left, rearward, pull outward, counterclockwise
Turn or move right	Right, clockwise
Turn or move left	Left, counterclockwise
Move upward	Up, rearward
Move downward	Down, forward
Retract	Rearward, pull, counterclockwise, up
Extend	Forward, push, clockwise, down
Increase	Right, up, forward, clockwise
Decrease	Left, down, rearward, counterclockwise

separated. Jamming often occurred when typing was done too rapidly, thus resulting in a need for a keyboard that would reduce typing rate. As technology improved and the typing ball and word processor were developed, the QWERTY arrangement remained as a standard design. Under the QWERTY arrangement 57% of letters typed are performed with the left hand for typing in English. About 80% of the user population is right-handed. The impact of this imbalance in hand assignment upon overall productivity is reduced when the right hand can also be used alternately for writing or other tasks.

Dvorak (1943) developed an alternative to the QWERTY arrangement. The QWERTY and simplified Dvorak arrangements are shown in Figure 5. The Dvorak arrangement, designed according to hand and finger capabilities has been shown experientially to result in somewhat higher productivity, but not enough to justify a widespread change to the system along with the need for retraining. Norman (1983) concluded that a 5–10% improvement in productivity is likely. It is possible, but not yet popular, for keyboard manufacturers to offer an optional Dvorak arrangement in their product lines.

To improve productivity in the use of keyboards, there are issues that are more important than key arrangement. These include multifunction keys, chording for common letter combinations (ion, ing, ary, ory, ere, the, an, and, ph, est, qu), and automatic return. Rochester *et al.* (1978) designed a small keyboard based on chording. Their design objective was to create one hand typing capability similar to typing with a stenotype keyboard. Their design included a thumb key plate that could be moved laterally for right or left hand preference. This special keyboard consisted of two rows of five keys each, operated using the first three fingers and an additional set of four keys operated by the thumb. The keys included special raised tactile dots used to press one, two or four keys. Up to three dots could be pressed at the same time, thus forming a chord as on a piano. The designers conducted experiments using several typing rates and found that with training, typists could increase their rates by 10% using this chord keyboard. The one-handed operation permits other activities to be done with the other hand, provided that information overload does not occur. Chord keystroke data entry has been used successfully by post offices to sort letters in the USA, Canada and UK. Earlier experiments with chorded keyboards suggest that it may be possible to increase data entry rates by as much as 50% over standard rates using chording. This could justify training costs for many applications with specialized dedicated information systems.

There are also keyboards with several functions for each key. When this feature is used, it is generally restricted to the top row of keys. This feature is useful when the keyboard is used for process control. Typically a row of unmarked keys is placed under the monitor screen. Each key is represented by a small square on the screen. Text corresponding to the functional operation being performed appears in separate windows showing the key functions for a given frame on the screen. When several rows of unmarked keys are used, corresponding pictures can be projected from the screen onto the keyboard using mirrors to show the current key functions. Since fewer keys can be used on this type of keyboard, fewer hand and arm motions are needed. Errors also tend to be reduced. Another method involves providing lights inside the keys so that they light up for each individual function

when the appropriate key is pressed. Such a keyboard provides feedback that can cue and guide the operator through a sequence of operations. Non-illuminated keys cannot cause errors if inadvertently pressed since they are deactivated. This results in a very low error rate in addition to a higher data entry rate. A warning signal is needed when bulbs in the keys fail.

Pressure sensitive keys are also possible. For example, light pressure on key could cause the function to be displayed on the monitor whereas heavy pressure could result in immediate action being taken. Maximum key depression could signal needed changes in the process. When this happens, a display on the monitor indicates a new set of keys.

Pre-programming can virtually eliminate sequential errors in keyboard commands to a process. This has advantages in terms of system safety, but operators may resent being over controlled. Programming requirements can be extensive and may utilize more computer capacity than is feasible, considering other real time process control demands. The capabilities and flexibility of a standard keyboard can be greatly enhanced through pre-programming, providing many options in the type and level of control over the process.

When a standard keyboard is used and the hands are placed near the center of the keyboard, ulnar deviation (outward lateral bending of the wrist) occurs, causing stress in the wrist joint. Cumulative effects of this stress over time have produced cumulative trauma disorders such as tendonitis. Splitting and angling the keyboard appropriately for the given text eliminates this source of stress. Rotating the two halves of the keyboard inward from the edge of the desk or work table (outer ends in) is known as slanting. Lowering the outer ends of each half below the work table surface is referred to as tilting. Elevating the rear (forward) end of each half is referred to as sloping. Adjustability of these angles is desirable to accommodate a variety of tasks and operator preferences. Ilg (1987) recommends a 30° symmetrical slant separation angle between the halves of the keyboard and tilting of the outer edges of the keyboard halves to a depth of 5.3 cm. The slope angle of the keyboard can be between 0 and 15°. If the operator is standing, the slope should be between 0 and −30° (raised on the front edge to reduce wrist extension).

The size of the key surfaces must meet several criteria. First, the area of the keyboard should be kept as small as possible to minimize hand travel distances. Second, the key surface area and separation distances between keys must be adequate to accommodate the fingertips and prevent simultaneous key depressions. To fulfill these requirements, a square keytop surface 12–15 mm in length is recommended. For shop floor use or when gloves are worn by the operator, large key sizes and spacing are needed. Ordinary spacing between keytop centers has been standardized at 19.5 mm, satisfying the ergonomically derived recommendation of 18–22 mm.

The key depression resistance force should be common for all keys between 0.25 and 0.50 N with a displacement of 0.8 to 1.0 mm for skilled operators. For unskilled operators more feedback is needed. This can be accomplished by providing a resistance force between 1 and 2 N and a key displacement between 2 and 5 mm.

Feedback in keyboard operation occurs as tactile (touch) or contact feedback and also as kinesthetic feedback from finger muscle activity. In addition, there is visual feedback from the

keyboard or monitor screen display as well as auditory feedback from the keyboard. For skilled operators, the auditory feedback is not important. For operators who are learning or unskilled, auditory feedback is needed. It should be possible for operators to remove auditory feedback if it is not desired.

A contrast ratio of 3:1 should be provided for labeling keys. Coloring is normally not important. If the associated monitor displays are light on a dark background, the same relationship should exist for the keys. Keys should not reflect off the monitor screen. Matte finishes should be used whenever possible. Letters and numbers on keys should be at least 2.5 mm. Black lettering on white keys is recommended. The International Electrotechnical Commission (1975) produced color standards applicable to keys as well as illuminated information. These standards are recommended for use in keyboard design.

3.3.3.2. *Keypads*

Keypads are used when only numeric digital entry is required. The most common keypad entry device is used on the touchtone telephone. Other common applications occur in calculators, adding machines (printing calculators used in accounting and bookkeeping). Some keypads are included on keyboards as supplementary data entry devices. There are two accepted standard designs for keypads, one for the telephone and one for calculators. These are shown in Figure 6.

Both standard keypad layouts have three rows with three keys per row plus the zero button under the bottom row and in the center. The telephone layout obeys the ergonomic sequence principles of left to right and top to bottom. Conrad and Hull (1968) found the telephone arrangement to be slightly faster but with a more significant advantage in accuracy (8.2 versus 6.4% with the calculator arrangement). Alternating between the two keyboard arrangements caused an additional decrease in accuracy. The telephone layout is likely to be the standard for all telephones. It is recommended, from an ergonomics perspective, that all calculators and calculating machines, as well as keypads for numerically controlled machines, be fitted with this arrangement to prevent the transfer errors noted earlier.

3.3.4. Touchscreens

A touch screen is a keyless matrix-based touch-sensitive surface that overlays either a computer screen or projected beams (infrared or acoustic) that are interrupted when surface contact is made. A contact based screen typically has a thin metal net containing

Design of controls

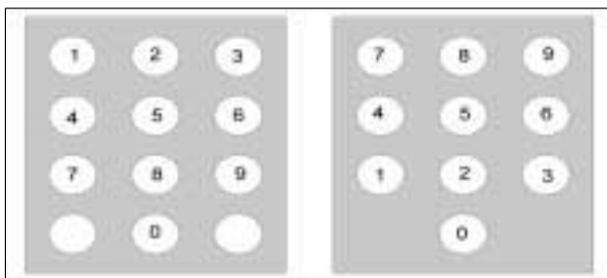| Push-button telephone layout | Adding machine layout |

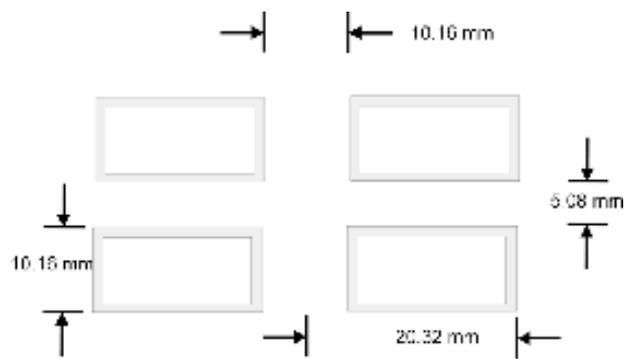**Figure 6. Layout of numerical keyboards**

Figure 7. Dimensions of touch-screen keys preferred by subjects and yielding the smallest number of errors. Source: Sanders, MS and McCormick, EJ, 1993, *Human Factors in Engineering and Design* (7th ed.) (New York: McGraw-Hill Inc.) p. 369, reproduced with permission of the McGraw-Hill Companies.

electrical circuits that are activated when the net is touched. Some designs use a transparent pressure sensitive material with special measurement bridges that detect variations in pressure distribution.

Pointing with the finger on a touch screen is often faster than pointing with a light pen, depending on target size and selection factors. The reliability of the touch screen is reduced by inaccuracy of pointing. Beringer and Peterson (1985) found that people tend to contact a touch screen between 3.2 and 6.4 mm below the intended target image. In their study, accuracy was highest at the bottom of the screen and lowest at the top. Reasons for this could possibly be attributed to excessive movement of the arm resulting from its resting position and from parallax in viewing the screen.

The size of the key targets to be touched on the screen also affects performance. Beaton and Weiman (1984) investigated the effects of horizontal and vertical target dimensions as well as target separation. Only the vertical dimension affected performance in terms of the number of errors. Larger target keys resulted in better performance than smaller target keys. The optimum target dimensions and arrangement resulting in the fewest errors and also preferred by subjects who participated in the study are shown in Figure 7.

The touch screen has an advantage over the traditional keyboard in its ability to present limited selections to untrained users. If messages or commands must be varied, the traditional keyboard should be used. Pfauth and Priest (1981) summarized the operational advantages and disadvantages of touch screen devices as shown in Table 6.

Factors influencing the acceptance of touch screen devices by users are summarized in Table 7.

The advantages to be realized from using touch screen devices depend upon the net gain achieved through simplicity of operation and reduced training over the loss of flexibility in user–computer dialogue and the additional dedicated design and programming.

3.3.5. Light pens

As suggested by its name, the light pen is shaped like a ballpoint pen and operates using a photocell to detect light radiated from a computer monitor screen. An electrical cord attaches the light pen to the computer. When the pen is pressed against the screen,

Table 6. Advantages and disadvantages of touch screen devices (from Pfauth and Priest, 1981)

Advantages

Direct visual-to-tactile control (input/output to one location)
Fast data entry and control for certain tasks
Minimal training (software flexible for different jobs and operator skill levels)
Only valid options are available
High operator acceptance
Immediate feedback
Symbolic graphic representation
Minimal operator memorization required
Minimal eye-hand coordination problems

Disadvantages

Potentially high initial cost
Increased programmer time
Not as flexible for some types of input (unless used in combination with keyboard)
Parallax affecting touch locations
Screen glare
Physical fatigue from reaching to screen
Finger visually blocking the screen
New method of programming interface software

Table 7. Factors in the acceptance of touch screen devices (from Pfauth and Priest, 1981)

Potential Users

Those in extremely high-stress enviroments
Completely naive users
Users who have a predetermined limited interface with the computer syster (e.g., management personnel)
Where available workspace does not allow a separate keyboard

Advantages

No prior training needed; no need to learn how to use a keyboard or memorize commands
Dialogue technique can guide system user toward the desired information
The user is likely to view the system as a partner in the problem-solving process
User does not need to understand the specifics of how the system works

Disadvantages

Data entry through the TSD will be considerable slower than through the traditional keyboard
High cost to implement
Flexibility of system is dependent upon software capability, i.e., limited number of display options

it senses the cathode ray tube (CRT) scanning beam as the beam illuminates the pixel directly under the pen. In this way the light pen serves a similar function as the touch screen. Since the computer receives the screen location coordinates of the pen from the screen and a simultaneous signal also passes from the pen through the cord into the computer, the computer can track the position of the pen on the monitor screen. The light pen can be used to query menu items or cause animated variation of a display such as opening or closing a valve. The light pen can also be used to move a cursor on the screen but it appears to have no advantage over other methods such as a mouse in this role.

Biomechanical fatigue can occur during prolonged use of a light pen. A properly designed armrest will prevent stresses causing this fatigue and its associated discomfort.

Since the light pen must be held close to the monitor screen a reflective shield or light filter cannot be placed on the screen. Thus, glare from outside light sources can present a problem. A light pen also requires accurate positioning. This can result in high visual demands and poor working postures. Care should be taken to assure that the entire workstation is designed for use of a light pen when this option is chosen.

3.3.6. Mouse and trackball

A mouse is a small hand positioned and hand-operated device typically 9–10 cm in length, 5–6 cm in width and ~3 cm high containing two-to-three control buttons, used to manipulate and select menu items and operations on a computer monitor screen. The mouse contains a built-in trackball on its lower side that rolls on a rubberized pad to produce corresponding cursor movements on the monitor screen. The mouse is an extremely popular data selection and manipulation device for desktop computers. The control/response (CR) ratio can be changed for selected applications or can be made velocity dependent instead of displacement dependent. From an ergonomics perspective, the mouse is generally very user friendly. As with any prolonged muscular tension, continuous intense gripping of the mouse can cause biomechanical stress and fatigue in the hand. The mouse is not suitable for drawing since it is moved by the wrist and arm whose extrinsic muscles lack the fine control of the intrinsic muscles located in the hand which control fine finger movements. When computer-aided drawing is desired, a graphics tablet (or digitizing tablet) should be used. These receive input from finger movement or from a stylus on the surface of the tablet.

The trackball is a small sphere inserted into a round hole in the keyboard or separately on a table using a separate mounting. Rolling the ball with the fingers effects two-dimensional positioning of a cursor on the computer monitor screen. There is a direct proportional relationship between the surface displacement of the trackball and the motion of the cursor on the screen. A joystick with two-dimensional movement is sometimes used in place of a trackball.

There has been no conclusive evidence favoring the mouse, trackball or joystick for controlling cursor movement. In general, the mouse is preferred whenever a clear space exists for the contact pad. Otherwise, the trackball is generally preferred for confined operating space.

4. RECOMMENDATIONS

Design of a control system and its associated human-operated controls should proceed according to the following sequence:

1. Design the overall system through interactive meetings between the engineering designer and the system user/operators and determine tasks to be performed by machines and tasks to be performed by people.
2. Define carefully these factors and system variables that must be controlled.
3. Consider human workload in terms of information processing and control operation. Determine task intensity and duration requirements and crew size. Determine necessary and desirable human functions in system operation.
4. Modify human and machine functions according to system performance requirements and human physical and psychological limitations.

5. Select controls and design control panels and operating stations according to ergonomic guidelines, operator preferences and accepted stereotypes. Stereotypes should not compromise safety or operational effectiveness.

6. Design the control/display layout and the human–machine interface environment (control cab, cockpit, workstation). Arrange controls according to function and by sequence within function. Locate important functions closer to the operator.

7. Consider anthropometric requirements for the user population. Include allowances for clothing, and movement within, entering and exiting the control area.

8. Design and simulate operator workstations using computer graphics. Build mockups for new workstations or control involving new technology.

9. Conduct operational simulation using ergonomic human operator software and/or conducting user testing in a mockup to reveal potential operating problems. Modify the design developed in step 8 according to the operational information obtained in step 9.

10. If warranted by technological complexity, time and cost limitations and designer/user agreement, develop and build a dynamic simulator for critical control stations to investigate time dependent human–machine control functions and relationships. Modify system hardware and software as needed to solve problems encountered in using the simulator.

11. Build the design prototype control station. Involve the user/operators in final design evaluations and changes.

REFERENCES

BEATON, R. and WEIMAN, N., 1984, *Effects of Touch Key Size and Separation on Menu-Selection Accuracy*. Technical Report no. TR 500-01 (Beaverton: Taktronix, Human Factors Research Laboratory).

BERINGER, D. and PETERSON, J., 1985, Underlying behavioral parameters of the operation of touch-input devices: biases, models and feedback. *Human Factors*, 27, 445–58.

BOFF, K.R. and LINCOLN, J.E., 1988, *Engineering Data Compendium* (OH: Wright-Patterson Air Force Base, Human Engineering Division, Armstrong Aerospace Medical Research Laboratory).

CHAFFIN, D.B. and ANDERSON, G.B.J., 1991, *Occupational Biomechanics* (New York: Wiley).

CONRAD, R. and HULL, A.J., 1968, The preferred layout for numerical data-entry keysets. *Ergonomics*, 11, 165–73.

FITTS, P.M. and JONES, R.E., 1947, *Analysis of Factors Contributing to 460 'Pilot-Error' Experiences in Operating Aircraft Controls*. In H.W. Sinaiko (ed.), *Selected Papers on Human Factors in the Design and Use of Control Systems* (New York: Dover, 1961), 332–58.

ILG, R., 1987, Ergonomic keyboard design. *Behavior and Information Technology*, 6, 303–9.

INTERNATIONAL ELECTROTECHNICAL COMMISSION, 1975, *Colours of Indicator Lights and Push Buttons* (Paris: IEC).

MCMILLAN, G.R. and EGGLESTON, R.G., 1997, *Nonconventional Controls*. In G. Salvendy (ed.), *Handbook of Human Factors and Ergonomics* (2nd ed.) New York: John Wiley & Sons.

MEISTER, D., 1971, *Human Factors: Theory and Practice* (New York: Wiley-Interscience).

NORMAN, D., 1983, *The DVORAK Revival: Is it Really Worth the Cost?* (La Jolla: Institute for Cognitive Science and San Diego: University of California).

PFAUTH, M. and PRIEST, J., 1981, Person–computer interface using touch screen devices. In *Proceedings of the 25th Annual Meeting of the Human Factors Society* (Santa Monica: Human Factors Society), 500–4.

ROCHESTER, N., BEQUAERT, F.C. and SHARP, E.M., 1978, The chord keyboard. *Computer*, 11, 57–63.

SANDERS, M.S. and MCCORMICK, E.J., 1993, *Human Factors in Engineering and Design* (7th ed.) (New York: McGraw-Hill).

SHERIDAN, T.B., 1992, *Telerobotics, Automation and Human Supervisory Control* (Cambridge, MA: MIT Press).

SWAIN, A.D. and GUTTMANN, H.E., 1980, *Handbook of Human Reliability with Emphasis on Nuclear Power Plant Applications*. NUREG/CR-1278 (Washington, DC: US Nuclear Regulatory Commission).

Keyboards

D. A. Thompson

Department of Industrial Engineering and Engineering Management, Stanford University, Stanford, California, USA

1. INTRODUCTION

Keyboards allow informational interaction with machines (computers, calculators, telephones) instruments, process control, vehicles, etc. In this manner human users can communicate their instructions, requests, preferences, and knowledge manually to machines, which in turn display the results of their activities back to the human operators, closing the feedback loop. Keyboards remain the most versatile of the ways people communicate with machines, although other alternatives are becoming more available in specialized applications (e.g., mice, voice, tablets, scanners). This article will discuss the most frequently occurring keyboards in some detail, and by inference refer to other, less frequently encountered keyboards and their function.

2. TYPES OF KEYBOARD

2.1. Computer

In the late twentieth century, the word "keyboards" generally means computer keyboards. And the operation of computer keyboards has become an important manner in which vast amounts of information are stored, retrieved, analyzed, and shared with others over networks. Keyboards also represent a significant source of employment for millions of people who design, build, and use them. The following sections refer primarily to the use of computers for skilled text and data entry on detachable keyboards.

2.1.1. Types of keys

Typical full-sized computer keyboards have over 100 discrete keys, falling into seven key categories (Table 1). Portable, notebook, and hand-held computers generally follow this same pattern but with a reduced number of keys and keyboard space, and more compact key spacing.

As each key is struck, a unique ASCII signal is generated and sent to the computer's processors for interpretation and display on a visual display screen. Some function keys do not alter the computer display when pressed but rather switch, or toggle, between two states for other keys. For example, pressing and releasing *Insert* changes the manner in which characters are inserted into text—with replacement or without replacement.

Num Lock changes the entries on the numerical keypad from numerals to function keys for cursor movement. In both cases, a misunderstanding about which state the keyboard is in leads to confusion and errors. Therefore, each function switch key should have an unambiguous visual indicator of its state, typically a light on or adjacent to the key, or in a central location on the keyboard with proper labeling.

Control keys function in concert with other keys to their function. For example, *Alt + F* keys may pull down a file menu when struck together, or chorded, and *Ctrl + S* may save the current file when chorded. Other control keys relating to specific proprietary software may also be included.

2.1.2. Layout of keys

Computer keyboard design is an offshoot of typewriter keyboard design, keeping much of the same key layout proposed by Sholes in 1873 and first built by Remington & Sons in New York City. It is standardized in ANSI X4.23-1982. And although alternative key layouts have been proposed, including the well-known innovation by August Dvorak in 1936 (the Dvorak Simplified Keyboard), and a simple alphabetic sequence arrangement, the QWERTY key arrangement proposed by Sholes has become the dominant pattern. (QWERTY refers to the key sequence on the row above the left hand's home row.) Retaining this layout from typewriters into computer generations has allowed trained typists to move from one to the other and from one computer model to other computer models.

The present computer keyboard layout typically consists of a QWERTY layout of alphabet and punctuation keys, with a row of numerals across the top, and all surrounded by various function keys. To the right are additional function keys and a rectangular numeric keypad, duplicating the numeric keys above the alphabet keys for convenience. For additional comments on key layout, see ANSI/HFS 100-1988.

Since many typing tasks involve both numerical data entry and text entry, it is useful to the keyboarder to have the keys all arranged as close to one another as reasonable, avoiding open, unused space on the keyboard. This will minimize the movement of hands and arm rotation from one part of the keyboard to another. Differing types of keys may be color-coded to differentiate

Table 1. Categories of key characteristics on computer keyboards.

Key Type	Purpose	Examples	How activated
Printing keys:			
Alphabetic	Enter text	a through z	Strike key
Numeric	Enter numbers	0 through 9	Strike key
Punctuation	Format other entries	~ ' ! @ # $ % ^ & * ( ) _ - + = { } [ ] | \ / : ; " ' < > , . ?	Strike key
Non–Printing keys:			
Function keys:			
programmable	Perform specific software activity	F1 through F12	Strike key
preprogrammed	Manipulate other entries	*Enter, Tab, Home, End, Delete, Page Up/Down, Escape*	Position cursor, strike key
function switch	Alternate between two specific states	*Caps Lock, Shift, Num(eric) Lock, Scroll Lock, Insert*	Strike key to toggle between states
Control	No unique function	*Ctrl, Alt*	Chorded with one or two other keys to change their effect.

among them. Key tops are slightly concave for a more comfortable touch by the fingertip. Full-sized keyboards generally have square keys with a strike surface 12 mm on a side and 19 mm key spacing (center to center). Portable or hand-held computers often have a tighter spacing, making keyboarding slower and more difficult.

The early computer keyboards frequently had function keys off to the sides of the alphabet keys, where they were typically struck with the fifth, or weakest, finger. More recently, the programmable function keys have been moved above the alphabet keys, where they are easily accessible by all of the fingers. Unfortunately, many keyboard layouts still have some frequently used preprogrammed function keys off to the right of the alphabet keys closest to, and often struck by, the fifth finger of the right hand.

Virtually all keyboards have a row of numeric keys directly above the alphabet keys, numbered 1 to 0 from left to right. General-purpose computer keyboards also have a numeric keypad on the right end of the keyboard, arranged in a 3×4 matrix similar to that of a digital calculator. (The calculator layout is an inverted telephone keypad layout.)

2.1.3. Shape

The early computer keyboard shape uniformly followed the typewriter pattern of straight rows of keys in a stairstep fashion. Most still do, but since the early 1980s various ergonomic shapes are becoming available. The purpose of these improved geometries is to eliminate the keyboarder's ulnar deviation (hands rotated outwards) and pronation (palms downward), both of which require deviations from the hand's neutral posture (a thumbs-up handshake position).

One of the new geometries is a fixed slope upward from each side of the keyboard toward the center, and a fixed "opening angle", or rotation of each half of the keyboard into somewhat of an A pattern. A typical fixed geometry ergonomic keyboard, the Natural Keyboard made by Microsoft, is illustrated in Figure 1.

The other approach uses a variable geometry, allowing the ultimate shape of the keyboard to be adjusted to the most comfortable and useful shape by the keyboarder, and then locked into position. One of the more adjustable variable geometry keyboards, the Comfort Keyboard made by the Health Care Corporation, is illustrated in Figure 2. Both of these approaches reduce to varying degrees the musculoskeletal stresses on the hands, arms, and wrists by allowing more of a neutral wrist and arm posture. Because of the use of the familiar QWERTY key layout for both of these ergonomic keyboards, adapting to the new keyboard shape is generally not difficult for most users.

Figure 1. **Example of a fixed geometry ergonomic keyboard.**

Figure 2. **Example of a variable geometry ergonomic keyboard.**

A radical type of keyboard has been designed in recent years, and is favored by some users. It involves "chording" by the fingers from a stationary position on two hand-shaped keyboards. Chord keyboards are operated by striking a single key or chords of two or more keys simultaneously for each character desired. The hands rest in a single position that is seen by many users as more comfortable than the hand movements required by a conventional keyboard, but the higher information content of each keystroke tends to result in slower keying rates.

The height and slope of the keyboard should be such that neither significant extension nor flexion of the wrist occurs during keyboarding. Wrist extensions of 0-15° degrees are preferred, and extensions higher than this may be unhealthy. The wrist posture depends somewhat on the relative position of the keyboarder with respect to the keyboard. Some recent studies indicate that a negative slope (sloping downward and away from the user) minimizes wrist extension and its accompanying wrist strain.

2.1.4. Key travel

The maximum vertical displacement of keys when pressed should be 2 to 4 mm, but it should not be less than 1.5 mm or more than 6 mm. Most standard keyboards are in the range 3 to 4 mm. Ideally, there should be a soft, slightly cushioned bottoming of the keys at the end of their travel so as to minimize fingertip microtrauma.

2.1.5. Force

Many if not most keyboarding jobs are highly repetitive, applying small amounts of force many times a day. Text entry and data entry operators often average between 14 000 and 15 000 keystrokes over a 7-hour day, or roughly 100 000 keystrokes per day (60 words per minute for only six hours is 108 000 keystrokes). Because of this high repetition rate, the force of each keystroke must be kept to a minimum (but not below 0.25 N in order to avoid unintended activation). The ANSI/HFS 100-1988 standard puts forth a maximum allowable key force of 1.5 N (150 grams force), and a preferred key force of 0.5-0.6 N. Other keyboard literature recommends a maximum key force of 0.8 N Moreover, recent research indicates that typical keyboarders strike the keys with 3-5 times as much force as is required, depending on the design of the keyboard and their personal typing style. Clearly, the required force to depress the keys designed into the keyboard must be kept to a minimum when the number of keystrokes is very large.

Generally, required key forces should be uniform over the whole keyboard. This eliminates the need to remember which keys require extra force, or a lighter touch. Where the key activation force is highly variable among keys, keyboarders tend to use more force for all the keys, increasing their fatigue and possibly discomfort levels. And where a heavy stroke is applied to a key requiring only a lighter touch, the key bottoms out more sharply and may contribute to fingertip microtrauma.

2.1.6. Tactile feedback

Tactile feedback is the subtle but detectable change in fingertip force during the keystroke that tells the keyboarder the stroke has registered and the key can be released. It is typically described subjectively by keyboarders as the "feel" of the keys, and it is quantified as the force-displacement relationship of the keystroke. It is recommended by ANSI/HFS 100-1988 for all keyboards, but many keyboards built in the 1980s do not possess it. Its importance is that the presence of tactile feedback, detected as a snap or breakaway by the fingertip, reduces the amount of force that users naturally apply to the keys during typing.

Figure 3 illustrates tactile feedback as the relationship between force applied by the finger and the resulting key displacement for two typical key mechanisms. One key (linear spring) behaves much like a linear spring in compression; its displacement is essentially proportional to the finger force until it hits bottom at around 4 mm displacement. The first sense the keyboarder has that the keystroke is complete is the bottoming resistance starting at 135 cN (135 grams force).

The other key (tactile feedback) depresses under force to 60 cN, but at a displacement of 1 mm this force peaks out and the key begins to require less and less force to depress it. This force declines to an inflexion point of only 30 cN at 2.3 mm, and then rises to the key's bottoming point of 75 cN at about 3.5 mm. This pattern for the falling off of key force early in the keystroke represents the tactile feedback characteristic of the key and it is sensible to the keyboarder's fingertips.

The "early warning" effect of the force inflexion allows the keyboarder to *begin* reducing their finger force on the keys and to *start* withdrawing their fingers in mid-stroke. And even though their follow-through of the keystroke generally results in the key bottoming out, their having begun finger withdrawal does lower the finger forces at the bottom of the keystroke. The first indication to the keyboarder that the key has been pressed sufficiently hard for the tactile feedback key is the peak force of 60 cN, but
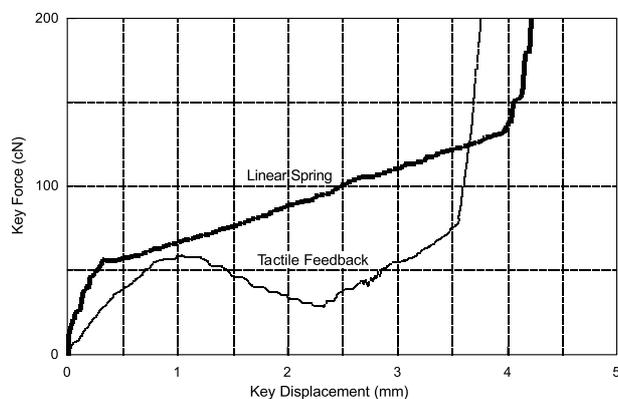
Figure 3. Examples of force-displacement curves for keyboard keys with and without tactile feedback.

for the linear spring key it is the key's bottoming out at 135 cN. The difference between these two values contributes to the differences in "feel" and in workload.

Some early keyboards possessed auditory feedback as keys were struck, but this is not a useful substitute for tactile feedback. And because it may be rather distracting in a large office, there needs to be a way of turning it off.

2.2. Health Effects of Keyboarding

Within the last few decades, some negative health consequences have accompanied. the use of computer keyboards. Besides the design and use of the keyboard itself, other contributing factors are the design of the workplace, the manner and pace of the job, and the psychosocial job environment.

The combined work characteristics of excessive key force, high repetition, poor posture, and extended work duration have been identified as playing a role in the incidence among computer keyboarders of upper extremity musculoskeletal disorders (also known as repetitive strain injuries, cumulative trauma disorders, cervicobrachial disorders, occupational overuse syndromes, etc.). Keyboard designers can directly influence two of these four primary factors—key force and hand/wrist posture. They should also understand the health importance of the other two factors—repetition and duration—and alert keyboard users to practices that may relieve some of the musculoskeletal stress which accompanies keyboarding, such as the proper use of wrist rests, document holders, and ergonomic chair positioning.

3. OTHER KEYBOARDS

A wide variety of keyboards exist in out culture today, because of the wide range of machines and instruments we communicate with. Telephone keyboards (desk, cordless, and cellular) have increased their complexity and sophistication, as have those on television sets and VCRs (and their accompanying remote controls), desk and pocket calculators, automobile dashboards, machine controls, cash registers, and so on.

Generally, the computer key design characteristics can be considered as appropriate for these noncomputer keyboards as well, since they are often used by the same people. However, to the extent that we generally use these alternative devices and machines with less frequency than we use computers, the issues of key layout, key force, tactile feedback, and wrist postures have a lesser health and ergonomic design significance. Nevertheless, the keyboard designer is cautioned that a somewhat reduced keying frequency but with substantially higher key forces and possibly poor wrist postures (e.g., cash registers) may still result in health hazards for the users.

4. REFERENCES

ANSI/HFS 100-1988, *American National Standard for Human Factors Engineering of Visual Display Terminal Workstations*, Human Factors and Ergonomics Society, PO Box 1369, Santa Monica, CA 90406.

KROEMER, K.H.E.,1972, Human engineering the keyboard. *Human Factors*, 14, 51-63.

POTOSNAK, K.M., 1988, Keys and keyboards, In Helander (ed.) *Handbook of Human-Computer Interaction* (New York: North-Holland), pp. 475-94.

SMITH, M.J., and COHEN, W.J., 1997, Design of computer terminal workstations. In Salvendy (ed.) *Handbook of Human Factors and Ergonomics* (2nd ed.) (New York: John Wiley), pp. 1638-88.

Manual Control Devices

T. Ivergaard
National Institute for Working Life, Östersund Branch, Sweden1.

1. INTRODUCTION

This chapter covers the design of the more traditional manual control devises, and the design of specific controls for communication with computers. Design of other forms of controls is dealt elsewhere, e.g. speech generation and recognition and human-to-human control.

The first step in the design of controls is to carry out a functional description of the task and to define the conceptual structure of which the task is a part. It is also important to analyze the logical functional relationships between the control devices and the information available to the operator.

Traditional control panels have the advantage that they give feedback to the operator of the maneuvers, which have been carried out. Examples of the advantages and disadvantages of each type of control are presented with design recommendations for these controls.

Keyboards are the traditional input devices used in data processing (DP) applications to communicate with computers. In the process control situation, it can often be advantageous to use other types of control such as multi-way joysticks or light pens. The advantages and disadvantages of each type of control are described.

2. FUNCTIONAL ASPECTS OF CONTROLS

The control device is the means by which information on a decision made by man is transferred to the machine. The decision may, for example, be taken on the basis of previously read information devices or on the basis of information from other sources, or from some form of cognitive process. Functionally, controls may be divided into the following categories:

1. Switching on/off start or stop.
2. Increase and reduction (quantitative changes).
3. Spatial control (e.g. continuous control upwards, downwards, to the left, or right).
4. Symbol/character production (e.g. alphanumeric keyboards).
5. Special tasks (e.g. producing sound or speech) (not included here).
6. Multifunction (e.g. controls for communicating with computers).

Examples of control type (1) include the starting or stopping of motors, or switching lamps on or off type (2) may consist of an accelerator pedal to increase and reduce the flow of fuel to the engine. Traditionally the best known example of spatial control (3) is the steering of a car. Examples of character production (4) include typewriting and telegraphy.

Of particular importance in control room design are the types of control (6) used in conjunction with computers. Controls operated by the hand are of particular use where great accuracy is required in the control movement. Hands are considerably better at carrying out precision movements than feet. Where a very high degree of accuracy of movement is required, it is best for only the fingers to be used.

Because the power available from the leg is considerably greater than that from the hands, foot controls are suitable for maneuvering over long periods or continuously. Foot controls are also valuable where very large pressures are needed, as the body weight can be added to the force of the strong leg musculature. It may also be necessary to use foot control devices where the hands are occupied in other tasks.

When designing traditional types of control, it is possible to design them in such a way that they naturally represent the changes one wishes to bring about in the process. For example, a lever that is pushed forwards may determine the forward direction of movement of a digger bucket. Or the flow in a pipe can be stopped by turning a knob which lies on a line drawn on the panel. In this way the design of the control increases the understanding of the current state of the process.

For communication with computers, the keyboard is often chosen for carrying out all the different control functions. Technically, it is often easy to connect a keyboard to a computer system. Other control devices also exist for communicating with computers, such as light pens. However, a particular failing of this type of multipurpose control device is that the control movements in themselves have no natural analogy with the changes, which they aim to bring about in the process.

3. ANATOMICAL AND ANTHROPOMETRIC ASPECTS OF CONTROL DESIGN

Some of the principal anatomical and anthropometric aspects will be considered (Grandjean 1988).

One important limitation of the recommendations available today is that they are based on the Caucasian races. For the Japanese population, for example, the measurements must be adapted for their proportionally shorter leg lengths. The following rules can be applied in the design of all types of control.

1. Maximum strength, speed, precision or body movement required to operate a control must not exceed the ability of any possible operator.
2. Number of controls must be kept to a minimum.
3. Control movements that are natural for the operator are the best and the least tiring.
4. Control movements must be as short as possible, while still maintaining the requirement for "feel."
5. Controls must have enough resistance to prevent their activation by mistake. For controls only used occasionally and for short periods, the resistance should be about half the maximum strength of the operator. Controls used for longer periods must have a much lower resistance.
6. Control must be designed to cope with misuse. In panic or emergency situations very great forces are often applied, and the control must withstand these.
7. Control must give feedback so that the operator knows when it has been activated, even when it has been done by mistake.
8. Control must be designed so that the hand/foot does not slide off or lose its grip.

3.1. Push–buttons and Keys

These are suitable for starting and stopping, and for switching

on or off. This type of control is also suitable for foot control, where it should be operated by the "ball" of the foot. The following recommendations apply to both hand- and foot-operated controls:

1. Resistance of the push-button should increase gradually, and then disappear suddenly to indicate that the button has been activated.
2. The top of the button should have a high coefficient of friction to stop the fingers/feet from sliding off. Where push-buttons are to be activated by the fingers, the concave form is preferable.
3. To indicate that the button has been activated, a sound should be emitted if the workplace has low light levels.

3.2. Toggle Switches

Toggle switches can show two or three positions. Where there are three positions, one should be up, the middle one straight out and the other one downwards. Toggle switches take up very little room. The following recommendations also apply to toggle switches:

1. A sound should be heard to indicate activation of the switch.
2. If a few switches are used, they should be placed in a row. Vertical positioning requires more space to avoid accidental operation.

3.3. Rotary Switches

These can be divided into two categories — cylindrical and winged. The primary difference is that the winged version has a pair of "wings" above the cylindrical part. The wings function both as a positional marker and as a finger grip. Rotary switches may have from three to 24 different positions. They require a relatively large amount of space because the whole hand has to have room to turn around the switch. However, where multiple position switches are used, they take up less space than the number of push buttons or toggle switches required to fulfil the same function. Rotary switches can either have a fixed scale and moving pointer or a moving scale and fixed pointer. A variant on the moving scale is to have a window, which only shows a small part of the scale. The following recommendations apply to rotary switches:

1. In most applications rotary switches should have a fixed scale and moving pointer.
2. There should be a detent in every position.
3. The turning resistance should steadily increase and then suddenly decrease as the next position is approached.
4. Cylindrical switches (knobs) should not be used if the resistance has to be high. In these cases, wing knobs are preferable.
5. Where only a few positions (2–5) are needed, they should be separated by 30–40°.
6. Where < 24 positions are used, the beginning and end of the scale should be separated by a greater amount than between the different positions.
7. Where the workplace has low lighting levels, a sound should be made to denote that the switch has been activated. In these cases there should also be a definite stop position at the beginning of the scale, so that the positions can be counted out.
8. The scale should always increase clockwise.

9. The hand should not shield the scale.
10. The surface of the switch should have a high coefficient of friction so that the hand does not slip.
11. The distance between panel and knob should be at least 3 mm.
12. The maximum amount of slope on the sides of the knob should be 5°.

3.4. Levers

Levers are activated either by the whole hand or just by the fingers. In general, where fine control is needed, only the fingers should be used. The following recommendations apply to levers:

1. The maximum resistance (force) for push–pull movements with one hand, with the control placed centrally in front of the body, is between 12 and 22 kg, depending on how far from the body the control is positioned.
2. The maximum resistance for push–pull movements for two hands are double that for one hand.
3. The maximum resistance for one hand moving in the left–right direction is ~9 kg, and is considerably lower in the opposite direction.
4. The maximum resistance for two-handed movements in the left–right direction is ~13 kg.
5. The lever movement should never be greater than the arm's reach without moving the body.
6. Where precision is required, a supporting surface should be provided for the part of the body used; an elbow rest for large hand movements and a hand rest for finger movements.
7. When levers are used for step-wise control (e.g. gear levers), the distance between positions should be one-third of the length of the lever.
8. Where the lever also acts as a visual indicator, the distance between positions can be reduced. The critical distance is then the operator's ability to see the markings.
9. The surface of the lever handle should have a high friction coefficient, so that the hand does not slip,

3.5. Cranks

These are suited to continuous control where there are high demands for speed. Cranks can be used for both fine and coarse control depending on the degree of gearing selected. The following recommendations apply to cranks:

1. Cranks are preferable to wheels where two or more revolutions are to be made.
2. For small cranks < 8 cm in radius, the resistance should be at least 9 N and a maximum of 22 N when rapid movement is required.
3. Large cranks of 125–220 cm radius should have a resistance between 22 and 45 N.
4. Large cranks should be used when precision is required (accuracy between a half to one revolution), with the resistance between 10 and 35 N.
5. The handle should have a high surface friction to prevent the hand slipping.

3.6. Wheels

Wheels are used for two-handed operations. Identification of the position is very important if the wheel can be rotated through

several revolutions. In addition, the following recommendations apply:

1. The turning angle should not exceed + 60° from the zero position.
2. The diameter of the ring forming the outside of the wheel should be between 18 and 50 mm, and should increase as the size of the wheel increases.
3. The wheel should have a high surface friction so that the hand does not slip.

4. CONTROLS FOR COMMUNICATION WITH COMPUTERS

The traditional controls in administrative computer systems are various different types of keyboard. There is often a numerical keyboard as well as the traditional typewriter keyboard. These types of keyboard have been tested over a long period and can be well specified. They are also thought to be well suited to most forms of administrative computer system.

For a more specialized application, for example, computer systems for control and monitoring of process industries, the situation is often very different. The requirement for control devices is unique to each type of process industry, depending on the process to be controlled and the type of computer system installed. It is thus impossible to give any specific guidelines for the control devices on computerized control systems. However, some overall guidelines may be given. The advantages and disadvantages of different control devices for computerized systems will be examined. Many of the devices are new and have not yet been subjected to ergonomic evaluation. It is therefore difficult to give detailed guidelines for all controls. Finally, some more detailed design recommendations for the different types of keyboard will be presented.

4.1. Advantages and Disadvantages of Different Controls for Computers

The more traditional types of control can of course also be used for computerized systems. Controls that have been produced specifically for communicating with computers include:

- keyboard with predetermined functions for various keys;
- keyboard with variable functions for the keys;
- light pen;
- touch screen;
- electronic data board;
- voice identification;
- trackball and joystick (multi-position lever); and
- mouse.

4.1.1. Keyboards with predetermined functions for the keys

Keyboards with predetermined functions for all keys normally have two main parts — an alphanumeric or a numeric part, and a function key part. The function key part has different keys for different predetermined tasks, such as starting, stopping, process a, b, c, etc. The keyboard works by the operator pressing the keys in a certain order, which he either remembers, or with the aid of some form of crib-sheet.

This type of keyboard is characterized by the need for a large number of keys, usually one per function. Where there are many functions and several subfunctions within every main function,

problems arise with grouping the keys in the proper way and in positioning the keys in a mutually logical way, which is consistent in terms of movements.

It is unusual to be successful with this at the first attempt; the keyboard will need to be redesigned when it has been operational for long enough to build up enough experience to determine its optimal design. Making changes to the keyboard are often costly, but if the keyboard is not redesigned at a later stage, it means that large and frequent arm movements become tiresome and time-consuming. The advantage of this type of keyboard is that it needs relatively little computer programming and, to a certain extent, a standard board can be used, at least for the alphanumeric part.

4.1.2. Keyboards with variable functions for the keys

Keyboards with a variety of functions for each of the different keys are relatively uncommon, but they often exist as part of the more traditional keyboard (e.g. the top row of keys on the keyboard). Keyboards with a variety of functions per key are often particularly useful in process industries. A common form is to have a row of unmarked keys under the monitor screen, and to have squares representing the different keys directly above them on the screen. Depending on the picture being shown on the screen, text appears in different windows showing the functions that the keys have for each frame.

Another application for this type of keyboard is to build lights into the keys. The relevant keys light up for each particular function. The lights in the keys are lit or extinguished when particular keys are pressed, depending on the sequence of operations required. In this way the operator is guided through the correct operation sequence.

Non-illuminated keys are then disconnected from the system. The risk of errors occurring with this type of system is very small, and work on this keyboard is also faster, particularly if the operator is not accustomed to the work. It is important, however, that if the lamps in the keys break a warning signal is produced.

There are also applications where keys can be pressed with different pressures. A light pressure on the key causes the function associated with that key to be written out on the screen, and the action is taken if the key is pressed harder. If it is fully depressed a signal is sent to the computer dictating changes to be made to the process. The operator can also receive new information on the screen, which informs him which new keys can be used.

Depending on its design, the keyboard can be pre-programmed to lead the operator naturally through the work. This type of programming of the keyboard functions may be an advantage for especially important types of operation, where errors could have serious consequences. A major disadvantage from the operator's perspective is that they may feel their work is being too highly controlled. Another disadvantage with this type of keyboard is that it requires a lot of programming, and this takes up a large part of the computer's capacity. An advantage is that the hardware does not need to be changed (re-building or extending the keyboard, etc.) to any great extent even if a major change is to be made in the function of the control. In other words, this form of control is very flexible.

4.1.3. Light pens

The light pen consists of a photocell that senses the light radiated

from the phosphor on a CRT screen. The light pen reacts every time a pixel on the screen is lit up by the electron ray within the tube. The signal passes from the light pen to the computer, which at the same time receives information on where and when the spot passes different places on the screen in this way the computer can identify where the pen is on the screen

The light pen can point to parts of the screen one wishes to know more about. It can also be used to activate different functions. If for example, one pointed to a valve, and at the same time pressed a button on the side of the pen, this may cause the valve to close. The light pen is suitable for moving cursors on a screen. However, it is difficult to see any operational advantages of light pens over other controls.

If a light pen is used over a long period, it is necessary to have a specially designed armrest to prevent discomfort. The light pen has to come close up against the screen, which means that it is impossible to have any form of reflection shield or filter on the screen and this can give rise to visual problems. Positioning of the light pen must be exact, which makes considerable demands on vision and also contributes to bad working posture in many instances.

4.1.4. Touch screens

Touch screens involve moving the finger, a pen, a pointer or some other object within an active matrix placed over the screen. This active matrix may be designed in several ways. It could be composed of a thin metal net, for example, on which an electrical circuit is made when it is touched. Electrical bridges and infrared beams can also be used to determine touch on the screen. Another type of touch screen is based on the use of a transparent material, which senses the pressure of the touch on the screen. Special measurement bridges are used to determine how the pressure field is distributed over the screen, and the position of the touch is deduced from this.

Functionally, the touch screen is very similar to the light pen and has similar advantages and disadvantages, although an additional disadvantage is that the screen becomes dirty. An advantage is that it is sometimes faster to point with the finger than with a light pen; however, the technical reliability of the touch screen is usually considered to be lower than that of the light pen. If it is necessary to send a large number of different types of words and information to the computer, the traditional keyboard is preferable.

4.1.5. Electronic data boards

Electronic data boards consist of a rectangular plate, which represents the surface of the screen. Some form of electric field is created over the plate. When a sensor is run over the board's surface, it "senses" its position on the board. One common form of board is placed directly onto the screen, and in this case functions very like a light pen or touch screen. Another form of board is placed beside the screen, and one can work with a transparency of a picture. One of the advantages of the electronic data board is that one can very quickly make drawings or change them.

4.1.6. Trackball

The trackball is a mounted sphere that can be rotated in all directions and can be placed on a table or special fixing. The ball is usually used for moving the cursor on a screen. The cursor moves a certain distance (x/y directions) or with a speed proportional to the movement of the ball.

4.1.7. Joystick

The joystick, which is a lever movable in all directions, has a similar function to the trackball.

4.1.8. Mouse

The mouse is a small device with wheels or a ball mounted on the underside. If it is moved to the left or right, this represents a corresponding movement on the screen. The mouse is especially suited for moving the cursor and for transferring graphic information.

There is no conclusive evidence to produce recommendations for the use of the trackball, mouse or joystick. In practice most people seem to prefer the mouse if one has access to a free table surface, otherwise the trackball is generally preferred.

5. KEYBOARD

The keyboard is still the most common computer input device. The design of the keyboard has a significant effect on the operator's performance in terms of speed and accuracy.

The most common keyboard layout is the QWERTY layout. Where two hands are used on the keyboard, 57% of the workload is on the left hand, even though 80% of the population is right-handed. This is advantageous for the type of job where the right hand alternates between handwriting and typing. It is also important to have one standard keyboard layout. Although the QWERTY layout is not the most efficient (it is said to have been designed for slowness, so that early mechanical typewriters did not become overloaded), it is now the best compromise as it has become the standard keyboard.

Keyboards are usually designed so that the alphanumeric section is in the center, with the cursor, editing keys and numeric keypad to the right. Function keys may be placed anywhere on the keyboard, but to give an aesthetically pleasing design they are often placed on the left-hand side. Lateral hand movements also require less energy than longitudinal (front to back) one.

There are no specific recommendations for keyboard layout, as their design is extremely sensitive to the task being carried out. However, a degree of flexibility must be incorporated in the design to cater for all variations in user requirements. One solution to this would be to develop a modular keyboard consisting of several units. Each unit would be made up of a different set of keys. The units could be arranged in the desired layout based on the results of the task analysis. However, care must be taken in using a flexible keyboard configuration due to the risk of a negative transfer of training. For example, if an operator carries out a number of different tasks, and different keyboards are used for each of the different tasks, then high error rates must be expected.

Chord keyboards are a combination of a keyboard and a coding system. In a similar way to keys on a piano, one can press several keys at the same time. The advantage of this type of keyboard is that key-pressing speed compared with a standard typewriter is considerably > 50% faster. There are, however, no special design recommendations for this type of keyboard. In

general, it may be said that this type of keyboard needs further study before any firmer recommendations can be given regarding suitable areas of use and suitable design.

The numeric keyboard appears in two different designs. The accepted layout is a 3 _ 3 plus a key set, but there are two alternatives within this. Adding machines have the 7, 8 and 9 keys on the top row, while push-button telephones have the 1, 2 and 3 keys at the top. Nearly all telephones use the 123 keypad. It would be advantages that all numeric keyboards are of this design. Uniformity is important, and the user should not have to switch from one keyboard design to the other while working.

The height of the keyboard is largely determined by its physical design, for example, electrical contacts and activating mechanism. Thicker keyboards (> 30 mm thick) should be lowered into the table surface to ensure a correct user posture. This unfortunately does not allow for flexible workplace design. Ideally, keyboards should be as thin as possible (< 30 mm thick from the desk surface to the top of the second key row and not need to be lowered into the surface

Keyboards can be stepped, sloped or dished. There is no evidence on the relative advantages of any of these profiles. The most important factor is for the keyboard to be able to be angled between 0 and 15° up at the back and, if the keyboard operator is standing, it is advantageous if it can be raised at the front from 0 to 30°.

The size of the key tops is a compromise between producing enough space for the finger on the key, while at the same time keeping the total size of the keyboard as small as possible. Key tops should be square and 12–15 mm in size. This size is quite sufficient for touch typing, but in cases where keyboards are used for other tasks, for example on the shop floor, key sizes can be larger. The spacing of keys is standardized to 19.05 mm between key top centers. This is within the ergonomic recommendations of between 18 and 20 mm. The force required for key displacement should be the same on all keys.

For skilled users, the actuating force should be 0.25–0.5 N. and the key displacement (travel) 0.8–1.0 mm (from rest to activation of system). For unskilled users the force should be 1–2 N and the displacement 2–5 mm. The user requires feedback to indicate that the system has accepted the keystroke. This is an important keyboard characteristic, although the exact requirements vary according to the individual levels of user skill.

In normal typing and other key-pressing tasks, there is

kinesthetic (muscle) and tactile (touch) feedback from the actual depression of the key, auditory feedback from the key press and/or activation of the print mechanism, and visual feedback from the keyboard or from the output display. For skilled operators, feedback from the keyboard (sound and pressure change) is of little importance. When learning, and for unskilled operators, this feedback is important. The operator should be able to remove the acoustic feedback.

The color of the keys is not generally regarded as important. A dark keyboard with light lettering is preferable when used in conjunction with light-on-dark image displays, and care should be taken not to cause any distracting reflections on the screen by light key colors. Matt finishes should be used where possible.

Care should be taken when using colors to code various function keys. Attention should not be drawn to a red key or a group of red keys if their importance in the system is minimal. These principles concerning color may also be applied to any information lights found on the keyboard.

REFERENCES

GRANDJEAN, E., 1988, *Fitting the Task to the Man* (4th ed.) (London: Taylor & Francis).

GRANDJEAN, E. and VIGLIANI, E., 1980, *Ergonomics Aspects of Visual Display Terminals* (London: Taylor & Francis).

IVERGÅRD, T., 1989, *Handbook of Control Room Design and Ergonomics* (London: Taylor & Francis).

McCORMICK, E.J. and SANDERS, M.S., 1982, *Human Factors in Engineering and Design* (5th ed.) (New York: McGraw-Hill).

MURRELL, K.F.H., 1958, *Fitting the Job to the Worker: A Study of American and European Research into Working Conditions in Industrial Situations* (Paris: Organisation for European Economic Co-operation — EPA).

PHEASANT, S.T., 1990, Anthropometry and the design of workspaces. In J.R. Wilson and E.N. Corlett (eds), *Evaluation of Human Work: A Practical Ergonomics Methodology* (London: Taylor & Francis), 557–73.

PORTER, J.M., FREER, M., CASE, K. and BONNEY, M.C., 1990, Computer aided ergonomics and workspace design. In J.R. Wilson and E.N. Corlett (eds), *Evaluation of Human Work: A Practical Ergonomics Methodology* (London: Taylor & Francis), 574–620.

Multivariate Visual Displays

C.M. Carswell and S.E. Waters
Department of Psychology, University of Kentucky, USA

1. INTRODUCTION

Human understanding of any object, concept, or process usually requires the observation and measurement of multiple variables. The economic status of a nation, for example, is typically described by a number of financial indices. Likewise, the operational status of a process, such as the conversion of solar energy into electricity, depends on system inputs and component efficiencies. The present article discusses the progress made, and problems encountered, in using multiple attributes of a visual display as analogs for the multiple measures used to describe such physical and conceptual entities. The article also describes some of the principles used in the design of such displays.

2. EVOLUTION AND FUNCTIONS OF MULTIVARIATE DISPLAYS

2.1 Statistical Graphics: the Third Dimension and Beyond

The earliest multivariate displays used spatial variation in two dimensions to represent two variables of scientific interest. For example, in 1686 Sir Edmund Halley created a plot of barometric pressure as a function of altitude. This early scatterplot used the vertical and horizontal positions of points to represent altitudes and pressures, respectively, of different observations.

Representing a third variable on the printed page proved to be more challenging. One of the first solutions was offered by William Playfair (1759–1821), the Scottish economist and political scientist, who developed and promoted many of the graphical formats still in use. Rather than trying to use the artist's illusions of depth to create variation in perceived distance, Playfair used transformations of the various features of abstract objects. Figure 1 presents an early example of this technique. Here, each object represents a different country, with the length of the left ray representing population, the length of the right ray representing total revenues, and the circle size representing geographic area. Contemporary statisticians have used the same approach, with both abstract and familiar objects (e.g. faces), to represent three or more variables.

Russia

Great Britain and Ireland
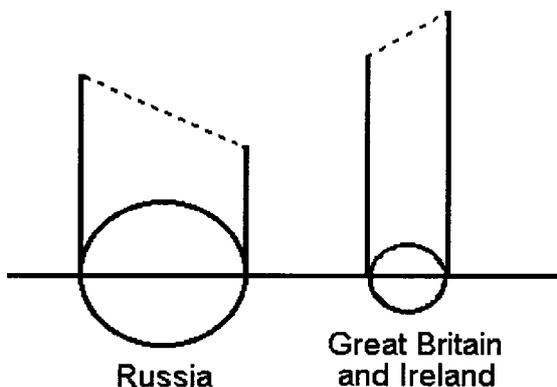
2.2 Dynamic Displays of Two or More Variables

The main goal of statistical graphics is to facilitate the exploration of data sets, allowing for comparisons across specific variables and observations. Such graphics are also used to communicate findings to relevant audiences. Thus, statistical graphics aid both discovery and disclosure. Transportation and industrial technology added new requirements to the design of multivariate displays. These displays were designed to facilitate the supervision and control of the dynamic systems they represented. They were frequently updated, introducing the appearance of constant motion. This motion added several new dimensions, such as speed and direction, to the designer's repertoire, thus allowing the presentation of still more variables. Recently, statisticians have begun to incorporate some of the dynamic properties of industrial graphics into statistical graphics (Cleveland and McGill 1988).

3. PRINCIPLES BORROWED FROM UNIVARIATE DESIGN

Many of the design principles applied to univariate displays are equally valuable for the design of multivariate displays. While the success of a multivariate display is certainly more than the sum of its component codes, the following guidelines for representing individual variables are still valuable.

3.1 Visual Analogies

When possible, designers should choose display dimensions that are direct visual analogs of the variable being represented. In Figure 1, Playfair used the area of different circles rather than the lengths of rays to represent the areas of different European countries.

3.2 Linguistic Conventions and Population Stereotypes

Even when describing abstract concepts, it is common to use directional or geometric terms. Visual displays should conform to these linguistic conventions whenever possible. For example, when graphing the mean affective state of a sample of subjects, the designer should keep in mind that negative affect (e.g. depression) is often referred to as "feeling low," while euphoria is associated with "things looking up." Thus, affect should be represented by varying vertical position, with higher positions representing happier moods.

3.3 Comparisons versus Classifications

The visual attributes that best support classification are sometimes not the same attributes that support accurate comparisons. For example, if an absolute judgment is required, then the hue of a light would be preferable to its brightness, and geometric shape would be preferable to area. If comparisons are more important, then dimensions close to the top of the following list, especially the top three, should be favored (Carswell 1992; Cleveland and McGill 1985):

> position on common aligned scales
> position on common nonaligned scales
> length
> slopes/angles
> area
> volume/density/color saturation
> hue

3.4 Redundancy

For particularly important variables, designers should use more than one visual dimension to represent their values. For example, if information about temperature is crucial, then temperature might be coded as (1) position along a vertical scale (keeping in mind that higher positions should stand for "higher" temperature) and as (2) colors (with higher temperatures coded as "warmer" colors such as orange and red).

4. PRINCIPLES SPECIFIC TO MULTIVARIATE DISPLAYS

When more than one variable must be represented within a display space, the arrangement of these representations becomes as critical as their individual designs. The viewer's subjective organization of a particular arrangement is referred to as the *perceptual organization* of the display. A viewer's perceptual organization determines which elements appear to go together to form groups, configurations, or objects. This, in turn, determines how accessible (or inaccessible) specific variables will be and may produce new visual dimensions, called *emergent features*, which can sometimes be useful to the viewer.

4.2 Influencing Perceptual Organization

The gestalt laws of perceptual organization have been applied to display design to ensure that some displayed variables appear to go together while other variables appear quite separate. Figure 2 illustrates how these laws might be applied to hypothetical variables A, B, and C, each of which is represented by the height of a bar. In this example, design manipulations are intended to encourage the grouping of variables B and C while separating the B–C group from variable A. A variety of techniques, alone or in combination, can be used to achieve the desired effect:

As shown in Figure 2, variables B and C can be effectively grouped by increasing their *spatial proximity* to one another while increasing their average distance from variable A. The perceptual grouping of variables B and C can also be accomplished by giving them a *similar color or fill pattern* and ensuring that this shared feature is sufficiently distinct from the color/fill used for variable A. Variables B and C may also be included in a *common region* by enclosing them in a framework. In addition, *contours* may sometimes be added to connect the parts of the individual representations that actually vary, in this case the bar tops. The
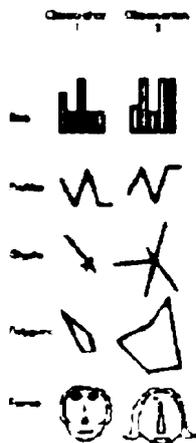
orientation or overall shape of the contours can sometimes serve as additional cues to the viewer. This design strategy is the basis for profile and polygon displays, discussed below. Finally, in order to more strongly group variables B and C, several of these strategies may be combined, as in the bottom display in Figure 2.

4.3 Object Displays

The most extreme technique for promoting perceptual grouping is to represent each variable as a different attribute of a single object — for example, Playfair's "sunburst" object in Figure 1. When considering object displays, it is important to note that the attributes or dimensions of an object can interact with one another in different ways. The most important distinction for design purposes is the difference between homogeneous and heterogeneous displays.

Homogeneous object displays are those composed of repetitions of a single dimension, for example the repeated use of point position in a line graph. Such displays are likely to produce emergent features. Alternatively, *heterogeneous object displays* mix and match visual dimensions, as when ray length and circle area are combined in Playfair's "sunburst." Such display dimensions are less likely to interact, fuse, or create emergent features. For example, a colored light that indicates temperature may be integrated with an analog clock that shows elapsed time by making the clock's hands vary in color. The clock is perceived as a single object, but the perception of the hand's position is not influenced by its color (or the color by the hand position). However, the combined value of both may be easier to perceive because of the spatial integration.

According to Carswell and Wickens (1996), homogeneous object displays should generally be used when mathematical integrations or comparisons are required by the viewer. For example, the horizontal and vertical extents of a rectangle are homogeneous dimensions that interact to produce its overall shape and size. This display would be useful if the viewer needs to know the relative value or the product of the two variables represented by height and width. Heterogeneous object displays should generally be reserved for situations where conjunctions of values (or value ranges) along several variables are necessary for the user to appropriately recognize specific conditions or states.

4.4 The Proximity Compatibility Principle (PCP)

The PCP assumes that display elements that form strong perceptual groups allow viewers to more easily divide attention across the component variables. Divided attention will be useful if all the variables in the group must be used, in conjunction, to perform the viewer's task. However, there can also be costs associated with perceptual grouping, particularly when grouping is strong as in object displays. The cost comes when the viewer must "zoom in" or focus attention on one variable in the display without distraction from the others.

The general guideline that follows from the PCP is that the designer should try to fully understand how information from the displayed variables would be used. If the 2-, 3-, or k-displayed variables must be used in conjunction, then arrangements that promote strong perceptual grouping are indicated. However, if the viewers must deal independently with the variables across

time, then arrangements that promote a looser, even isolated perceptual organization will result in better performance. The general guidelines have, overall, gained empirical support. However, the advantages of unified or integrated displays for divided attention tend to be greater than the costs associated with integration when focused attention is required (Wickens and Carswell 1995).

4.5   Representation Aiding (RA)

This approach is particularly appropriate when applied to displays that must support the supervision of processes or systems that have well-defined physical constraints and for which the goals of the viewer are well understood. Clearly outlining these constraints and goals is a fundamental step in using RA. The designer's task is to ensure that these constraints and goals are adequately reflected in the perceptual organization of the displayed variables, especially in the perceptual interaction and relative salience of display attributes (Bennett *et al.*1997).

Research on both RA and the PCP has revealed that one way perceptual grouping leads to enhanced performance is through the production of salient emergent features. However, rather than allowing these emergent features to emerge by chance, RA emphasizes the intentional use of emergent features to represent important system constraints or relations. Emergent features should be designed to adhere to the normal standards for individual codes discussed above — they should make use of spatial and linguistic analogies, they should support either classifications or comparisons, as appropriate, and, if especially critical, their salience should be increased by the use of redundancy.

5. MULTIVARIATE FORMATS: FROM THE COMMON TO THE CONTROVERSIAL

Since the early 1970s, a variety of multivariate formats have found their way into scientific journals, business magazines, intensive care units, and nuclear power plant. Some of the more common (or more controversial, in the case of the face display) are illustrated in Figure 3. With the exception of the bar charts, all of these formats are considered to be object displays.
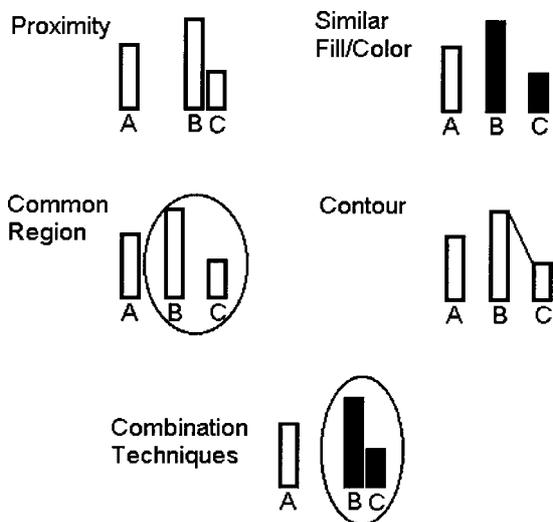
5.1   Bars and Profiles

These formats are familiar to most people because they are used extensively in the presentation of univariate and bivariate data. For example, the daily newspaper often shows a series of bars that each represent different instances of the same variable (e.g. temperature measured on different days). However, each bar can also be used to represent a different variable (e.g. temperature, pressure, or humidity on a single day). If meaningfully scaled, these variables can be compared with one another, or compared as a unit with other observations.

Bar charts are among the most common and flexible of formats. Functionally, they balance the demands to focus on one variable with the need to integrate or compare all of the variables. The individual values remain relatively easy to locate because they are represented by separate objects. However, bar graphs can also produce emergent features such as colinearity and symmetry that can communicate a more integrative message.

If bar tops are replaced by points, and these points are then connected by line segments, a profile plot similar in form to the familiar line graph is created. Profile plots increase the salience of the emergent features found in bar charts, but they may make the use of the individual data values more difficult. Such plots should be reserved for situations where the overall configuration rather than the individual variable values are of prime importance.

5.2   Glyphs and Polygons

These formats are similar to bars and profiles except that they are plotted in a polar coordinate space. The prime advantages of these displays over linear forms are twofold. The individual variables may be easier to locate because they are each presented at a distinct orientation and location, unlike bars and profiles which vary only in location. In addition, emergent features such as overall size and shape may be available in the polygon display.

5.3   Face Displays

Face displays resulted from the observation that humans can perceive and remember small differences in faces. In addition, faces have emotional significance which is familiar to everyone (e.g. sadness, happiness, fear).

Seeking to capitalize on these assets, Chernoff (1973) developed a caricature face that could display a maximum of 18 variables by using separate facial features.

The face display is a heterogeneous object display that is best used when specific feature conjunctions rather than configurations must be identified. However, special care must be taken in assigning variables to features since some, such as the mouth, seem to be more salient than others. The mouth should thus be reserved for variables that are themselves most critical to the viewer's task. In addition, the natural meaningfulness of faces can be a tremendous asset, allowing a potential viewer to have a general idea of what the data mean even without extensive training. However, if care is not taken in assigning facial features to variables in a manner consistent with facial stereotypes (as when a negative financial variable is assigned to a smiling mouth) then miscommunication is almost a certainty. Face displays have remained the most controversial multivariate format largely due to the possibility of such errors.

5.4 Other Multivariate Formats

A variety of other formats have been designed, including trees and castles for the display of hierarchical variables, perspective displays, and animated graphics. These and other formats are discussed in Chambers *et al.* (1983).

REFERENCES

BENNETT, K.B., NAGY, A.L., and FLACH, J.M., 1997, Visual displays. In Salvendy (ed.), *Handbook of Human Factors and Ergonomics* (New York: Wiley), pp. 659–96.

CARSWELL, C.M., 1992, Choosing specifiers: an evaluation of the basics tasks model of graphical perception. *Human Factors*, 34, 535–54.

CARSWELL, C.M., and WICKENS, C.D., 1996, Mixing and matching lower-level coes for object displays. *Human Factors*, 38, 1–22.

CHAMBERS, J.M., CLEVELAND, W.S., KLEINER, B, and TUKEY, P.A., 1983, *Graphical Methods for Data Analysis* (Monterey, CA: Wadsworth).

CHERNOFF, H., 1973, The use of faces to represent points in k-dimensional space graphically. *Journal of the American Statisical Association*, 68, 361–8.

CLEVELAND, W.S. and McGILL, R., 1985, Graphical perception and graphical methods for displaying scientific data. *Science*, 229, 828–33.

CLEVELAND, W.S. and MCGILL, R., 1988, *Dynamic Graphics for Statistics* (Belmont, CA: Wadsworth).

WICKENS, C.D. and CARSWELL, C.M., 1995, The proximity compatibility principle: its psychological foundation and relevance to display design. *Human Factors*, 37, 473–94.

Pointing Devices

C. Baber

University of Birmingham, UK

1. INTRODUCTION

The notion of "pointing" at objects on the computer screen, using some form of pointing device, is central to the "direct manipulation" paradigm of contemporary computing. Rather than having to recall obscure commands or ambiguous file names, the user of a direct manipulation interface is able to look for and recognize the object or function that they are seeking (Shneiderman 1982). The purposes of this entry are to provide an overview of pointing devices and their performance.

2.0 TYPES OF POINTING DEVICES

Having established that a pointing device allows the user of a computer to select objects on the screen, we can now consider different ways in which this task can be performed. The obvious approach would be to find the object required and simply reach out and touch it, i.e. users would select an object on the screen by touching it with their forefinger (alternatively, selection could be performed using a stylus or pen). These are direct pointing devices: the user is able to touch directly or otherwise select the object required. In comparison, indirect pointing devices — such as mice, trackballs, joysticks — require the user to manipulate a control in order to move a cursor in order to select the object.

2.1. Touchscreen

With a touchscreen, selection can either occur when the finger "touches" the object, or can occur when the finger "touches and then lifts off" the object. The advantage of the latter approach is that the user is able to drag the object around the screen. There several ways in which touching can be detected. The finger can interrupt a field above the screen, e.g. beams of infrared light or acoustic waves. The point of interruption is detected and used to signal the co-ordinates of selection to the computer. One potential problem with this approach is that accuracy depends on the relative distance between screen and overlay. As the distance between the two surfaces increases, so the possibility of parallax increases. There are three ways of reducing parallax:

1. make the object at least 20 mm² to give a reasonable surface area for the finger to touch;
2. tilt the screen to around 30° to the horizontal, so that the user's line of sight is on a similar plane to their hand movement;
3. bring the touch surface as close as possible to the screen. This is the idea behind overlay touchscreens, e.g. resistive and capacitive. In these designs, pressure from the finger changes the capacitance or resistance of the overlay and this change is used to indicate coordinates of touch. Comparative studies of the different types of touchscreen suggest that capacitive overlay and infrared beam produces slightly better performance than the other types.

2.2 Pens

Rather than using a finger to interrupt a touch-sensitive surface, it is possible to use a pen. Early versions of pens tended to rely on light, e.g. the pen contained a probe that detected bursts of light from the computer screen which could then be used to determine the coordinates of the touch. Alternatively, a pen or stylus could be used with a touch-sensitive tablet. In resistive tablets, pressure from the pen onto the tablet led to two separate sheets being brought together at a specific point, thus indicating the point of touch; in magnetic tablets, the pen incorporated a probe which detected changes in magnetic field of the tablet. Some of the contemporary pen-based Personal Digital Assistant (PDA) devices tend to use a plastic stylus which is used to press on surface overlay on the screen. The majority of PDA systems use a touch to select approach (as opposed to touch and lift off — see 2.1). Recent developments in pen-based technology have meant that PDA can incorporate some form of handwriting recognition so that the pen can be used for drawing or writing as well as pointing.

2.3. Mice

The mouse contains a ball mounted against several sensors which can detect movement around 360°. While the majority of mice use leads to communicate with the computer (to either a mouse port or the RS232 port), there are mice which use either infrared or ultrasound. The mouse, like other indirect pointing devices, is primarily a cursor-control device; in order to select objects on the screen, an additional button is required. Typically the gain of a mouse should be set to around 1:1, i.e. movement of the cursor on the screen corresponds to movement of the mouse. Rate-controlled mice relate the velocity of cursor movement to the distance the mouse travels, but it is unlikely that this will benefit performance; providing acceleration during the ballistic phase is likely to lead to overshooting of targets. There have also been efforts to improve the homing phase through provision of additional feedback. The body of the mouse can be modified to incorporate actuators which can press against the user's fingers as an object is approached. Objects on the screen require some form of force field (defined by their relative position on the screen), so that as the cursor enters a field the user will receive additional feedback. Usually, incorporation of additional tactile feedback tends to improve positioning times (although accuracy may be compromised, which increases the likelihood of over-shooting the target).

2.4. Trackballs

The trackball can be thought of as a mouse inverted in its socket. The movement of the ball against the sensors corresponds to movement of the cursor on the screen. The direction and velocity of cursor movement mimics the movement of the ball. This means that a fast spin of the trackball can produce fast movement of the cursor over the screen. Thus the trackball offers additional functionality over, say, a mouse. On the other hand, the task of dragging objects across the screen can be difficult with the trackball because one finger needs to hold down a button while the rest of the hand manipulates the trackball. This problem can, of course, be handled by either having a button which stays down or by using two hands.

2.5. Joysticks

On a displacement joystick, the stick moves proportionally in the direction of force supplied by the user and remains in the

end position. This means that the user receives kinesthetic feedback from the actual movement of the joystick, in addition to visual feedback received from watching the moving object. On the other hand, an isometric joystick remains fixed in response to user action; gauges sense the magnitude and direction of pressure exerted by the user and move the object accordingly. In this design, the user has similar visual feedback but quite different tactile feedback to that obtained from the displacement joystick. The rate-controlled joystick provides the additional function of increasing rate of movement relative to increasing force against the stick, i.e. the harder the user presses, the faster the object moves. A spring-loaded joystick, which moves in response to user action, but moves back to a central position once the action is complete, can be seen to represent a combination of the other two types. Spring-loaded joysticks have been introduced into laptop computers to provide a convenient pointing device that can be situated in the keyboard (usually in the center of the home row of the keyboard).

3.0 DEVICE PERFORMANCE

System designers are primarily interested in which device would be best for a given application. This brings us to the question of how to compare pointing devices. The most common form of comparison between devices involves using the devices to perform specific tasks and then comparing their performance on some measure. The measures used are typically speed and accuracy, with some measure of user preference included. As noted in the discussion of device characteristics, pointing devices differ on several dimensions so that finding a task which does not unduly favor any one of these dimensions can be very difficult. For example, if we compared a mouse against a trackball on a task involving fast movement of the cursor from one part of the screen to the other, then we might find in favor of the trackball; on the other hand, if we compared the same devices on a task involving dragging an object onto a small target then we might find in favor of the mouse. Thus, in addition to measuring gross performance, such as speed, we ought to think of measuring (or at least defining) "task-fit".

3.1. Speed

Cursor movement is rarely performance for its own. Usually one moves the cursor as part of another task, such as object selection or movement. It is this point at which "task-fit" can be a significant variable in performance. For example, cursor keys are universally shown to be an inferior means of moving the cursor, in comparison with mice or trackballs, but this finding is reversed if the cursor only has to move a small distance or if the cursor can be made to "jump" to the next object in a given direction. For the most part, direct pointing devices, such as touchscreens, tend to produce the fastest selection time. As far as indirect pointing devices are concerned, the mouse tends to outperform other devices, such as joysticks and trackballs, and has been found superior to some direct pointing devices, i.e. pens. Comparison of devices in terms of speed have tended to flounder on differences between device characteristics. This means that while device A is better than device B for task X we cannot say with certainty that the same will hold for task Y. This has led several researchers to consider alternative ways of defining performance time. One approach which has proved popular over the past decade or so is

Fitts law (see MacKenzie 1992 for an excellent review of this work). If the values of a and b were "constants", then one might anticipate that they would hold for a given device over different experiments (or at least be quite similar). Unfortunately, work to date does not show this to be the case, e.g. Baber (1998) reports that for mice, one study found that a = 108, b = 392, while another study, by different researchers, found a = 107, b = 223. An alternative perspective, put forward by MacKenzie (1992), is that one should look at the Index of Performance (IP), i.e. the ratio between Index of Difficulty and Movement Time. By taking the ratio of IP between two devices, one might get an idea of relative difference. Using this measure, he showed that one study yielded an IP of 2.3 while another study yielded an IP of 2.2 for the mouse over the isometric joystick. While this looks impressive, Baber (1998) suggests that it does not hold for other device comparisons. This suggests that comparisons based on performance times will be suitable for specific comparisons but might not be suitable for generalization to other applications or tasks.

3.2. Accuracy

Accuracy usually refers to whether the user can move a cursor onto the target. Inaccurate movements could consist of over- or under-shooting the target. As one might expect from the discussion of Fitts law above, accuracy is also affected by target size and movement distance. Often when one looks at the "hits" on a target, one sees a clustering of "hits" around the center. This indicates the active target width at which people aim. Where comparisons have been conducted in terms of accuracy, researchers have tended to find that direct pointing devices do not fare as well as indirect pointing devices. However, this might relate the fact that driving a cursor onto an object involves quite different psychomotor skills to simply reaching out and touching the object. One might find that the principle differences between direct and indirect devices lie not so much in time taken to move toward a target as in the time available to make corrections to the movement.

3.3. User Preference

In many studies, users are asked to provide an indication of which device they prefer. One should bear in mind here that preferences are typically recorded by asking people to rank the devices. Consequently, one needs to be aware that ranking between devices A, B, and C might produce quite different results to ranking between devices A, B and F. However, one interesting pattern that emerges from these studies is that user preference typically follows a similar trend to speed; users tend to rate touchscreens higher than other devices, even when touchscreen accuracy is sufficiently low to require users to make repeated attempts to select an object.

4.0 FUTURE DEVELOPMENTS

Buxton (1986) has noted that the majority of pointing devices require one-handed operation. He argues that future computer systems ought to allow people to exploit bi-manual activity. Fitzmaurice and Buxton (1997) demonstrate a computer-aided design package which allows the user to combine activity with both hands using different devices. Design was also the focus of a device proposed by Kameyama and Ohtomi (1993); a

manipulable object which can be modeled and shaped as if it were made of modeling clay. This means that the physical activity of the person can be captured directly for use by the computer. An extension of this concept is the use of gestures as a means of interaction, e.g. through physical pointing at an object in space, or through "miming" the use of a specific tool, or through the use of sign language. An excellent introduction to these ideas can be found in Bolt (1984). These developments can be applied equally well to conventional desktop computers or to virtual reality.

REFERENCES

BABER, C., 1998, *Beyond the Desktop: Designing and Using Interaction Devices* (San Diego: Academic Press).

BOLT, R.A, 1984, *The Human Interface: Where People and Computers Meet* (Boston: Lifetime Learning Publications).

BUXTON, W., 1986, There's more to interaction than meets the eye: some issues in manual input. In D.A. NORMAN and S.W. DRAPER (eds), *User Centered System Design* (Hillsdale, NJ: Lawrence Erlbaum), pp. 319–38.

CARD, S.K., MORAN, T.P., AND NEWELL, A., 1983, *The Psychology of Human-Computer Interaction* (Hillsdale, NJ: Lawrence Erlbaum).

FITZMAURICE, G.W and BUXTON, W., 1997, An empirical evaluation of graspable user interfaces: towards specialized, space-multiplexed input. *Proceedings of CHI'97*, New York: ACM, pp. 43–50.

KAMEYAMA, K. and OHTOMI, K., 1993, A shape modelling system with a volume scanning display and multisensory input device. *Presence*, 2, 104–11.

MACKENZIE, I.S., 1992, Fitts' law as a research and design tool in human-computer interaction. *Human-Computer Interaction*, 7, 91–139.

SHNEIDERMAN, D., 1982, *Designing the User Interface* (Reading MA: Addison-Wesley).

Systematic Control of Exposure to Machine Hazards

J. Etherton

Division of Safety Research, National Institute for Occupational Safety and Health, J. E. NIOSH/Division of Safety Research, 1095 Willowdale Rd., Morgantown, WV 26505.

1. INTRODUCTION

Engineering controls to prevent human contact with hazardous machine motion are a vital element in working safely with machinery. In the hierarchy of steps that can be taken to prevent injury, engineered controls are less effective than eliminating the energy source entirely. Nevertheless, they are more effective than personal protective equipment or training. Typically, the safety professional's job, as a machinery builder, user, or modifier, is not to design machine safety devices, but to effectively apply purchased devices. The user should understand the limitations of particular devices and should select and use only those devices that eliminate task risks or reduce them to an acceptable level. This chapter is organized under the following headings: (1) introduction; (2) hazardous machine motion; (3) injury risks; (4) safeguarding; (5) safeguard reliability; and (6) safeguarding and mode of actuation.

2. HAZARDOUS MACHINE MOTION

Prime movers impart motion to the working elements and tooling of industrial machinery. By timely interruption of power to prime movers, motion can be stopped before a human is in the motion path. Machinery motion can result from: (1) normal operation; (2) inadvertent human actuation of control switches; or (3) a failure within the machine control.

For each of these circumstances, the way that components and devices are designed and arranged in the machine control system is critical to injury avoidance.

3. INJURY RISKS

For 1992–95, the US Bureau of Labor Statistics (1997a) reported 588 occupational fatalities involving caught-in-machinery. Fosbroke *et al*. (1997) identify three machinery occupations among the 50 highest fatal injury risk occupations. Their listing is based on lifetime risk/1000 workers, with an assumed 45-year working life. For millwrights, 9.17 of 1000 millwrights die of a fatal work injury before their working life is completed. For machinery maintenance personnel the rate is 7.32 deaths per 1000 workers, and for other precision metal workers it is 4.01. The highest occupation is timber cutting and logging with a lifetime risk of 62.7.

As for non-fatal injuries, for 1992–95 there were between 34 000 and 36 000 non-fatal injuries per year involving caught-in-machinery according to the US Bureau of Labor Statistics (1997b). The incidence rates (cases involving days away from work/10 000 full-time workers) for that same time period for all of private industry ranged from 5.0 to 4.2. The two highest rates were in the manufacturing industry (14.4–12.2) and agriculture (12.1–

5.3). The reasons for human exposure to machine hazards are diverse, but a recurring scenario involves clearing and cleaning.

4. SAFEGUARDING

Machinery safeguards control the actions of the operator or they control the machine as operators move toward hazards. To qualify as a valid safety control, standards of performance must be met. Performance requirements can be found in the standards of the Occupational Safety and Health Administration (OSHA), American National Standards Institute (ANSI) and International Standards Organization (ISO).

4.1. Barrier Guards

Barrier guards are a fixed, protective fence between the operator and the machinery hazard and are used during normal operation. This is the most basic form of protection because the person is kept away from machine motion. The machine controls operate independent of this safeguard. Barriers are suited to hand feeding or automatic feeding. On general-purpose machines, adjustable guards may be needed to accommodate the size of various work pieces.

4.2. Interlocked Guards

For guards that need to be moved from their protective position, safety interlock switches can be placed such that they send a stop signal when the guards are out of position. Interlocked guards are appropriate when frequent adjustment or cleaning is performed during normal operation. Replacing the movable guard should not result in machine motion immediately restarting. A separate, deliberate action should be required to initiate action. Mechanical mounting of switches must be such that they cannot be easily taped or wedged closed, bypassing their intended purpose. Protective covers are an important feature for interlocks used in contaminated environments.

4.3. Devices

Machine safeguarding devices rely on a machine's control circuit to stop motion before human contact. Human approach or entry into dangerous areas can be detected by sensors, constrained by hand attachments, kept at a safe distance by location of manual controls or impeded by gates that open and close automatically.

4.3.1. Presence-sensing devices

These devices generate a sensing field that detects the location of a person. They are installed and adjusted to initiate stopping of hazardous motion before it can be reached inadvertently. They only are applicable on machines that can be stopped in mid-cycle. There are two main types: electro-optical and radio-frequency (RF). Both types should establish a safety distance from the hazard, which is the distance at which they signal the machine to stop. If a person can step completely through a light plane, the device should issue a stop command. Care must be taken to see how far fingers can penetrate the plane of a light curtain and what effect this has on safety distance. Likewise, fluctuation in the sensitivity of a RF field should be adjusted for. These fluctuations can be due to other objects entering the field or changes in the grounding of the protected person.

4.3.2. Two-hand control

This type of device usually is comprised of machine activation palm buttons that are a safe distance from the point of operation. After releasing the buttons, time is not sufficient to reach the point of operation before it ceases to be hazardous. The safety distance that is used equates a reach time based on hand speed (1.6–2.0 m/s depending on waist-level or shoulder-level button location; Pizatella and Moll 1987) to machine response time (control signal processing time + braking time). Both buttons should be pressed almost at the same time, e.g. within 500 ms of each other, and the release of either button should stop the machine. Two-hand trip is used on machines that cannot be stopped in mid-cycle. In two-hand trip safeguarding, the distance between buttons and hazards must be great enough to prevent hazard contact before the machine's cycle is completed. A variation on the use of buttons is using two sensing devices that actuate the machine when each hand is on or near its respective sensing pad.

4.3.3. Safety mats

A safety mat detects a person standing on or stepping onto it and signals hazardous motion to stop. They should be fixed to the floor; large enough not to be stepped over; and initiate stopping before someone stepping toward the hazard at a normal speed can reach the hazard.

4.3.4. Automatic gates

Gates that automatically open and close are another safety device. Some automatic gates will lock closed until machine motion has stopped. Others will lock closed while tooling is closing and then open for parts feeding during non-hazardous machine motion away from the operator.

4.4. Emergency Stops

Emergency stops are manually operated stop controls intended for use when a person detects that they or a co-worker have become exposed to a hazardous condition, or to prevent further injury after being caught. The common human interfaces are large mushroom-shaped buttons, safety bars and trip wires. Hard-wiring these into the control circuit is recommended. The effectiveness of emergency stops depends on how easily they can be reached, how much resistance they have before actuating and how quickly the machine actually stops after they are used. Since they are not used very often, a periodic check that they are functional is appropriate.

4.5. Safeguarding by Distance

Some machines, such as press brakes, require humans to hold work at a safe distance from tooling. Workers do not come in contact with machine motion because the workpiece is between them and the hazard. The administrative controls and training needed are high and more effective safeguards should be used when possible. Safeguarding by location means that machinery hazards are out of reach overhead or at locations where only servicing or maintenance tasks are performed, using lockout/tagout procedures.

5. SAFEGUARD RELIABILITY

The level of risk for which safeguarding is most effective varies with respect to: (1) mechanical reliability, (2) control reliability and (3) accommodating human reliability.

5.1. Mechanical Reliability

Mechanical reliability is a safety control characteristic that relates to both machine components and to machine guarding. Brakes, clutches and mechanical power transmission components that are involved with the timing of machine motion will need to be of sufficient strength to be reliable. Mechanical strength is also a defining aspect of barrier guard effectiveness (European Committee for Standardization 1996a). Verification may be required for the resistance of guards to impact from persons, parts of tools, high-pressure liquids, etc. Before carrying out this verification it is necessary to identify the foreseeable impact hazard to which the guard may be subject, e.g. low velocity impacts from persons, high velocity impacts from broken parts of tools, impact from high-pressure fluids. When verifying the impact strength of a guard, it is necessary to take account of the properties of the materials from which the guard is constructed. This examination should include the strength of joints used and the strength of fixing points, slides, etc. by which the guard is attached to the machine or other structure.

Factor of safety (FS) is a common way to specify a design so that it will have adequate strength. Factor of safety for a guard can be defined as FS = S_u/S_a, where the ratio of the ultimate strength (S_u) of its material and fasteners to an allowable, working stress (S_a). Using a factor of safety permits the design to not waste material, yet be strong enough to prevent failure in case loads exceed expected values or other uncertainties react unfavorably. In general, the ductility of the material determines the property upon which the factor should be based. Materials having an elongation of > 5% are considered ductile. Factor of safety can be based upon the yield strength or the endurance limit. For materials with an elongation < 5%, the ultimate strength must be used because these materials are brittle and fracture without yielding (Avallone and Baumeister 1986). Factors of safety based on yield are often taken between 1.5 and 4.0. For more reliable materials or lower-risk design and operating conditions, the lower factors may be appropriate.

Guarding for agricultural equipment (ASAE 1993) is required to retain adequate strength under expected climatic and operational conditions for their intended use. The reference force that the guard must withstand a 123 kg (270 lb) person leaning or falling against it.

5.2. Control Reliability

Control reliability means the capability, by design, for a safety control to provide a protective function even when its design has been compromised due to component failure. OSHA's only control reliability standard applies to hands-in-die operation of mechanical power presses. OSHA 1910.217 (b)(13) requires that the control system will be constructed so that a failure within the system does not prevent the normal stopping action from being applied to the press when required, but does prevent initiation of a successive stroke until the failure is corrected.

The failure will be detectable by a simple test, or indicated by the control system. This requirement does not apply to those elements of the control system which have no effect on protection against point of operation injuries.

The logic embedded in a safety control system derives from a combination of relays, microprocessors, and/or pneumatic components. The failure modes of these components are quite

different. Electrical noise and power supply spikes can cause problems with microprocessors, but not with relays. Design should include: active redundancy; comparison between outputs of redundant components; and regular exercising of components to ensure their functionality. Shielding against electromagnetic interference can also be appropriate.

Five categories of safety device control reliability are set forth in EN 954, *Safety of Machinery; Safety-related Parts of Control Systems* (European Committee for Standardization 1996b). The categories in order of increasing fault tolerance are: controls that simply operate to perform the intended product function; well-tried safety control components; a fault in the safety circuit is detected on start-up (start-up check); a single fault in the safety circuit does not lead to failure of the safety circuit (redundancy); and a single fault is detected at or before the next call on the safety system. The interested reader should refer to the actual standard.

5.3. Design that Accommodates Human Reliability

People do not always do the safe thing. Sometimes they will reach toward hazardous locations, they will not put a safeguard back in place because it is too burdensome or they will simply bypass a safeguard. Industrial anthropometry is the science concerned with determining the dimensions of the human body and defining how these dimensions apply to work tasks. Percentile (%ile) is a convenient way of indicating how much of the range of human variability is accommodated by a given dimension, or the fraction of the population who measure less than this dimension. For example, the 95%ile US civilian male hand length is 20.6 cm; with 5% of the population having longer hands (Kroemer 1987). For hand-reach dimensions through long parallel openings, Villaincourt and Snook (1995) give the most current US dimensions.

The agricultural machinery standard ASAE S493 (1993) provides a set of rules for openings that will impede exposure of hands, arms, feet and the whole body to dangerous machine areas. The functional anthropometry of whole-body reaching should also be considered when installing barriers that can be leaned over, or for standing reaches (Etherton 1991).

The weight of a guard that can be completely removed from a machine is a factor in its easy replacement. Yet, guarding must be substantial enough for rugged use. Thus, there may be a design tradeoff if a guard is made of materials that are too dense.

6. SAFEGUARDING AND MODE OF ACTUATION
6.1. Foot Controls and Barrier Guards

Trump and Etherton (1985) describe the likelihood of inadvertent actuation with foot controls. When a foot control is used, safe-

guarding should be especially difficult to circumvent, or redundant safeguarding may be considered. When safeguards have been removed for cleaning, clearing or set up, users of foot controls should be particularly cautious. These guidelines also apply to foot-controlled press brake applications that are safeguarded by work holding at a safe distance.

6.2. Presence-sensing Device Initiation

Presence-sensing devices can be installed and operated automatically to cycle some machines once hands have been withdrawn from the sensing field. Because of the sensitivity limitations of some devices, this option should never be used without a thorough safety review of the entire machine and safeguarding system.

REFERENCES

AMERICAN SOCIETY OF AGRICULTURAL ENGINEERS, 1993, *Guarding For Agricultural Equipment*. S493 (St Joseph: ASAE).

AVALLONE, E. and BAUMEISTER, T., 1986, *Mark's Standard Handbook for Mechanical Engineers* (9th ed.) (New York: McGraw-Hill).

ETHERTON, J., 1991, Anthropometry applications for machine guarding. In Mital, A. and Karwowski, W. (eds) *Workspace, Equipment, and Tool Design* (New York: McGraw-Hill) pp. 205–17.

EUROPEAN COMMITTEE FOR STANDARDIZATION, 1996a, *General Requirements for the Design and Construction of Fixed and Moveable Guards*. EN 953 (Brussels: ECS).

EUROPEAN COMMITTEE FOR STANDARDIZATION, 1996b, *Safety Related Control Systems – Safety of Machines*. EN 954 (Brussels: ECS).

FOSBROKE, D.E., KISNER, S.M. and MYERS, J.R., 1997, Working lifetime risk of occupational fatal injury. *American Journal of Industrial Medicine*, 31,459–67.

KROEMER, K., 1987, Engineering anthropometry. In Salvendy, G. (ed.) *Handbook of Human Factors* (New York: Wiley-Interscience).

OCCUPATIONAL SAFETY AND HEALTH ADMINISTRATION, 1992, *General Industry Safety and Health Standards*. 29 CFR 1910 (Washington, DC: US Department of Labor).

PIZATELLA, T.J. and MOLL, M.B., 1987, Simulation of the after-reach hazard on power presses using dual palm button actuation. *Human Factors*, 29, 13–22.

TRUMP, T. and ETHERTON, J., 1985, Foreseeable errors in the use of foot controls on industrial machines. *Applied Ergonomics*, 16, 103–11.

US BUREAU OF LABOR STATISTICS, 1997a, *Census of Fatal Occupational Injuries, 1992–1995* (Washington, DC: USBLS).

US BUREAU OF LABOR STATISTICS, 1997b, *Occupational Injuries and Illnesses: Counts, Rates And Characteristics, 1992–1995* (Washington, DC: USBLS).

VILLAINCOURT, D. and SNOOK, S., 1995, A review of machine guarding recommendations. *Applied Ergonomics*, 26, 141–5.

Visual Display Technology

D. L. Post and H. L. Task

Air Force Research Laboratory, Wright-Patterson Air Force Base, OH45433-7022, USA

1. INTRODUCTION

Many of the design problems encountered by human factors professionals involve electronic visual displays. Although modern human factors training often covers the psychophysical, perceptual and cognitive aspects of information display, education about display hardware — which is often needed to understand and solve display design problems and work productively with display engineers — is rare. This problem is exacerbated by the fact that most of the literature on display hardware is written for engineers and many aspects of the field are evolving quickly. This entry summarizes the major human factors parameters that are relevant to electronic visual displays, explains the principles of operation for the display technologies of greatest importance today and describes their characteristics insofar as they affect visual performance, and discusses the most common types of display systems that incorporate these technologies.

2. DISPLAY HUMAN FACTORS PARAMETERS

2.1. Photometric Characteristics

One of the most fundamental display characteristics is brightness. Brightness is assessed usually by measuring its psychophysical correlate, luminance, with a photometer. Another basic characteristic is contrast, which is a ratio of on and off pixel luminances; Table 1 lists common measures. Ambient illumination often reduces display contrast by adding reflected light to that coming from the display. Display screens are usually treated to reduce this problem, although most treatments reduce display luminance.

Grayscale is the number of luminances the display can produce at a fixed setting of the brightness and contrast controls, expressed usually as the base-2 logarithm of that number with units called "bits." For example, "8 bits" of grayscale denote 256 levels. The visual system's response to differing luminances is compressive, so technologies having exponentially increasing video-input-to-luminance response functions come closest to providing equally discriminable grayscale levels (although single steps may not actually be discriminable).

For most display technologies, grayscale limits are not inherent and result instead from built-in digital drivers. The exceptions are technologies having pixels that can only be fully on or off. A few technologies — notably CRT — have analog grayscale, which means their luminance can vary continuously and grayscale is limited, if at all, only by their video input signal. Grayscale can be produced in technologies having only on and off pixels by temporal multiplexing, which involves turning pixels on for varying fractions of a frame period, or by dithering, which is a type of spatial multiplexing that involves turning groups of adjacent pixels on and off so the group average has the desired luminance. Usually, dithering is used only as a supplement to temporal multiplexing. Emissive on–off technologies can be dimmed only by using up some of their grayscale (non-emissives can be dimmed by adjusting the illuminant), so they may exhibit contrast artifacts if they are dimmed.

The angular distribution of light emission for some display technologies is approximately Lambertian, which means their luminance is essentially independent of viewing angle; others depart from this ideal in ways that usually make the display harder to read as one moves off axis. LCD often exhibit additional anomalies off axis, namely reduced or even reversed contrast and color shifts, although manufacturers are using increasingly sophisticated methods to reduce these problems.

2.2. Spatial Characteristics

"Resolution" is used widely, but inconsistently, to describe the amount of spatial detail a display can produce. Display designers often specify resolution in cycles or pixels per unit distance on the display (e.g. cycles/cm), where a cycle is one complete sine wave. Sometimes, the unit distance is implicitly the display's width or height, in which case the specification gives the total number of cycles or pixels that can be produced. Specifications for television sets often state only the number of raster lines that can be displayed. Although conventional television broadcasting provides only 525 lines in the USA (only 483 are visible, though) and 625 lines in the UK and Europe (575 visible), the capacity to display more implies higher resolution within each line.

Table 1. Common measures of contrast.

Name	Defining equation[1]
Contrast ratio[2,3]	L_{max} / L_{min}
Relative contrast[2,3]	$(L_{max} - L_{min}) / L_{min}$
Luminance contrast[3]	$(L_{max} - L_{min}) / L_{max}$
Luminance modulation or Michaelson contrast	$(L_{max} - L_{min}) / (L_{max} + L_{min})$

[1] L_{max} is the greater of the two luminances; L_{min} is the lesser.

[2] The contrast ratio and relative contrast are undefined when $L_{min} = 0$.

[3] For symbology, L_{max} and L_{min} are replaced sometimes by the symbol and background luminances, respectively, in the first three equations, which creates the possibility of contrast ratios < 1 and negative values in the other two cases.

Table 2. Contemporary digital resolution standards.

Label	Aspect ratio[1]	Resolution (H $\times$ V) in pixels
QVGA	4:3	320 $\times$ 240
VGA	4:3	640 $\times$ 480
SVGA	4:3	800 $\times$ 600
XGA	4:3	1024 $\times$ 768
SXGA	5:4	1280 $\times$ 1024
UXGA	4:3	1600 $\times$ 1200
SDTV[2]	4:3 or 16:9	704 $\times$ 480
HDTV[3]	16:9	1920 $\times$ 1080
	16:9	1280 $\times$ 720

[1] Aspect ratio is the ratio of image width to height.

[2] SDTV approximates the resolution of the 525-line analog system used currently in the USA. It requires rectangular pixels in one of two possible sizes to achieve the stated aspect ratios. All other entries in Table 2 use square pixels.

[3] HDTV offers two possible resolutions.

Psychophysicists usually prefer cycles or pixels per unit angle subtended at the viewer's eye (e.g. cycles/degree) because these measures account for the effects of viewing distance and display size on image appearance. Thus, two displays having equal resolution from a display designer's perspective may have different resolutions from a psychophysicist's perspective and vice versa.

Additional confusion is introduced sometimes by failing to distinguish between the resolution of the display and that of the video system driving it or the image it is displaying. For example, using an SXGA (Table 2) graphics card does not give a display SXGA resolution, nor does showing a VGA image yield VGA resolution. Another poor practice is confusing resolution with the amount of visual information in an image. For example, a photograph may be described as having higher "resolution" than a drawing of the same scene.

CRT produce images by drawing on a screen with a continuously moving electron beam. Their resolution is mainly a function of their video bandwidth, refresh rate and electron-beam spot size. These parameters' effects are described most completely by the spatial modulation transfer function, which graphs the response to sinusoidal inputs as a function of their frequency, in a manner analogous to the frequency–response plots one often sees for audio equipment.

Other display technologies consist of a matrix of discrete pixels that are set to the desired luminances by addressing them one or possibly several rows at a time, with all columns loaded simultaneously — methods called "line by line" and "multi-line" addressing, respectively or, more generally, "matrix addressing." For these technologies, resolution is simply a function of the number of horizontal and vertical pixels. Another addressing method is to connect each pixel to its own driver. This "direct" addressing technique allows each pixel to be driven throughout each frame, but is practical only for simple applications, requiring only a few dozen pixels or segments (e.g. watches, calculators, and home appliances) because otherwise the number of connections and drivers becomes prohibitive.

Matrix displays come in two types: passive, which have only a pair of electrodes at each pixel, and active, which add a switch (typically, a thin-film transistor, or "TFT," but sometimes a diode) at each pixel so voltage can be maintained throughout the frame period. Passive-matrix displays are cheaper but can apply voltage to the pixels only while their row is selected. Therefore, their luminance and contrast tend to be inferior to otherwise identical active-matrix displays and diminish as the number of rows increases, thus limiting the number of rows that can be provided.

2.3. Temporal Characteristics

It is obvious that changing a display image requires redrawing it, but it is less obvious that, for most display technologies, the image fades unless it is refreshed. (The exceptions are bi-stable technologies, which have pixels that remain on or off until they are switched.) In the USA, television refreshes at 30 Hz by alternately drawing all the odd- and then even-numbered lines at 60 Hz. Each set of lines is a "field," both fields constitute a "frame," and this alternating system is called "2:1 interlacing." (In the UK and Europe, the field rate is 50 Hz, yielding a 25-Hz frame rate.) Interlacing reduces bandwidth requirements while avoiding flicker for typical television images, but tends to produce jitter in text. Therefore, computer images are usually non-

interlaced (this method is known also as "progressive scanning") and refresh at 50 Hz or more. The rate needed to prevent flicker increases as display luminance, contrast, and angular subtense increase. Sometimes, the rate must also be adjusted to prevent beats with fluorescent lighting from becoming visible.

Display luminance is greatest if the pixels emit throughout each frame, but this can cause image smear if the observer tracks a moving object. Therefore, motion rendition is best for display technologies that produce adequate time-averaged luminance while emitting only for small fractions of a frame period. A display having pixels that cannot switch completely during one frame will exhibit contrast losses for moving images, so it is also important that the pixels respond quickly. Most LCD technologies are troublesome in this regard. Speed is even more crucial for technologies that rely on temporal multiplexing to achieve grayscale or use temporal integration to produce color, because in these cases the pixels must switch more than once during each frame.

2.4. Colorimetric Characteristics

The range of colors a display can produce is called its "color gamut" and is determined by its red, green and blue (RGB) primaries. The closer the primaries come to being monochromatic, the more saturated they become and the larger the gamut becomes; however, no finite set of primaries can reproduce the entire range of colors humans can perceive. Furthermore, increasing a primary's saturation requires reducing its spectral bandwidth, which typically reduces its luminous efficacy; therefore, primary selection often involves a tradeoff between gamut and luminance.

Television broadcasting uses a standard set of primaries that have been chosen for compatibility with available CRT phosphors. Developers of other color display technologies generally try to match or exceed television's gamut. For emissive technologies, one approach uses a white emitter with a patterned RGB color filter in front; another uses RGB emitters, possibly enhanced by a patterned color filter. For non-emissive technologies, the common approach uses a white illuminant and color filters. Luminous efficacy can often be increased by selecting an illuminant having its radiance concentrated in three RGB wavebands and matching the filters' transmittances to those wavebands.

Every color display has a "white point," which is the chromaticity it produces when its RGB channels are driven equally and may or may not be user-adjustable. Television broadcasting and a few other applications use standard white points, defined in terms of Commission Internationale de L'Éclairage (CIE) chromaticity coordinates. Differences in white points are often at least partly responsible for problems in obtaining accurate color transfer when images are viewed on different displays, printed, or scanned.

Most displays produce color by spatial integration: The screen is divided into RGB subpixels, which subtend a visual angle so small that the diffraction patterns they produce on the retina overlap and mix. Varying the subpixel luminances produces what appear to be full-color pixels having any color an RGB mixture can produce.

A similar method, used by most projection displays, is addition: Separate RGB images are produced by three

monochrome displays and superimposed internally or at the screen. Addition can yield higher resolution than spatial integration because every pixel can take on any color, but it requires three displays and is therefore bulkier and more expensive.

Color can also be produced by temporal integration — a technique often called "field sequential": the display generates spatially superimposed RGB fields in rapid sequence, within a single frame period, and the visual system integrates them temporally to form a full-color image. Temporal integration has the same resolution advantage as addition, but requires fast displays, sacrifices luminance because each color field is displayed for no more than one-third of each frame, and the images can break up into their RGB components if there is relative motion between the image and the viewer's eye.

Finally, a subtractive process can produce color: white light passes through a stack of three transmissive displays, and each display modulates the R, G, or B component of the light so a full-color image emerges at the front. Subtraction also has a resolution advantage, but requires three displays, careful design to minimize transmission and contrast losses as the light proceeds through three apertures, thin displays to avoid parallax in direct-view applications, and may require special optics to focus the layers simultaneously in head-mounted and projection applications.

2.5. Size and Weight

CRT tend to be the heaviest and bulkiest of the display technologies because they are funnel-shaped glass vacuum tubes. Their shape results from the need to provide space for the electron beam to scan the screen. Their weight increases exponentially with screen size because the glass must be thickened to withstand the atmospheric pressure resulting from the internal vacuum. Most other technologies consist only of thin, lightweight layers sandwiched between two flat pieces of glass, so they are more compact and their weight is proportional with screen size. LCD usually require a backlight, though, which can increase their weight and depth substantially.

3. EMISSIVE DISPLAY TECHNOLOGIES

3.1. Cathode-ray Tube (CRT)

A CRT produces images by modulating an electron beam as it scans across a phosphor screen. The electrons are produced by heating a cathode, accelerated toward the screen, and focused and scanned by dynamic electromagnetic or electrostatic fields. Dynamic focusing compensates for the changing distance to the screen as the beam scans. Most CRT scan the beam in a repeating "raster" pattern, consisting of closely spaced horizontal lines; however, oscilloscope CRT — and some that are used exclusively to show symbols and line graphics — use a "stroke" pattern that moves the beam as if it were a pen. Stroke writing provides increased luminance (because the phosphor can be excited more times per s) and resolution.

Most color CRT scan three electron beams across a patterned screen containing RGB phosphor dots or stripes, thus producing color by spatial integration. A metal structure ("shadowmask") behind the screen, having one opening for each RGB triplet, helps ensure that each beam strikes only the appropriate phosphor; adjustable magnets also assist. Older, "delta-gun" CRT arrange the beams in a triangular pattern and include user-adjustable

circuitry to ensure a condition called "convergence" that, if violated, produces color fringes that are noticeable especially around white objects on the screen and reduce resolution. Modern, "in-line gun" CRT arrange the beams in a row and use circuitry and magnets, which are set at the factory and rarely adjusted afterward, to achieve convergence. Color can also be produced by temporal integration, using a white-phosphor CRT in conjunction with a switchable or rotating color filter — this approach is used sometimes in head-mounted systems.

3.2. Vacuum Fluorescent Display (VFD)

The VFD is a type of flat CRT that excites a low-voltage phosphor using electrons produced by heating a cathode grill consisting of fine, heated wires. The flow of electrons is controlled by a wire grid that lies between the cathode and a phosphor-coated opaque anode array. The photons then pass through the grid, cathode, and a cover glass. Another arrangement, which deposits the phosphor on the cover glass, is also used sometimes. The anode array can be matrix addressed or, more commonly, arranged in segments. Usually, a blue-green emitting phosphor is used because it has the highest luminous efficacy, but a good range of hues up to red are available.

VFD are used widely as simple numeric, alphanumeric, and bargraph displays measuring a few square centimeters in products like VCRs, but it has not been practical to > ~160 cm^2 for more general-purpose use because this would require adding internal support to control the gap between the cathode and anode, which would interfere with the electron flow. Grayscale and dimming can be produced by varying the anode voltage. Motion rendition is good due to low response time and the emission is Lambertian, so viewing angle is not a problem. Color can be produced by spatial or temporal integration, but multicolor units consisting of differently colored monochrome areas are far more common.

3.3. Field-emission Display (FED)

The FED is another type of flat CRT, in which a phosphor-coated glass screen is excited by electrons produced by an addressable matrix of tiny cathodes. There are basically two types: low-voltage (40–100 V), which require special phosphors and have not yet matched the luminances available from CRT, and high-voltage (5–10 kV), which use conventional phosphors and can match CRT luminances but are vulnerable to damage and failure caused by arcing. Grayscale and dimming can be produced by temporal multiplexing, cathode modulation, or dithering, although multiplexing supplemented by cathode modulation usually yields the best result. The response time is low and the emission is Lambertian. Color is produced by spatial integration, using RGB phosphors, although temporal integration is also feasible.

3.4. Image Intensifier Tube (I²T)

The I²T is a unique combination of sensor and display in one package. Contemporary third-generation I²T consist of three primary elements: (1) a photo-cathode screen, which converts photons to electrons; (2) a micro-channel plate (MCP) that accelerates and multiplies the electrons; and (3) a phosphor screen — coated on a fiber-optic or plain glass faceplate — that converts the multiplied electrons to photons. Typically, a power supply is included, to convert battery power to the relatively high voltage needed by the MCP. In operation, an image is focused onto the

photo-cathode, which then emits electrons in the general direction of the adjacent MCP. The MCP contains millions of tiny tunnels (micro-channels), which amplify the incident electrons in proportion to their number. After exiting the MCP, they are accelerated toward the nearby phosphor screen by an electric field. The phosphor nearest each micro-channel glows with a luminance that is proportional to the number of incident electrons, thus producing a monochrome image that replicates the image focused on the photo-cathode, except that its luminance is greatly amplified. See the section on NVG for information on typical uses of I²T.

3.5. Light-emitting Diode (LED)

An LED is basically a solid-state material — typically gallium phosphide or gallium arsenide — that has been treated so it consists of two layers that form a P-N junction diode, and passing a current across the layers causes them to emit light. A reflector is placed behind the diode to direct rear-emitted light toward the front, and a lens is often incorporated at the front to further shape the output. The diodes can be arranged in segments to form numeric, alphanumeric, or bargraph displays or laid out as an addressable matrix. A full range of colors, including RGB, is available, so color displays using spatial integration can be created. Grayscale and dimming can be produced by varying the current or temporal multiplexing. The reflectors and lenses are usually designed to produce non-Lambertian emission, to increase on-axis luminance. Large matrices of LED for use as general-purpose desktop color displays have not proven to be practical, mainly because of the difficulty of assembling discrete devices at the required densities (producing different colors on one substrate has proven impossible, thus far), but are used widely as message signs. Numeric units — which were once common in wristwatches and calculators until being replaced largely by LCD — are used in clocks and, along with alphanumeric and bargraph units, as instrument readouts.

3.6. Organic Light-emitting Diode (OLED)

An OLED display consists of a thin transparent layer of organic semiconductors or polymers, sealed usually between two glass or plastic substrates bearing orthogonally oriented electrodes on their inner surfaces. The front electrode is transparent and the rear is reflective. Applying a current across the electrodes causes the organic layer to emit visible light. Plastic substrates may make it possible to produce flexible displays, which can be rolled up or folded. Silicon substrates have also been used in experimental, miniature active-matrix OLED designed for head-mounted applications. Currently, this technology is under development, and although it is regarded widely as very promising, there are few commercial products. Grayscale and dimming are produced by varying the electrode current. Viewing angle is good due to Lambertian emission. OLED materials capable of producing RGB light are available, so full color can be produced by spatial integration or, if the substrates are stacked, by addition; however, in the latter case, the stack must be very thin to avoid parallax.

3.7. Electroluminescent (EL) Display

The most common type of EL display today uses alternating current to excite a thin-film EL phosphor, yielding the "ACTFEL" structure. The phosphor is sandwiched between thin insulating layers and this multi-layer structure is deposited typically on a glass substrate bearing transparent electrodes. Reflective, orthogonally oriented electrodes are then applied and the phosphor is viewed through the glass. Thus, only one substrate is required and the display can be substantially thinner and lighter than other flat panel technologies. Ceramic and silicon (supported by ceramic) substrates are also used, although these designs require a cover glass. Miniature devices using active-matrix addressing have been produced for head-mounted use. The angular emission is Lambertian. Grayscale and dimming for EL use temporal multiplexing, enhanced sometimes by dithering. Color is achieved by spatial integration, using either patterned phosphors or a white phosphor plus color filter array.

3.8. Plasma Display Panel (PDP)

A PDP consists mainly of a gas that is sealed between two glass substrates bearing orthogonally oriented electrodes on their inner surfaces and, usually, a reflective coating at the rear. Passing a current across the electrodes ionizes the gas, producing a plasma that emits light. Monochrome PDP use neon, which emits reddish-orange light. In color displays, an ultraviolet-emitting gas is used and the ultraviolet excites phosphors that are coated on the inner surface of one of the substrates. Thus, a color PDP is similar to a fluorescent light bulb. Either alternating or direct current can be used, but the former is more common today. PDP are inherently bi-stable, so grayscale and dimming are produced by temporal multiplexing, enhanced sometimes by dithering. Response is very fast, so motion rendition is good. Color is produced by spatial integration. The emission angle is Lambertian for color panels but is isotropic for monochrome panels, causing luminance to increase with viewing angle.

4. NON-EMISSIVE DISPLAY TECHNOLOGIES

4.1. Illumination Sources

Non-emissive displays require a light source. For some applications, such as LCD wristwatches, ambient illumination reflected off the display may be sufficient, but usually a built-in source is needed. Most direct-view LCD use one or more tubular fluorescent lamps, mounted in a box containing a reflector or coupled to a light guide. Other approaches that provide more uniform illumination have been tried, though, including EL panels, flat and serpentine-shaped fluorescent lamps, and CRT. Projection systems require more light than can be provided compactly by a fluorescent source; they use xenon or metal-halide arc lamps. Xenon has a broad spectrum that must be filtered heavily to produce acceptable RGB primaries, thereby reducing overall luminous efficacy. Metal-halide lamps emit mainly in narrow wavebands dictated by the choice of halides, thereby reducing the need for spectral filtering; the fill gas is unavoidably excited too, though, adding at least some broad-spectrum light. Compact, solid-state lasers producing RGB wavelengths have become available recently. Their narrowband spectra are advantageous for many display purposes, and their prices, efficiencies, lifetimes, sizes, weights and cooling requirements are moving rapidly to a point that may make them attractive for projection uses.

4.2. Liquid Crystal Display (LCD)

Liquid crystals (LC) are elongated molecules that flow like liquids

at typical ambient temperatures and tend to align with one another. There are many types of LC and ways of using them to make displays, but the most common today are nematics, sealed between two pieces of glass that have been treated so the LC lie parallel with the glass surfaces and undergo a rotation top to bottom, forming a helix — a configuration called "twisted nematic" (TN). In a transmissive TN-LCD, polarizers and transparent electrodes are added to the glass surfaces, and light passing through one polarizer rotates with the twist until it strikes the second polarizer, which either passes or absorbs it, depending on whether the polarizers are aligned with each other or crossed. Applying a voltage across the electrodes rotates the LC so they are perpendicular to the glass, destroying the twist and thereby reversing the second polarizer's effect. Intermediate voltages produce intermediate transmittances and, hence, grayscale. Reflective TN-LCD have a reflective surface at the rear and only a front polarizer, but work basically the same way. The electrodes can be segmented or arranged in a matrix.

Most passive-matrix LCD today use a twist ranging from 180 to 270°, called a "supertwisted nematic" (STN) configuration, which increases the number of rows that can be addressed but reduces viewing angle, response speed, and grayscale. Active-matrix LCD (AMLCD) provide the best viewing angle, speed, and grayscale, but are more costly. Other important types of LCD today are ferroelectrics (FLC), which are bi-stable and respond to control voltages very quickly but have difficulty producing grayscale, polymer-dispersed (PDLC) and nematic curvilinear aligned phase (NCAP) LCD, both of which vary between transparent and light-scattering states and use no polarizers but have trouble producing high contrasts, and polymer stabilized cholesteric texture (PSCT) LCD, which are bi-stable and therefore do not produce grayscale readily.

4.3. Digital Micromirror Display (DMD)

A DMD is an addressable array of tiny mirrors, each of which is mounted on a flexible stalk. Each mirror can be rotated very quickly (e.g. 10 ms) to either of two orientations, thereby deflecting incoming light out to a projection lens or an absorber. Thus, each mirror is a pixel that can be either on or off. Grayscale is produced by temporal multiplexing; dimming is accomplished by modulating the illumination. Color is produced usually by temporal integration using a filter wheel, but triple-DMD systems using addition and dual-DMD hybrids are also available. DMD do not lend themselves to direct-view applications because of the need for a projection lens; furthermore, it is impractical at the moment to make them bigger than a few square centimeters in area. They are used widely in projectors, though, and head-mounted applications are also possible.

5. OPTICAL DISPLAY SYSTEMS

5.1. Projectors and Screens

Projection displays use one or more lenses to enlarge and focus an image of an internal display onto a screen. Rear projectors usually incorporate the screen into the housing and therefore tend to be bulky, whereas front projectors use a separate screen and are therefore more compact but require the screen to be distant from the projector. Rear projection screens are usually treated to reduce reflections from room lighting and incorporate lenses to produce a non-Lambertian light distribution that reduces the image's viewing angle but provides greater luminance within the intended angle — a feature called "gain." Front projection screens can also have gain.

Some projectors use CRT to create an image; others, termed "light valves," use non-emissive displays. The least expensive, lightest, and most compact design for a color projector uses a single display and lens system; field-sequential color is most common in self-contained versions, whereas spatial integration is the norm in LCD overlays for overhead projectors (subtractive overlays have also been produced). Greater luminance can be obtained, though, by using three displays to produce separate RGB images that are combined additively; this design is more common in self-contained systems. Most triple-display projectors use dichroic beamsplitters, which reflect one band of wavelengths while passing the rest, to add the images internally and, in light-valve projectors, to first separate white light into R, G, and B so each component can illuminate the appropriate display. Most CRT projectors, however, use a separate projection lens for each CRT and add the RGB images at the screen.

The most important human factors issues for projectors are resolution, luminance, contrast, optical distortion, viewing angle, size, and weight.

5.2. Head-up Display (HUD)

As the name implies, a HUD is intended to permit the operator of a vehicle (originally an aircraft and, more recently, an automobile) to view a display without looking down toward the instrument panel. It does this by superimposing a display image on the outside world. The original intent was only to assist weapon aiming, but another advantage has been recognized since: The virtual image is normally at or near optical infinity, so the operator does not have to refocus when switching from viewing the world scene to viewing the display.

A HUD consists of three major components: (1) an image source (e.g. a CRT); (2) optics to produce a virtual image of the image source; and (3) an optical combiner, which combines the virtual image with the directly viewed exterior world scene. There are basically two types of HUD: those that produce a real exit pupil and those that do not. If the HUD uses relay optics, it has a real exit pupil, which means at least one eye must be within the exit pupil to see any part of the display. If the HUD uses simple magnifying optics, it does not have a real exit pupil and eye position is less critical, although at least one eye must still be positioned within a limited volume of space to see the entire display.

The most important human factors issues for HUD are the combiner ratio (which affects the relative luminances of the symbology and outside world scene), viewing location, exit pupil size or viewing volume, symbology size, type, luminance, and color, stray reflections, and the size and shape of the field of view.

5.3. Head- and Helmet-mounted Display (HMD)

An HMD consists typically of three major components: (1) a miniature display; (2) optics to convey a virtual image of the display to the eye(s); and (3) some (usually adjustable) means of mounting the display and optics to the head. The optics may or may not include a combiner: If the HMD is meant to provide an image that can be superimposed on the external world scene,

the optics must include a see-through combiner; if the HMD is for "immersion" only, a combiner is unnecessary and the optics are easier to design. Most HMD use miniature CRT or LCD, which must have very high resolution (in terms of cycles/mm or pixels/mm) to produce a satisfactory image after magnification by the optical system. One exception is the "virtual retinal display," which produces an image by scanning a modulated laser beam directly across the retina in much the same way a CRT scans an electron beam across a phosphor screen. Another exception scans an image of an LED array back and forth across the retina.

HMD can be monocular, bi-ocular, or binocular, monochrome, multi- or full-color, see-through or non-see-through, adjustable focus or fixed focus, and real exit-pupil forming or non-pupil forming. The most important human factors issues associated with them are weight, comfort, size, luminance, binocular alignment, resolution, optical distortion, optical distance of the image, size and shape of the field of view, exit pupil size or viewing volume, and the distance the eye can be from the optics ("eye relief").

5.4. Night Vision Goggle (NVG)

NVG are a special type of HMD. Most NVG are binocular, with each ocular consisting of three major components: (1) the objective lens, which produces an optical image on the input side of an (2) I^2T; and (3) an eyepiece lens to produce a virtual image of the I^2T's output. NVG typically include assemblies that permit them to be adjusted and worn on the head. NVG containing second-generation I^2T are most sensitive to wavelengths from 400 to 750 nm; third-generation devices are most sensitive to 650–900 nm.

The most important human factors issues associated with NVG are ocularity (binocular, bi-ocular, or monocular), output luminance, signal to noise ratio, optical distortion, eye relief, gain, adjustability (fore/aft, interpupillary distance, focus, tilt, up/down), weight, size, and comfort.

5.5. Three–dimensional (3D) or Stereoscopic Display

"3D" is used sometimes to refer to images seen on a conventional display that include depth cues such as perspective, surface shading and hidden lines. A stereoscopic 3D display, however, provides binocular disparity, so each eye sees a slightly different image that corresponds to what it would see in the real scene. There are several ways to produce binocular disparity. One is to use two independent image sources (e.g. CRT) and an optical system that presents a different image to each eye. Another approach is to use a single display and encode or multiplex the left- and right-eye images so each eye sees only the appropriate one. A common encoding technique, used with projection displays, polarizes the left- and right-eye images oppositely and

superimposes them on the screen; the observers wear oppositely polarized lenses, so each eye sees only the appropriate image. A common multiplexing technique, used with direct-view displays, places electronic shutters in front of the eyes and alternates the left- and right-eye images on the display in synchrony with the opening and closing of the shutters.

The most important human factors issues associated with stereoscopic displays are the accuracy of the binocular disparities and hence depth cues, visibility of the images to the wrong eyes ("crosstalk"), frame rate, luminance, color, binocular alignment, and the effects of mismatch and lack of normal covariance between the images' optical distance (which is usually fixed) and the varying vergence cues and requirements produced by the binocular disparities.

6. FURTHER READING

For further information about display human factors issues, see Tannas (1985), Travis (1991), and Widdel and Post (1992). For display technology, see Csorba (1985), Tannas (1985), Sherr (1993, 1998), MacDonald and Lowe (1997) and Nelson and Wullert (1997). For optical display systems, see Weintraub and Ensing (1992) and Melzer and Moffitt (1997).

REFERENCES

CSORBA, I.P., 1985, *Image Tubes* (Indianapolis: Howard W. Sams).

MACDONALD, L.W. and LOWE, A.C., 1997, *Display Systems Design and Applications* (Chichester: Wiley).

MELZER, J.E. and MOFFITT, K. (eds), 1997, *Head-mounted Displays* (New York: McGraw-Hill).

NELSON, T.J. and WULLERT II, J.R., 1997, *Electronic Information Display Technologies* (River Edge: World Scientific).

SHERR, S., 1993, *Electronic Displays* (2nd ed.) (New York: Wiley).

SHERR, S., 1998, *Applications for Electronic Displays* (New York: Wiley).

TANNAS, L.E., Jr (ed.), 1985, *Flat-panel Displays and CRT* (New York: van Nostrand Reinhold).

TRAVIS, D.S., 1991, *Effective Color Displays: Theory and Practice* (San Diego: Academic Press).

WEINTRAUB, D.J. and ENSING, M., 1992, *Human Factors Issues in Head-up Display Design: The Book of HUD* (OH: Wright-Patterson AFB, CSERIAC).

WIDDEL, H. and POST, D.L. (eds), 1992, *Color in Electronic Displays* (New York: Plenum).

Visual Fatigue and Stereoscopic Displays

T. Takeda

Faculty of Engineering, University of Tokyo, 7-3-1 Hongo, Bunkyo, Tokyo, 113-8656, and Core Research for Evolutional Science and Technology (CREST), Japan Science and Technology Corporation (JST), Japan

1. INTRODUCTION

The use of new visual informational displays such as head-mounted displays for virtual reality, heads-up displays for automobiles and stereoscopic displays for industrial design, medical training and entertainment has been steadily growing. Along with this increased use, there are scattered reports (Peli 1996) that visual fatigue might be induced by prolonged use of those displays. Although accommodation and vergence have a close relationship with such phenomena, few objective studies exist.

This chapter firstly explains specific characteristics of accommodation when looking at two-dimensional (2D) drawings with ample depth cues. Then it presents the difference between visual responses when subjects are looking at a moving object in a realistic or virtual three-dimensional (3D) space. The fact that accommodation can be evoked by 2D drawings from an acquired depth sensation and the remarkable difference in the accommodative responses in the real or virtual 3D space are attributed to be a cause of visual fatigue while using these new stereoscopic visual informational displays.

The control mechanism of the eyes consists of accommodation, eye movements and pupil reflex subsystems. These subsystems interact with each other considerably and, hence, are called the "near triad" system. Primary characteristics of each subsystem have been studied objectively over 40 years. However, the dynamic features of accommodation in realistic visual environments and especially the interactions between the subsystems have not been thoroughly clarified. The main reason has been that accommodation could not be measured while subjects were looking at real objects while moving their eyes.

In a pioneering study of accommodation to apparent distance, Ittelson and Ames (1950) reported that (1) with monocular vision, subjects showed a mean accommodation of 0.46 D (diopter, which is defined as the reciprocal of focal distance) and vergence toward the apparent distance indicated by the continuously varying size of a card presented by a projector, and (2) with binocular vision, the subjects showed no vergence.

Kruger and Pola (1989) recorded definite accommodative responses for sinusoidal size changes of the stimulus with an infrared optometer. They analyzed the role of blur, chromatic aberration, and size change of a target and measured the influence of the size change on accommodation. However, the stimuli were presented in an optical system that required subjects to gaze into the instruments monocularly.

Enright (1987) measured the vergence response monocularly with two TV monitors when subjects scanned line drawings. He found that the vergence changed if subjects shifted fixation within the drawings while one eye was covered, but the vergence did not change with binocular vision. However, he could not measure accommodation simultaneously.

Takeda *et al.* (1990) measured accommodation and eye movement of the right eye while subjects were looking at stimuli binocularly under normal viewing conditions. Then they measured the influence of line of sight and found that (1) vertical shifts of eye direction had a slight influence depending upon focus distance, and (2) horizontal shifts had virtually no influence (Takeda *et al.* 1992). Afterwards, they identified objectively the influence of size change of a spotlight in a totally dark room. It was found that accommodation was really influenced by the depth sensation induced by the change of spotlight size.

However, they could not measure vergence in those experiments because the apparatus used could not measure left eye movement. They improved the apparatus to enable vergence measurement. The results showed that accommodation was sensibly influenced by many kinds of depth sensation but vergence was less influenced by such sensation under binocular viewing conditions. Then the characteristics of accommodation were explored when looking at stereoscopic images and some peculiar responses were found and thought to be a cause of visual fatigue.

2. METHODS

2.1. Experimental Design

The principle of measurement system TDOIII (3D optometerIII) and the stimulator TVS (3D visual stimulator) are explained elsewhere in this volume.

In the first two experiments, two different drawings with ample depth cues (*Christina's World 1948* by A. Wyeth; *Mt Fuji Viewed Through Waves off the Coast of Kanagawa* by K. Hokusai) were used to examine what kind of influence on accommodation might be posed by watching them while subjects are moving their eyes. The measurement was performed under normal lighting conditions. The drawings were placed 40 cm from the subjects' eyes.

In the third experiment, a Maltese cross image formed by the TVS was used to examine the influence of a stereoscopic virtual image on accommodation. The visual angles of the target were 0.67° at 1 m (1D) and 0.67 or 2.00° at 33 cm (3D). The Maltese cross was white with a black background. The luminance of the cross was set at 6 cd/m², which was determined to avoid any unexpected influence of stray reflections, to lessen the influence of changes in light amount on pupil size and also to increase the pupil size to make measurement easier. As the TDOIII and the TVS were covered with a black cloth, the target was the only visible light and, hence, was subjectively rather bright.

2.2. Subjects

Subjects were two males and three females who had visual acuity of > 1.0 with correction. Their age was 31.8 ± 5.9 years and accommodative amplitude 5.2 ± 1.2 D (mean ± SD). S5 was −5.5 D myopic and wore a soft contact. All subjects were trained as observers for visual experiments generally, but they had no prior experience of experiments with the TDOIII. They were not informed as to the purpose of the experiment in which they were participating. Their only defined task was to obtain as clear an image as possible.

Table 1. Amount of evoked accommodation and overshoot when subjects gazed at the three different stimuli. Upper rows of respective stimuli are the evoked accommodations (EA) and lower rows are overshoots (OS); SD, standard deviation. (a) Wyeth; (b, c) Hokusai; (d) Realistic target, condition R; (e) stereoscopic target, condition S. Apparent depth sensation of 2D stimuli evoked real accommodation. Larger evoked accommodations resulted from stimuli associated with binocular parallax; overshoots were evoked by stereoscopic images.

Stimuli		S1	S2	S3	S4	S5	Mean	SD
a	EA	0.59	0.68	0.68	0.65	0.78	0.68	0.07
Wyeth	OS	0.07	0.32	0.23	0.21	0.33	0.23	0.10
b	EA	0.25	0.29	0.41	0.36	0.44	0.35	0.08
Hokusai (Left)	OS	0.08	0.17	0.10	0.12	0.11	0.12	0.03
c	EA	0.29	0.26	0.29	0.32	0.31	0.29	0.02
Hokusai (Up right)	OS	0.10	0.10	0.08	0.09	0.14	0.10	0.02
d	EA	1.24	1.52	1.51	1.59	0.82	1.34	0.32
Condition R	OS	0.26	0.17	0.28	0.09	0.28	0.22	0.08
e	EA	0.33	0.79	0.75	0.50	0.57	0.59	0.19
Condition S	OS	0.13	0.81	0.34	0.26	0.58	0.42	0.27

2.3. Procedure

Visual conditions were changed stepwise every 10 s by computer control (beep or image change) to get 60 single responses. One session consisted of five presentations of each state and 12 sessions were done with 30-s interval recesses, during which subjects were instructed to relax their eyes at their own discretion. All the experiments were carried out binocularly and the 60 responses were averaged. Steady-state accommodation was defined by the average of the last 3 s of each response to eliminate transient responses. Then the difference between the two steady-state accommodations was defined as the evoked accommodation level in the two different visual conditions (EA in Table 1). The peak of accommodation was determined by the initial peak value. Overshoot (OS in Table 1) was defined to be the absolute value of the difference between the peak and the evoked steady state accommodation. Vergence was calculated by subtracting the horizontal eye positions of both eyes.

3. RESULTS

3.1. Actual Artwork Experiments (Experiments 1 and 2)

Subjects were instructed to gaze binocularly at predetermined points in a reproduction of *Christina's World 1948* (Figure 1a). The photograph was 31 x 23 cm and was placed 40 cm (i.e. 2.5 diopters) from the subject's eyes. They gazed alternately at N and F for 10 s each. Gaze positions were not recorded on the picture, but their location was specified verbally. Averaged responses of subject S1 are shown in Figure 1b in which the subject shifted his eye position at 5 and 15 s. If accommodation was purely controlled by blur there would be no need to change accommodation while looking at the drawing because it is kept at a fixed distance. However, when the subjects shifted their eye position from the shoulder of Christina (indicated by N) and looked at the horizon near the small hut (F), accommodation level (Acc) clearly receded. Inversely, when they looked back at the shoulder of Christina (N) an apparently near target, the accommodation level increased. The mean averaged
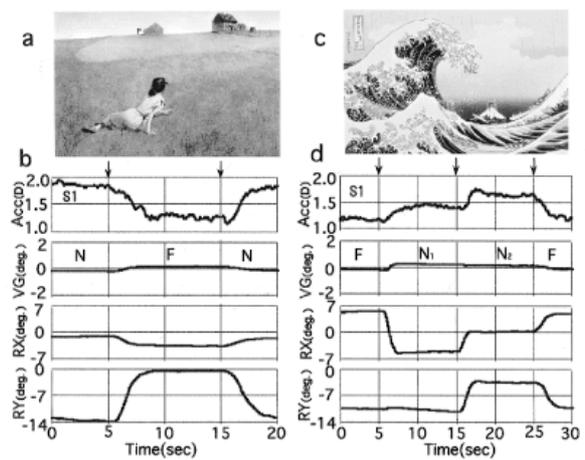

Figure 1. Drawings used for stimuli and resultant evoked accommodation. Both apparent depth and vertical gaze direction evoke accommodation without vergence. Accommodation evoked by the apparent depth sensation was larger than that due to raising eye position. Acc, accommodation; VG, vergence; RX, horizontal right eye position; RY, vertical right eye position. Gaze shift occurred at 5 and 15 s as indicated by the arrows.

accommodation of the five subjects was 0.68 ± 0.07 D (Table 1a). It is noteworthy that virtually no vergence (VG) was induced by this response.

By using a subjective measurement, Ripple (1952) reported that accommodation level decreased after raising the line of sight. Thus, there is a possibility that the accommodation measured might be induced by that effect. To check directly whether accommodation is really evoked by depth sensation independent of vertical eye movement, the second experiment used a reproduction of a woodcut made by K. Hokusai (*Mt Fuji Viewed Through Waves off the Coast of Kanagawa* from his "Thirty-six Views of Mt Fuji"; Figure 1c). The photograph was 37 x 25 cm and also was placed 40 cm from the subject's eyes. They gazed at F, N1 and N2 sequentially; these were separated by 7 cm each (visual angle was 8∞) and again the symbols were not only displayed on the picture, but only specified verbally. Subjects were instructed to look at the top of Fuji (F) in the middle of the picture, and then at two different points on the wave (N1, N2); the points were all at about the same distance from each other, but had different relative vertical positions.

Figure 1d shows averaged responses of subject S1. Abbreviations are the same as in Figure 1b. Gaze shifts occurred at 5, 15 and 25 s. A small but significant amount of near accommodation was recorded on shifting eye position from F to N1; N1 and F were at the same vertical level. Then an additional small amount of near accommodation was induced when eye position moved obliquely from N1 to N2. The difference between the accommodation levels when looking at F and N1 can be considered as the amount of accommodation evoked purely by apparent depth; the difference in accommodation levels when looking at N1 and N2 can be considered to be accommodation evoked by apparent depth minus the accommodation evoked by raising the line of sight of 8°. The accommodation due to the depth cue was larger than the one due to moving the gaze direction vertically (Table 1). Also the vergence remained virtually constant.
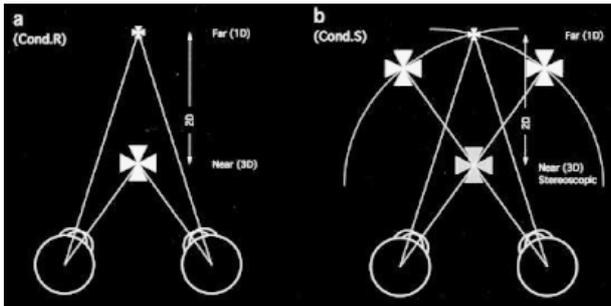
Figure 2. Representation of stimuli in experiment 3. (a) A realistic image of a Maltese cross was presented; (b) a stereoscopic image of a Maltese cross was presented in which the real images were presented at 1D with proper binocular parallax and, hence, formed a stereoscopic image at 3D.
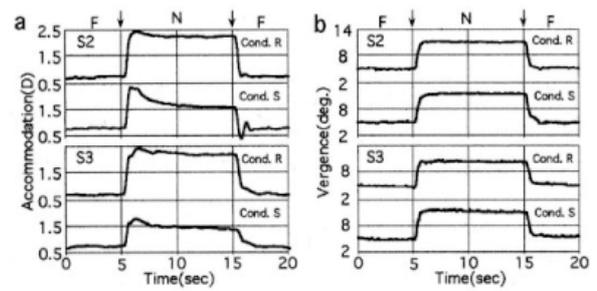

Figure 3. Accommodation and vergence responses of subjects S2 and S3 when they gazed at the Maltese cross in the TVS under conditions R (distance, binocular parallax and size were changed) and S (binocular parallax and size were changed but distance was kept constant).

3.2. Three-dimensional Experiment (3)

As there are many types of stereoscopic displays using binocular parallax and these displays are the easiest way to provide free stereoscopic images, it is interesting and necessary to know how accommodation is influenced by such devices.

Target distances of the Maltese cross were changed stepwise between 100 cm (1D) and 33 cm (3D) as shown schematically in Figure 2. In condition R, binocular parallax and target size were changed together with distance as if a real target had moved between the two points (Figure 2a). Then binocular parallax and target size were changed similarly to condition R, but the optical distance was kept fixed at the far point (1D) simulating stereoscopic displays (condition S; Figure 2b). Condition R simulated a real target movement between the two points and condition S simulated a virtual target movement.

Figure 3a and b shows the averaged accommodation and vergence responses of subjects S2 and S3 respectively. They showed larger accommodation shifts than they did for the previous stimuli (Figure 1). The amount of accommodation toward the realistic target (Table 1d) was about double that toward the stereoscopic target (Table 1e). Vergence responses were roughly equal for both conditions in both subjects, which roughly coincided with the amount expected from geometrical calculation.

Accommodative responses in condition S showed an additional interesting feature. The accommodation of subject S2 exhibited considerable overshoot (OS, defined as the absolute value of the difference between the peak and the evoked accommodation) then receded to a steady-state value. The overshoot of subject S3 was somewhat moderate. As for S2, S4 and S5, the overshoots in condition S were more than twice those in condition R (Table 1).

By expanding the time scale to between 5 and 7 s with subject S2's responses, it was found that vergence responses started 115 ± 19 and 120 ± 21 ms earlier than accommodation responses in the respective conditions (Figure 3c). Vergence was clearly driven by the change in binocular parallax; likely in condition S, accommodation was partially driven by the vergence and partially by perceived depth sensation. However, blur increased because the target was presented at a fixed distance of 1 diopter. Most likely, accommodation receded due to that increased blur. Accommodation recession was not clearly noted with the 2D picture stimuli (Figure 1, Table 1a–c).

4. VISUAL FATIGUE VIEWING STEREOSCOPIC DISPLAYS

It was shown that larger accommodation associated with vergence was evoked when looking at a stereoscopic image of the Maltese cross formed by the TVS binocularly (condition S in experiment 3). The accommodation evoked by the image should have been caused by the vergence and the perceived depth sensation. The amount of evoked accommodation reduced considerably from 1.34 ± 0.32 D in condition R to 0.59 ± 0.19 D in condition S.

The overshoots of accommodation in condition S were clearly larger than in condition R with S2, S4 and S5. It was moderately large in S3 and not large enough with S1. The hypothesis that the overshoot was larger in condition S than in condition R was not statistically significant. However, if we define the ratio of overshoot by overshoot/static evoked accommodation, the ratio in condition S was larger than the ratio in condition R.

From the above analyses, the overshoot in condition S seems somewhat larger than that in condition R. It was supposed that the vergence and strong depth sensation evoked larger accommodation in condition S compared with the previous two experiments (experiments 1–2), and the blur induced by the large accommodation reduced the accommodation in the condition S.

As the use of new visual informational displays has been steadily growing, there are emerging reports (Peli 1996) that visual fatigue might be induced by prolonged use of these displays. Although accommodation and vergence might have a close relationship with such phenomena, there has been no objective research on this issue. It is common to try to induce very strong depth sensations in stereoscopic movies so that audiences are attracted. Stimuli with considerable depth are presented with stereoscopic designs because they use displays at relatively close positions. The screens of head-mounted displays are very close to the eyes and it is easy to present stronger depth sensations even though the image is located optically farther than the physical distance with the aid of lenses. The strong 3D sensation evoked by those displays should produce large amounts of accommodation. Although such accommodation responses are natural when looking at real objects, it is not desirable from the standpoint of the visual system when viewing display because it increases the amount of blur and imposes an unnatural visual environment. Hence, this unnatural visual environment may produce a heavy visual burden and severe visual stress. Removal

of unnecessary wide-scale problems may occur in the near future, by achieving necessary improvements in modern informational displays.

In conclusion, it has been shown that (1) accommodation was evoked by the apparent depth sensation created from fine drawings or motion parallax, (2) the evoked accommodation was separated from the accommodation evoked by line of sight, (3) vergence was not affected by depth sensation with 2D stimuli under binocular viewing conditions, (4) stereoscopic stimuli (condition S) evoked larger accommodation compared with 2D stimuli, but evoked smaller accommodation compared with a realistic target (condition R), and (5) the near accommodation evoked by the stereoscopic condition receded considerably after an initial peak of near accommodation, probably because evoked near accommodation produced excessive amounts of blur. The results imply that the growing number of complaints in using the various stereoscopic displays or head-mounted displays largely come from this aspect of accommodation.

REFERENCES

ENRIGHT, J.T., 1987, Perspective vergence: oculomotor responses to line drawings. *Vision Research*, 27, 1513–26.

ITTELSON, W.H. and AMES, A. Jr, 1950, Accommodation, convergence, and their relation to apparent distance. *Journal of Psychology*, 30, 43–62.

KRUGER, P.B and POLA, J., 1989, Accommodation to size and blur changing in counterphase. *Optometry and Vision Science*, 66, 455–8.

PELI, E., 1996, Health and safety issues with head-mounted displays (HMD). *IDW* (International Display Workshop) '96, Kobe, 493–6.

RIPPLE, P.H., 1952, Variation of accommodation in vertical directions of gaze. *American Journal of Ophthalmology*, 35, 1630–4.

TAKEDA, T., FUKUI, Y. and IIDA, T., 1990, Dynamic eye accommodation induced by depth sensation. *Optometry and Vision Science*, 67, 450–5.

TAKEDA, T., NEVEU, C. and STARK, L., 1992, Accommodation on downward gaze. *Optometry and Vision Science*, 69, 556–61.

Warning Design

A. S. Adams

School of Psychology, University of New South Wales, Sydney NSW 2052, Australia

1. INTRODUCTION

A warning, for present purposes, is a safety communication device that should be designed to attract attention to itself, to give information about a hazard, its degree of seriousness, its probable consequences and of the way in which those consequences can be avoided. This chapter will discuss what is known about the principles that determine the effectiveness of a warning. It will not deal with legal questions such as requirements for particular wording, nor, except indirectly, with the question of when a warning should be used. In many countries there is legislation governing the design and use of warnings and such legislation will always take precedence. National standards associations, workplace safety organizations and consumer organizations are a good source of relevant information.

It may come as a surprise to some to hear that warnings often do not work, but that is indeed so. Why do we not obey warnings? Some claim that it is the severity of the consequence that determines behavior and that if the consequence warned about is severe, even though of low probability, we will be likely to obey a warning. Others claim it is the probability of injury that is critical, so that if a warning is about a low probability event, even if life-threatening, it is likely to be ignored. The details of this equation are still unclear. On a pharmaceutical product, for example, a warning about a very small probability of a life-threatening side effect will often cause a consumer to avoid the product. On the other hand, a workman who feels he can complete a job much faster by ignoring a warning and not using protective equipment may take the shortcut to save time even if the hazardous event to which he thereby lays himself open is life-threatening. There is evidence that such people see that the warning as being for other, possibly less skillful people, or simply that the accident will happen to others, not to them.

2. WARNINGS AS COMMUNICATION DEVICES

Although we may at times choose to ignore a warning of which we are well aware, the warning as a communication device can nevertheless perform more or less effectively. Many variables that determine the effectiveness of warnings have been studied. To place these variables in context one must first examine how a warning achieves its aim. Various models have been proposed, but broadly speaking if a warning is to be complied with it must first be seen and understood, and then, after that, there must be appropriate attitudes and motivation to comply on the part of the receiver.

The first two processes of seeing and understanding each require something of the warning device on the one hand and of the receiver of the warning on the other. If either factor in any sense "scores a zero" there will be no communication. Most of this article relates to the design of the warning device, but it is important to note that if the intended receiver of the warning is not in some sense receptive to it, the warning, no matter how well designed, will be ineffective. Thus, someone who is familiar with a situation may not see a prominent warning. Even if the warning is seen, someone who assesses the situation and feels that that they can function without heeding the warning may well do just that, possibly judging that the warning is there for the benefit of others who lack their skill or knowledge.

Action in the presence of a warning has many determinants, the information received from the warning itself being only one of them. Because this article is concerned with warning design, variables relating to the receiver of the warning or the circumstances in which the warning occurs will not be discussed, except indirectly. These include the age and sex of the receiver of the warning, the cost of compliance – for example how easy it is to obtain and use any equipment referred to in a warning – and social variables such as whether or not others are seen to comply with the warning. Suffice it to say that all these variables have been found to be important determinants of compliance.

3. DESIGN VARIABLES AFFECTING THE PERFORMANCE OF WARNINGS

Some 55 papers concerned with the ergonomics of warnings have been gathered in Laughery et al. (1994). Laughery and Wogalter (1997) also provide more details than can be covered here. What follows is a brief overview of the principles that have emerged from this research.

3.1. Signal Word, Color and Size

A warning must first attract attention to itself. One way most written warnings do this is by using a prominent signal word that both draws attention and provides an indication of the level of urgency of the warning. The words DANGER, POISON, DEATH have been shown to connote the highest level of urgency, followed by CAUTION and WARNING.

The way in which the wording on a warning is displayed is of obvious importance. A general principle governing the perception of written material is that high contrast is required for good legibility. The use of the color red is of assistance in drawing attention to a warning as it has been shown to be associated with higher urgency. However, this will only apply when red is not used extensively in the area immediately surrounding the warning.

Size is also of importance, although practical considerations usually set a limit to size. In a consumer survey on product warnings a householder commented on a large warning found on a household product by saying, "It's big, they must mean it." However, size or other features designed to ensure prominence will have a diminishing effect if they are used too extensively. We have already reached a point with road warnings where inordinate size is now needed to draw attention to a sign warning of some new and possibly unexpected hazard.

Finally, the font used is important, particularly where space is at a premium and a warning is designed to be seen from some distance – road signs, for example. Recent research has resulted in the development of a new font called Clearview that provides better legibility on a sign of a given size than existing fonts. In general, where more than a few words are involved, all upper case should be avoided.

3.2. Hazard Information, Consequences and Instructions

The signal word used with a warning draws attention and provides some information about the severity of the hazard – DANGER usually signifies a potentially life-threatening hazard, and CAUTION or WARNING something less significant. The remainder of the warning must then provide sufficient detail to enable those to whom the warning is directed to determine the severity of the hazard, the consequences of not complying with the warning and the behavior needed to avoid the stated or implied consequences.

In many work situations only a minimal amount of information is required to impart the appropriate message. In a steel foundry, for example, a sign saying "DANGER, Molten Steel" may be quite sufficient to remind workers of a whole range of possible hazards, consequences and appropriate avoidance behaviors. In other circumstances a warning may be required precisely because a given hazard is not well appreciated. Hazards associated with chemicals or solvents, for example, are often ill appreciated.

3.3. Use of Pictorial Devices such as Symbols

Many warnings use symbolic devices as part of the warning. There are a number of reasons for this. The most obvious is that a warning imparted by means of a symbol can communicate without the use of language and may therefore obviate the need for a multilingual presentation. Another reason is that the use of symbols may enhance the attention-getting aspects of the warning as a whole and may also make the warning visible from a greater distance. Finally, a warning presented in the form of a symbol may be understood more rapidly than one that requires the reading of a complete phrase or sentence.

A rather different reason for using a symbol is that even though it may not convey sufficient information, the symbol may serve the alerting function otherwise achieved with a signal word such as DANGER. For example, it is a convention embodied in a number of national and international standards that warnings should be symbolized by using a black symbol on a yellow triangular or diamond background with a heavy black border.

Figure 1. The poison warning symbol: an example of a warning symbol that is not well understood.

Figure 1 shows two warning signs using these conventions. Often the diamond is used within the road context and the triangle, as shown, within industry.

The warning shown in Figure 1 has been chosen because it is not well understood. However, the effect of using these symbols on a product, in a brochure, or at a work site is to draw attention and suggest danger, all without words. Often the exclamation point is used in conjunction with a worded sign primarily as a device to draw attention, and only secondarily to convey to at least some people the idea that the wording contains a warning.

Symbols are not an easy solution for problem warnings, nor for the language problem. Any warning, whether conveyed through the medium of words or of a symbol, must be clearly understood. Let one assume that 75% of a population as a whole can understand a poison warning written in English on a pesticide product, and that the 25% who cannot is made up of a significant proportion who can read only some other language. A judge in a court case suggested that in these circumstances a skull and crossbones poison warning symbol (Figure 1b) would have been an appropriate alternative to a multilingual warning. However, in a random-sample survey in the UK in 1977 this symbol was shown to 479 respondents. The symbol was shown within the triangle shape commonly used for warnings and was tested together with other symbols. On a very lenient criterion only 48% of the respondents correctly identified the symbol as indicating a poison. Thus the symbol used alone would have communicated successfully with a smaller proportion of the population than the English language warning used alone.

When designing a warning symbol there are guidelines that can improve legibility and comprehension. These are embodied in a number of national and international standards dealing with the use of symbols in signs in general, and in warnings in particular. International Standard ISO 9186:1989 (currently under revision) and Australian Standard 2342 – 1992 give guidelines for designing legible symbols and methods for testing their comprehension. ANSI Z535 from the USA gives extensive guidelines. For common warning symbols, such as "No Smoking", "No Swimming" and "Danger, electrical hazard," standardized symbols exist that have been subject to testing. Designing successful symbols for new concepts, however, has proven more difficult. The standards mentioned above require between 66 and 85% comprehension among a representative sample of users. In general these targets have been difficult to meet for complex concepts such as, "In case of fire do not use lift" or "Danger, freak waves — do not fish off the rocks."

In some circumstances it may be possible to convey a concept by means of several symbols in succession, much like a cartoon strip. This method has been used successfully to produce a warning for use on packets of candles to indicate that lighted candles should not be left unattended. If color is available for a symbol such as this the negating cross should be in red.

3.4. Location of Warnings

Much research has gone into the design details of warnings, but rather less into the question of where a symbol sign should be located to enhance its attention-getting properties. From the small pharmaceutical vial that provides insufficient space, to the household product on which the manufacturer deliberately wishes to minimize the prominence of a warning, each product presents

its own challenges. For a warning to be useful it must be appropriately attention getting and it must be present at the time the user needs it. This may be at the time of purchase, possibly to alert the user to the fact that certain protective equipment is required, or at the time of use to remind the user to make use of that equipment or to follow certain safety instructions. It is easy to say that the warning must be prominent, but how is adequate prominence to be judged? There is no easy answer as few situations requiring a warning are so standardized that a ready-made solution is available. The only recourse is to appropriate testing.

4. TESTING

Whenever a solution is proposed to a warning problem, the question, "How do we know it will work" should always be asked. Traditionally, a number of alternative solutions are proposed and each is examined experimentally to determine the best. This approach, however, is highly resource intensive; for example, it requires that each proposed alternative be designed to the completed stage. ISO 9186 presents methods for testing the comprehension of symbol signs that might be used as warnings. Over the years the methods proposed in the various drafts and revisions of that standard have become more and more efficient. A recently proposed method simply requires a group of respondents to record their judgements of how effective they think a given symbol will be. A formalized method of aggregating this data is presented in the standard so that the method can be used as a useful preliminary check. Adams (1998) describes the use of efficient testing-as-you-go methods. These methods are based on an iterative "process" testing method rather than the single-shot "outcome" experimental method often used to compare alternatives.

5. AUDITORY WARNINGS

Finally, nothing so far has been said about auditory warnings, as an entire article would be required to do justice to the problems they present. Briefly, the informational properties of an auditory warning must be entirely learned. An auditory warning such as an alarm in an intensive care unit does not of itself provide information about the hazard, its consequences or about how to avoid the hazard. However, the relative urgency of the situation signaled by an auditory warning can be indicated by the judicious choice of the warning's sound. Where a number of such warnings can occur in the same space, such as in an aircraft cockpit or an intensive care unit, care must be taken that interference, or masking, of one sound by another does not occur. For further information about problems such as these, and appropriate design solutions, see Edworthy and Adams (1996).

REFERENCES

ADAMS, A.S., 1998, The role and methodology of usability testing in information design. In H. Zwaga, T. Boersema and H. Hounhout (eds) *Visual Information for Everyday Use: Design and Research* (London: Taylor & Francis), pp. 3–20.

ANSI Z535,1997, *Standards for Warning Signs, Labels and Symbols.*

AS 2342 – 1992, 1992, *Development, Testing and Implementation of Information and Safety Symbols and Symbolic Signs* (Sydney: Standards Australia).

EDWORTHY, J. and ADAMS, A.S., 1996, *Warnings: A Research Prospective* (London: Taylor & Francis).

ISO 9186: 1989, 1989, *Procedures for the Development and Testing of Public Information Symbols* (Geneva: International Standards Organisation).

LAUGHERY, K.R., WOGALTER, M.S. and YOUNG, S.L., 1994, *Human Factors Perspectives on Warnings* (Santa Monica: Human Factors and Ergonomics Society).

LAUGHERY, K.R. and WOGALTER, M.S., 1997, Warnings and risk perception. In G. Salvendy, (ed.) *Handbook of Human Factors and Ergonomics* (2nd ed.) (London: Wiley), pp. 1174–97.

Part 6

Workplace and Equipment Design

Active Safety in Car Design

S. Doi

Human Factors Division, Toyota Central R&D Labs., Inc., Nagakute, Aichi, 480-1192, Japan

1. INTRODUCTION

In car design, it is indispensable that there is compatibility between safety, comfort, and energy consumption and environmental safety along with traffic smoothness and efficiency. In future transport systems, driver/vehicle/environment interactions are major issues in active safety of vehicle operations.

Human factors are the major cause of vehicle accidents. Almost all the causes of accidents are related to human factors involved in the driver's maneuvers: especially on the errors of recognition, judgement and operation. Human response and sensitiveness are very individual matters and completeness of driving safety is not merely obtained not only from the averaged value of the operational characteristics, but also from the accomplishments of the grasping the limits of the performances. Improving the maneuverability of the vehicle is an urgent problem for vehicle-based technologies like collision avoidance systems. In this field of studies it is necessary to discuss from the point of view of the man-in-the-loop to consider active safety performance of the vehicle.

2. FUNDAMENTALS OF VEHICLE SAFETY

2.1. Classification

Vehicle safety is usually classified into active and passive safety according to the situations before and after accidents. In active safety emergent operations of drivers and human errors are the main causes of accidents. Just before an accident emergency, avoiding performance is needed for vehicle. Prior to the accidental situations, driver-aid systems are necessary for maintaining safety driving and for preventing accidents. On the contrary, in passive safety, various driver protections are equipped in the vehicle and life-saving infrastructure.

2.2. Roles of Vehicle, Driver and Traffic Environments

2.2.1. Vehicle

The role of vehicle for active safety is:
- Principal performance of fundamental functions (stopping and steering).
- Driving visual cues and visibility.
- Maneuverability of steering and other operations.
- Free from stress and fatigue for long-term driving.
- Emergency-avoiding operations.

2.2.2. Driver

Depending on driver performance, the role of the driver is demanded as:
- Consciousness of safety driving.
- Adequate operation for smooth driving.
- Concentration for driving manners.
- Reflective operations for emergency avoidance.

2.2.3. Environment

Regarding traffic information, the role of traffic environments is:
- Suitable traffic infrastructure.
- Fully traffic information.
- Safety road environments.
- Roadside infrastructure and environments.

2.3. System Architecture

2.3.1. Intelligent transport systems

The goal of the ITS (intelligent transport systems) is multimodal effectiveness in every traffic system. Above all, vehicle safety is the key technology uniting with drive/vehicle/environment in these movements. System architecture is most important for a worldwide standardization of the apparatus.

2.3.2. Road traffic system

The next subsystem is the road traffic system whose elements are constructed with road information systems and the communicate systems between cars and roads. Each subsystem has an interface with other systems. Between these interfaces, the human–machine interfaces are more complicated for designing the equipment.

2.3.3. Vehicle safety system

Another subsystem is vehicle control and driver-aid, which consists of an information system like navigator and steering/braking operational systems. These components should be organized in human–machine system for the active safety of the vehicle. Both system technologies are composed with various systems:

1. Warning system technologies:
- Warning system for emergent driver conditions.
- Warning system for dangerous vehicle conditions.
- Vision enhancement system.
- Vision enhancement in nighttime driving and nighttime object detection.
- Blind area monitoring/warning system.
- Surroundings warning system.
- Road environment information acquisition/warning system.
- Inter-vehicle communication/warning system.
- Driver workload warning/reduction system.

2. Avoidance system technologies:
- Intelligent vehicle control system.
- Driver's hazardous condition detecting system.
- Blind area accidental avoidance system.
- Collision avoidance system with surrounding obstacles.
- Collision avoidance system utilizing road information.

2.4. Tools for Investigations

2.4.1. Driving simulator

Supposing a road traffic situation and information system in the car, driver/road/information interactions are used to investigate in the virtual environment of driving. The remarkable tools for such kinds of research are so-called driving simulators. Examples used in these studies are illustrated in Figure 1. The simulator is constructed with the moving base of 2 degrees of freedom, namely roll and lateral motions. On the moving base, the systems are mounted with driving cockpit and visual screen. Driving seat is
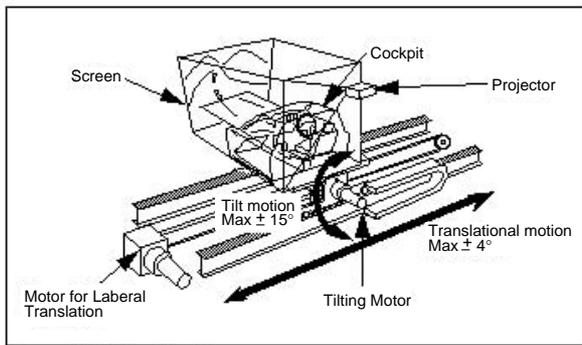
Figure 1. Driving simulator.

installed in the half-cut vehicle body and the visual scenes given by computer graphics systems according to the driver's operations.

2.4.2. Elements of simulator

Both of the motion and visual systems are controlled by computational vehicle model. The vehicle model has 3 degrees of freedom: longitudinal, lateral and yawing motions. The tires are modeled by an equation with combined status of lateral and longitudinal acting forces. The dynamic vehicle motion is calculated with a computer from the input signals of the driver, namely gas and brake pedal stroke, and steering wheel angle. The projected visual angle is 60° wide at the screen located at 2.5 m ahead from the driver. Inside the cockpit, imitating systems are equipped for steering efforts and interior noise/vibrations corresponding to the state of the running conditions.

The motion base is moved laterally by chained belts along the pair of lateral rails and rotating by another motor installed under the bottom of the cockpit. In the systems, the lateral motion is controlled proportionally by the calculated lateral displacements of the vehicle model and the rotational motion is also controlled according to the centripetal acceleration by the vehicle speed and the road curvature of supposed situations. Using the simulator, driver's behavior and vehicle performance are measured quantitatively in the same time.

3. DRIVER AND VEHICLE SYSTEMS

3.1. Emergency Avoidance Maneuvers

3.1.1. Emergent situations

Several kinds of experimental tasks for emergency avoidance are used to examine driver and vehicle systems. One is a stationary obstacle avoidance task, which can regenerate the same situations in the proving ground. The next case is a sudden emerged moving obstacle during running on the road. The other is an unexpected road surface change during running on curved roads. These tasks are supposed to be ordinary roads and the experimental subjects are used to be ordinary drivers, including female and elderly drivers. In more detail, these driving behaviors are examined by imitating obstacle emerging behind walls beside the way and sudden change to slippery surface on cornering.

3.1.2. Reaction time

From the results of these obstacle avoidance, the database of driver's operational characteristics depended on the modes of avoiding maneuver is derived. Comparing with the individual

operational characteristics, the difference of reaction time between the case of success and fail (collision) and the amount of each operation is clarified. According to the reaction time, in the emergent behaviors, driver response time for releasing the gas pedal is ~0.6 s and braking pedal reaction time is ~0.98 s. On the contrary, the steering reaction time is ~0.64 s on average. Comparing with these data from observation of driving managing behaviors leads to the relation between reaction time of braking/steering in various situations.

3.2. Driver Behavior Model

3.2.1. Model structure

A mathematical driver-operating model is derived from experimental data of emergent situations. The model is constructed with steering (lateral control) and the braking (longitudinal) operation.

3.2.2. Steering model

The parts of steering operation are based on the variable preview time and variable preview point. The model is constructed with three parts: prediction, target course generation and steering control block. In the prediction block, it is assumed that the driver involves the simple vehicle model and predicts the future vehicle position and vehicle state values. In the target course generation block, the driver's operating model generates the course of before and after the obstacle emerging. In the steering control block, steering angle is derived as a type of integral control that minimizes the both of corrected steering angle and the deviation between generated and predicted courses.

3.2.3. Braking model

The braking operation is expressed with the first-order delay system with lag time from the basis of experimental results. The lag time is equivalent to the braking reaction time. The gain of the first-order delay system is the maximum brake pedal stroke. The time constant corresponds to the time from the beginning of the braking operation to reach the maximum brake pedal stroke.

3.3. Evaluation of Active Safety Performance

3.3.1. Evaluation methods

Combining the driver model with the vehicle model, using the database of driver behaviors, driver–vehicle performance in the so-called man-in-the-loop is calculated for various cases of emergency. An example of a schematic diagram of performance evaluation methods is shown in Figure 2.
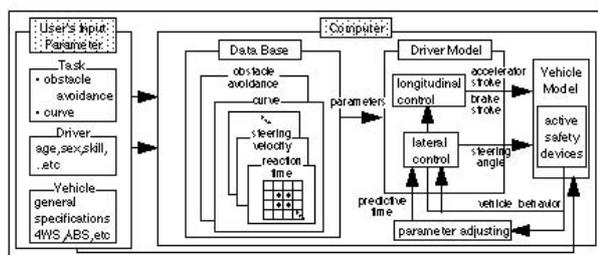

Figure 2. Schematic diagram of driver–vehicle performance evaluation system.

3.3.2. Flow of evaluation

First the user chooses the certain emergent task depending to the road situation and driver/vehicle specifications. Next, he selects the driver's characteristics from the database. As these supposed parameters are fixed, the last part of vehicle performances on the emergency is estimated by calculating the combined models.

3.4. Verification of the Effect of Control for Safety Driving

On design of vehicle control, using the above evaluation methods, the estimated effect is verified with the experimental behaviors in the driving simulator, for instance whether the trajectories of the vehicles would be similar with each other and steering maneuvers would be also improved. The subjective judgements of the drivers are also checked to be better operational feeling and the recovery of vehicle direction would be fairly improved. Thus, the driver-managing performance of the emergent situations is evaluated in advance with estimation of the effect of assistant control of vehicle.

4. DESIGNS FOR ACTIVE SAFETY PERFORMANCES

4.1. Stopping with Brake Assistance

4.1.1. Braking task

The design of braking characteristics is executed by the simulations with obstacle avoidance performance of the driver. In the case of emerging obstacle, supposing vehicle speed is 60 km/h, the gap length to the obstacle is 40 m in time of obstacle start to emerge, the avoiding performance is evaluated for associated behaviors with braking and steering. The driver's skill is supposed to be below the averaged, namely maximum steering velocity is 200°/s. The obstacle avoidance performance is shown with braking stroke as the functions of brake and steering reaction time.

4.1.2. Performance estimation

Performance is classified in three cases: avoidance, namely success to avoid the obstacle, the collision to the obstacle and the road-departure from the road edge. According to the results of avoiding ability concerned with both of brake and steering reaction time, the drivers who operate in a greater than steering reaction time of 1.0 s would came into collision to the obstacle. On the contrary, even if for rapid steering, delayed braking action over 0.8 s leads to the road-departure from the road edge. According to the amount of brake stroke, the avoiding ability would be increased with the stroke and the braking assistance control system is expected to be effective for poor power drivers.

4.2. Cornering with Vehicle Steering Control

4.2.1. Emergency on a curved road

Another emergent situation is sudden surface change to a slippery condition during a curved road. In the case that frictional coefficient of road surface is changed from 1.0 to 0.5 in rear wheel, supposing vehicle speed is 70 km/h and the curvature of the road is 100 m, the managing performance of the vehicle is evaluated in three kinds of steering reaction time. In these cases the prediction time in the driver model is supposed to be changed from 0.7 to 0.5 s before and after entering the low frictional area. The estimated results are shown that there are three patterns for managing performance on slippery curved roads.

4.2.2. Effect of steering control

Driver's assistance in emergent situation would be of necessity to improve driver–vehicle closed loop performance. Supposing the assistant steering controls, the effects of the control on the performance are evaluated.

In the cases of differential steering control and yaw-rate feed back control, the estimated vehicle trajectories, steering angles and front steer angle at that time are calculated. From the computational results every controlled case is effective for preventing the divergent phenomena of the vehicle after entering the low frictional area. The controlled case is fairly stable in the trajectory of the vehicle and would be estimated to stay within the lane. Thus, the driver-managing performance of the emergent situations is evaluated with estimation of the effect of assistant steering control.

4.3. Running with Mental Workload

There are many factors including environmental, mechanical and human factors that cause driving fatigue. Among these factors, vibration is one of the major factors because drivers are usually exposed to sustained mechanical vibrations including mental as well as physical stress.

Design of suspension is related to fatigue of the driver with mental workload during long-term driving. The influence of vibrations at low frequency, corresponding to the resonance of human head, is derived from basic statistical analyses of the relations between subjective ratings of fatigue and physiological variables. The suspension control capable of reducing the vibration prevented driver fatigue at a significant level in actual driving.

5. DESIGN OF HUMAN–MACHINE INTERFACE FOR ADVANCED TRAFFIC ENVIRONMENT

The new century is the age of advanced traffic environment organized by information as well as the cooperative relations of driver and vehicle. It would be of great issues how to construct
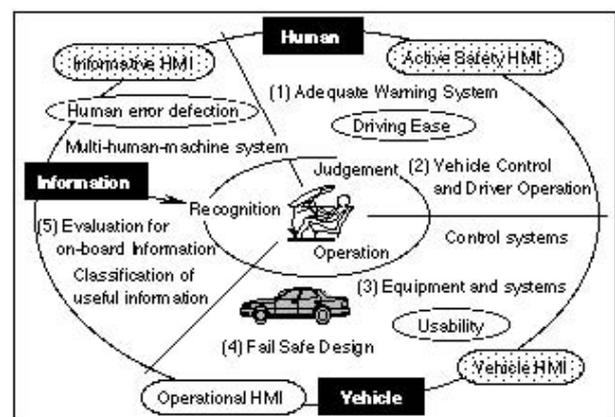

Figure 3. Human–machine interface (HMI) in active safety.

the human–machine interface toward driving support equipment and on-board information systems maintaining and improving the optimal and interactive relations with drivers.

In advanced active safety, there are four categories of the human–machine interface and five aspects of design problems as followings and they are illustrated in Figure 3 concerning to driver functions.

- Timely and adequate warning for driver.
- Sharing the responsibility with vehicle control and driver's operation.
- Equipment and system for understandable limit of vehicle critical performance.
- Fail safe design for total man–vehicle system.
- Setting criteria for on-board information system.

To make better solutions for these problems, it is more effective and necessary to take the considerations for analysis of physiological aspects for driver recognition and judging mechanism. A more psychological approach would be expected for human operational behaviors corresponding to the mental workload due to information during driving.

6. INTERACTIVE DESIGN WITH THE DRIVER

There are new approaches to develop the system technology to make the driver environment suitable to individual driving behavior. Aiming at better interactive design with driver, the behavior-based safety driving technology is consisted with three steps:

- Real-time measuring of personal behavior and learning the pattern of the individuals.
- Comparing with the accumulated database of behaviors to

understand the driver's situations and making a model-based approach to follow the individual manners.
- Fitting with individual personal behaviors by driver-aid systems.

According to our senior designer's remarks, preferable features of vehicle in the future are:

- Behind the wheel and willing to turn.
- Enjoyable to run about.
- Feel freely and operationally with ease.
- With comfortable sound and vibration.
- Affective and thrilling to drive.

Active safety design is still on the way to realizing the goal, but much remains to be done. Finally, taking these evaluation criteria into account, it should be more effective and adequate when designing the driving safety devices to make a better human interface in future transport systems.

REFERENCES

EVANS, L., 1991, *Traffic Safety and the Driver* (New York: John Wiley).

McLEAN, J.R. and HOFFMANN, E.R., 1973, The effects of restricted preview on driver steering control and performance. *Human Factors*, 15, 421–30.

McRUER, D.T. and JEX, H.R., 1967, A review of quasi-linear pilot models. *Transaction on Human in Electronics*, HFE-8, 231–49.

ROCKWELL, T.H., 1972, Skills, judgement and information acquisition in driving. In T.W. Forbes (ed.), *Human Factors in Highway Traffic Safety Research* (New York: John Wiley), 133–64.

WICKENS, C.D., 1984, *Engineering Psychology and Human Performance* (Columbus: Charles E. Merill).

Analysis of Office Systems

M. M. Robertson

Liberty Mutual Research Center, Hopkinton, MA 01748, USA

1. INTRODUCTION

Intensive use of computer and information technology for long periods of time in the office workplace calls for an examination of employee performance and possible work-related health and stress problems. The nature of these technologies and the work environment influence employee musculoskeletal fatigue, discomfort, and pain (e.g. Cakir, Hart, and Stewart 1978; Smith and Carayon 1989; Sauter, Dainoff, and Smith 1990), and may have an adverse effect on work effectiveness, stress and work-related musculoskeletal disorders (WMSDs) (e.g. Bongers, De Winter, Kimpier, and Hildebrandt 1993; Kuorinka and Forcier 1995).

Office systems, by their nature, are complex and multivariate. Understanding the potential causal factors of problems arising from poorly designed office systems, requires a macroergonomic approach using systems analysis and processes. The traditional *microergonomic* approach to office work environments emphasizes the microscale aspect of the work environment at the individual workstation (e.g. keyboard, screen image, manuscript placement, etc.) (see, for example, Sauter, Dainoff, and Smith 1990; Robertson and Rahimi 1990). Broadening this approach to a macroergonomic level, other office system components including physical environmental variables (e.g. layout, storage, and adjustability), work tasks, work organization, organizational structural, technology characteristics, and psychosocial variables are recognized as impacting individual, group, and organizational performance (Smith and Sainfort 1989; Robertson and Rahimi 1990; O'Neill 1998; Swanson and Sauter 1999). Understanding the interrelationships of these elements and their effects on health, stress, work, and organizational goals is necessary to provide a comprehensive or systems perspective of office environments.

With a macroergonomics approach, systems tools and processes are used to develop strategic criteria for the design of an office work system where the social and technical sub-systems are congruent with the organizational mission. Directed by strategic level criteria, problem-solving tools such as ergonomic analysis for workstation or job design enhances organizational effectiveness. Prior research suggests that there is the potential for translating the findings from a macroergonomic approach into relevant office system design and organizational planning interventions (e.g. O'Neill 1998; Hendrick 1997; Robertson and Robinson 1995).

This article provides an overview of a systems analysis tool using a macroergonomic approach to understand and identify problems and probable causal factors related to such office environments. Systems analysis also provides a *process* for developing strategic and systematic solutions for solving problems arising in a computer intensive office environment. The interventions and solutions presented here are based on 20 years of research and case studies drawn from a global sample of office technologies environments. Systems analysis also provides for the development of alternative solutions with techniques for evaluating the cost/benefits of each alternative, selecting and implementing solutions, and providing feedback and measurement of improved worker performance. By applying this systems analysis tool, the interaction and fit of the office worker, the job tasks, the organization structure, and the physical environment is improved, resulting in a more effective office as well as healthful work environment.

2. APPLYING A SYSTEMS ANALYSIS TOOL: UNDERSTANDING OFFICE WORK SYSTEMS

In applying systems analysis, the level of analysis is targeted at the business unit or departmental level, where specific objectives and issues are identified. Individual and group needs are also identified.

The following systems analysis tool is one frequently used in business and industry and is a modification of one proposed by Mosard (1982), which is based on earlier work in systems engineering. There are six analytic steps:
1. Defining the problem: The problem factor tree (see figure 1).
2. Setting objectives, developing an evaluation criteria table, and developing alternatives: The objectives/activities tree (see figure 2).
3. Modeling the alternatives: The input–output flow diagram (see figure 3).
4. Evaluating alternatives: The criteria scorecard: cost/benefit analysis (see table 1).
5. Selecting an alternative: The decision table: selecting an alternative based on future conditions (see table 2).
6. Planning for implementation, evaluation and modification: Scheduling and management project flow.

3. STEP 1: DEFINING THE PROBLEM: THE PROBLEM FACTOR TREE

Complexity is inherent in performance problems, therefore a systematic analytic approach to issue definition is necessary. From data previously collected or recently observed, a problem factor tree is constructed which identifies the problem, sub-problems, and causal factors, including their interrelationships (Mosard 1982).

A completed problem factor tree depicts a hierarchical, logical structure identifying the problematic elements (see figure 1). To develop a problem factor tree, work issues and problems are precisely stated and linked together through an iterative process (Mosard 1982). The lower level causal factors in the tree contribute to the major problem. Feedback loops may also be incorporated. The problem in an office work system is defined as: "An increase in turnover, lost work days, and claims disabilities, and a decrease in performance and effectiveness of office workers caused by occupational stress from office technologies and office system design" (see figure 1). It is hypothesized that the problem is caused by occupational stress from office technologies and workplace design. Further research revealed two contributory sub-problems: (1) psychosomatic stress, and (2) physiological stress. In addition, two sub-problems of psychosomatic stress were found: (1) psychosocial disturbances, and (2) perceived lack of environmental control.

Figure 1 also shows other potential causal factors which contribute to the defined problem: lack of job content, poor job design, and lack of flexible workstation design. These are depicted
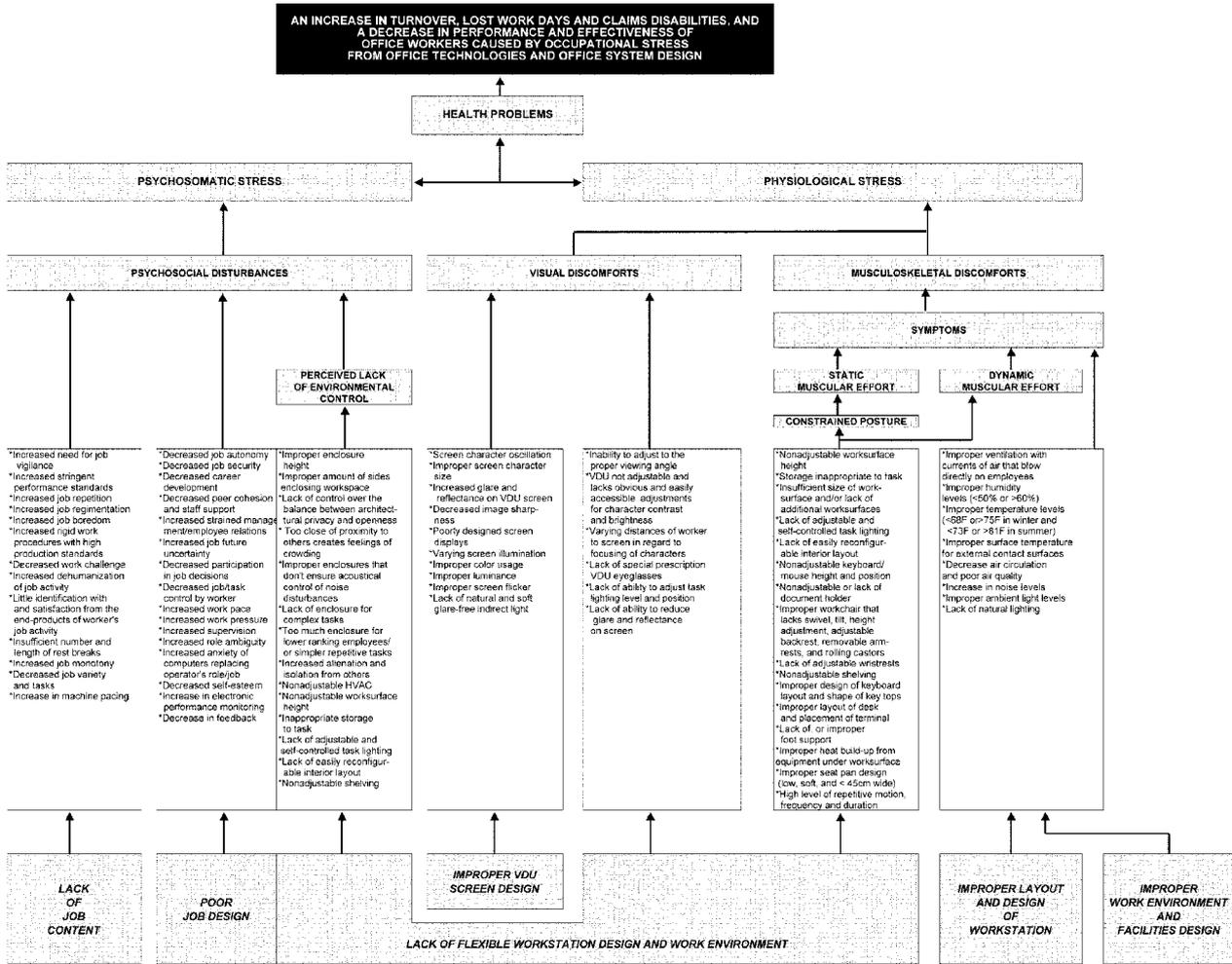
Figure 1. Increase in turnover, lost work days and claims disabilities, and decrease in performance and effectiveness.

at the base of the hierarchical structure. For the potential causal factors in job content and job design, several individual and group level factors are identified and shown in the middle of the tree. Poor job design components may include teamwork or collaborative issues at the departmental level, such as cross-functional teams, or at the individual level of a small, informal group gathering. Job content and job design should be viewed as the main element of the social sub-system. Potential casual factors in the technical sub-system are primarily located in the workstation design, layout configuration of the workspace, and screen design. Of the many sub-problems identified in regard to physiological stress, two are of paramount importance: visual and musculoskeletal discomforts. Numerous potential factors are shown, contributing to both of these sub-problems, including lack of flexible workstation design, improper VDU screen design, and improper layout and design of the workstation.

The problem factor tree depicts the causal factors and sub-problems that represent the integration of *microergonomic* and *macroergonomic* office system components and sub-systems, including organizational and job design issues. This initial step of the system analysis lends itself to developing an understanding of the technical and social sub-systems as well as the work environment sub-system. Other robust and valid systems tools

based on sociotechnical processes can be applied together with this particular system approach and model (e.g. Smith and Sainfort 1989; Hendrick 1997).

4. STEP 2: SETTING THE OBJECTIVES AND DEVELOPING ALTERNATIVES: THE OBJECTIVES/ACTIVITIES TREE

Objectives and evaluation criteria are developed for use in selecting the best alternative to address the causal factors. An "objective tree" is a hierarchical, graphical structure of objectives that address previously identified problems (Mosard 1983). The tree is created by identifying major needs, goals, objectives, and sub-objectives. In figure 2, the objective tree is depicted in the upper half of the figure and shows the major goals as: (1) "Decrease lost work days, claims disabilities, and turnover", and (2) "Increase performance and effectiveness of office workers by alleviating occupational stress from office system design and technologies."

The objective tree shown in figure 2 addresses the inherent psychological and physiological health problems. Alternatives are defined as a specified set of activities, tasks or programs designed to accomplish an objective (Mosard 1982). The objective/activity tree in figure 2 illustrates four alternatives, A through D, and
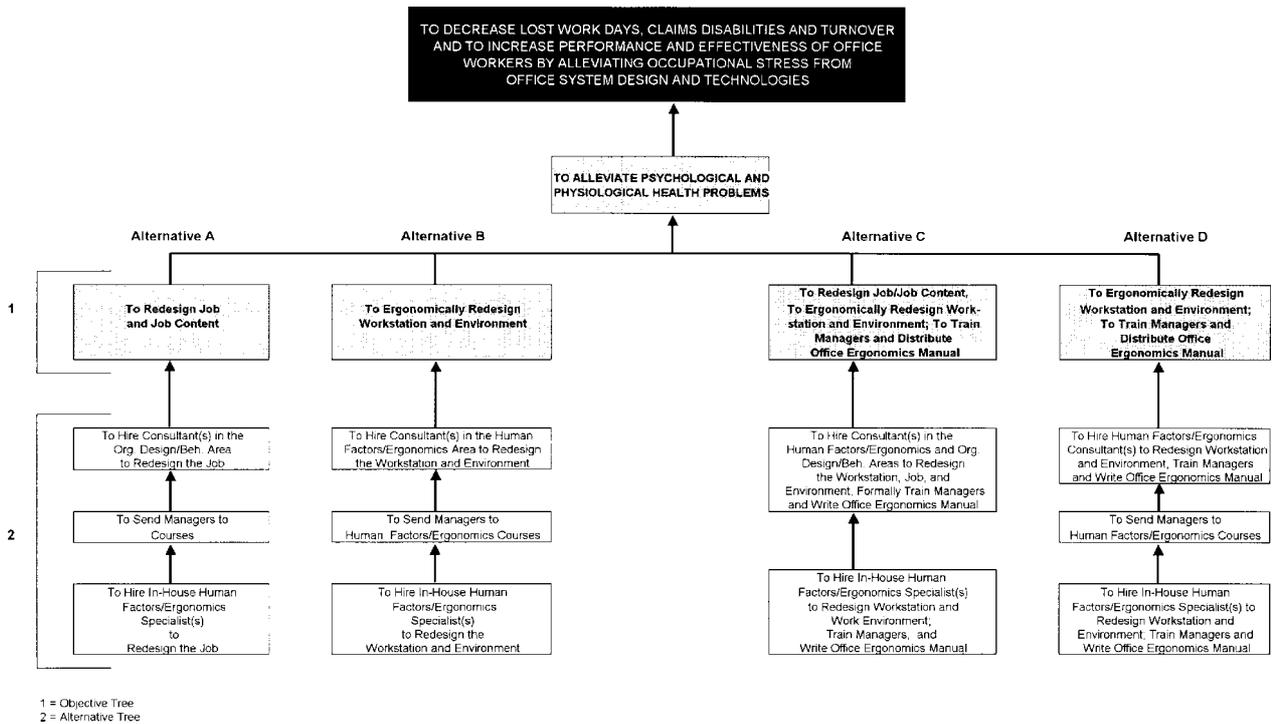
Figure 2. Ways of decreasing lost work days, claims disabilities and turnover, and increasing performance and effectiveness.

their appropriate set of activities. Four sub-objectives are defined: (A) redesign the job and job content; (B) ergonomically redesign the workstation and environment; (C) redesign the job, and (D) ergonomically redesign the workstation and environment and to train managers in job and workstation redesign and awareness and to write an office ergonomics manual. The degree of interaction between objectives, constraints, and the persons/ groups involved in the process should also be analyzed (Mosard 1982). Hybrid alternatives may be created which incorporate the best features of any of the initially identified alternatives. Alternative C is a hybrid alternative representing one of the many possible combinations of common activities. The four alternatives depicted were derived from case studies, field research, and longitudinal studies representing typical approaches implemented by companies to achieve the objective listed at the top of figure 2 (Smith and Sainfort 1989; Robertson and Robinson 1995; Robertson, Robinson, O'Neill, and Sless 1998; O'Neill 1998).

After the objectives and alternatives are selected, a preliminary decision criteria table is developed. This table is used to evaluate the "usefulness" of each of the alternatives as methods for accomplishing the objectives. Decision criteria typically include risks, costs, expected benefits, and measure of effectiveness, based on short-term and long-term perspectives (Mosard 1982).

5. STEP 3: MODELING THE ALTERNATIVES: THE INPUT–OUTPUT FLOW DIAGRAM

A descriptive or predictive model representing either each alternative set of activities or representing the entire system is developed. This is designed to allow alternative configurations to the systems to be analyzed. The system element interrelationship and/or gross resource requirements are depicted in order to determine the effectiveness of each alternative set of activities (Mosard 1982). Modeling techniques such as flow charts,

simulation, and systems dynamics modeling may also be used in this step.

The model used in this analysis is an input–output flow diagram (Mosard 1982). This modeling technique allows alternative configurations of the systems to be analyzed. In utilizing the input–output flow diagram, the inputs consist of people, resources, and information, and the outputs are the results and products. These outputs can, in themselves, become the sources of inputs to other sub-systems, and thus extend the diagram to fully represent the entire system being analyzed. Figure 3 illustrates the two phases of this model: the redesign phase and the operation phase. Inputs for the redesign phase include contributions from two general areas, human resources and financial resources. In the human resources component, individuals such as industrial psychologists, managers, employees, human resource managers, ergonomists, facility operations, trainers, and health and safety managers are included. The activities for each redesign project or program listed for phase 1, are shown in the left input box in figure 3. At the end of the redesign phase, the outputs become the inputs for the second phase, the operation phase. For example, in the job redesign program, managers and employees have acquired new skills and the jobs have been analyzed and redesigned. The managers and the employees will now interact within their own work systems and the results of these interactions are presented in the outputs (e.g. increase in performance, decrease in job stress, decrease in injury claims, and worker's compensation cases). Overall, these outputs fall into the categories of: changes in employee and group behaviors, organizational factors and reduction in business costs.
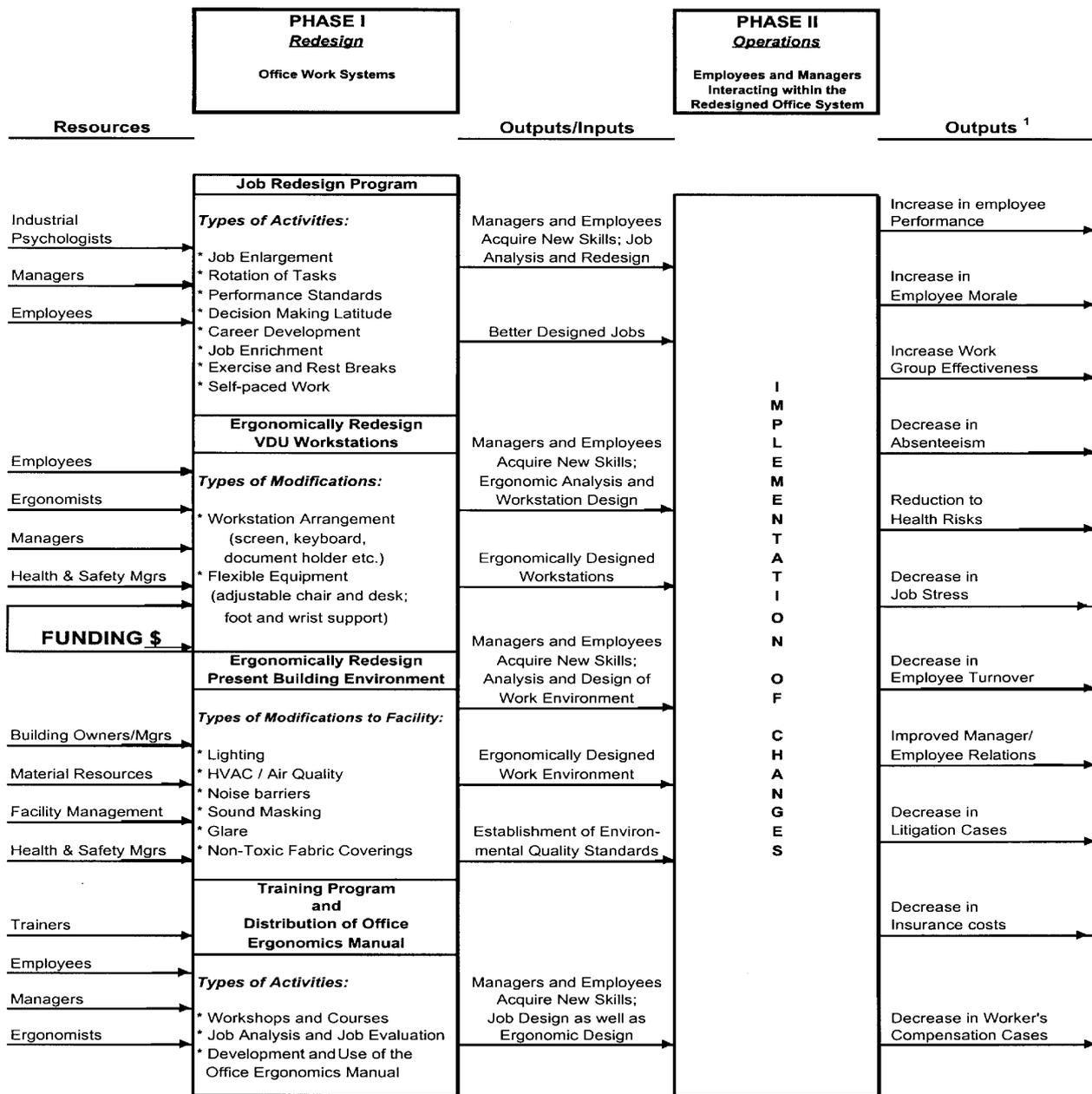
PHASE I *Redesign* Office Work Systems		PHASE II *Operations* Employees and Managers Interacting within the Redesigned Office System
Resources	**Outputs/Inputs**	**Outputs** [1]

Job Redesign Program

Industrial Psychologists

Managers

Employees

Types of Activities:

* Job Enlargement
* Rotation of Tasks
* Performance Standards
* Decision Making Latitude
* Career Development
* Job Enrichment
* Exercise and Rest Breaks
* Self-paced Work

Managers and Employees Acquire New Skills; Job Analysis and Redesign

Better Designed Jobs

Ergonomically Redesign VDU Workstations

Employees

Ergonomists

Managers

Health & Safety Mgrs

Types of Modifications:

* Workstation Arrangement (screen, keyboard, document holder etc.)
* Flexible Equipment (adjustable chair and desk; foot and wrist support)

FUNDING $

Managers and Employees Acquire New Skills; Ergonomic Analysis and Workstation Design

Ergonomically Designed Workstations

Ergonomically Redesign Present Building Environment

Building Owners/Mgrs

Material Resources

Facility Management

Health & Safety Mgrs

Types of Modifications to Facility:

* Lighting
* HVAC / Air Quality
* Noise barriers
* Sound Masking
* Glare
* Non-Toxic Fabric Coverings

Managers and Employees Acquire New Skills; Analysis and Design of Work Environment

Ergonomically Designed Work Environment

Establishment of Environmental Quality Standards

Training Program and Distribution of Office Ergonomics Manual

Trainers

Employees

Managers

Ergonomists

Types of Activities:

* Workshops and Courses
* Job Analysis and Job Evaluation
* Development and Use of the Office Ergonomics Manual

Managers and Employees Acquire New Skills; Job Design as well as Ergonomic Design

IMPLEMENTATION OF CHANGES

Increase in employee Performance

Increase in Employee Morale

Increase Work Group Effectiveness

Decrease in Absenteeism

Reduction to Health Risks

Decrease in Job Stress

Decrease in Employee Turnover

Improved Manager/ Employee Relations

Decrease in Litigation Cases

Decrease in Insurance costs

Decrease in Worker's Compensation Cases

Figure 3.   The two phases of an input–output flow diagram.

6. STEP 4: EVALUATING THE ALTERNATIVES: THE CRITERIA SCORECARD: COST/BENEFIT ANALYSIS

Evaluating alternatives is accomplished by measuring and comparing each alternative set of activities utilizing major decision criteria. These criteria generally include: project cost, risk of failure, effectiveness, and benefits for all appropriate future conditions. An evaluation criteria scorecard is developed for use in evaluating and comparing alternatives shown in table 1. The scorecard is based on a long-term perspective and incorporates the preliminary decision criteria previously defined in step 2.

Four alternatives were selected based on experiences that many organizations have had regarding activities for the implementation of office technologies (Robertson *et. al.* 1998; Sauter, Dainoff and Smith 1990). To complete a comprehensive cost-effective analysis for these alternatives, an economic advantage analysis is conducted. The economic advantage analysis identifies costs and benefits to provide an economic analysis for each alternative (Robertson and Rahimi 1991). This analysis further provides the detail which links the costs for each internal *micro* and *macro* work environment program and/or features of the physical work environment to the potential leverage in the investment in annual compensation for the employees in the improved environment.

The economic advantage analysis identifies costs and effectiveness metrics, including: (1) human resource costs (e.g. compensation, salary, turnover, and absenteeism, workers' compensation costs, injury costs); (2) facilities costs (work environment), (e.g. rentable space, operating costs, annual facility costs, furniture investments, technology and information investments, work environment strategy costs, construction costs); and (3) effectiveness measures (organizational (e.g. process

Table 1. Criteria Scorecard:Cost/Benefit Analysis[1]

Alternatives	Project Cost	Risk of Failure	Effectiveness	Benefits	Overall Rating
ALTERNATIVE A: Redesign job/job content Ratings:	–3 ($175,000)[2]	–2	+5	+6 (10%)[3]	6 (moderate)
ALTERNATIVE B: Ergonomically redesign workstation and enviroment Ratings:	–4 ($370,000)[2]	–3	+5	+6 (10%)[3]	4 (moderate)
ALTERNATIVE C: Redesign job/job content, ergonomically redesign workstation and enviroment, train managers and distribute office ergonomics manual Ratings:	–8 ($590,000)[2]	–1	+9	+9 (26%)[3]	9 (High)
ALTERNATIVE D: Ergonomically redesign workstation and enviroment, train managers and distribute office ergonomics manual Ratings:	–6 ($440,000)[2]	–2	+7	+8 (17%)[3]	7 (Moderate– high)

[1] In the four categories of Effectiveness, Risk of Failure, Cost and Benefits each alternative (A thru D) was subjectively rated on a scale ranging from 0 to 10 or 0 to –10. Since Risk of Failure and Project Cost are potentially negative characteristics, a rating scale ranging from 0 to -10 was used. A 0 rating indicates no cost; a -10 rating indicates a high cost. Effectiveness and Benefits are positive outcomes and therefore were rated on a 0 to 10 scale. A 0 rating indicates a low level of benefit/effectiveness. The scores from each of the four categories were summed to determine the "Overall Rating" for each alternative. Along with the numerical ratings, table 1 shows subjective descriptors of the ratings in parenthesis below each number in the "Overall Rating" column. Along with the numerical ratings table 1 shows the potential percentage increase in employee performance for each alternative in parenthesis below the ratings. Along with the numerical ratings, table 1 shows the approximate "project cost" in dollar amount for each alternative in parenthesis below the ratings.

[2] Costs are calculated per 100 employees.

[3] Benefits are evaluated in terms of % increase in worker and group performance.

efficiency, work environment changes, customer satisfaction, space utilization, unit/department (e.g. product development time, successful projects, number of customers; group and individual (e.g. error rates, amount of completed work, quality). These costs/benefits metrics are used for each proposed alternative to determine the economic advantage of each alternative or program/activities and may be expressed as a percentage of annual compensation demonstrated over "X" years.

7. STEP 5: SELECTING AN ALTERNATIVE: THE DECISION TABLE

The importance or weight as specified for a future condition is established for each evaluation criterion as determined by the system analyst. These weighted values are based on objective and subjective measures. A decision table similar to table 2 is developed structuring the evaluation information with alternatives listed on the "Y" axis and the future conditions statements on the "X" axis. A probability is determined for each of the stated future conditions. An example of future condition could be the probability and level of funding for the program. Each alternative would be evaluated in terms of level of funding and the probability of being funded.

8. STEP 6: IMPLEMENTATION, EVALUATION AND MODIFICATION

A schedule and sequence of tasks, responsibilities, and

requirements is developed for the implementation activities. This schedule might include a contingency plan with scheduled decision points and decision responsibilities. There are several scheduling techniques that are available that can be used as well as various database software programs for creating new databases containing effectiveness measures, costs, and other pertinent metrics collected during the systems analysis. Several other activities also occur in this step in order to define, establish, and develop the evaluation processes of providing feedback to the appropriate strategic decision-maker in the company regarding the results of the work environment–macroergonomic program. Using information gathered from the evaluation and feedback processes, selected modifications and changes to the program also occur. This process is a continuous feedback loop of applying the systems analysis approach to solve problems and to measure the effectiveness of any selected program.

9. CONCLUSION

The systems approach identifies the salient variables which influence employee health, stress, and performance within office systems. Described in this chapter is the use of a systems model incorporating the findings of empirical and field research, as well as case studies collected from a global sample related to office systems. Typically, companies focus on simplistic, microscale solutions for reducing occupational stressors associated with

Table 2. Decision Table: Selecting an Alternative Based on Future Conditions

Future conditions (eg.funding) Probability of funding	High level of funding 0.5	Moderate level of funding 0.25	Low level of funding 0.25
Alternative A: Redesign job/job content Ratings[1]	4	2	2
Alternative B: Ergonomically redesign workstation and enviroment Ratings[1]	5	1	3
Alternative C: Redesign job/job content, ergonomically redesign workstation and enviroment; train managers and distributeoffice ergonomics manual Ratings[1]	1	6	7
Alternative D: Ergonomically redesign workstation and enviroment; train managers and distribute office ergonomics manual Ratings[1]	2	4	5

[1] The numbers indicate the selection preference rankings based on the weighted criteria and overall rating score from table 1. Each alternative was subjectively rated on a 0 to 10 scale where a rating indicates a low preference and a 10 rating indicates a high preference.

intensive computer work. Such solutions yield limited, short-term gains and create an imbalance among the office sub-systems. The important benefits of the systems model is the understanding of office systems, and the effective integration of *micro-ergonomic* and *macro-ergonomic* approaches for solving organizational problems related to office environments for the longer term. This analysis assists ergonomists, health and safety professionals, and strategic organizational planners to evaluate alternative office environment intervention programs designed to address the identified objectives and problems based on direct and indirect

economic metrics. Due to rising healthcare costs, operating costs, as well as recent state and pending federal legislation, companies are now determining how to allocate corporate resources to address issues associated with employee problems related to the introduction of office technologies. Applying a systems analysis approach enhances an organization's competitive edge by improving the interaction and fit of the office worker, the job, the physical environment, and the organizational structure. The result is a healthy and effective office system work environment.

REFERENCES

BONGERS, P.M., De WINTER, C.R., KOMPIER, M.J., and HILDERBRANDT, V.H., 1993, Psychosocial factors and work and musculoskeletal disease. *Scandinavian Journal of Work, Environment and Health* 19(5), 297–312.

CAKIR, A., HART, D.J., and STEWART, D.F.M., 1978, *Research into the Effects of Video display Workplaces on the Physical and Psychological Function of Person*. Federal Ministry for Work and Social Order, Bonn, West Germany.

HENDRICK, H. 1987, Organizational design, In Salvendy, G. (ed.) *Handbook of Human Factors*, (New York: Wiley), pp. 470–94,

KUORINKA, I. and FORCIER, L., 1995, *Work Related Musculoskeletal Disorders (WMSDs): a reference book for prevention* (London: Taylor & Francis).

MOSARD, G. 1982, A generalized framework and methodology for systems analysis, *IEEE Transaction Engineering Management*, Vol. EM-29, 81–7.

O'NEILL, M.J. 1998, *Ergonomic Design for Organizational Effectiveness*, Lewis Publishers, New York.

ROBERTSON, M.M., and RAHIMI, M. 1990, A systems analysis for implementing video display terminals. *IEEE Transactions Engineering Management*, 37(1), 55–61.

ROBERTSON, M.M. and ROBINSON, M., 1995, Enhancing user control of VDT work environments: Training as the vehicle. *Proceedings of the Human Factors and Ergonomics Society 39th Annual Meeting*, Santa Monica, CA, 417–21.

ROBERTSON, M.M., ROBINSON, M., O'NEILL, M.J., and SLESS, J., 1998, Measuring the impact of work environment change programs: a systems approach. *Proceedings of the Human Factors and Ergonomics Society 42nd Annual Meeting*, Santa Monica, CA, 984–8.

SAUTER, S.L., DAINOFF, M.J., and SMITH, M.J. 1990, *Promoting Health and Productivity in the Computerized Office* (London: Taylor and Francis).

SMITH, M. and SAINFORT, P., 1989, A balance theory of job design for stress reduction. *International Journal of Industrial Ergonomics*, 4, 67–79.

SWANSON, N.G. and SAUTER, S. L., 1999, Psychosocial factors and musculoskeletal disorders in computer work. In Karwowski and Medras (eds), *The Occupational Ergonomics Handbook* (New York: McGraw-Hill), pp. 1813–19.

Anthropometry for Design

E. Nowak

Institute of Industrial Design, Warsaw, Poland

1. INTRODUCTION

The aim of this article is to present anthropometry as a set of measuring techniques and methods, and to prove its usefulness for the needs of design. Anthropometry originates from anthropology and is directed towards obtaining measurements of the man. Anthropology is the science about man. It deals with the changeability of the physical characteristics of man in time and space, and particularly with race differentiation, individuals' differentiation, ontogenesis and phylogenesis. In the English and US approaches, anthropology embraces the complete knowledge of man and can be divided into physical (biological) anthropology and cultural anthropology. In addition, there are other types of anthropology such as social anthropology and criminal anthropology.

Physical anthropometry is particularly useful for practical purposes. According to the accepted divisions most often used, physical anthropology can be separated into the following three basic parts:

- *population anthropology* (known earlier as race anthropology), which is the study of the intraspecies differentiation of man — including living conditions, history, and the present state;
- *ontogenetic anthropology*, which is the study of the ontogenesis of man;
- *phylogenetic anthropology*, which deals with the phylogenesis of man — that is, the origin of our species.

Anthropology is related in its research to biology and the humanities (archeology, prehistory, psychology, and pedagogy) as well as the technical sciences. Ergonomic anthropology deals with man as the basic unit of the man–technique system and, as such, has evolved and developed in tandem with technology and engineering. Together with other disciplines (including physiology and psychology), it aims at obtaining the best conditions for this system to function.

Ergonomic anthropology makes use of the scientific output of phylogenetic, ontogenetic, and population anthropology, and as a result of the problems it concentrates on, it benefits from various sections of medicine and psychology. An anthropologist dealing with ergonomics makes use of the classical anthropological science and develops this science for the needs of technology and engineering. Thus, he or she utilizes classical anthropometry, that is, the basic research methods applied to anthropology.

2. AIMS AND TASKS OF ANTHROPOMETRY

2.1. Measuring Methods

The basic anthropometric measurements of humans include:

- linear measurements
- angular measurements
- circumferences
- force measurements

Linear measurements include: breadth, height and length measurements. These are measured between recognized anthropometric points. Angular measurements are carried out between planes and lines that cross the human body. Body movements in the sagittal plane are termed flexions and extensions; back and head movements in the sagittal planes are termed bending to the right and to the left, and extremities' movements are termed adductions and abductions. According to these extremities, movements in the transverse plane are termed pronations and supinations, and back movements are termed left turns and right turns.

Circumferential measurements of the body are mainly carried out for the purposes of clothing design and for physical assessment. The basic measurements include: head, neck, chest, hips, arms, thigh, and shin circumferences. Force measurement is done in order to define the physical predispositions of man. In general, force is defined in relation to that exerted by the hand and foot. Moments of forces are used as data applied to the design of hand and foot control systems.

The basic aim of classical anthropometry is to provide objective and precise data on the somatic structure of man. In population anthropology, for example, anthropometry is used as a set of methods applied to defining biological differences that occur between human populations. Anthropometry in ontogenetic anthropology serves to assess the ontogenetic development of man, and provides data for defining the development process, the process of puberty, and the aging process.

Following the development of the body segments in particular ontogenetic periods, anthropometry describes changes in the proportions of the human body. During the physical development of humans, head measurements increase twofold, trunk measurements increase threefold, and limb measurements increase between four- and fivefold. The aim of anthropometry is not only to define differences in human body structure in relation to age, but also in relation to sex and the type of somatic structure. Significant differences can be found in somatic characteristics in men and women. In general, women are about 8–15 per cent smaller than men and are physically weaker.

Following the development of successive generations, anthropometry assists in defining and foreseeing developmental trends of populations. These concern such phenomena as development acceleration and secular trend. These phenomena result in significant differences between generations in somatic, morphological, and functional characteristics. Anthropometry describes these differences and provides data of somatic changes that occur in given populations. Pediatrics and ergonomics are able to make use of this data: pediatrics applies population data as biological standards to the evaluation of individuals; on the basis of somatic characteristics of a given population, ergonomics creates products adjusted to the body structure. This development in ergonomics has resulted in the production of methods and measuring techniques applied to anthropometry, and new techniques aimed especially at the needs of ergonomics, termed ergonomic anthropometry, have gradually evolved.

The main aim of ergonomic anthropometry is to provide data describing physical predispositions of humans in order to aid the design of work and living environments. This objective has resulted in the modification of existing methods and the development new ones. For example, in ergonomics many anthropometric measurements are performed on the basis of

classical anthropometric points and new fixed references. In classical anthropometry, The main reference basis for height measurements is the horizontal place of footrest — Basis (B). An additional vertical plane basis — Basis dorsalis (Bd) - has been introduced to meet the needs of ergonomics. This basis is mainly applied in the determination of body dimensions in the sagittal plane. These dimensions include depth measurements and reaches. To measure the body in the sitting position, two additional reference planes have been introduced. These are the horizontal seat plane — Basis sedilis (Bs) and the vertical plane — Basis sedilis dorsalis (Bsd). Ergonomic anthropometry has frequently developed reference systems to solve specific problems in respect of the requirements of constructors and designers of technical machines and appliances. The application of this method was used to determine the dimensions of the spatial zone of upper limb reaches (Damon *et al.*1966; Bullock 1974; Nowak 1978).

Initially, during the development of ergonomic anthropometry, adults were the main subject since at that time ergonomics dealt with the work environment of man. As ergonomics developed, its interests were extended to the life environment of man and included home ergonomics and leisure ergonomics. As the role of ergonomics increased, it was able to benefit from the use of data that contemporary anthropometry could supply on the various stages of human development. Assuming ontogenetic periods as the criteria for division, we can distinguish anthropometry for children and young people, adults and the elderly. Anthropometry applies adequate measuring techniques for investigating each of these group. For example, the length measurements of children up to one year old are done in the prone position by means of a special type of liberometer. The same measurements of the adult population are performed in the sitting position by means of the vertical anthropometer. The majority of methods and measuring techniques applied in anthropometry can be used for measuring disabled people. However, some of these are modified or simplified in view of the difficulties in obtaining measurements. For example, special measuring chairs are constructed to study people with lower extremities dysfunction as they can be studied only in the sitting position.

It is not easy to study the disabled using the methods applied by classical anthropometry. It requires a great deal of experience from those performing the experiments since the measurements need to be taken very quickly. The measuring methods that make it possible to carry out investigations at a distance are the most convenient, both for the subject and the investigator; these are termed non-tactile methods.

This type of method was applied by the Swedish researcher Thoren (1994) in carrying out measurements on a group of the disabled. A set of mirrors and cameras properly arranged and interrelated to CAD/CAM software made it possible to obtain anthropometric data in a spatial system very quickly. Photogrammetric methods are most often used in the measurements of disabled people, as these make it possible to assess deformities and changes in the body structure, dislocations of bones segments. It is also possible to define the shape and dimensions of the body regardless of the body position and its changes in time (Das and Kozey 1994). It should be mentioned that these kinds of methods are relatively expensive and not all research units can afford to use them.

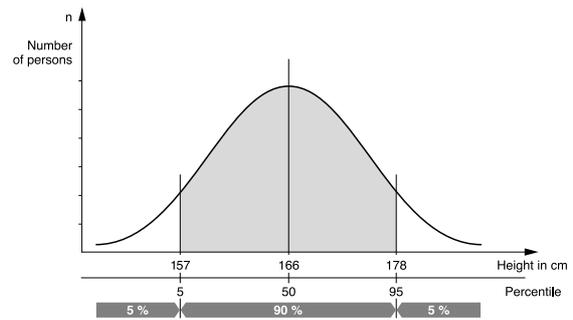Summarizing the above, it can be said that the sets of

Figure 1. Distribution of body height (stature) in Polish girls (18 years). About 90 per cent of all girls are between 157 and 178 cm tall; about 10 per cent are either shorter or taller.

measuring techniques and methods of the classical and ergonomic anthropometry can be applied to the measurements of the disabled. Some of these methods, however, require verification from the point of view of the arduousness of investigations.

2.2. Statistical Methods

Individuals with various body dimensions (tall and short) and body proportions (long or short extremities, or long or short trunks) exist in every population. In order to characterize a given population — that is, evaluate it in terms of numbers — anthropometry utilizes the basic statistical characteristics. Usually anthropometric features have the normal distribution and are arranged according to the "Gaussian distribution". Figure 1 illustrates such distribution.

It presents the body height of Warsaw girls aged 18 (Nowak 1993). The values of this feature were presented on the frontal axis, and the frequency of occurring (possibility) on the horizontal axis. Two basic statistical parameters are used to determine the distribution of features in a population. One of these parameters is the mean (m). It indicates where the distribution is located on the horizontal axis. The other is a quantity known as the standard deviation (S), which is the index of the degree of variability in the population under study — the "width" of the distribution or the extent to which individual values are scattered about or deviate from the mean. Mean values (m) and standard deviations are used to determine the statistical characteristics called percentile. They are useful both in developing biological standards and preparing standards for the needs of ergonomics. Assuming that the features investigated in the random test of the population have the normal distribution, as seen in figure 1, we can call percentile (Cp) the value of the feature that does not exceed p% of individuals. The values of particular percentiles (Cp) are calculated according to the following formula:

Cp = m + Sz

where Cp is the characteristics value on the level of the *p* percentile; *m* is the mean; S is the standard deviation; z is the constant for the percentile concerned (see statistical tables).

In order to obtain a better percentile interpretation we need to come back to figure 1. The height measurement of the investigated population of boys aged 18 distributes in a symmetrical way (Nowak 1993). Its highest point is the average stature, otherwise known as the mean. Since the curve is symmetrical, it follows that 50% of the population of girls are

shorter than average and 50% are taller. In this distribution the mean is equal to the 50th percentile. Other percentile values are also marked on the horizontal axis. The 5th and 95th percentile are used for designing purposes. The 5th percentile located closer to the frontal axis means that the 5% of the boys are shorter. Similarly, an equal distance from the mean towards the right of the chart is a point known as the 95th percentile. Hence, we can say that only 5% of the girls are taller. Ninety per cent of the population are between the fifth and 95th percentile in stature. Thus, using the values of the 5th and 95th percentile and applying the rules of ergonomics, products for 95 per cent of the population can be designed.

3. APPLICATION OF ANTHROPOMETRY

Examples of the applications of anthropometric investigations are given below; they are designs for work spaces, handles and holders, and working clothes.

Work space is determined mainly by the reach zone of the upper extremities defined in relation to three planes: sagittal, transverse and frontal. The first attempts to define the upper extremities reach zone concerned one-dimensional or two-dimensional configurations determined in frontal or transverse planes. Barnes (McCormick 1964) defined the upper extremities reach zone determining the so-called the maximum and normal zones. These zones were determined by the radii, which constituted the length of the arm or forearm. Further investigations aimed at determining the reach zone concerned three-dimensional configuration. Reach zones were defined on the basis of experimental research. Research works conducted in the USA (Damon *et al.* 1966) and Australia (Bullock 1974) should be mentioned here. The results of both experiments provided data for military purposes. The results of investigations conducted in Poland (Nowak 1978), thanks to the fact that the measuring system was "suspended" on acromial points, can be used while designing all kinds of work stands intended for work in both the standing and sitting positions. Work space of the upper extremities was defined based on investigations of 226 men and 204 women aged from 18 to 65 (Nowak 1978).

In these investigations, a spatial measuring system consisting of three inter-perpendicular axes is assumed to determine the arm reach area. The measuring system of the arm reach area was determined by the intersection of the following planes: the frontal plane, tangential to the vertical plane of the seat back, the sagittal–median plane, the transverse plane, crossing the acromial points.

The intersection of these planes marks the origin of the polcor co-ordinate axes of the measuring system (point C). The "C" point is fundamental to defining the reach areas for any working plane. The whole reach space of arm reach was divided into ten horizontal measuring planes. That which crosses the acromial points was accepted as the basis and marked "O". From this plane upwards four others follow every 120 mm and five follow downwards. The reach ranges of the left and right arms were recorded on each plane in a polar system with its center being the "C" point.

The investigations resulted in obtaining ten arm reaches for the left and right extremities. These reaches determined in a polar system define spatial area of arms. This space is a basis for designing spatial structures of machines, installations, and workplaces. Figure 2 shows an example of utilizing the data on

Figure 2. An example of utilizing the data on the reach area of arms in the design process.

the reach area of arms in the design process of a real workplace situation.

For the correct design of all kinds of handles and holders it is necessary to determine the functional capabilities of the hand and its anthropometric dimensions. The human hand belongs to specific and highly precise work tools. The grip capabilities of the hand are supported by the opposition of the thumb. This enables the hand to perform a wide range of manipulation tasks.

Figure 3. Hand grip named according to different classifications: power grip, hammer grip or clench grip palmar.

Almost all types of activities require different hand arrangements, which means using different types of grips. The performance of a handle — that is, its shape and form — should be adjusted to the grip most convenient to perform a specific activity. There are many grip classifications. One of them assumes, as a criterion of division, the form of hand arrangement or its parts in relation to the object and direction of effective force used (Nowak 1993). Figure 3 shows one of these hand grips.

Following Nowak (1993), this is a hand clamp grip making use of the thumb opposition. According to other authors, this grip is defined as a cylindrical or coiled grip. This grip is applied while using hand tools including saws, drilling machines or hammers, as well as when opening a door with the use of a door handle. When we know the dimensions of the space occupied by the hand grasped on a cylindrical handle and a handle diameter, we can easily determine the distance of a door handle in relation both to the door plane and the door frame.

Anthropometric measurements taken for the purposes of clothes design embrace a different set of characteristics from measurements made for the purpose of designing work spaces, machines, and tools. First of all measurements of the following circumferences are needed: neck, chest, waist, hip, wrist, and thigh. For the design of working clothes angular measurements of the extremities movement ranges are also essential, as well as measurements defined as arcs: for example, the arcs of the front length of the trunk, the length of the back, shoulders, upper and lower extremities are all very important. The dimensions of the arcs are determined in a motionless standing position and while changing body position — for example, moving to squatting, kneeling, and standing positions, for the latter, both bending forwards and bending backwards. Additionally, the construction of garments for wheelchair users is based on data obtained in a sitting position with consideration given to the trunk movement forward and movements of the lower extremities. Figure 4 shows selected anthropometric characteristics and their application in clothes design (Dabrowska-Kielek *et al.* 1993).
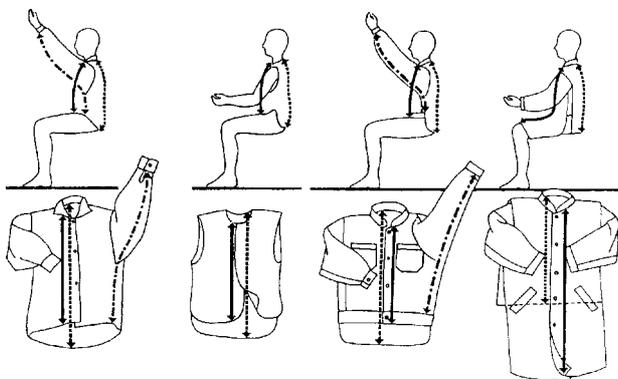
REFERENCES

BULLOCK, M., 1974, The determination of functional arm reach boundaries for operation of manual control. *Ergonomics*, 3, 375–88.

DAMON, B., STANDT, H.W., and McFARLAND, R.A., 1966, *The Human Body in Equipment Design* (Cambridge: Harvard University Press).

DAS, B. and KOZEY, I., 1994, Structural anthropometry for wheelchair mobile adults. *12th Triennial Congress of International Ergonomics Association,. Toronto,* 3, 63–5.

DABROWSKA-KIELEK, W., SZYMANSKA-PETRYKOWSKA, J., and SOCHA-DUDEK, E., 1993, Requirements for clothes design for the disabled using a wheelchair. *Conference: ''The Human Dimension: Ergonomics and Design''* (Warsaw: Institute of Industrial Design) pp. 68–9 (in Polish).

McCORMICK, E.J., 1964, *Human Engineering. Adjustment of Machines and Appliances Construction to the Man* (1st edn) (Warsaw: Panstwowe Wydawnictwo Techniczne) (in Polish).

NOWAK, E., 1978, Determination of the spatial reach area of the arms for workspace design purposes. *Ergonomics*, 7, 493–501.

NOWAK, E., 1993a, *Anthropometric data for designing a pupil's work place.* (1st edn) (Warsaw: Prace i Materialy IWP), p.148.

NOWAK, E., 1993b, Hand grips classification. *Conference: ''The Human Dimension: Ergonomics and Design'.* (Warsaw: Institute of Industrial Design), pp.30–1 (in Polish).

THOREN, M., 1994, Clothing made to fit the disabled users. *12th Triennial Congress of International Ergonomics Association, Toronto,* 3, 187–9.

Figure 4. Utilizing of anthropometric data for the construction of clothes meant for the disabled using permanently a wheelchair.

Anthropometry: Definition, Uses and Methods of Measurement

R.E. Herron

Professor of Ergonomics, Department of Environmental Health, Colorado State University, CO 80525, USA

1. DEFINITION

The word "anthropometry" was coined by the French naturalist Georges Cuvier (1769—1832). It was first used by physical anthropologists in their studies of human variability among human races and for comparison of humans to other primates. Anthropometry literally means "measurement of man," or "measurement of humans," from the Greek words *anthropos,* a man, and *metron,* a measure. Although we can measure humans in many different ways, anthropometry focuses on the measurement of bodily features such as body shape and body composition ("static anthropometry"), the body's motion and strength capabilities and use of space ("dynamic anthropometry").

2. ORIGINS

The origins of anthropometry can be traced to the earliest humans, who needed information about body parts for many of the same reasons which apply today — to fit clothing, design tools and equipment, etc. No doubt they also used body measurements for other, "non-design" purposes, such as footprints to estimate the body size of potential adversaries. These and other early applications called for the measurement or estimation of height, as well as the shape and size of hands, feet, and other body parts. Such needs gave rise to the very early use of the terms, "span," "cubit," and "canon," which connote extended arm width, height, and a standard, respectively.

Body proportions were of great interest during classical times, which is clearly evident in the work of artists and sculptors of the period. Around the year 15 BC, the Roman architect Vitruvius wrote about the potential transfer of harmonious body proportions to the design of beautiful buildings. However, it was the work of Renaissance artist-anatomists, including Alberti, Pierro della Francesca, Leonardo da Vinci, and especially Albrecht Durer, that ushered in the scientific beginnings of anthropometry. Durer's four-volume publication on human proportions was the first serious attempt to systematize the study of human size and shape.

3. USES OF ANTHROPOMETRY

Today, anthropometric measurements are used in a remarkably wide variety of scientific and technical fields, ranging from genetics and nutrition to forensics and industrial design. Within the field of ergonomics, there are myriad applications of anthropometry, primarily associated with different aspects of design for human use.

The goal of ergonomics is to design tools, workplaces, and environments in such a way that humans can function most effectively — in other words, to *optimize* human performance by achieving the best possible fit between the human operator, the equipment (hardware and software), and the working environment (physical and psychosocial). This fit is often referred to as "the human–machine interface." Anthropometry can and does play a major role in achieving this goal because variations in bodily features, such as shape, size, strength and reach, affect the way people perform tasks and, thus, have an important influence on whether the human–machine interface is a good one. The breadth of possible applications of anthropometry for improving the human interface is remarkably wide-ranging, from industrial equipment, clothing and furniture, to surgical tools, farm implements, aircraft controls, and virtually every item in the environment with which humans interact.

Over the years, engineers, designers, architects and others who design products or processes have increasingly recognized the need for anthropometric data on the users of their creations. Of course, the need for anthropometric information and the type of data required varies greatly from one application to the other. In some areas, the fit is "soft," as in a loose garment such as a bathrobe; in other areas, the fit is "hard," for example, in a respirator for protection against breathing toxic fumes. The fit of the bathrobe can be an approximation and still serve its intended purpose, whereas the respirator mask must conform closely to the geometry of the face in order to maintain adequate contact and prevent leakage. In the case of the bathrobe, data on height and a few body girth measurements for the prospective user population may be all the information needed to ensure adequate body coverage for a good interface. However, for the respirator mask, it may be necessary to obtain detailed three-dimensional measurements of individual facial geometry to ensure a satisfactory fit.

4. METHODS
4.1 Traditional methods

The earliest methods and tools for making anthropometric measurements were very simple, but they were quite effective and some of these rudimentary devices, such as measuring sticks and calipers, have endured to the present day. The types of data that result from the simple tools shown in figure 1 are quite varied and include numerous length measurements — e.g. height, various widths, such as the shoulder and pelvis, and circumferences, such as the waist and chest.

These tools appear to be deceptively easy to use but, in fact, when used for scientific or engineering purposes, they require a high level of care in order to achieve acceptable levels of validity
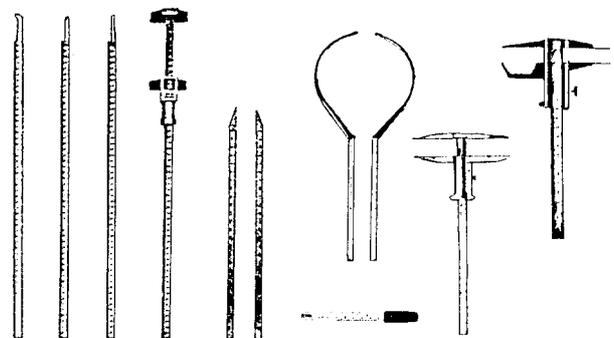

Figure 1. An array of traditional anthropometric measurement tools

and reliability. For example, it is important to know where to locate and how to align the measuring device (tape, caliper, measuring stick or whatever) on the body surface and to do so in a consistent, reproducible manner. In using a measuring tape, the tension in the tape and the degree of tissue compression must be suitably controlled.

Over the last hundred years, many volumes have been written about how to perform traditional anthropometric measurements and the reader is referred to the list of publications given below for further details.

4.2   Modern methods

Over the last thirty years the tools of anthropometry have changed dramatically, propelled by advances in computer and shape-sensing technology. Traditional linear measures of the body and body parts have gradually given way to 3-D and 4-D measures, computer models and, most recently, fractals, using new multi-dimensional sensing devices which capture considerably more of the subtle variations of human form and function.

The transformation of traditional anthropometry took a major turn in the 1960s with the growing recognition that the human body is an irregular, three-dimensional, dynamic organism, which calls for different strategies of mathematical abstraction and different measuring instruments than those used for obtaining traditional linear dimensions. This development led to greater use of mapping approaches, using contours and coordinates, and mathematical approaches based on polynomials, nurbs, b-splines, and other strategies for representing the irregular, multi-dimensional bodily features. A recent conference report (Vannier, Yates and Whitestone 1992) provides a valuable overview of these 3-D anthropometric methods and their future prospects.

Traditional anthropometry has been largely confined to surface measurements of the body, with the exception of a rather limited use of X-rays. The recent rapid growth of Computed Tomography (CT), Magnetic Resonance Imaging (MRI), Positron Emission Tomography (PET), and other new medical imaging devices has exposed the potential for generating previously inaccessible anthropometric data on internal body structures and functions. Obviously, knowing the dimensions and the motions of internal body parts can provide invaluable information for many ergonomic needs. For example, it would certainly be helpful

Figure 2. Modern anthropometric measurement methods

to know what happens to the geometry of internal organs and systems during the performance of various tasks and when the body adopts different postures.

Other recent developments in anthropometric recording and instrumentation strategies include the use of 3-D images for matching body parts in reconstructive surgery, automated hip surgery, measurement of heart and other organ volumes for transplants, stereometric brain surgery, and dose radiation measurements. Although, it may not be immediately obvious, all of these recent applications can translate into new potentials for advancing the use of anthropometry for ergonomic design and fitting purposes.

The use of 3-D and 4-D computer models to represent human form and function is still in an early stage, but already over 200 models have been developed for various ergonomic applications. Interest in creating more accurate and more versatile computer models of the human body has received a major impetus from the growing use of animation in the motion picture industry. This line of activity can be expected to grow dramatically in the near future as better (more comprehensive and statistically valid) multi-dimensional body data become available and the power of micro-computers continues to grow. Having a wide range of human models covering myriad details of human form and function stored in an immediately accessible form would seem to be the preferred *modus operandi* for helping engineers and designers to extend their uses of anthropometry for ergonomic purposes.

5. WORK SPACE–ENVELOPE MEASUREMENT

The measurement of the work space-envelope for different occupational activities is an important aspect of ergonomic design. The space-envelope occupied by the body while performing a task is larger than the space taken up by the body itself. The first measurements of this type were made using simple mechanical devices but, over the last twenty years, this approach has been superceded by non-contact, three-dimensional video imaging techniques. The video methods are portable, inexpensive, and easy to use.

6. STRENGTH MEASUREMENT

The measurement of human strength is important for the design of tools and equipment and other ergonomic applications. The growing availability of modern electronic dynamometers and strain gauges linked to computers has made the taking of human strength measurements a relatively simple operation; however, the acquisition of valid and reliable strength data requires considerable care and skill. Details about the different methods for measuring human strength can be found in the publications listed below.

7. RANGES OF ANTHROPOMETRIC DATA

In many ergonomic applications, it is necessary to know the ranges of pertinent anthropometric measurements found among a particular population, e.g. the prospective users of a new piece of equipment. The words "average" and "percentile" often appear in discussions about such matters, and the way these terms are used in anthropometry requires special attention. Although a single bodily characteristic, such as height or weight, can be expressed as average or at a particular percentile level, e.g. at the
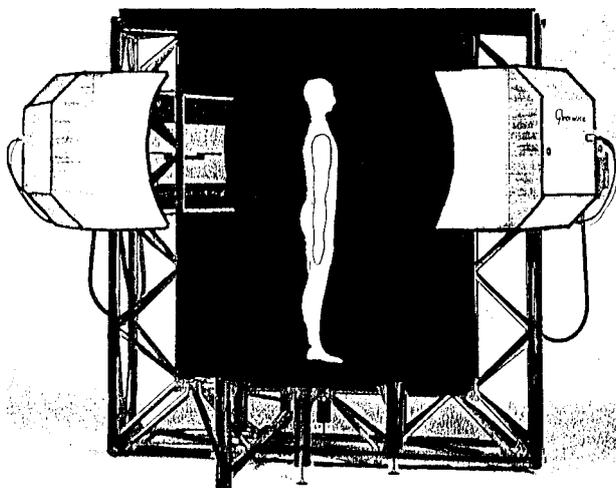
50th percentile level, there is no such thing as an "average" human or a "50th percentile human." This stems from the fact that the position of each individual relative to an average value or a percentile scale varies from one bodily feature to the other. An individual's height may be at the 75th percentile level (i.e. 75 % of the population are the same or of lesser height), but his weight may be at the 60th percentile and his chest girth at the 50th percentile. Thus, there is no such thing as a 95th percentile human — except as it relates to a specific bodily feature — because combining several 95th (or any other single) percentile values for various dimensions on the same body produces an unrealistic human form. Anthropometric variables can be combined, but this requires the use of multi-variate and other statistical methods, which are described in one or more of the publications listed below. These references also discuss the correct use of means, medians, standard deviations, standard errors, and other statistics for ergonomic purposes.

Another major issue in obtaining or producing representative anthropometric data relates to problems of population sampling. For example, if the goal is to design a piece of equipment for the general US population and there is a need to know the range of body shapes and sizes, there is no extant database of such information. The available anthropometric databases do not include a statistically valid sample of the US population. Therefore, it is often necessary to find a compromise, which might involve combining anthropometric information from various sources, such as military populations, and from limited samples of the general population. Furthermore, the ergonomic needs of special populations such as the elderly, disabled, infant/child and ethnic groups often call for anthropometric data which are representative of the particular segment or segments of the population.

Sampling strategies and tactics are explained in more detail in the references given below. However, the current problems in this area will not be alleviated in any significant way until a comprehensive anthropometric survey of the US population is completed. Recently, several countries have conducted national anthropometric ssrveys and others plan to follow suit in the near future. The cost and logistical difficulties involved in such surveys will probably limit them to the more prosperous nations for the foreseeable future, so that the development of a truly global anthropometric data base is unlikely for many years to come.

8. THE FUTURE OF ANTHROPOMETRY

Although the tools of modern anthropometry have already become quite sophisticated, no doubt the future will bring even more dramatic "high-tech" advances. These will include a wide range of new sensors (non-contact, multi-dimensional sensors, etc.) to measure the changes (static and dynamic, as well as internal and external) in body shape and size which accompany everyday activities, as well as those which are associated with growth, disease, aging, etc. Computer body models will reach a new level of realism based on more comprehensive and valid population data on a wider range of useful measures.

Another noteworthy development that illustrates what the future holds is the National Library of Medicine project launched in 1992 to create 3-D "visible humans." Each visible human is graphically reconstructed from the images of a series of fine cross-sections, taken from head to foot, of two "representative" human

cadavers — one male and one female. The first dissections of this type were conducted by Eycleshymer and Schoemaker in 1911 when they selected a "representative" series of cross-sections from a sample of fifty individuals. In 1974, a group at the Biostereometrics Laboratory, Baylor College of Medicine, similarly sliced a cadaver into 92 cross-sections, using a specially designed saw. The images of successive cross-sections were then used to reconstruct the 3-D geometry of selected internal organs and systems for use by army ballistics researchers studying what body tissues would be affected by different missiles entering the body at various locations and angles. Already, the National Library of Medicine visible humans have been widely used and have met important needs for teaching and other purposes, but their representativeness of anthropometric variation among humans is obviously quite limited.

Future anthropometric methods will cope better with the fact that humans are not fixed objects like statues, but are dynamic organisms whose structures and postures change daily and throughout life — these methods will recognize that there are no fixed points on the body, just ever-changing irregular forms. We must exploit the potential of more "holistic" parsimonious mathematical abstractions for representing, measuring, analyzing, and interpreting the subtleties of human form and function, both internally and externally, from conception to death. Comprehensive multidimensional human models based on statistically valid anthropometric data for a wide variety of populations and specialized groups will be instantly available for use by engineers, designers, and others, in their computers and virtual-reality simulators. Such a development will help to elevate the science and technology of anthropometry to a new level of precision and utility.

REFERENCES

Further details about the above methods can be found in the following publications:

ANTHROPOLOGY RESEARCH PROJECT STAFF (eds), 1978a, *Anthropometric source book. 17 Volume 1. Anthropometry forDesigners* (NASA Reference Publication 1024), (Houston: NASA Scientific and Technical Information Office).

BADLER, N.I., PHILLIPS, C.B., and WEBBER, B.L., 1993, *Simulating humans computer graphic animation and control* (New York, Oxford University Press).

CHAPANIS, A. (ed.), 1975. *Ethnic variables in human factors engineering* (Baltimore: Johns Hopkins University Press).

CHEVERUD, J., GORDON, C.C., WALKER, R.A., JACQUISH, C., KOHN, L., MOORE, A., and YAMASHITA, N., 1990, *1988 Anthropometric Survey of U.S. Army Personnel: Correlation Coefficients and Regression Equations,* Parts 1–5 (Tech. Report Natick/TR-90/033-36) (Natick, MA: US Army Natick Research, Development and Engineering Center).

DAMON, A., STOUDT, H.W., and McFARLAND, R.A., 1966, *The Human Body in Equipment Design* (Cambridge, MA: Harvard University Press).

DANIELS, G.S., 1952, *The "Average Man"?* (Tech. Note WCRD 53-7), Wright-Patterson Air Force Base (Dayton, OH: Wright Air Development Center).

HERRON, R.E., 1973, Biostereometric measurement of body form. *Yearbook of Physical Anthropology 1972*, vol. 16 (New York: American Association of Physical Anthropologists), pp. 8–121.

HERTZBERG, H.T.E., 1955, Some contributions of applied physical

anthropology to human engineering. *Annals of the New York Academy of Science,* 63, 616–29.

HUMAN FACTORS AND ERGONOMICS SOCIETY, 1988, *American National Standard for Human Factors Engineering of Visual Display Terminal Workstations* (ANSI/HFS 100-1988). (Santa Monica, CA: Human Factors and Ergonomics Society).

KELLY, P.L., and KROEMER, K.H.E., 1990, Anthropometry of the elderly: status and recommendations. *Human Factors,* 32, 571–95.

KENNEDY, K.W., 1964, *Reach Capability of USAF Population. Phase I: The Outer Boundaries of Grasping —Reach Envelopes for the Shirt-Sleeved, Seated Operator* (AMRL-TDR64-59) (Wright-Patterson Air Force Base OH: Aerospace Medical Research Laboratory).

KROEMER, K.H.E., 1985, Office ergonomics: workstation dimension. In D.C. Alexander and B.M. Pulat (eds), *IndustrialErgonomics* (Norcross, GA: Institute of Industrial Engineers), pp. 187–201.

KROEMER, K.H.E., KROEMER, H.B., and KROEMER-ELBERT, K.E. (1994). *Ergonomics* (Englewood Cliffs, NJ: Prentice-Hall).

LOHMAN, T.G., ROCHE, A.F., and MARTORELL, R., 1988, *Anthropometric Standardization Reference Manual* (Champaign, IL: Human Kinetics Books).

McCONVILLE, J.T., ROBINETTE, K.M., and CHURCHILL, T.D., 1981, *An Anthropometric Database for Commercial Design Applications* (Phase I, Final Report, NSF Grant DAR8009861) (Yellow Springs, OH: Anthropology Research Project).

McCONVILLE, J.T., TEBBETTS, I., and ALEXANDER, M., 1979, *Guidelines for Fit-Testing and Evaaluation of USAF Personal-Protective Clothing and Equipment* (AMRL-TR-79-2) (Wright-Patterson Air Force Base, OH: Aerospace Medical Research Laboratory).

MEINDL, R.S., ZEHNER, G.F., and HUDSON, J.A., l993, *A Multivariate Anthropometric Method for Crew Station Design (Statistical Techniques)* (AL/CF-TR-1993-0054) (Wright-Patterson Air Force Base, OH: Aerospace Medical Research Laboratory).

MIL-STD 1472C, 1981, *Human Engineering Design Criteria for Military Systems, Equipment, and Facilities* (Washington, DC: U.S. Department of Defense).

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION (NASA), 1986, *Man-Systems Integration Standard, Volumes I and 11* (NASA-STD-3000) (Houston: NASA).

PHEASANT, S., 1986, *Bodyspace: Anthropometry, Ergonomics and Design* (London: Taylor & Francis).

ROE, R.W., 1993, Occupant packaging. In J. B. Peacock and W. Karwowski (eds), *Automotive Ergonomics: Human Factors in theDesign andUse of Automobiles* (London: Taylor & Francis), (pp. 11–42).

ROEBUCK, J.A., KROEMER, K.H.E., and THOMSON, W.G., 1975, *Engineering Anthropometry Methods* (New York: Wiley).

SNYDER, R.G., SPENCER, M.L., OWINGS, C.L., and SCHNEIDER, L.W., 1975, *Physical Characteristics of Children as Related to Death andInjury for Consumer Product Safety Design* (Final Report UM-HSRI-BI-75-5) (Ann Arbor: University of Michigan, Highway Safety Research Institute. Prepared for Consumer Product Safety Commission).

TILLEY A.R., 1993, *The Measure of Man and Woman* (New York: Whitney Library of Design).

US DEPARTMENT OF JUSTICE, 1991, *Americans With Disabilities Act: Title III,* Federal Register (28 CFR Part 36), vol.56, no. 144 (Washington, DC: US Government Printing Office).

WOODSON, W.E., 1981, *Human Factors Design Handbook* (New York: McGraw Hill).

Biomechanics of Wrist in Computer Keyboarding

R. W. Marklin† and G. G. Simoneau‡

†Department of Mechanical and Industrial Engineering, and ‡Marquette University, Physical Therapy Department, 346 Walter Schroeder Complex, PO Box 1881, Marquette University, Milwaukee, WI 53201-1881, USA

1. INTRODUCTION

Within the past three decades of published literature upper extremity WMSD has often been attributed to mechanical and electronic keyboard usage (Tichauer 1978, Kroemer 1972, Armstrong 1986, Smith *et al.* 1981, Sauter *et al.* 1991, Bergqvist 1995). Based on published literature, it appears that the design of the conventional computer keyboard is implicated in the etiology of upper extremity WMSD among keyboard users for the following two reasons: (1) the often-cited occupational risk factors of repetitive movements and deviated posture of the wrist in the flexion/extension and radial/ulnar planes are an inherent part of typing on a computer keyboard; and (2) there are several cross-sectional studies that have shown a strong positive relationship between musculoskeletal discomfort or pain and keyboard usage (Duncan and Ferguson 1974, Smith *et al.* 1981, Sauter *et al.* 1991, Bergqvist 1995).

1.1. Design of Conventional Computer Keyboard

The conventional flat computer keyboard requires the operator to hold the hands and forearms in a relatively awkward position. With this keyboard, the operator needs to pronate the forearms substantially in order to hold his/her palms almost horizontal. In addition, the operator must deviate both wrists in the ulnar direction in order to rest his/her fingers on the home keys. Ulnar deviation from subjects typing on a conventional keyboard is typically 10° or more. Computer keyboard users typing on conventional keyboards also extend their wrists ~20° (Simoneau *et al.* 1999).

1.2. Design of Split Computer Keyboards

Typically, the design of split computer keyboards differ from the conventional keyboard in their slant angle (Figure 1). The typical conventional keyboard has a slant angle of 0°, a slope ranging from 0 to 15°, and a tilt angle of 0°. Fixed-angle split keyboards typically have 12.5° of slant (25° opening angle) and the adjustable-angle split keyboard can be adjusted from 0 to 90°. (Some adjustable-angle split keyboards can be separated at pivot point connecting the halves.) Studies have investigated whether commercially available split keyboards place the wrist in a more neutral posture than a conventional keyboard (Honan *et al.* 1995, Marklin *et al.* 1999). In summary, it appears that if a commercially available split keyboard has an opening angle of ~25° (12.5° slant angle), then wrist ulnar deviation is reduced to almost a neutral position in the radial/ulnar plane. Marklin *et al.* (1999) conducted a study on 90 experienced office workers to determine how commercially available split computer keyboards af-
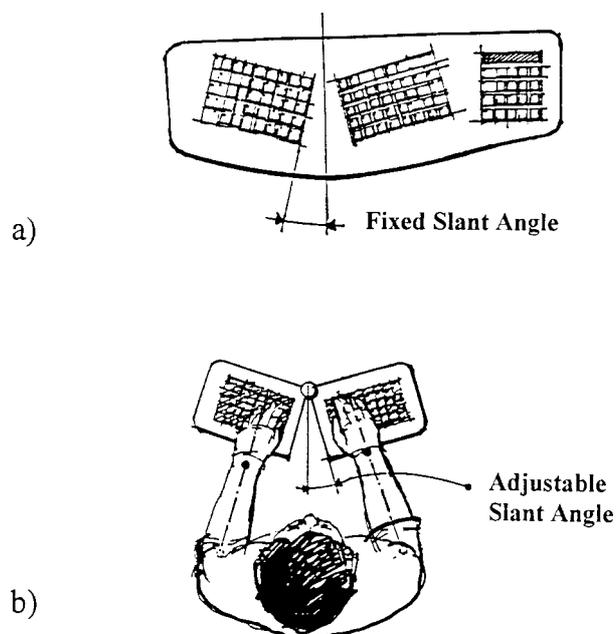

a)

Fixed Slant Angle

b)

Adjustable Slant Angle

Figure 1. (a) Sketch of a commercially available split fixed-angle keyboard. (b) Sketch of a commercially available split adjustable-angle keyboard

fected ulnar deviation of the wrist. As illustrated in Figure 1, the split keyboards tested had the QWERTY layout of keys and were of two designs — split fixed-angle and split adjustable-angle. The slant angle of the split fixed-angle keyboard was 12.5° (25° opening angle). The halves of the split adjustable-angle keyboard were adjusted to each individual so that each subject's wrists were aligned with the forearms, resulting in an approximately neutral radial/ulnar angle. The mean slant angle for all subjects who typed on the adjustable-angle split keyboard was 10.5° (SD 3.6°; range 5–20°). Therefore, the opening angle on the split adjustable-angle keyboard varied from 10 to 40° with a mean of 21°.

Compared with a conventional keyboard, the results from Marklin's *et al.* (1999) study indicate that split fixed-angle and split adjustable-angle keyboards reduced mean ulnar deviation of the right and left wrists. The split adjustable-angle keyboard reduced mean, maximum and minimum ulnar deviations by ~8° compared with the conventional keyboard. Compared with the conventional keyboard, the split adjustable-angle keyboard decreased mean ulnar angle from 13.3 to 5.7° for the left wrist and 10.7 to 2.5° for the right wrist. These differences in mean, maximum and minimum ulnar deviation were statistically significant. The split fixed-angle keyboard reduced mean, maximum and minimum ulnar deviation angles significantly by ~10° (mean left wrist: 16.5 to 5.8°; mean right wrist: 7.9° ulnar to 1.2° radial). Note that for all keyboards ulnar deviation of the right wrist was significantly less than the left wrist. This difference ranged from 3 to 7°.

The finding that split keyboards place the wrist closer to a neutral posture in the radial/ulnar plane substantially reduces one occupational risk factor of WMSD, namely ulnar deviation of the wrist. However, the two alternative keyboards had minimal influence on wrist extension. Both split keyboards resulted in wrist extension of ~15° (no significant difference between split fixed-angle and adjustable-angle keyboards), which was ~5° less

wrist extension than when typing on a conventional keyboard (Simoneau *et al.* 1999). Therefore, these two types of split keyboards did not fully address the posture component of wrist extension, which has been shown to have more effect on carpal tunnel pressure than radial/ulnar deviation. However, with respect to biomechanical factors other than carpal tunnel pressure, reduction in ulnar deviation is theoretically beneficial. This article explores in depth these biomechanical factors, namely carpal tunnel pressure and resultant reaction force on the tendons and median nerve.

2. CARPAL TUNNEL PRESSURE

Compared with a conventional keyboard, the goal of split keyboards is to decrease ulnar deviation of the wrist during typing. Although the ergonomics literature is replete with recommendations to reduce ulnar deviation of the wrist, experimental studies on carpal tunnel pressure have not shown an increase in carpal tunnel pressure from wrist ulnar deviation up to 15° (compared with a neutral wrist position). Weiss *et al.* (1995) used needle catheters to measure the effect of wrist position in the radial/ulnar and flexion/extension planes on carpal tunnel pressure. Although these researchers found the lowest carpal tunnel pressure of ~8 mmHg when the wrist was deviated ~2° ulnarly and 2° extension, the range of 10–15° ulnar deviation was the position where the greatest number of their 20 subjects recorded the lowest carpal tunnel pressure. Compared with neutral wrist posture, 10° of ulnar deviation does not increase carpal tunnel pressure, as supported by data from Rempel *et al.* (1997a, b). With six newtons of force applied by the fingers and the absence of fingertip loading, Rempel *et al.* found that carpal tunnel pressure for 10° of ulnar deviation was 36.1 and 15.4 mmHg respectively, while pressure for neutral position in the ulnar plane was 44.6 and 19.7 mmHg respectively. At 20° ulnar deviation, the carpal tunnel pressure increased to 40.9 and 21.5 mmHg respectively for 6 N and no loading on the fingertips. The effect of wrist extension on carpal tunnel pressure was more evident as contrasted to the effect from ulnar deviation (Rempel *et al.* 1997a, b). Compared with 15° wrist extension, carpal tunnel pressure increased from 41.1 and 18.5 mmHg respectively to 53.5 and 27.7 mmHg respectively for 6 N and no loading on the fingertips when the wrist was extended 30°.

In physiological studies on animals, pressures even as low as 20 mmHg could result in damage to the neuron, as demonstrated and reported by Dahlin and Lundborg (1990). Axonal transport decreased 75% when pressure applied to the vagus nerve of a rabbit increased from 10 to 20 mmHg. When the pressure increased to 30 mmHg, the nerve showed marked morphological changes, such as displacement of the neuron's nucleus and changes in the neuron's metabolism. If one were to generalize Dahlin and Lundborg's (1990) findings of detrimental changes to nerves from pressures as low as 20 mmHg to the human wrist, then wrist ulnar deviation up to 20° with no loading on the fingertips would not be a risk factor of WMSD. However, whether the wrist was in a neutral position or ulnarly deviated 20°, loading on the fingertips of 6 N could possibly expose one to WMSD because carpal tunnel pressure is > 30 mmHg. In studies where forces on computer keys were measured during typing tasks, peak key forces ranged from 2.5 to 4.4 N (Sommerich *et al.* 1996, Rempel *et al.* 1997a, b). Research is needed to determine the effect of repetitive applied forces in the above range on carpal tunnel pressure across a range of wrist ulnar deviation postures.

3. MODELING OF RESULTANT REACTION FORCE ON TENDONS AND THEIR SHEATHS

Although 10° of ulnar deviation of the wrist appears to have negligible effect on carpal tunnel pressure when compared with neutral posture, the greater ulnar deviation would increase theoretical resultant forces exerted against the flexor tendons passing through the carpal tunnel (Armstrong and Chaffin 1978, Schoenmarklin and Marras 1990). The increased resultant forces on the tendons and their sheaths could contribute to inflammation, possibly causing tenosynovitis. Swelling due to tenosynovitis could result in compression against the median nerve (carpal tunnel syndrome).

Modeling of tendons wrapping around a deviated wrist joint provides a basis for the etiology of WMSD affecting the wrist joint. As a forearm muscle contracts and moves its tendon accordingly, the tendon passing through the wrist can rub against its adjacent surface, usually a bone or ligament, as a rope rubs against a non-rotating pulley. Likewise, when the muscle lengthens, the tendon moves in the opposite direction against its adjacent structures. In the wrist area, the repetitive rubbing of the tendons against the carpal bones and flexor retinaculum (carpal ligament) could cause tenosynovitis, which is inflammation of the tendon's sheath.

Based on Landsmeer's (1962) model, Armstrong and Chaffin (1978) developed a static model of a tendon wrapping around a joint. Landsmeer's model of a tendon is analogous to a rope bent around a non-rotating pulley, and Figure 2 illustrates Armstrong and Chaffin's (1978) model as a reasonable representation of Landsmeer's model. When the wrist is flexed, the flexor tendons
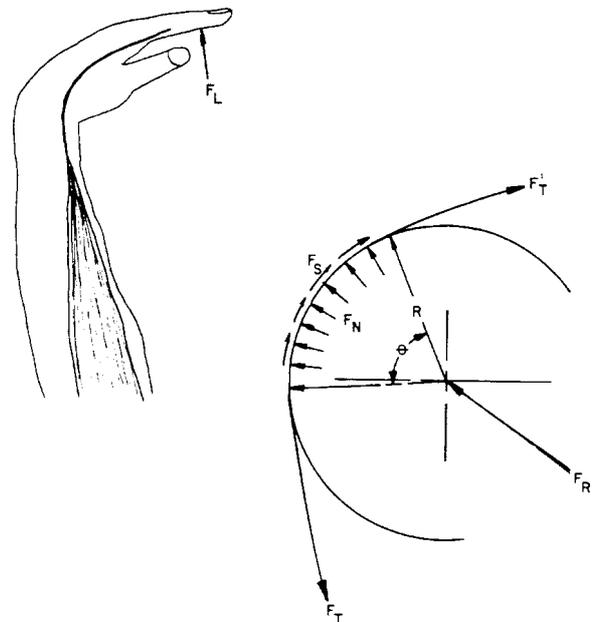

Figure 2. Armstrong's and Chaffin's (1978) biomechanical model of a flexor tendon wrapping around the flexor retinaculum. F_t, tendon force; F_r, resultant reaction force exerted against the tendon (Chaffin and Andersson 1991)

bend around the flexor retinaculum that is assumed to have a constant radius. When the wrist is extended, the flexor tendons are supported on the dorsal side by the carpal bones that are also assumed to have a constant radius. Armstrong and Chaffin (1978) found that the radius in a flexed posture is larger than in an extended posture.

The arc length of the tendon wrapping around the pulley is defined in equation (1) (Chaffin and Andersson 1991):

$$X = R \bullet q \tag{1}$$

where

X = tendon arc length around pulley (mm)
R = radius of curvature of supporting tissues (mm)
q = angle of deviation of wrist from neutral (in radians).

The resultant reaction forces acting normal to the tendon are shown in Figure 3 and defined in equation (2) (Chaffin and Andersson 1991):

$$F_n = (F_t \bullet e^{(m \bullet q)})/R \tag{2}$$

where

F_n = normal supporting force per unit of arc length (N/mm)
F_t = average tendon force in tension (N)
m = coefficient of friction between tendon and supporting synovia
q = wrist deviation angle (radians)
R = radius of curvature of supporting tissues (mm).

Since m is considered small (~0.0032; Fung 1981), it can be approximated by zero. This changes equation (2) to equation (3) (Chaffin and Andersson 1991):
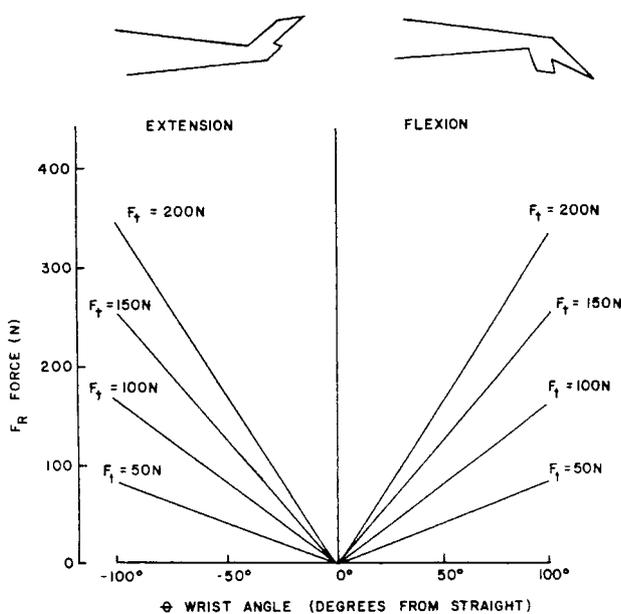
$$F_n = F_t/R. \tag{3}$$

Figure 3. **Resultant reaction force (F_r), as modeled by Armstrong and Chaffin (1979), that is exerted against the flexor tendons as a function of wrist angle and tendon force (Chaffin and Andersson 1991)**

Equation (3) reveals F_n is a function of the tendon force and radius of curvature. As the radius of curvature decreases, the normal supporting force per unit of arc length increases. The normal supporting force for women would be greater than for men because women have smaller wrists. Also, as the tendon force increases, the normal supporting force increases.

The resultant reaction force, F_r in Figure 2, is the force of the carpal bones pressing against the flexor tendons. F_r is defined in equation (4) (Chaffin and Andersson 1991):

$$F_r = 2 \bullet F_t \bullet \sin(q/2) \tag{4}$$

where

F_r = resultant force exerted by adjacent wrist structures on the flexor tendons (N)
F_t = tendon force (N)
q = wrist deviation angle (in radians).

Equation (4) indicates that F_r is a function of the tendon force and wrist deviation angle, but is independent of radius of curvature. Figure 3 illustrates this relationship in that as the tendon force and wrist angle increase, the resultant force F_r increases linearly.

Dynamic movements that accelerate and decelerate the tendons around a non-rotating pulley could exacerbate the trauma imposed on the tendons. Schoenmarklin and Marras (1990) developed a dynamic model of a flexor tendon bent around the carpal bones or flexor retinaculum, taking into account the acceleration and deceleration of a tendon's movements. This model analyzes the effects of peak angular acceleration on the resultant reaction force that the wrist bones and ligaments exert on tendons and their sheaths in the flexion/extension plane. Like the Landsmeer (1962) and Armstrong and Chaffin (1978) models, Schoenmarklin and Marras (1990) model the tendon as a rope bent around a fixed pulley.

The quantitative effects of the wrist's peak angular acceleration on resultant reaction forces were based on the free body diagram (FBD) and mass·acceleration diagram (MAD) approach in engineering dynamics. The maximum tendon force ($F_{t\text{-max}}$) was computed as a function of five peak angular accelerations (3000, 6000, 9000, 12 000 and 15 000°/s²). Based on empirical data from normal subjects, 15 000°/s² was ~50% of peak wrist acceleration in the flexion/extension plane (Schoenmarklin and Marras 1993). $F_{t\text{-max}}$ is derived from equation (5) (LeVeau 1977):

$$F_{t\text{-max}} = F_{t\text{-min}} \bullet e^{(m \bullet q)} \tag{5}$$

where

$F_{t\text{-max}}$ = maximum force in flexor tendons, which is the force that the extrinsic flexor muscles in the forearm exert on their tendons (N)
$F_{t\text{-min}}$ = minimum force in flexor tendons, which is the force that the flexor tendons transmit to the hand and fingers (N)
m = coefficient of friction between tendons and their sheaths
q = wrist deviation angle (radians).

Since the coefficient of friction for human synovial joints bones is estimated to be very low (0.0032 according to Fung 1981), then the calculation of the $F_{t\text{-max}}$ force is very close to $F_{t\text{-min}}$. F_r expressed in equation (4) was calculated as the resultant force necessary to resist $F_{t\text{-max}}$ and $F_{t\text{-min}}$.

5. RESULTANT REACTION FORCES FROM ULNAR DEVIATION

As shown in Figure 4, the resultant reaction force exerted against the tendon (F_r) increases approximately linearly as wrist angle or angular acceleration increases, resulting in a curved plane that signifies an interactive effect between wrist angle and angular acceleration. The greatest F_r occurs when wrist is accelerated at a deviated wrist posture. The large peak reaction forces exerted on the flexor tendons and their sheaths are due solely to deviated wrist position and motion *without* any externally applied load in the hand. If loads were applied in the hand (e.g. power grip or pinch grip) while the hand was accelerated in deviated postures, then F_r would increase even more, resulting in even more stress on flexor tendon tissue.

Compared with an approximately neutral wrist posture from typing on a split keyboard, ulnar deviation of 10° or more from typing on a conventional keyboard would result in increased resultant reaction forces on the tendons passing through the wrist (see equation 4). The resultant reaction force would increase even more if the wrist were accelerating (or decelerating) (Figure 4) or if the fingertips were applying a load (see equation 4). Since typing requires acceleration (and deceleration) of the wrist in the ulnar plane and forces exerted by the fingertips, then the resultant reaction force on the tendons would be greater than a static ulnar position of the wrist. The resultant reaction forces on the tendons could cause the tendon and its sheath or the fibrous sheath moored to bone or ligament to hypertrophy or inflame, which could result in tendinitis or tenosynovitis. The occurrence of either tendinitis or tenosynovitis would most likely increase m in equation (5), thereby increasing F_{t-max} and F_r even more (equation 4). Inflamed tendons or their sheaths would occupy more space in the carpal tunnel, thereby exerting more pressure on the median nerve and theoretically contributing to the etiology of carpal tunnel syndrome.

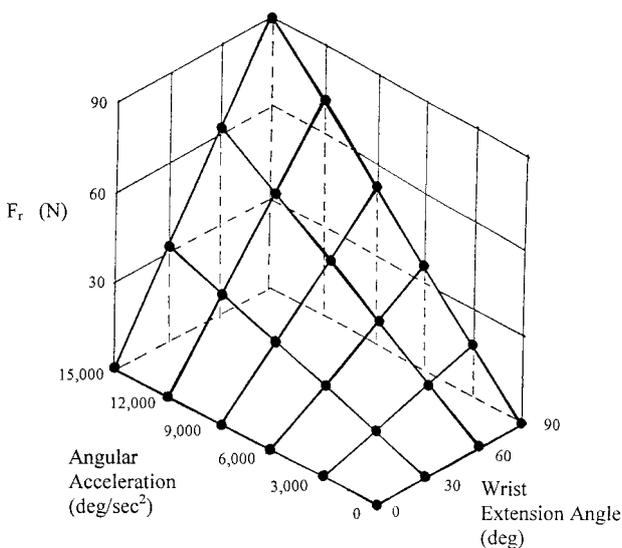Although the results from modeling of tendons in Figures 2–4 are theoretical (and to the authors' knowledge there are no published literature on measurement of resultant reaction forces), histological experimental studies on tendons passing through the carpal tunnel have shown degenerative changes in tendon structure from wrist deviation. Armstrong *et al.* (1984) found hypertrophy and increased density in the synovial tissue of the tendons passing through the carpal tunnel. These authors suggested that biomechanical factors, such as repeated exertions with a flexed or extended wrist posture, could have partially caused degenerative changes in tendon tissue. In an investigation of the viscoelastic properties of tendons and their sheaths, Goldstein *et al.* (1987) found that flexion/extension wrist angle increased the strain difference between tendons, their sheaths, and bones and ligaments that form the anatomical pulley. When the wrist was extended ~10°, the strain in the flexor digitorum profundus (FDP) tendons, which pass through the carpal tunnel and move the fingers, was ~10–15% lower on the side distal (hand side) to the flexor retinaculum than the proximal side (forearm side). This difference in strain within a tendon creates shear traction forces, which are magnified when the wrist angle is deviated to 65° flexion or extension.

Although Armstrong *et al.* (1984) and Goldstein *et al.* (1987) discussed degenerative changes to the tendons passing through the wrist with respect to wrist flexion/extension, their theories could possibly apply to ulnar deviation. The strain difference on the tendons from 10° ulnar deviation could possibly cause hypertrophy of tendon tissue in the carpal tunnel, thereby leading to the etiology of tenosynovitis or carpal tunnel syndrome from inflamed tendons and their sheaths pressing against the median nerve. Experimental research is necessary to validate whether ulnar deviation of 10–15° would generate a magnitude of resultant reaction force detrimental to the health of tendon tissue.

6. CONCLUSION

The reduction of ulnar deviation to almost a neutral position from typing on split keyboards minimizes one of the occupational risk factors of WMSD associated with typing, namely, ulnar deviation. The findings from Simoneau's *et al.* (1999) study and others reported in the literature that conventional computer keyboards consistently require ~10° or more of ulnar deviation could explain why typing on conventional keyboards has been problematic with respect to WMSD. Theoretically, as the wrist angle approaches neutral, the resultant reaction forces from the carpal bones and carpal ligament on the tendons and their sheaths passing through the wrist decreases (Armstrong 1986). Less net reaction force pressing against the sides of the tendons and their sheaths would theoretically decrease the incidence of tenosynovitis (Tichauer 1978). However, the theoretical benefits of reduced ulnar deviation from alternative keyboards are mitigated by carpal tunnel pressure studies that have shown that ulnar deviation of 10° to 15° degrees does not increase pressure in the carpal tunnel compared with neutral position (Rempel *et al.* 1997). Further reductions of WMSD risk factors from typing on split keyboards could be achieved by addressing wrist extension.

Figure 4. **Resultant reaction force (F_r) exerted by the carpal bones or flexor retinaculum against a flexor tendon and its sheath as a function of wrist angle and acceleration (Schoenmarklin and Marras 1990)**

REFERENCES

ARMSTRONG, T.J. and CHAFFIN, D.B., 1978, An investigation of the relationship between displacement of the finger and wrist joints and the extrinsic finger flexor tendons. *Journal of Biomechanisms*, 11, 119–28.

ARMSTRONG, T.J., CASTELLI, W.A., EVANS, F.G. and DIAZ-PEREZ,

R.D., 1984, Some histological changes in carpal tunnel contents and their biomechanical implications. *Journal of Occupational Medicine*, 26, 197–201.

BERGQVIST, U., 1995, Visual display terminal work — a perspective on long-term changes and discomforts. *International Journal of Industrial Ergonomics*, 16, 201–9.

CHAFFIN, D.B. and ANDERSSON, G.B.J., 1991, *Occupational Biomechanics* (2nd ed.) (Chichester: Wiley).

DAHLIN, L.B. and LUNDBORG, G., 1990, The neuron and its response to peripheral nerve compression. *Journal of Hand Surgery*, 15B, 5–10.

DUNCAN, J. and FERGUSON, D., 1974, Keyboard operating posture and symptoms in operating. *Ergonomics*, 17, 651–62.

FUNG, Y.C., 1981, *Biomechanics: Mechanical Properties of Living Tissues* (Berlin: Springer).

GOLDSTEIN, S.A., ARMSTRONG, T.J., CHAFFIN, D.B. and MATTHEWS, L.S., 1984, Analysis of cumulative strain in tendons and tendon sheaths. *Journal of Biomechanics*, 20, 1–6.

HONAN, M., SERINA, E., TAL, R. and REMPEL, D., 1995, Wrist postures while typing on a standard and split keyboard. In *Proceedings of the Human Factors and Ergonomics Society 39th Annual Meeting*, 366–8.

KROEMER, K.H.E., 1972, Human engineering the keyboard. *Human Factors*, 14, 51–63.

LANDSMEER, J.M.F., 1962, Power grip and precision handling. *Annals of Rheumatic Diseases*, 21, 164–70.

LEVEAU, B., 1977, *Biomechanics of Human Motion* (Baltimore: Williams & Lissner).

MARKLIN, R.W., SIMONEAU, G.G. and MONROE, J.F., 1999, Wrist and forearm posture from typing on split and vertically inclined computer keyboards. *Human Factors* (submitted).

REMPEL, D., BACH, J.M., GORDON, L. and SO, Y., 1998, Effects of forearm pronation/supination on carpal tunnel pressure. *Journal of Hand Surgery*, 23A, 38–42.

REMPEL, D., KIER, P.J., SMUTZ, W.P. and HARGEN, A., 1997a, Effects of static fingertip loading on carpal tunnel pressure. *Journal or Orthopaedic Research*, 15, 422–6.

REMPEL, D., SERINA, E., KLINENBERG, E., MARTIN, B.J., ARMSTRONG, T.J., FOULKE, J.A. and NATARAJAN, S., 1997b, The effect of keyboard keyswitch make force on applied force and finger flexor muscle activity. *Ergonomics*, 40, 800–8.

SAUTER, S. L., SCHLEIFER, L.M. and KNUTSON, S.J., 1991, Work posture, workstation design, and musculoskeletal discomfort in a VDT data entry task. *Human Factors*, 33, 151–67.

SCHOENMARKLIN, R.W. and MARRAS, W.S., 1990, A dynamic biomechanical model of the wrist joint. In *Proceedings of the 34th Annual Meeting of the Human Factors Society*, Orlando, FL, 805–9.

SIMONEAU, G.G., MARKLIN, R.W. and MONROE, J.F, 1999, Wrist and forearm posture of users of conventional computer keyboards. *Human Factors* (accepted).

SMITH, M.J., COHEN, B.G.F. and STAMMERJOHN, L.W., 1981, An investigation of health complaints and job stress in video display operators. *Human Factors*, 23, 387–400.

SOMMERICH, C.M., MARRAS, W.S. and PARNIANPOUR, M., 1991, A quantitative description of typing biomechanics. *Journal of Occupational Biomechanics*, 6, 33–55.

TICHAUER, E.R., 1978, *The Biomechanical Basis of Ergonomics: Anatomy Applied to the Design of Work Stations* (Chichester: Wiley).

WEISS, N.D., GORDON, L., BLOON, T., SO, Y. and REMPEL, D.M., 1995, Position of the wrist associated with the lowest carpal tunnel pressure: implications for splint design. *Journal of Bone and Joint Surgery*, 77A, 1695–9.

Consumer Product Design

A. Hedge
Department of Design and Environmental Analysis
Cornell University, Ithaca, New York, USA

The creation of consumer products drives much of the world's industrial economy. Consumer product usability provides a competitive advantage to companies in today's market place. Human factors deal with the application of human–system interface technology to enhance usability (Hendrick 1996) and it plays an important role in all phases of the consumer product design and development process. The label "ergonomically designed" is applied to a diverse array of products, from simple products, such as erasers and pens, to more complex products, such as computers and cars. Ergonomic design is used to denote improved product usability, whether the label is applied to hardware products, or to software. The label "ergonomically designed" is also used to differentiate between the depth and quality of thinking in the design. By definition, an ergonomically designed product should be one that is most appropriate to the task to be performed, that best fits the range of product users in terms of their dimensions, abilities, expectations and skills, and that is the easiest to use in an efficient way. Ergonomically designed products are not inherently more expensive to produce than non-ergonomic counterparts. What truly differentiates ergonomically designed products is that they are the outcome of a systematic application of human factors principles and knowledge to the product design and development process.

1. DEFINITION OF A CONSUMER PRODUCT

The label "consumer product" is very broad, and it encompasses a diverse array of products, ranging from comparatively simple products, such as kitchen utensils, to much more complex technologies, such as a car. Figure 1 shows a useful taxonomy in which products are categorized along two dimensions, design complexity, which ranges from simple to complex, and cognitive demands, which range from low to high. Products shown in this figure are organized according to these two dimensions, for example, a television (TV) is a high complexity product in terms of its electronic design and manufacture, but it poses a low cognitive demand on users, often requiring little more than pressing an on/off switch, channel changer and volume control for successful operation. By way of comparison, a car is also a complex product, but driving a car places a high cognitive demand on users, and consequently users need to be trained in how to operate this product successfully. The fundamental goal of human factors in consumer product design is to reduce the cognitive demands as much as possible so that even the most complex of products is inherently easy to use by the broadest possible range of people.

2. PRODUCT DESIGN PROCESS

Consumer products are the outcome of a product design process. The length and complexity of this process varies according to
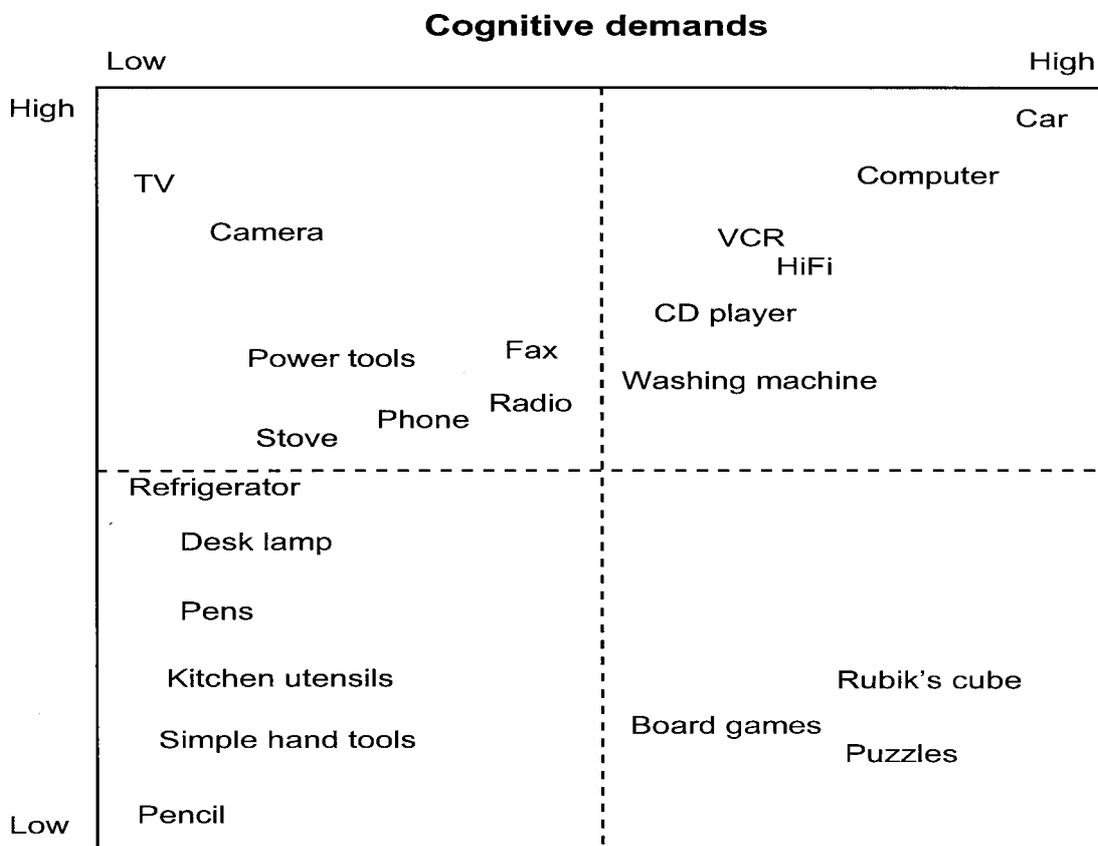

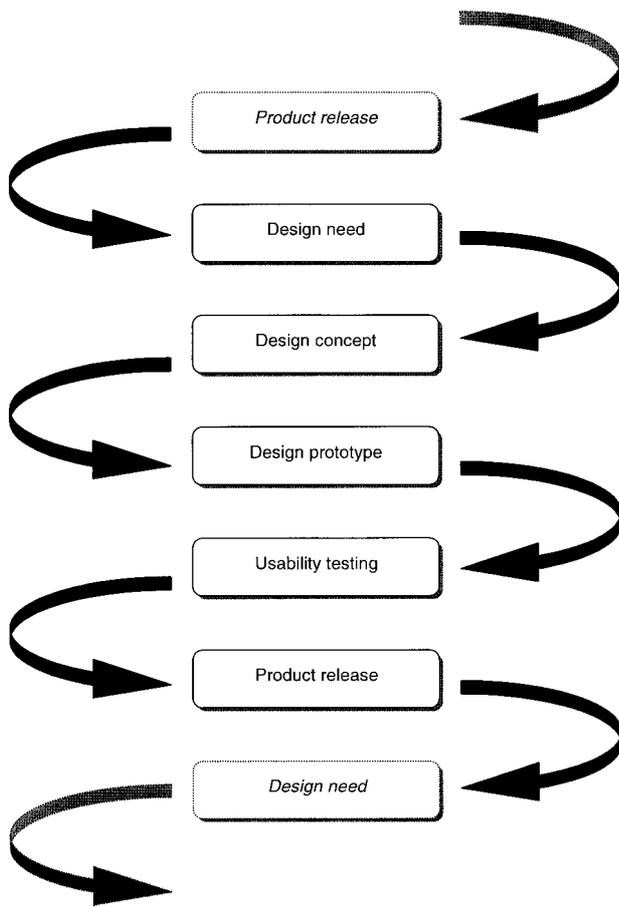Figure 1. Taxonomy of consumer products

Figure 2. Five functional stages in the design process for a consumer product

the complexity of the product. The design process is less complex and faster for a simple product, such as a can opener, where the design process may occupy only a few weeks, than it is for a complex product, such as a car, where the process may include consideration of a much wider variety of human factors issues and extend over several years. Human factors specialists are not designers, and typically they work with designers and other professionals on a product design team providing information input as appropriate.

The diversity of consumer products makes it difficult to define a universal design process. Also, the design process does not tend to be a linear sequence of activities and it may be an almost continuous activity. As soon as one model design is finished, work starts on revising this for the next model. Consequently, it is useful to think of the design process as a general sequence of five stages of functional activities, and good ergonomic design requires human factors input at each of these stages, regardless of the nature and complexity of the product (Figure 2).

2.1 Design Need

All product design starts with the identification of a need. This need may involve defining the capabilities of a new product or it may involve defining design changes required to improve an existing product. Human factors specialists use a variety of information sources to help in this stage of the design process. Developing a clear understanding of user requirements is critically important to the development of useful and usable products.

Sometimes this involves gathering information directly from users through interviews and surveys, and sometimes it involves processing data on user accidents, injuries, and errors in operating the product, which can help to pinpoint deficiencies in current product designs that can be rectified in future designs.

Often, new product development or product redesign rests on developing a clear understanding of the consumer's wants or needs. This requires the application of techniques that can profile consumer desires and behavior in a way that produces appropriate design-oriented information. Various methodologies are employed at this stage, and typically they include at least the following:

- *User Requirements Analysis* — determining what the user really needs. This may be done using focus groups, consumer interviews or consumer surveys. It may also involve applying techniques such as Delphi analysis and Brainstorming.
- *Competitive Analysis* — comparison of features and capabilities for existing products with a view to identifying gaps.
- *Product Safety Assessment* — assessment of safety and performance issues with current products. Ensuring product safety and designing for accident prevention is the designer's responsibility. To achieve this requires ideally anticipation of all possible patterns of use and misuse. The best designed products will enhance usability and, at the same time, prevent injuries and accidents. A common approach to improving product safety is to provide warnings, labels, and instruction manuals which inform and warn the user about appropriate and inappropriate product operation. There is a substantial human factors literature on these issues.
- *Task Analysis* — systematic evaluation of the tasks to be performed and an assessment of how well existing products meet the requirements of each task. Task analysis provides details of user expectancies, knowledge, and actions during product operation.
- *Allocation of function* — functional analysis of the existing tasks can help to identify gaps in current product offerings that may offer new market possibilities. For a particular functional requirement, an allocation of function exercise can clarify what functions are required of a product and what a person will perform. For example, the development of the autofocus capability for cameras has automated a once manual function, and this has helped to improve product performance by reducing photographic errors (blurry, out of focus images), and it has also simplified the use of even complex cameras.

The human factors professional can also help to define specific requirements for different user populations — for example, defining the different ranges of issues when designing for children, adults, differently abled people, pregnant women, the young or the elderly.

2.2 Design Concept

Once design needs have been agreed by a design team these must be translated into design concepts. Design concepts are design ideas that can be built if desired. The translation of design needs into design concepts may involve writing a document — the design program — which states the goals of a design and ways in which the concept will satisfy these goals. It may also involve

preparing preliminary design drawings, either sketches or computer simulations of the product. This allows the design team to gauge product appearance and to work with clients to determine which concepts will be pursued. Usually, several different design concepts can satisfy a design. Human factors help in the process of selecting between alternative design concepts by analyzing alternative product designs based upon detailed consideration of the interactions between the user and features of the product. This systems analysis of the user–product interaction is the unique domain of human factors and the information produced allows the designer to identify product strengths and weaknesses.

At this stage there is also a need to consider relevant safety and design standards for products, as well as potential product liability issues.

2.3 Design Prototype

Often it is difficult to select among alternative design concepts without building corresponding design prototypes to be able to hold and manipulate. Human factors can play an important role in testing the ergonomic design of alternative product prototypes. This testing can take various forms. In terms of product dimensions and shape the testing might involve comparing product dimensions against appropriate anthropometric norms for the range of user dimensions. Other relevant dimensions may be tested, such as reach distances, clearance dimensions, strength.

The testing of a design prototype may also include evaluating other design attributes:

- design of displayed information, either as text, symbols, images.
- use of color, shape, and texture, especially for the design of any controls.
- weight of the product and its center of mass.
- arrangement of displayed information and controls.
- spatial mapping relationships between controls and displays, to check for concordance with any population stereotypes.
- sight lines and user orientation to the product.
- user expectations about how product features will work.
- overall user reactions to the product appearance, features, and comfort, and especially to its ease of use.

The ease of use of a product depends on the goodness of fit between the design features and the user's characteristics. Product features must match the user's anthropometric, biomechanical, sensory, and cognitive capacities and limitations. This can be particularly challenging when a diverse user population will use a product. In such situations, the human factors specialist may test the prototypes with those at the extremes of the user population, as well as those around the average, to determine the performance limits of product performance.

Human factors testing of a design prototype may be done in several ways:

- on paper, where it might involve comparisons between designs using a scaled 2D anthropometric mannequin.
- on computer, using a 3D virtual mannequin.
- in focus groups where alternative prototypes can be evaluated and information collected on the group likes and dislikes of specific products features.
- as fitting trials in which people actually hold and handle the product prototype.

2.4 Usability Testing

Those design prototypes that show the most promise may be built as functional prototypes or as functional computer simulations. These functional prototypes can then be tested more extensively in either laboratory or real-world settings to establish their usability and performance. Usability testing of a product should be undertaken prior to market release because it provides invaluable information about the product in operation. Stanton, 1998) proposes that in consumer product design at least four factors determine usability:

- *Learnability* — the product should allow users to attain an acceptable performance level in a reasonable time.
- *Effectiveness* — the product should attain acceptable performance over a desired range of tasks, in a specified range of environmental settings when operated by a specific user group.
- *Attitude* — attaining acceptable performance must be possible without undue user fatigue, discomfort, stress, frustration, or dissatisfaction.
- *Flexibility* — the extent to which the product can deal with tasks beyond those originally specified.

2.5 Product Release

When a product is ready for release into the marketplace, human factors also play an important role in defining the learning and training requirements for appropriate product use. Often the correct use of a product requires the creation of appropriate documentation, such as user manual or on-line help system. If the product is sufficiently complex it may require actual user training, and human factors plays a role in instructional design to ensure efficient acquisition of the requisite knowledge and skills to be able to use the product in a proficient way.

3. EXAMPLES OF ERGONOMIC PRODUCT DESIGNS

There are numerous examples of products that have benefited substantially from human factors input into their design. Examples include:

- bent handle tools, such as hammers and pliers, designed to improve wrist posture and reduce upper-limb injuries (Dempsey and Leamon 1995)
- domestic appliances, such as washers and dryers (Martel 1998)
- assistive chair designs that help the elderly and those with disabilities to sit and rise with ease (Mizrahi *et al.* 1995).
- electric guitars (Marmaras and Zarboutis 1997)

4. BARRIERS TO THE ERGONOMIC DESIGN OF CONSUMER PRODUCTS

Human factors input into the product design process can be substantial when there is early involvement, but frequently this does not occur and the process is suboptimal for various reasons:

- *Professional barriers* — product design is frequently undertaken by industrial designers or engineers who may have only a rudimentary understanding of the discipline.
- *Organizational barriers* — top management are often unaware of the human factors , and consequently they fail to include a human factors professional as a design team member from the outset.

- *Process barriers* — often human factors input is not sought until a late stage of product design process. Early design decisions may be made based upon engineering and economic considerations, and by the time a human factors analysis is undertaken it may be too late to change the product design fundamentally in order to enhance usability.
- *Temporal barriers* — with economic pressures to shorten product development cycle times, human factors may be excluded when companies erroneously believe that human factors input will substantially lengthen the product design process.
- *Product barriers* — although human factors professionals can usually demonstrate performance differences between alternative product designs, users may prefer designs that perform suboptimally for a variety of non-ergonomic reasons, such as appearance and familiarity (Andre and Wickens 1995).

5. PROFESSIONAL ACTIVITIES

In 1998–9 the consumer products Technical Group was the fifth largest technical group of the human factors and Ergonomics Society with 496 members, which represents some 10% of society membership. The topic of consumer products was the largest single category in survey of ergonomics research publications between 1990–3, with some 4,500 papers, which represents over 20% of publications during this time (Stanton 1998). In 1993 the human factors and Ergonomics Society began publication of a design-oriented journal, *Ergonomics in Design*.

REFERENCES

ANDRE, A.D. and WICKENS, C.D., 1995, When users want what's not best for them, *Ergonomics in Design*, October, 10–14.

DEMPSEY, P.G. and LEAMON, T.B., 1995, Implementing bent-handled tools in the workplace, *Ergonomics in Design*, October, 15–21.

HENDRICK, H.W., 1996, Good ergonomics is good economics, *Proceedings of the 40th Annual Meeting of the Human Factors and Ergonomics Society*, Santa Monica, CA, pp. 1–16.

MARMARAS, N. and ZARBOUTIS, N., 1997, Ergonomic redesign of the electric guitar, *Applied Ergonomics*, 28, 56–67.

MARTEL, A., 1998, Application of ergonomics and consumer feedback to product design at Whirlpool. In Stanton, N. (ed.), 1998, *Human Factors in Consumer products* (London: Taylor & Francis) pp. 107–26.

MIZRAHI, I., ISAACS, B., BARNEA, T., BENTUR, N., and SIMKIN, A., 1995, Adapting the chair and adapting the user, *Ergonomics in Design*, April, 27–32.

STANTON, N. (ed.), 1998, *Human Factors in Consumer Products* (London: Taylor & Francis).

Design of Automobile Interiors

L. Bandini Buti

Società di Ergonomia Applicata, Milan, Italy

1. INTRODUCTION

The ergonomic aspects of motor vehicle interiors are not explicitly treated by the European Community directives on "approval of motor vehicles," which, on the other hand, define performance in relation to active and passive safety, to emissions and to the characteristics that various vehicle typologies must have.

The SAE normatives (Society of American Engineers) apply to the safety of motor vehicles, and define the anthropometric standards for the dimensional relationship between man and automobile. Car manufacturers have agreed on a common normative standard based on those of the SAE (European normatives ECIE — European Car Manufacturers Information Exchange Group) to diffuse and elaborate criteria for habitability and to define the car interior "package." The Community Directive for cars, which, on the other hand, explicitly requests that ergonomic principles be respected, is not, however, applicable to vehicles in movement, and therefore not to motor vehicles.

Industry has felt the need more and more to introduce ergonomic principles into the design of vehicle interiors, irrespective of the need to comply with laws and to adhere to regulations. In fact, the major car manufacturers have created bodies and laboratories concerned with ergonomics, some of which indeed have been active since the 1960s (Wisner *et al.* 1965).

2. EVOLUTION OF INDUSTRIAL PRODUCTS

The request for the introduction of ergonomic concepts into industrial products has been increasing since the beginnings of industrialization in proportion to factors regarding consumer expectations, developments in technology and the involvement of those carrying out the planning. In the evolution of industrial products there are homogenous periods.

2.1. First Period, up to the 1950s — Technology Engineers

In this first period design is characterized by the supremacy of technology and engineers. Products are "simple" from the point of view of use in that every function is performed by an easily identifiable component. One need only think of the structure of the mechanical typewriter, which is heavily conditioned by the levers that control the hammers. The structure of the product unequivocally indicates its function.

The automobile of the 1930s and 1940s was a product determined by technology. It is in this period, in fact, that those technological and mechanical innovations that will profoundly change it and give it the characteristics of a mature product are developed. Engineers conceived the functional aspects of the vehicle produced, and the vehicle was then "dressed up" by bodywork and parts design specialists and craftsmen. The purchaser of the automobile was interested above all in performance offered and technological innovations, which were often revolutionary. In this period ergonomics, in terms of an organic attitude of attention to the characteristics and needs of the user, may be considered absent.

2.2. Second Period, up to the 1990s — Image Designers

The second period, which continues to the present day, is characterized by the supremacy of image and designers. Products, materials and technology are still reasonably simple, and so can be easily dominated by designers by means of a general knowledge of their field. Technological aspects are considered as an important part of the design process but are, however, subordinate to the objective of providing products with original formal qualities.

In the 1970s the evolution of the market and various other factors, among which was the petrol crisis, began to exacerbate competition between car manufacturers who started investing, above all, in image. Also buyers, more frequently, choose according to image, which moreover responds very well to new life-styles. This is why many products with a fairly poor or traditional technological content are renewed by means of design inventiveness, which is also the invention of new modes of use that are more coherent to the evolution of society. One may take as an example portable televisions that destroy the myth of the television as a status symbol.

Respect for human characteristics in the design of motor vehicles comes from the sensitivity of the designer. Ergonomics begins to be introduced, but as an interventional option near the end of the design process, with the aim of adding further qualities, which are considered, however, as accessories.

2.3. Third Period, from the 1990s — Quality Work Groups

In the past few years, the organizational model that left the development of mass-produced motor vehicles to the creativity of designers and their reliability to technicians has found itself in crisis because it has been seen that the product, as realized by the technicians, ended up corresponding less and less to the synthetic and ingenious images of the artist. This has been proven by the constantly increasing complexity of the design and realization process, owing to norms and regulations (for safety and comfort), to the request for ever increasing levels of performance (usability), to new trends (ecology and energy efficiency), and to obligations dictated by marketing, technology and Information Technology (IT). An ever higher level of specialization in design and production is also called for, along with a new role for component suppliers who end up conditioning overwhelmingly the development of the design concept.

Thus, we arrive at the third phase, which represents the tendency currently in development, and which is characterized by quality and work groups.

Users take it for granted that a product will meet expected levels of performance, they appreciate more and more the qualities of use of a product, and they evaluate characteristics of comfort and pleasantness which, in the end, influence and direct their

choice of purchase. To meet these needs, many areas of manufacturing, but in particular that of motor vehicle production, are developing new organizational models that have quality as their objective and propose a design phase organized in work groups, in which specialists from various sectors relative to the development of the project and to various disciplines participate. Among these, those who study usability and pleasantness are becoming ever more important. At this stage, the application of ergonomics to design becomes a necessary phase, not just a verification of the finished product.

3. ERGONOMIC QUALITY IN MOTOR VEHICLES

In the automobile industry, the focus on ergonomics in the design of vehicles is, these days, automatic in that it is foreseen and imposed by procedure. Ergonomics is therefore present right from the start of the design process, and the possibility of making changes is still great, which allows ergonomics to have a great influence right from the initial stages of conception of the product. In fact, if one waits until a prototype is available to carry out ergonomic verifications, as still happens with some cases, the possibility of making significant changes becomes almost zero and any changes would generate extremely high costs of intervention.

Analysis and research of automobile interiors during the design process vary in terms of means, methods and instruments depending on the phase of development of the project, from the non-material idea to the functional prototype.

In the first, non-material phase, the design of the vehicle is presented purely as an idea, in that it is made up of briefing documents and graphic or virtual suggestions. In this phase there is very little possibility of getting the opinions of people who represent the actual user in order to obtain directions and ideas relative to the use of the product (which does not yet exist), while it is possible, on the other hand, to involve experts.

Following this, the idea can be represented making use of mock-ups (as a whole or in parts) which faithfully represent the dimensional aspects, but tell one very little, if anything, about finishes and functioning. Mock-ups allow one to carry out verifications of use on subjects, but they do not provide a complete experience of use at all (for example, changing gears, even though this cannot be done when stationary; evaluating the comfort of seats in the absence of the vibrations associated with the movement of a vehicle; the quality of feel of a steering wheel which is related to the surface finish and to the "grip" of the materials used; qualities of noise and force which can only be evaluated meaningfully on finished products fitted to the actual structure they will be used on). The sample of subjects interviewed cannot be made up ordinary people, but rather it is necessary to obtain the services of selected subjects who can mentally simulate real conditions of use, people, in other words, who can operate in the abstract. When it is possible to create fully or partially functional vehicles, analyses can be carried out with real subjects under real conditions of use. Experience has shown that even when functional prototypes are available it may not be easy to carry out tests using ordinary people, for reasons relating to the classified nature of a product not yet on the market or for reasons of safety relating to products that are not yet approvable.

The ergonomic quality of motor vehicles regards the quality of the relationship between the user and the product during use;

it is, therefore, the result of the interaction of the ergonomic characteristics of the product with specific users in specific environments and in specific contexts of use.

To define ergonomic qualities it is necessary to compare the reactions of users (either single users or homogenous groups of users) to the ergonomic characteristics of the product in order to define a scale of quality. Ergonomic characteristics are therefore aspects, of themselves neither positive nor negative, that have to be studied during analyses and the definition of design.

In relation to motor vehicle interiors, the parameters able to influence psychophysical well-being or pleasure of use are habitability, accessibility, comfort, pleasantness, usability and external visibility.

3.1. Habitability
This is the ability of the motor vehicle to accommodate the user.

3.1.1. Postural comfort
The driving arrangement refers to a posture conditioned by the constraints imposed by the act of driving (awareness of the road, awareness of the dashboard and displays, operating the steering wheel and other controls, operating the pedals).

Passenger arrangement (both front and back) refers to postures that are very different with respect to that of the driver, as they are only slightly constrained by the criteria of safety norms and regulations (the use of safety belts, normatives relating to the carriage of children).

3.1.2. Spaciousness
The spaciousness required by the upper body refers to the restrictions, both while moving and stationary, of the trunk and arms. The spaciousness required by the lower body refers to the restriction of the legs and their movement.

3.1.3. Perceived habitability
The perception of space is a determining factor in the general sensation of comfort and it is provided by the complex relation of the physical dimensions of the inside of the vehicle to the perception of that which is external to the vehicle (by means of the windows and windscreen), and to ease and liberty of movement inside the vehicle and the image that forms and colors create.

3.2. Accessibility
This is the absence of restrictions of movement in getting into and out of the vehicle. It is the capacity to allow these movements to be carried out with a minimum of postural change and the maximum possible naturalness.

For the upper body, this may be conditioned by the thickness of clothes, by the mobility they allow, and by the presence of objects being carried (bags, umbrellas, etc.), while for the lower body, the posterior, legs and feet, it can be conditioned by the clothes (skirts, trousers) and type of shoes being worn.

3.3. Comfort
This is the general state of well-being that derives from the reduction or absence of perceived disturbances. It is a passive and sensorial concept, synonymous with convenience.

3.3.1. Vibrational comfort

This regards the effects of the mechanical vibrations induced by the motor and those induced by the road surface, transmitted through the suspension system.

3.3.2. Acoustic comfort

This regards the effects of noise produced by the mechanical parts and of the noise of movement (air turbulence, road surface). It is influenced by the mechanical characteristics of the vehicle and by the degree of soundproofing of the vehicle interior.

3.3.3. Tactile comfort

This regards the quality of contact with the surfaces of the vehicle interior. It is conditioned by the intensity (touch only, operating, operating with force) and duration of contact.

3.3.4. Comfort of vision

This regards the quality of forms and colors.

3.3.5. Thermal comfort

This regards the quality of the microclimatic conditions of the interior and the thermal sensations of the contact surfaces. These can derive from the influence of the external climate on a stationary vehicle (a hot or cold vehicle), and from the microclimate owing to the systems of air treatment in the vehicle.

3.3.6. Comfort relating to smell

This regards the effects of odors given off by materials and surface treatments.

3.4. Pleasantness

The sensorial pleasantness of products regards qualities perceived through the human senses (touch, sight, hearing, smell, kinesthesia) that cannot be absolutely measured, qualities that are valid for all and for always. It is an active and cognitive aspect that responds to expectations coming from a mental model, recalled by the product and re-projected onto the product itself.

3.4.1. Acoustic pleasantness

This regards the quality of perception of the noise produced by the vehicle while in movement, during the operation of its parts (e.g. the opening and closing of doors) and of the artificial noises of signaling and warning.

3.4.2. Tactile pleasantness

This regards the pleasure that can derive from bodily contact with surfaces. It is influenced by the quality of the surface and the intensity and duration of the contact.

3.4.3. Visible pleasantness

This regards the pleasure that can be derived from the sight of forms and colors that are perceived as harmonic or meaningful.

3.4.4. Thermal pleasantness

This regards sensations of warmth or cold that can derive from contact with the materials that go to make up the interior (heat isolators or conductors).

3.4.5. Pleasantness relating to smell

This regards the pleasure that can be derived from the presence of odors that are perceived as pleasant.

3.5. Usability

This is the effectiveness, efficiency and satisfaction with which specific users can achieve specific aims in particular environments (ISO 9241, which regards the design of equipment with video terminals for use in offices). It is the ability of the system to perform its task, when operated by the user in the operative environment. It is measured in terms of cognitive charge, sturdiness and accessibility of the system.

3.5.1. Visibility

The ability to distinguish a stimulus from a background. It regards the capacity to identify the subsystems (and in particular the controls) interior to the vehicle.

3.5.2. Legibility

This regards the visual quality of symbols, signs and indicators of the control system, and the graphic or descriptive indications of the operation of the controls.

3.5.3. Intelligibility

This regards the quality of the message transmitted by the control and its interpretation on the part of the user. It depends on the user's knowledge and experience, on their cultural characteristics, and on their individual ability to perceive the symbolic messages.

3.5.4. Ease of reach

This regards the ease of reach of the controls. It depends on the anthropometric characteristics of the subjects and postures assumed.

3.5.5. Ease of operation

This regards the characteristics of the controls as a whole, that allow the user to achieve the result expected in an effective, efficient, and satisfactory manner. It regards:

- the mode of operation (number of fingers employed, direction of movement, characteristics of handling and holding during operation and on letting go);
- prehensility (dimensions, shape, surface design, grip, softness of parts to be handled and gripped);
- perceived weight and effort of operation (the effort necessary for carrying out the movement from one position to another on the part of the user, expressed as maximum force required and the relation between force and movement);
- form (the semantic message transmitted by form to its operational characteristics); and
- the space that allows function in movement to take place.

3.5.6. Feedback from controls

This is the quality of the message given off by the completed operation. It regards:

- visible feedback (warning and indicating lights and symbols);
- acoustic feedback (significant mechanical noise, artificial noises and sounds); and
- immediate feedback (perception of the completed operation, variations in the force required for the operation, perception of the end of the operation.

3.6. External visibility

This is the external field of vision perceived from the positions assumed by the occupants of the vehicle.

3.6.1. Fields of visibility

The direct field of visibility is the quantity of the external environment visible from the seating positions through the windows and windscreen of the vehicle. The indirect field of visibility is the dimension of the external environment visible through mirrors or videocameras.

3.6.2. Quality of vision

- Transparency characteristics of the windows and windscreen.
- Characteristics of the removal of rainwater and mist.
- Reflections on the windscreen caused by the illumination of the instrumentation and incidental light on the dashboard.

4. OPERATIVE ASPECTS

Ever increasing competition demands a reduction in the time to market and a response to the needs of the user, and imposes stringent constrictions on the designer. Today, the buyer is much more pragmatic than in the past and so is prepared to pay when faced with an adequate equivalent, which he or she also identifies with convenience and ease of use. As a result of this, even style has to adapt to the presence of ergonomic obligations as well as normative and technological ones.

In the automobile sector ergonomics aims to introduce the concepts of habitability, accessibility, comfort, usability and pleasantness into vehicles in such a way that they respond to the needs and desires of the user. Ergonomic intervention must not, under any circumstances act as a substitute for stylistic creativity. It will provide, along with technology, marketing and production, important elements for a formulation based on the incoherent obligations and propositions that go to make up the project "input," that the designer, with his creativity, will have to transform into a functional, coherent and purposeful product.

To achieve these results ergonomic principles must be introduced during the vehicle design process by an ergonomics body that will also carry out verifications on the finished product. This body's objective is to provide the ergonomic input necessary for the stylistic definition of the design plans, so that any modifications that would have to be made after the definition and freezing of the style (both interior and exterior) are reduced to a minimum or eliminated. Ergonomics bodies in companies carry out both autonomous and applied activities.

4.1. Autonomous Activities

- Strategic activities — above all these are speculative and experimental activities for the definition of the methodologies, instruments and means required for achieving the expected results. They refer to the particular articulation in bodies and to organizational structure proposed for the development of the products typical of the motor vehicle sector and of the specific manufacturer.
- Normative activities and those of standardization — consist of the diffusion of the normatives relating to the relationship between man and machine, of the definition of specific ergonomic norms and standards, and of their application and verification.

- Preparation and up-dating of an ergonomics data base — consists of the definition of ergonomic parameters for comparing different motor vehicles, and of the up-dating of a data base which all designers have access to.

4.2. Applied Activities

These are the planned and regulated activities that are developed by ergonomics bodies in collaboration with platforms for the development of a specific model. The introduction of ergonomics into the standard design procedures avoids the episodical nature of intervention and has substituted the traditional activity of "consultation on request" on behalf of those in charge of the models or the platforms. The activities of the ergonomics body within a work group that is developing a model foresee:

- the definition of the ergonomic objectives for the specific model. This is based on analysis of the competition, on studies made with panels and on specific choices for the model;
- the definition of the packaging and the graphic representation of the ergonomic aspects predicted for the interior and exterior styling of the specific model. Values are represented as ranges of accessibility in such a way as to direct, yet not constrict, the activities of the designers and stylists; and
- collaboration on the introduction and verification of ergonomic parameters in the project development of the model (continuous feedback).

5. MEANS AND INSTRUMENTS

The introduction of ergonomic principles into projection and design requires the definition of standards, ranges of values and methods of verification that must be obtained from tests with subjects on physical objects or with virtual tests.

5.1. Tests with Subjects

- *Movement analysis* — carried out by means of the video recording of predefined or free tasks with products, prototypes or mock-ups. The analysis of the filmed material based on accepted standards and the elaboration of the data collected, allow one to formulate conclusions. The complexity of the tests tends to limit the number of subjects to between 20 and 30.
- Survey of the pressure of contact between the human body and the seats — carried out in a static environment or with tests on the road using sensors, making use of subjects of various sizes.
- Physiological tests — carried out in order to evaluate the physiological parameters of subjects undergoing road tests.
- Usability tests — carried out on the finished product or on prototypes using subjects with differing psychophysical characteristics depending on the nature of the product being studied. Evaluations are carried out with one or more methods of measurement: error analysis, time and operational analysis or analysis of the strategies employed by subjects in order to achieve their objective.
- Pleasantness tests — carried out on finished products, on prototypes or on mock-ups, using subjects to evaluate the sensorial qualities perceived by the user through the senses (touch, sight, hearing, smell, kinesthesia). Pleasantness is subjective and heavily linked to variable aspects of the user relating to their ethnic or social background, to different

individual perceptive characteristics, and to the evolution of society's general or individual "tastes."

5.2. Virtual Tests

- Modeling — use of three-dimensional mannequins/dummies to evaluate certain ergonomic parameters of the vehicle. They are not intelligent and therefore do not require preliminary calibration with subjects. They are applicable to the analysis of posture, visibility and usability. They can be effectively applied when graphic representations of the project are available but solid or operational representations are not.
- Simulation — use of computer packages to simulate the movements of entry to and exit from the vehicle and the movements carried out when inside the vehicle. As they are simulations of human behavior, they require the definition, through testing, of the strategies and mathematical models of behavior (biomechanical model).

6. CONCLUSIONS

The attitude that left to the intuition and sensitivities of the designer, the response of a product to the needs and characteristics of the user, no longer seems to satisfy the complexities of user demand, technological specialization and the complex organization of production. Common sense, which may still preside over the design of simple products, must give way to methods and means, and to suitable and replicable concepts, which yield results that can be evaluated.

More and more, ergonomics is becoming a body of specific knowledge that proposes sophisticated methods of investigation and study. Confirmation of this is provided by the fact that, in ergonomic centers in the motor industry the presence of specialist mathematicians and bioengineers, as well as cognitive psychologists, has become indispensable. It should be added that the formation of "on the job training" and "apprenticeships" providing contact between apprentices and those with greater expertise, is rendered difficult by an elevated staff turnover, and so formation and continuity of company discipline relies more on the memory of the body itself than on that of the staff. Another observation relates to the professionalism of designers, which is becoming ever more accentuated in the direction of specialization in formal aspects, an inclination of the evolution of academic studies that tend to favor aesthetic and formal aspects over humanistic ones.

REFERENCES

BANDINI BUTI, L., 1998, *Ergonomia e Progetto* (Milan: Maggioli).

BANDINI BUTI, L., 1993, Intorno all'automobile. *Ergonomia*, no. 1.

BONAPACE, L., BANDINI BUTI, L. and TARZIA, A, 1997, Sensorial quality assessment: a method to incorporate perceived user sensation in product design. In *Proceedings of the IEA 13th Congress: Application in the Field of Automobiles*, Helsinki.

ISO 9241 — *Ergonomic Requirements for Office Work with Visual Display Terminals — Part 11.*

PEACOCK, B. and KARWOWSKI, W., 1993, *Automotive Ergonomics* (London: Taylor & Francis).

WISNER, A. *et al.*, 1965, Etude Biomécanique de l'homme soumis a des vibration de basses frequences. *Le Treavail Humain*, 28.

Design of Visual Displays for Teleoperation

S. H. Park† and J. C. Woldstad‡

†School of Industrial and Mechanical Engineering, Hannam University, 133 Ojeong-dong, Taeduk-gu, Taejon 306-791

‡Department of Industrial Engineering, Texas Tech University, Lubbock, TX 79409-3061, USA

1. INTRODUCTION

Human exposures to hostile work environments can be reduced when tele-operated systems are used to remotely control manipulators in these environments. Tele-operation systems were initially developed by the nuclear industry, but are currently applied in areas including surgery, construction, mining, warehousing, firefighting, undersea exploration, military operations, and space. The role of human operators in tele-operated systems varies with the level-of-automation of the system. Some systems require only supervisory control from the operator, while many others require direct manual manipulation through a controller. However, even for highly automated systems, human operators play an important role as a backup when the system fails.

Telepresence can be achieved through sensing appropriate information about the tele-operation task environment, and providing this information to the human operator at the remote site (Sheridan 1992). A great deal of human engineering information concerning the design of workstations for telerobotic systems is closely related to meeting telepresence requirements. In this regard, human depth perception (as a third dimension) based on the information provided by the two-dimensional (2D) surface of a video display terminal (VDT) very much determines the quality of a tele-operation system.

2. DEPTH PERCEPTION

A visual display has a basic limitation in that the surface of the display screen is two-dimensional. However, when designing a visual display for tele-operation tasks, it is necessary to represent depth or distance as a third dimension along the line of sight. Depth and/or distance perception are achieved through the combination of several depth cues. The term, cues, has been utilized to formalize the specification of stimulus conditions for space perception (Carr 1935).

In the study of perception, an object and its properties in the world are called distal stimuli. They only stimulate our nervous system by patterns of energy (e.g. light energy, sound pressure wave). We call the patterns of energy that reach and affect our sense organs proximal stimuli. Therefore, a depth cue is a pattern of proximal stimulation that contains information about the spatial location of distal objects. Depth cues can be classified into two types: monocular and binocular. Monocular depth cues require the activity of a single eye; binocular depth cues require the use of two eyes.

2.1. Monocular Depth Cues

2.1.1. Relative size

The same-sized objects produce smaller tracings when they are at greater distances: this is called the cue of relative size, or size perspective. By comparing the apparent size of a distant object with that of a similar, much closer object, the relative distance of the distant object can be approximated.

2.1.2. Interposition

A nearer (overlapping) object interrupts the outline of a farther (overlapped) object. This is an effective depth cue, but it can only indicate which object is in front, not the distance separating them.

2.1.3. Linear perspective

If the size of a distal object is fixed, the visual angle will be inversely proportional to the distance from the object: this is called linear perspective. A constant distance between points subtends a smaller and smaller angle at the eyes as the points withdraw from the eye. For example, telephone wires appear to approach each other (i.e. the retinal images of the lines converge) as the distance from the eyes increases although they are parallel. Therefore, converging lines are a cue that they are parallel and receding in depth.

2.1.4. Monocular movement parallax

When a subject's eye moves with respect to the environment, or vice versa, there exists a differential angular velocity between the line of sight to an object (fixated) and the line of sight to any other objects. For example, far objects move with the direction of movement, while near objects move against the direction of movement (Graham 1965). By observing the amount and relative direction that a given image moves on the retina, its distance can be approximated.

2.1.5. Familiar size

We can use objects known size to infer relative depth. For example, we know a man is taller than a boy. However, if they produce the same size of retinal image, then we deduce the man is located farther away than the boy. This is often a weak or ineffective cue (Hochberg 1978).

2.1.6. Light and shadow distribution

Shadows provide some information about the orientation of the objects and their three-dimensional (3D) shapes. Objects may appear to lie at different distances and have different dimensions as combinations of shadow and highlight change (Graham 1965). If objects have a light source from one direction, they will have shadows unique to their shape and orientation.

2.1.7. Gradient of texture-density

A gradient is the rate of some measured property changing over a continuous, extended stimulus. The surface of most objects is likely to be covered with a reasonably uniform texture or pattern. When looking straight ahead at a textured surface, the gradient of texture-density is zero; as the slant increases, the gradient increases. The gradient of texture-density can provide precise and relatively unambiguous information about the distances and sizes of surfaces and objects in the world (Hochberg 1978).

2.2. Binocular Depth Cues

2.2.1. Convergence

The eyes are capable of convergence, in which both eyes turn inward toward the medial plane. A large convergence corresponds to near objects and a slight convergence corresponds to far objects. The ocular muscles control the angle of convergence. The brain receives proprioceptive messages from the ocular muscles about the degree of convergence (Wickens 1992). By analyzing the information received, the brain can approximate the angle of convergence. In this way, convergence may serve as a depth cue. A large convergence may lead to the response "far-off", while a slight convergence may lead to the response "nearer."

2.2.2. Stereopsis

The retinal image of a distal object in the right eye is different from that of the same distal object in the left eye. The difference in the retinal image plays a great role in spatial discrimination. In the 3D world, the view each eye receives is somewhat different because the two eyes see the object from slightly different positions. Differences in these views give two possible depth cues: double image (each eye contributes different image of far object when viewing near, and vice versa) and binocular disparity (Hochberg 1978). The disparity is the difference between where a target falls on the right eye and the left eye.

The disparity can generate a powerful depth cue. This depth cue can be obtained by taking two photographs of a scene (a stereoscopic picture pair, stereogram), one from the position of each eye (65 mm apart average), and presenting each picture to its appropriate eye, and then viewing such photographs with special devices called stereoscopes. The stereoscope is composed of two conversing lenses and a supporting frame that simply separate right and left views.

3. VISUAL DISPLAYS

3.1. Principles of Display Design

3.1.1. Visual momentum

Visual momentum refers to the visual landmarks that film editors generate to reduce visual inconsistency among several scenes when editing a film. The concept of visual momentum can be applied to integrate information of one display into the other displays among multiple displays. For example, a visual interface for a telerobotic system can be composed of multiple 2D displays having different reference frames (e.g. a plan view and side views). In such a case, human operators must integrate information across multiple displays in order to control the robots arm. The concept of visual momentum can be applied by presenting a line originating from the face of the gripper of the robot. The line always aims the direction of the gripper, and therefore visual momentum is provided across the displays having different reference frames.

3.1.2. Object integrality and the principle of proximity compatibility

An object (or integrated) display is a display which integrates multi-dimensional information into one object. For example, the conventional attitude indicator in aviation represents the aircraft's pitch and roll information (2D) in one display. It should be noted that integrating two dimensions into an object will help performance if the information associated with those two dimensions is relevant, but decrease performance if the information is irrelevant.

The integrality of displayed was further expanded by Wickens and his colleagues, introducing the proximity compatibility principle — "To the extent that information sources must be integrated, there will be a benefit to presenting those dimensions in an integrated format (high mental proximity). In contrast, to the extent that information must be treated separately (low mental proximity), the benefits of object displays will be reduced" (Wickens 1992: 98). In summary, close proximity can be used to increase performance if parallel processing is guaranteed in the object display. However, it may cause problems and decrease performance if tasks require focused attention on a particular dimension in the display.

3.1.3. Working memory

New information is stored temporarily at a working memory until it is used or stored in a long-term memory. It is well known and accepted that human working memory is limited in size and time. The concept of chunk has been employed to define the working memory limit, where a chunk can be any set of information associated with long-term memory. Based on the concept of chunk, human working memory is limited by 7 ± 2 chunks of information (Miller 1956). When designing visual displays, it is desired to avoid exceeding the limit of 7 ± 2 chunks of information at the level of working memory.

3.1.4. Mental rotation

The time required to compare two visual images increases linearly as the angular disparity between the two images increases. For example, using multiple displays having different reference frames, human operators must mentally rotate a standard alignment in congruence with the alignments of other views. The mental rotation requirements of a task are more demanding if human operators must frequently scan different views.

3.2. Dimensionality

3.2.1. 2D displays

A 2D display is an orthogonal-view that provides spatial information about two dimensions. The only depth cue that can be provided using a 2D display is interposition, which only indicates which object is in front of the others, not the distance separating them. A multiple 2D display can be composed of a plan-view, a front-view, and/or side-views (i.e. all 2D orthogonal views). In the multiple 2D display, 3D spatial information is provided by the combination of at least two views; for example, the plan-view provides spatial information about two dimensions while the side-view or the front-view provides spatial information about a third dimension, such as altitude.

A 2D display does not provide any depth information in a third dimension. However, it provides unambiguous depth information in two dimensions and removes ambiguous pictorial depth cues. The 2D display, if provided with other 2D displays (i.e. forming the multiple 2D display), can provide unambiguous depth information in 3D. As compared with the perspective display, however, one of the main disadvantages of the multiple 2D display may be additional scanning and integrating effort among separate displays.

3.2.2. 3D displays

A 3D perspective display can be achieved by projecting an object onto the view plane (projection plane) and then mapping the view plan onto the display screen. The 3D perspective display provides more natural spatial information about 3D environments by creating the appearance of depth along an observer's line of sight, as compared with the 2D display. Most of the monocular depth cues can be achieved through the 3D perspective display. However, the monocular depth cues are inherently ambiguous. Based on the human stereoscopic vision capability of fusing two retinal images into one image, a stereoscopic display can be generated by adding a depth cue (called stereopsis) to the perspective display.

3.2.3. 2D versus 3D displays

It is hard to conclude that 3D displays are superior to 2D displays, or vice versa. A multiple 2D display can improve tele-operator performance as compared with the perspective display if the task requires frequent use of focused attention on a particular dimension. In general, the choice of display format is dependent on the specific task being performed (e.g. gross positioning tasks requiring global attention or dexterous tasks requiring focused attention).

3.3. Use of Visual Aids

Using computer generated visual enhancement cues can enhance the communicative purpose of a visual display. Figure 1 shows several visual enhancement cues that extend from the point of motion (gripper of the robot) and allow the user to orient the manipulator in space. In Figure 1b a solid reference line is placed orthogonal to the face of the gripper of the robot allowing the user to orient and aim the gripper and objects attached to it. The reference line is a useful cue that has been found to help operators orient and position the gripper when performing a telerobotic task. However, it does not provide appropriate depth information. A translucent reference cylinder or four solid lines (Figure 1c and d) are better enhancements. These cues have volumes in 3D space; therefore, they not only aim the orientation of the gripper but also provide depth information by overlapping other objects in the work environment.

4. CONTROL FACTORS

4.1. Display–Control Correspondence

One-to-one correspondence between display and control devices, ensuring that a control movement and a resulting change in the display are in the same relative location, should be provided as far as possible (Sheridan 1992). In general, lateral displacements < 15° can be adapted to by humans almost completely. However, the ability to adapt to mismatches is decreased for angles > 15°. Any type of inversed feedback is most disruptive.

4.2. Force Feedback

Tele-operation performance can be improved when force feedback is provided. The human body's joint, muscle, and tendon receptors sense the net reaction force and torque acting on the hand (i.e. resolved force sensing as defined by Sheridan 1992). Force feedback to human operators can be achieved by measuring the force and torque acting on the tele-operator, and then driving motors on the controller which provide the hand of the human operators with the same force and torque. However, there is little information about the effectiveness of this method. As an alternative, display of force feedback to the operator can also be
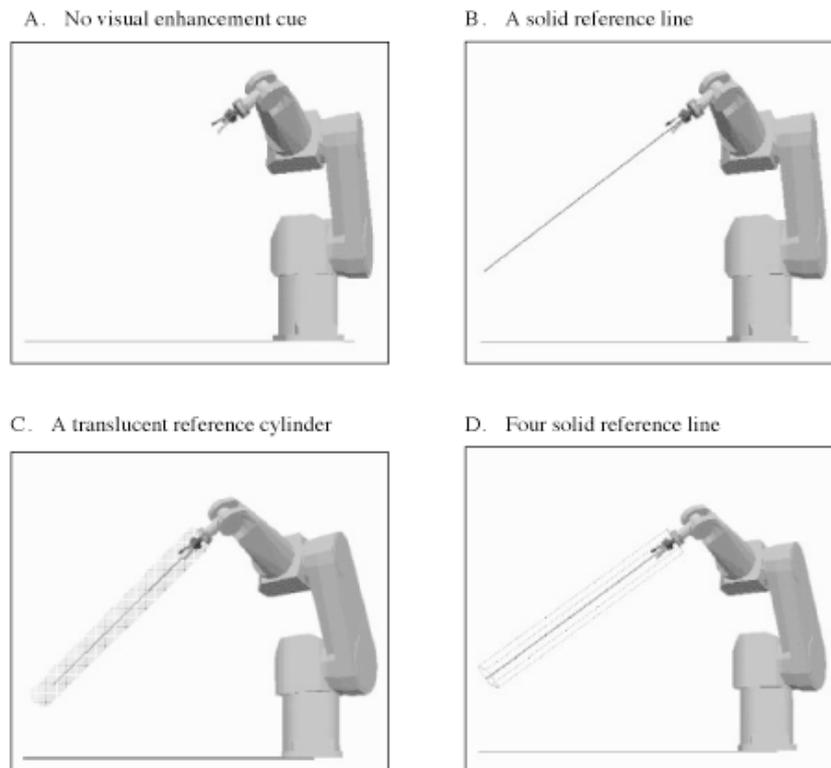
Figure 1. Level of visual enhancement cues for a telerobotic task. (A) No visual enhancement cue; (B) a solid reference line; (C) a translucent reference cylinder; (4) four solid reference line.

achieved through the use of computer graphic force-torque display.

4.3. Teletouch

Teletouch refers to the remote sensing of differential forces acting on the skin in time and space. Teletouch has been found to support depth perception, provide detailed information about surfaces of objects contacted by grippers, and provide information about surfaces visually occluded by other objects. A major problem associated with teletouch is the difficulty associated with displaying artificially sensed pressure patterns to the skin on the hand, when the hand is at the same time operating the control stick (Sheridan 1992). One useful way to present teletouch information to the operator is the use of visual display method (i.e. visually representing tactile data).

5. RECOMMENDATIONS

- Select display dimensionality based on the given task being performed. Consider a multiple 2D display format if the task requires frequent use of focused attention.
- For 3D perspective displays, provide visual enhancement cues to aid depth perception.
- As a visual enhancement cue, a single line is not sufficient enough to aid depth perception. Design a visual enhancement cue which itself has a volume in 3D space.
- Consider task difficulty when selecting the level of visual enhancement.
- For 3D perspective displays, ensure that the control movement and the resulting change in the display are in the same relative location. Maintain lateral displacements to < 15°.
- Provide the hand of the human operator with the same force of the manipulator (force feedback). Consider a display of force feedback as an alternative.

REFERENCES

CARR, H. A., 1935, *An Introduction to Space Perception* (New York: Longmans, Green).

GRAHAM, C.H., 1965, Visual space perception. In C. H. Graham (ed.) *Vision and Visual Perception* (New York: Wiley), 504–47.

HOCHBERG, J.E., 1978, *Perception* (2nd ed.) (Englewood Cliffs: Prentice-Hall).

MILLER, G.A., 1956, The magical number seven, plus or minus two: some limitations on our capacity for processing information. *Psychological Review*, 63, 81–97.

SHERIDAN, T.B., 1992, *Telerobotics, Automation, and Human Supervisory Control* (Cambridge, MA: MIT Press).

WICKENS, C.D., 1992, *Engineering Psychology and Human Performance* (2nd ed.) (New York: HarperCollins).

Ergonomic Design of Factory Buildings in Tropical Countries

R. Nath Sen

Ergonomics Centre, Multimedia University, 63100 SelangorMalaysia

1. INTRODUCTION

Those who build factories and office buildings seldom consider the use of the principles of ergonomics (the science, technology and art of people at work) with special reference to the effects that rooms, spaces and buildings have on people and their performance. A factory building may be suitable for today's needs but would almost certainly be inadequate for the needs of the future.

Many of the factories in the tropical countries are, unfortunately, mere copies of the design of the factories built earlier in the cold climate and hence not appropriate to what is required for such an environment. Many of the new types of factory buildings, therefore, are worse than those found with traditional factory buildings in hot climate.

The problems of the working environment, especially in hot tropical climate in the developing countries are even more numerous and of higher magnitude than those in cold climate. Unless correct measures are taken at the very beginning of the design of the working environment, so that the workplaces are physically and mentally satisfying as well as efficient and economic, it may be too late to do much at later stages or it may be too costly to effect modifications, if at all possible.

The ergonomist and architect should design the factory buildings as to bring out the best of the natural possibilities. The task of the environmental control is to ensure the best possible thermal conditions by relying on structural (passive) controls, which may obviate the need for any mechanical or electromechanical (active) controls; but even if these controls have to be used, their task will thereby be reduced to the minimum.

2. BASIC ELEMENTS IN DESIGN PROBLEM

Four basic determining elements in every design problem are function, form, fabricating material and finance. An ergonomist must bring a harmony between these elements in the design of factories for hot tropical climate.

The important factors in the proper design of the factories are :

(1) The site relation to human habitation, landscape, vegetation, altitude, and others, whether near the periphery of a town or in rural situations or in a valley, to protect from flood, earthquake, storms, insects, termites, to avoid inversion temperature (figure 1), pollution, etc. to have good ventilation, low humidity, etc.

(2) The orientation of the building, including roofs, walls, doors and stairs, windows, etc., in relation to the wind direction, and the angle of solar radiation, to have the best natural ventilation and minimal thermal heating of the buildings, in relation to human comfort, activity and efficiency.

(3) The insulation and thermal capacities of the building materials.
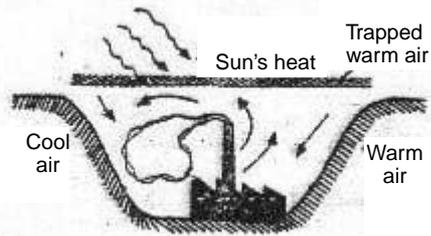
Figure 1. Possible effects of temperature inversion on a factory in a valley.

(4) The use of sound material and the form of the construction.

(5) Correct design of workspaces, windows, doors, stairs, corridors, etc., based on the static and dynamic body measurements and movements of the workers.

It is so obvious that the types of factory buildings and building materials for cold climates cannot solve the problems of the factories in countries where heat is the dominant problem, where, due to economic reasons, the consequences of poor design cannot be compensated by very costly electromechanical air conditioning, and where the workers differ in body form, thermal responses, etc.

3. DESIGN IN RELATION TO METEOROLOGICAL CONDITIONS, SITE, LOCATION AND LAYOUT OF SURROUNDINGS

It is very important to the location and design of factory buildings to consider the meteorological data over the years concerning rain, temperature, sunshine, humidity, speed and direction of wind, smog, etc. If the yearly data are not recorded and analyzed, it is essential to have at least data for 12 months for the selection of sites and design of factory buildings. Analyses of many of the elements of microclimate have a deciding role on the sites and location of the factory buildings, particularly in tropical climates.

The air temperature in a city can be about 8ºC higher than in the surrounding countryside and a difference of even 11ºC has been observed.

Local features, including the different types of buildings can substantially modify the air movement and the air temperature of working environment. In olden days, heavyweight stone or brick building with high thermal capacity and with high ceilings were being used in tropical climates to provide comfortable conditions during the summer months as well as in the winter. Nowadays, because of the cost, very lightweight buildings with low thermal capacity and with thin concrete slabs are being used, which would heat up quickly during the day and cool down quickly during the night. When large glass windows are provided to increase the natural daylight, it facilitates entry of solar radiation also, which warms the surfaces inside the room, which in turn re-radiate, and this results in the uncomfortable conditions experienced inside the building.

It is important to know the frequency, like duration and nature of some rare events such as dust storms, thunder storms, earthquakes, tornadoes, hurricanes, floods, etc., since the designer must classify these rare events into those which affect human comfort and those which may endanger the safety of the factory, buildings and the lives of the workers.

3.1 Wind Speed

Wind speed is very important for comfort in the humid climate of tropical countries. There is always a reduction in wind speeds near to the ground- down to 40 percent over rough terrain and 30 percent in urban centres. Under certain conditions with fewer slopes, funnelling effect, sharp ridges and solid obstructions, wind speed may increase due to suction, turbulence and vortexes. It is possible to locate sites, which would experience a considerable degree of natural wind. The wind speed on urban sites is reduced to less than half of that in the adjoining open country. The funnelling effect along the closely built-up street or through gaps between tall building blocks can be more than double the speed. At the leeward corners of obstructions, strong turbulence and eddies can also be set up.

Wind speed can be reduced by 50 per cent by a long horizontal barrier at a distance of 10 times the height and by 25 per cent at a distance of about 20 times the height. Trees reduce dust movement, give a "green" outlook from windows and provide desirable shades from solar radiation. The advantages and disadvantages of trees, hedges, shrubs, etc. should be carefully considered.

All the internal roads leading to the different buildings of a factory should have shade from the trees planted on the sides of the pedestrian pavement.

3.2. Humidity

The higher the temperature of the air, the more water vapour it can hold. Due to the lowest layer of air being heated by the ground surface during the day, its relative humidity (RH) is rapidly decreased and, as a result, the rate of evaporation is increased when water is available to be evaporated as with an open surface of water, with rich vegetation or with higher air movement. The situation is reversed during the night. On a clear night, especially with still air, the RH increases as the lowest layer of air cools.

3.3. Absorption of Solar Radiation, Materials and Form of Construction

The ground loses much heat by radiation, particularly on clear nights, and soon after sunset, its temperature falls below that of the ambient air. The direction of the heat flow is reversed from the air to the ground. The lowest layer of air becomes cooler.

The greatest source of heat gain can be the solar radiation entering through the windows. This could, in fact, increase the indoor temperature far above the outdoor air temperature. Overheating is a problem in all tropical climates. For the reduction of solar heat gain through windows, four variables are within the control of the designer :

1. Orientation and size of windows;
2. External shading devices;
3. Internal blinds, curtains, etc.;
4. Special glass.

3.4. Design of Shading Devices

In hot climates, it is very important to shade the outside walls of a building exposed to high levels of sunlight. This can be done by creating permanent screens or louver blades, a reinforced canopy or externally applied venetian blinds, or planting tall trees with thick leaves or shrubs. These are very effective when they shade the east and west walls of the building, which are exposed to the morning and evening low-level sunlight.

Shade is required not only against direct solar radiation but also against diffused radiation from the sky which, in tropical regions, may reach very high intensities (0.75 Kcal / cm^2 / day on a horizontal surface).

When horizontal adjustable louvers are used, they should be constructed so as to enable their opening at an angle of approximately 120°. Vertical shading devices consist of lower blades or projecting fins in a vertical position. Egg-crate shading devices are combinations of horizontal and vertical elements. The many types of grille-blocks and decorative screens may fall into this category. The construction of shading masks for moderately complex shapes is effective for any orientation depending on dimensions.

3.5. Design of Roofs and Walls

If a heavy-weight roof with an external layer of efficient insulating material, itself protected by a waterproof light coloured (whitewash) covering, is used, heat flow during the day from external layers is restricted by the insulation and reflection. Only a small portion of the potential heat is absorbed in the elements.

High heat capacity concrete walls externally insulated by rockwool or expanded plastic and covered by waterproofing materials are suitable for the purpose. All external surfaces should be as near to white as possible. The high thermal capacity of the concrete layer reduces the effect on internal temperatures of any heat, which thus penetrates.

The whole roof may be externally covered by polythene sheeting at a distance of 10-20 cm above the roof surface. Polythene (polyethylene) is transparent to restrict the radiative cooling of the roof at night. The disadvantage of the method is the deterioration of the polythene sheets due to exposure to the sun.

The alternative of double roofing at much less cost, especially in the factories in rural areas in hot climates, is to maintain vegetation on selected portions of the slanting roof.

3.6. Orientation and Design of Windows

In the equatorial location, the main windows should face north or south to avoid solar heat gain. At the higher latitude, though an orientation away from the Equator would receive the least sunshine, it may be desirable to have some solar heat gain in the winter when the sun is low, and so an orientation towards the Equator may be used where the workplace does not generate much heat. In both locations the minor openings at unimportant workplaces should be placed on the east and west sides. Solar heat gain in the west side can be particularly troublesome as its maximum intensity coincides with the hottest part of the day. If the wind to be captured or a pleasant view is to be utilised, etc., the opening of windows may at times override the solar consideration.

The air movement could be grossly influenced by the way the window blinds or sashes open. If the hinges on the windows are fitted properly depending on the direction of prevalence of the wind, the window blinds or sashes would act as deflectors to direct the wind through the windows, whereas if the hinges are fitted in the wrong way, the wind would be directed away from the room. In many of the factories in the tropical climates, just changing the hinges from one to the other side of the window frame may improve the climatic conditions greatly (fig. 2). This point has been overlooked in many factories in hot climates.
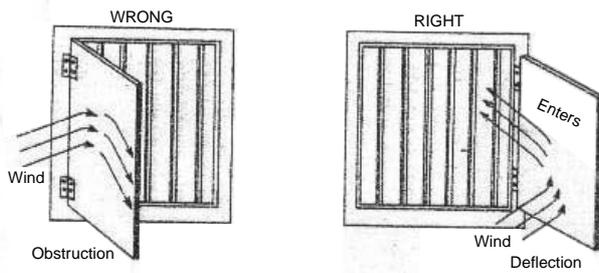
Figure 2. Correct method of fixing window blinds or sashes so as to facilitate entry of wind from the prevailing direction of wind.

The most effective height of the windows from the human comfort aspect is about 0.5 to 1.5 m above the floor. It is preferable to use horizontally pivoted windows with upper hinges which, when open, would direct the airflow downwards.

To raise the building on pillars is advantageous in a warm, wet climate because it enables better ventilation by locating the windows above the zone of maximum damping of wind by the surrounding vegetation, and also by enabling the cooling of the floor from below, which is particularly beneficial at night. In addition, the building is better protected from floods and from termites.

One of the basic needs of the human being is change and variation, a fact that has been ignored by early research workers. This point is particularly noticeable in mechanically controlled environments, such as in air conditioned buildings, where the environmental conditions can be and often are kept constant within very fine limits. What the designer should aim at is a range of comfort conditions within which considerable variations are permitted. It is quite interesting to observe that people enjoy natural, cool and fluctuating fresh breezes even when these stop for a few seconds at random, while people complain of the monotonous air movement at the same temperature and constant speed in an artificial climate. If these observations and causes are proved beyond doubt, in future the artificial climate may have to incorporate the random variation of air speed and air temperature within prescribed limits to provide the most comfortable conditions for workers.

The ordinary ventilation in the factories and workshops in hot climates should be at least 5.0 ft^3 (1.4 m^3) per person per minute. The air speed at the head level should be at least 100 cm / s (200 ft / min).

3.7. Design in Relation to Lighting, Colour and Noise

It is surprising that even today simple issues of heating, lighting and ventilation are too often inadequately considered and acoustic problems are not properly dealt with.

The effects of colour on people at work are to be considered for the scientific use of colours in the rooms and shop floors. In hot areas the "cool" blue or green colours, as against "warm" red should be used to give subjective sensations of coolness or impressions of reduced temperature.

The ceiling of a factory building plays an important role, particularly in reducing reverberant noise. Though people,

furniture, wall linings, soft flooring, etc., all act to absorb noise to some extent, a considerable proportion of any noise travels upwards to the ceiling. There are a wide variety of acoustically absorbent materials suitable for use in ceiling. The so-called "false" ceiling with sound absorbing material not only reduces noise but also helps to insulate and thus minimise transfer of heat from a hot roof to the shop floor.

3.8. Design in Relation to Safety, Health, Pollution and Welfare

Normally, factories should be so designed that the use of personal protective equipment against heat, dust, smoke, fumes, noise, accidental injuries, etc., is eliminated or at least minimal. If the hazards cannot be reduced at the source, then personal protective equipment has to be used, but one has to foresee that it might be impossible for the workers to wear protective equipment in hot conditions.

The normal psycho-physiological conditions of activity and rest with recovery from stresses are impeded by unfavourable climatic conditions and the resulting stress on body and mind causes discomfort, loss of efficiency and may eventually lead to a breakdown of health or even cause accidents. It is a challenge for the designer of the factory building to strive towards the optimum of total comfort, and health i.e. complete physical and mental well being.

A well designed working environment includes not only suitable physical conditions of ample ventilation, heat dissipation, illumination and other comfort standards, but also the tangible and intangible amenities that can transform discontent and boredom into interest and a sense of participation by the workers, as for example, the availability and use of shower facilities in hot climates, which are very much favoured.

For minimising accidents, the design criteria should take into consideration the major causes and frequency of different types of accidents from the record of similar industries in hot climates.

In hot, dry conditions, the chances of fire are much greater than in cold conditions, and hence greater precaution should be taken and better facilities provided for fire exists, structural fire barriers, safe internal and external access, especially in multistoried buildings. Industry has to count the cost of necessary precaution and control measures and the cost will in the long run have to be related to the total benefits to be expected from the process. The costs should be regarded as an essential parts of the process and not simply as an added burden to be carried on the back of the manufacturer.

4. FACTORIES

4.1. Air Pollution

The recent planning policy of most of the countries has been to encourage siting of industrial zones on the outskirts of new towns, or siting State or Government-sponsored industrial estates outside townships, mainly to reduce the effects of pollution on the people living in the towns. With the ever-increasing number of factories or industries, the threat of air, water and land pollution greatly increases.

To deal with the aspects of pollution directly related to the design and landscape and buildings for industries, one has to consider the effects of waste materials or surplus energy generated by various forms of human activity in industry threatening damage

to man's health, possessions, food supply, recreation and also to plants, animals and wildlife. In addition, there is pollution due to noise and other environmental nuisances generated from the factory buildings.

Air pollution arises from smoke, fumes and other gaseous emissions, dust and grit from the factory processes directly discharged into the atmosphere. Obviously, the design of the factory building, including the ventilation system, chimneys, insulation of furnaces in the shop-floor etc., must be done properly to cope with the maximum interference and pollution of the air by toxic and other substances and by heat. These ambient thermal pollution problems are enhanced by tropical climates. It is, therefore, very important that industrial plants and buildings be built to help in the effective control of pollution and its reduction.

The effect of a temperature inversion in trapping smoke and fumes near the ground is obvious. It is important to find out the height at which temperature inversion occurs. Very tall chimney stacks with correct height are able to penetrate the inversion layer to discharge their fumes in the rising air and thus avoid pollution. (Figure 2)

4.2. Factories and Water Pollution

Water pollution may occur from the falling rain passing through smoky and polluted air, by ways of ditches, ponds, streams, rivers, estuaries, etc. to the direct polluting of the sea itself. Many of the wastes of the factory are directly discharged into rivers, except in countries where there is strict enforcement of control measures to avoid pollution. Pollution of water may not necessarily render it toxic but often the effect is to deprive the water of oxygen and thus reduce its capacity to support life. The effects of thermal pollution are similar. The design of the factory buildings to avoid such pollution is an important necessity.

4.3. Factories and Land Pollution

Polluting of land arises from the solid wastes. Industrial chemicals destructive to insects and bacterial life may be leaked into the soil from the waste dumped upon it. The toxic and caustic effluents from chemical plants usually involve great risk. The disposal of this must be done properly and the design of the factory building helps this greatly.
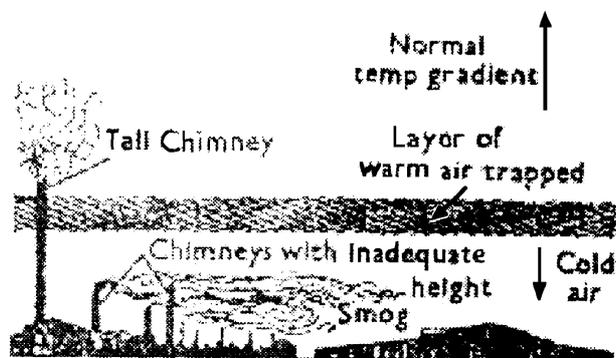

Figure 3. The effects of a temperature inversion in trapping fine dusts, smoke and fumes near to ground level. Very tall chimneys or stacks only are able to penetrate the inversion layer and discharge their fumes in the rising air.

4.4. Design in relation to Storage, Cleanliness and Maintenance

In tropical countries, the prevalence of hot and dry conditions leads to a lot of dusts. The factory should be so designed that the routine storage and maintenance of cleanliness are made easy. Hence, either the glass window should be at the lower level, or the glass windows near the top of a "saw-tooth roof" factory, for example, should have at least a low-cost bay with safety guardrails for safe, easy and regular cleaning. An ergonomically designed broom brush / wiper fitted on rollers so that the windows can be cleaned from the outside by pulling from one end to the other is also useful.

Easy accessibility for regular and proper maintenance of machinery must be considered in the design of the factory shop floor.

5. DESIGN OF INDUSTRIAL ESTATES

It is essential to have ergonomic considerations in designing industrial estates to reduce the cluster of small buildings and to improve the layout by various methods. A logical and scientific flow diagram for intake of raw materials and for output of finished products should be worked out for all the factories in the group. Efficient road, rail and conveyor systems should be made so that common points for packaging, loading and unloading, and common facilities for maintenance, security, safety, medical clinics, canteen, recreation, sports and other organisations, and common services for electricity, fuel, gas, water, steam, compressed air, refrigeration medium, telephone, modem, inter and intranets and a ring circuit of refuse and waste disposal could be ergonomically viable and useful. According to the suitability, any one of the different types of layout or plans such as linear or radial or ring types (figure 4) may be used.

5.1. Control of Factory Design

Unless the planning, location and design of the factory buildings are properly controlled from the very beginning, it will be extremely difficult to avoid grave situations in the years to come. In many countries, national laws, acts and local rules, regulations and restrictions determine location, construction and material usage in factory buildings. These acts, rules, regulations and restrictions should also be based on the principles of ergonomics and on appropriate guidelines so that the effective control can be established to humanise the environment and make for the proper development of the area and the progress of society as a whole.

An application to build giving an outline of the proposed factory should be made to the local planning authority. Permission to build should only be given with the condition that a detailed plan be submitted within a stated period. The local authority, through its appointed offices, should have the right of inspecting the work to ensure that the building is constructed according to the approved plan.

The factory inspectorate, like a "watchdog", should make good use of its loud bark and its nose for trouble, but should reserve the use of its sharp teeth for those rare occasions when they might be required to ensure the effective control of properly designed factories suitable for the climate.
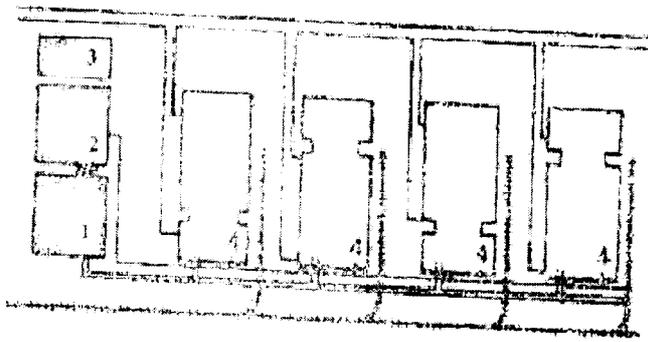
LINEAR PLAN

1. Energy and Heat service
2. Refuse and Waste disposal
3. Welfare services
4. Factory space

—— Service links
=== Road accesses
+++ Rail access
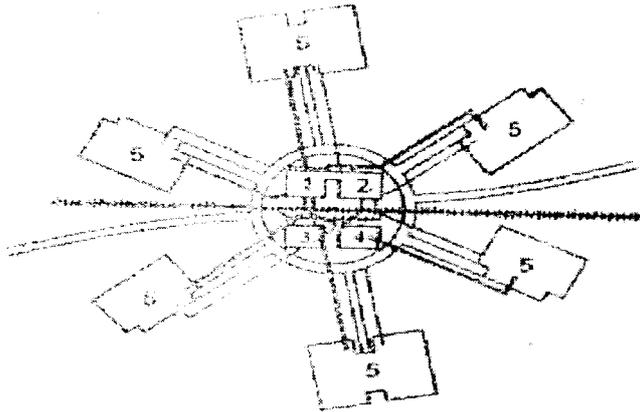
RADIAL PLAN

1. Energy and Heat
2. Refuse and Waste disposal
3. Rail goods storage and dispatch.
4. Welfare services
5. Factory space

—— Service links
=== Road accesses
+++ Rail access
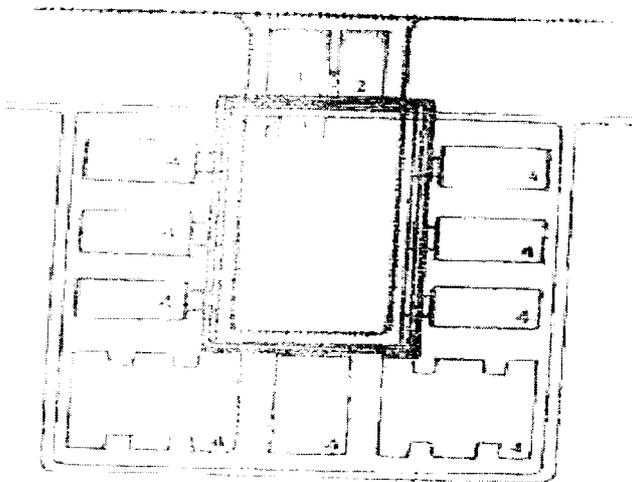
RING PLAN

1. Energy and Heat service
2. Refuse and Waste disposal
3. Welfare services
4. Factory space

—— Service links
=== Road accesses
+++ Rail access

Figure 4. Different types of layouts for factory buildings in industrial estates.

REFERENCE

SEN. R. N., 1982, *Certain Ergonomic Principles in the Design of Factories in Hot Climates, Proceedings of the International Symposium on Occupational Safety, Health and Working Conditions, Specifications in Relation to Transfer of Technology to the Developing Countries held from 23-27 November, 1981, International Labour Office, Geneva, p. 123-147.*

Ergonomic Product Design

L. Bandini Buti

Società di Ergonomia Applicata, Milan, Italy

1. INTRODUCTION

Ergonomic design has as an objective the realization of quality products adapted to their user. Manufacturers tend to consider that a product is high quality if it respects a certain set of technical specifications; there seems to be little distinction between purely commercial needs and those of the end-user. Ergonomics counters this way of thinking by bringing into consideration the concept of ergonomic quality, which is concerned with the reaction or response of a product to the functional needs of user in everyday life and to his or her psycho-physiological characteristics.

Ergonomic design is based upon the active definition of a product by means of analysis of use and of its characteristics and its other needs.

In recent years the debate over industrial design and ergonomics has revealed a tendency to attribute a growing importance to the integration of ergonomics with design and innovation. At the Financial Times Conference "Product Strategies for the '90s", London, 15–16 October, 1990, T. H. Thomsen, for example, argues that, among other things, good design renders a product useful, self-explanatory and in tune with its environment. He also cites other important design factors which emphasize the psychological aspects of the product/user relationship; the need for study from the ergonomic viewpoint, giving a feeling of uniqueness, easy identification on the part of the consumer, generating trust in the product itself, creating an ideal user–machine interface, taking psychological factors into consideration. Thomsen also considers certain practical qualities (properties of handling, ease of cleaning, noise produced, clarity of the instruction manual) as being of importance.

Umberto Eco illustrates another aspect linked to a correct usability of products when he writes "the product must also demonstrate its purpose and the way in which it is to be used … this is a fundamental aspect of design which is not always given due consideration …". He also states that "These phenomena must also form part of the panorama of Italian design, otherwise one understands neither that which constitutes Italy nor that which constitutes design" (Sartogo 1982). The need for a product to communicate its use is clearly indicated by the author in this extract

> the product must also demonstrate its purpose and the way in which it is to be used. Thus, we see that there is a communicative aspect to a product, which is an important part of its design. Scissors are a perfect example of this: their form clearly indicates where the fingers should be inserted to operate them, even to someone who has never seen them before. Scissors are rightly considered a masterpiece of design: not only do they cut, but also they tell one how they are to be handled and used.

> This fundamental aspect of design is not always taken into consideration. Indeed, it should be recognized that quite often "unconscious" or "uncivilized" design from other, more primitive times, was actually much more masterly than that

produced by well-known names in the field of design in our own time. This stems, perhaps, from the fact that often (for "aesthetic" reasons) new forms are invented which say little or nothing to those who must use the product. … There is, here, a certain element of paradox in that, in their attempts to realize functional products, designers have been apt to design products which communicate little more than some design philosophy. In other words, the design doesn't say "I may be used in this or that way" as much as "I am a perfect example of good design"….

Ergonomic design must be considered as a way of confronting the themes of planning, by placing man at the center of the process. It is developed using various types of research, planning projects, and checks which all interact to form a unified process geared towards the definition of products and systems adapted for human use. Form is an essential part of the relationship between man and the object, and is thus a field of ergonomic analysis and intervention.

2. ERGONOMICS, DESIGN AND INNOVATION

The field of ergonomics has been involved with the theme of the integration of ergonomics and design, providing specific contributions which propose a couple of peculiar concepts in this area, namely those of *participation* and *globality*.

In its relationship with industrial design and with innovation, the concept of *participation* means paying attention to the needs of the user, attention which must be given through all the various phases of a project. It is particularly important to make an analysis of the user's possible behavior patterns in relation to the product, whether these be proper or improper, so that information may be gathered on any possible difficulties they may encounter. In this way one is able to work from a number of interesting "starting points" which stimulate the introduction into projects both of technical innovations and design innovations capable of resolving such problems.

The concept of *globality*, on the other hand, signifies the need to single out the minimum level of material and logical dimensions to be taken into consideration when working on project research, below which the control of the effects which the intervention shall bring about in the entire system becomes impossible. In the context of such a hypothesis it is necessary to investigate the complex technological and use relationships of the system (made up of the product and its immediate logical environment) before carrying out any analysis of the product being researched.

Ergonomic methodology is particularly effective in the carrying out of analyses of products and complex systems where aspects linked to use are prevalent (e.g. an ashtray or a tray are minimally complex in terms of use but have maximum aesthetic impact).

2.1. Integrating Ergonomics into the Design Process

The principles outlined above may be considered as forming part of the "culture" of a project, where they make clear that the product and the systems have to be planned in such a way as to cater for the needs of the user. To apply these principles it is necessary to develop a specific operative methodology and to become more familiar with some basic concepts which allow for prediction, innovation and systemization in the approach taken.

In the planning of products and systems, applied ergonomics makes use of means which are based on the analysis of individual test subjects, an efficient way of investigating the individual response to a product in everyday use.

Obviously, in the early stages of a project, when work is still at a conceptual or non-material planning level, tests cannot be carried out with users. It is, however, both possible and important to gain information using subject samples with other, similar products, experimental products and prototypes, and virtual reality simulations, in order to understand, observe, predict, visualize, communicate, and only then implement (Wasserman and Moggridge (1990), in Learning for experience — An approach to design strategies for product success, from the Financial Times Conference "Product Strategies for the '90s").

The feedback, be it positive or negative, which a product receives from the user is a fundamental part of the process from which innovative ideas may drawn. As soon as models and prototypes are available (even if only partially functional) it is important to carry out tests using product samples with subjects in order to ascertain whether innovative ideas introduced solve problems identified at the beginning of the process effectively, or if indeed they introduce any new problems. Thus, it is also possible to quantify and evaluate any possible benefits obtained. From such evidence it is also possible to ascertain criteria which may be used successively in the design process of other products.

Following this tests may be carried out using fully functional prototypes which have been conceived giving due consideration to the entire sphere of planning aspects (ergonomics, design, technology, manufacture, ecology, costs), tests which, however, allow one to intervene only in the case of particularly grave circumstances, as, generally by this stage, significant investment has already been made and any changes are likely to be regarded as too much effort in terms of time and costs.

2.2. Ergonomics and Form

Those working in the field of ergonomics often feel that the form of a product, and any connotations thereof, can be disregarded. Rationalism maintained that "that which is right is beautiful" with the intention of confirming the priority of functional aspects over formal ones and that beauty is essential to the correctness of the solution.

This paradigm, which compares form and function may be translated as the conviction that the image of a product is a negligible entity in terms of its evaluation and, still more, that a product which is ergonomically correct is probably anti-aesthetic.

The world of science makes use of methodologies based upon the comparability of data and on the universality of assertions. It cannot in any way use these for the evaluation of formal aspects which it thus tends to ignore as being uninfluential. While it is true that beauty is beyond measurement and also that analyses of form and color always refer to the significance of these aspects in relation to the culture and history of those who perceive them, we cannot be simply permitted to maintain that the problem does not exist. It will be necessary to create rigorous investigative methods even if these are not "scientific" in the sense defined by the exact sciences.

2.3. Industrial Products

An industrial product may be considered ergonomic if, during the various phases of its life, conception, realization, use, disposal or recycling, it not only does not cause damage but also generates beneficial psycho-physical conditions in all those who enter into contact with it; thus, if a product is only or principally capable of not causing damage or illness, it may be said that it fulfils the objectives of hygiene and medicine in the work place, but not those of ergonomics.

The performance area relates principally to the dimensional aspects of the product, the safe-guarding of health and safety of all those who enter, for whatever reason, into contact with the product, and to respect for its surroundings during manufacture, use and disposal. A peculiar characteristic of the assembled topics in the performance area is that the parameters used to describe these are valid for the population as a whole regardless of the characteristics of the individual. In fact the need to respect the electric safety of a product, not to cause pollution during or after use or for dimensional correspondence to a large population, are not factors that depend on who uses the product or where and how it is used.

The area of well-being, on the other hand, is principally concerned with the usability and agreeability of the product. These aspects are variable depending on the psycho-socio-cultural characteristics of the user, and must, therefore, be evaluated in relation to the particular characteristics of the individual and in reference to a precise physical, social and cultural space. It is, in fact, very likely that that which may be pleasing to an Oriental may not be so to a European, just as those sounds which today are considered pleasing and harmonious to the ear, were considered, when first created, as unharmonious since they hadn't yet entered the collective consciousness.

3. ERGONOMICS PROJECT ANALYSIS

To render industrial products more ergonomic one needs to carry out an objective evaluation (using techniques for the definition of quantifiable data) and the collection and comparing of the subjective experiences of potential users (which provide information on trends).

3.1. Objective Data

Objective evaluation techniques refer to objective and quantifiable data (dimensional aspects, anthropometric data, noise levels, illumination, strain, etc.). These enable us to measure many parameters, to compare products in production, both with one another and with models and prototypes, and to formulate new/future projects from objective data. They permit the evaluation of quantifiable aspects and are indispensable for the definition of guidelines and regulations and for the verification of their compliance with such.
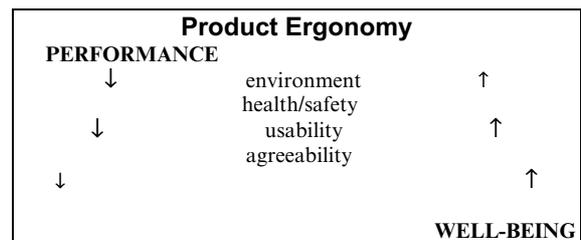
Figure 1. Characteristics that allow a product to be defined as ergonomic

TOTAL

PRODUCT
MODIFICABILITY

NIL

HIGH

USER
INVOLVEMENT

LOW

NON-MATERIAL PROJECTS

MODELS

PROTOTYPES

FINISHED PRODUCT
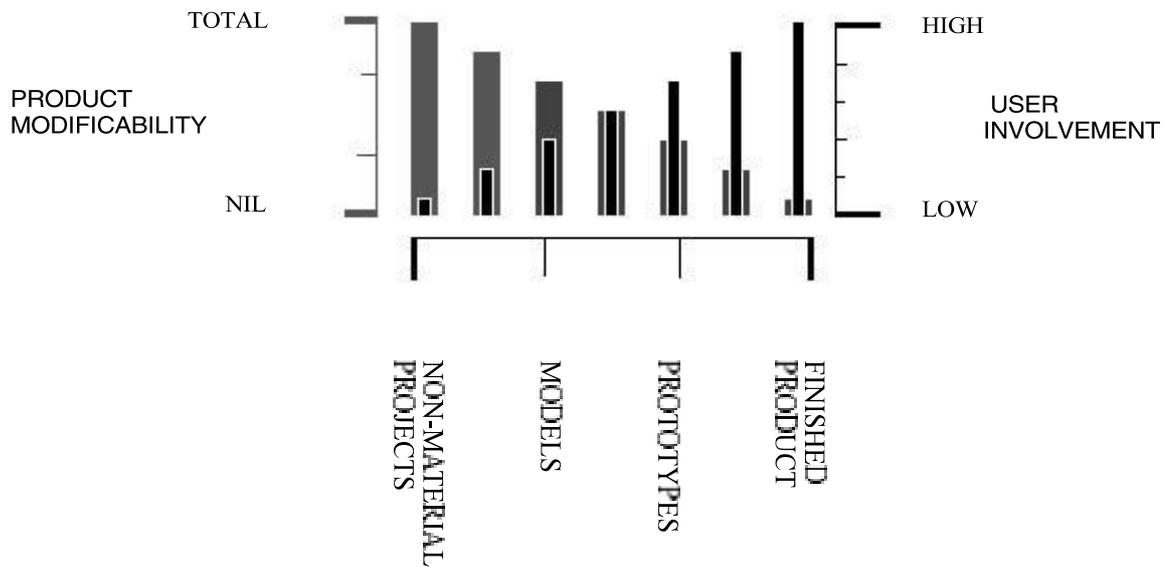
Figure 2. Ergonomic intervention and project development

3.2. Subjective Aspects

Objective data are of little use when analyzing individual behavior. The evaluation of factors such as the significance people give to messages and signals, the efficiency of tactile and visual perception of control units and displays, strategies of use, the effect of more than one stimulus being simultaneously present in the sensory receptors, is linked to a large number of aspects which vary greatly according to the user (aspects such as their perceptive abilities, their psycho-physical characteristics at the moment of use, their habits, their knowledge of and familiarity with the product or system, etc.). These factors, when considered together, indicate that in order to integrate ergonomics into the projection of products and systems, one must apply methods based on the analysis of significant subjects, which is an effective way of finding out what the individual response is under real conditions of use. Ergonomic intervention is characterized, for the greater part, by subjective aspects, as it means adapting environments and products to the person, not in an abstract manner but with respect to the specificity of each individual.

4. ERGONOMIC DESIGN

The various phases and activities involved in the development of a product, from the first ideas to the final prototype and production, may differ depending on the nature of the product being designed, on the dimensions of the series produced, on the level of technology and automation of the industrial process, on the dimensions of the proposed market, on the size and diversification of the population for whom the product is intended and on its degree of formation and knowledge of the product category, but a number of phases are both typical and indispensable in all cases:

4.1. Non-material Phase

In the initial phase the product to be designed is presented as an idea only, in as much as it is constituted of written material, images and graphic or virtual representations. During this phase there is little opportunity to study the idea with ordinary people in order to obtain their thoughts on the use of a product which has not yet been realized materially.

4.2. Make-up Phase

During the next phase the idea may be represented with three-dimensional models. These represent well the dimensional aspects of the product but tell one little or nothing about finishes and functioning. They do allow one to effect verifications of use on subjects, but the non-functional characteristics of the model does not allow for an experience of complete use (e.g. activities of use which require a complete model, handling qualities linked to the finish of materials and to the qualities of grip of materials, qualities of noise and strain which may only be evaluated from finished products and those assembled and fitted to the appropriate structures). A sample of interviewees cannot, in this case, be made up of members of the general public, as it is necessary to choose subjects who are able mentally to simulate real conditions of use, in other words to able to think and perform in the abstract.

4.3. Prototype Phase

During the third phase pre-series or functional prototypes may be created with which analyses may be carried out making use of real subject under real conditions of use. Experience has shown, however, that even when functional prototypes are available it may not easy to conduct trials with members of the public, for reasons of product security and confidentiality (for products not

yet on the market) or for safety reasons (with products not yet approvable or approved).

One should note here that a correct process of ergonomic planning is not effected from the beginning of the project, in as much as one waits until prototypes are available for carrying out verifications with subjects. Thus, the possibility of making changes to the product conception is greatly limited, if not altogether irrelevant. Subjective investigations therefore require the application of specific and differentiated techniques depending on the actual phase of project development.

5. THE USER

The principles of ergonomics applied to a project affirm that environments, products and systems have to be adapted to the psychophysical characteristics of the individual. This would seem to assert that every environment, product and system has to be conceived and realized specifically for one single person in so far as every person is unique. Such procedures do in fact take place in the sphere of artisan production, personal tailoring, or in the making of seats for Formula One drivers.

Such principles seem to be in clear contradiction of fixed, repeatable references and the standardization required by industrial production of environments, products and systems.

While remaining faithful to the general principle which affirms that an ergonomic product has to adhere to the needs of the single user, ergonomics recognizes that the aspects that characterize an individual can be studied and rationalized in that many of these are common to all people, others being common to large sections of the population, still others being linked exclusively to a person's individuality.

5.1. Characteristics Common to All

These consist of the characteristics that denote someone as human rather than any other living form, such as a capacity for vision, touch, hearing, walking, and emitting articulated sounds. The biological and psychosocial disciplines provide infinite data that can and must be used during planning.

5.2. Cross-sectional Characteristics

Other characteristics are, on the other hand, specific to large cross-sections of humanity. These regard, for example, the thermoregulation mechanisms of the body which vary in relation to skin color or cultural characteristics, so that sounds perceived as harmonious by certain populations are not perceived as so by others. One should note that such aspects also vary over time. The pictorial representations of the Impressionists, when first shown to the world, were judged as an offence to good taste, so much so that the adherents to that school were always excluded from official exhibitions, whereas today they are considered among the greatest painters of all time.

5.3. Individual Characteristics

Within this framework, which refers to all humanity or its large categories, more specifically individual variables (which may be cultural, biological or pathological) must be researched.

Individual anthropometric characteristics tell us that nobody corresponds perfectly to the anthropometric data assumed to characterize individuals of his or her size; the capacity for performing physical tasks which is linked to one's state of health, musculature and diet; problems of vision often related to age etc. These aspects may still be measured and statistically analyzed.

Other parameters in the areas of cognitive psychology and cultural background regard more properly subjective aspects such as habits, posture, ways of dealing with aggression from others and in general a person's attitude to life which leads to different modes of behavior when faced with the same stimulus depending on deeply rooted factors or momentary states such as anxiety, haste, the desire for self-affirmation, etc.

Individual aspects cannot easily be reduced to universal numeric values, and so their introduction into the framework of the project requires specific techniques based on specialist analysis and tests and verifications carried out on significant samples of the likely users of the product.

products and systems	ALREADY PRODUCED	BEING DEVELOPED
OBJECTIVES	- verify ergonomic quality - certify ergonomic quality	- introduce ergonomic quality into the design phase
PRODUCTS (material or computer simulated)	- existing and functional	- non-material idea (on paper or computer) - non-functional models - partially functional prototypes
OBJECTIVE ASPECTS	- data drawn from analysis and laboratory tests	- data drawn from analyses and partial tests on models and ptototypes - data drawn from analyses and tests on similar products
SUBJECTIVE ASPECTS	- expert evaluation for comparison with standards and regulations - use analysis with subjects	- use analysis of partial aspects on models and prototypes - extrapolation of use analyses on similar products
EMOTIONAL ASPECTS	- analysis of product agreeability with subjects, for comparison with other products of the same category	- analysis of product agreeability with subjects for the definition, description and evaluation of relevant parameters

Figure 3. Comparison of methods of analysis for products already realized and those in the planning phase -

5.4. Methods of Analysis

To be able to introduce ergonomic concepts into product and systems design, it is necessary to intervene right from the beginning of a project, before it becomes impossible to reverse any choices made. Analyses have, therefore, to be programmed in all stages of the project:

1. *General research of "needs"* and their evolution over time (projection for when the product will be on the market), carried out through market research and by consulting scientific literature.

2. *Specific research* carried out on a significant sample of potential users, employing plans, models and/or prototypes.

3. *Research into similarities* carried out on a significant sample of potential users, with products already on the market, which in various ways are significant of the intended product (through similarity, in various parts, etc.).

While there are no relevant conceptual differences in the analysis of products at the design phase and that of products already realized, there are many in the methods of such analysis. In the case of existing and functional products analyses may be carried out using the product under real conditions of use. One may evaluate both parameters measurable by physical means and those measurable by cognitive ones. With an adequate methodology it is possible to arrive at the ergonomic certification of a product. In the case of products which have not yet been realized, the analyses necessary for guiding the development of the project in an ergonomic sense, may be executed on parts of the product, on single aspects taken out of the product context, or on products similar to that being developed. In this case the analyses differ from those carried out on similar products, not as regards the ends sought, but as regards a different planning of the research and the need to develop an analysis concentrating on data.

REFERENCES

BANDINI BUTI, L., 1998, *Ergonomia e Progetto* (Milan: Maggioli).

SARTOGO, P., 1982, *Italian Re-evolution*, exh. cat. (La Jolla: Museum of Contemporary Art).

NORMAN, D.A., 1988, *The Psychology of Everyday Things* (New York: Basic Books).

RABARDEL, P., 1995, *Les hommes et les Technologie* (Paris: Armand Colin).

Ergonomic Workstation Design

B. Das

Department of Industrial Engineering, Dalhousie University, Halifax, Nova Scotia NS B3 J 2X4, Canada

1. INTRODUCTION

Ergonomic workstation design based on engineering anthropometry and occupational biomechanics can play a major role in the reduction of many risk factors of occupational injury (Grandjean 1982). Anthropometry and biomechanics are closely related, because occupational biomechanics provide the bases for the use of engineering anthropometry to the problems of workstation design (Pheasant 1989).

Anthropometry is the technology of measuring various human physical straits, primarily such factors as size, mobility, and strength. Engineering anthropometry is the effort in applying such data to workstations, equipment, tools and clothing design to enhance the efficiency, safety and comfort of the worker. In the context of the workstation design, engineering anthropometry is employed to develop design parameters or dimensions for such a design.

Biomechanics deals with various aspects of physical movements of the human body and body members. The human body is considered as a linked body segments subjected to the internal forces generated within the musculoskeletal system and external forces imposed by the work situation (Karwowski 1992). Biomechanics uses law of physics and engineering concepts to describe various body parts and the forces acting on these parts during work activities (Frankel and Nordin 1980). The biomechanics or anatomy applied to the work situations is often considered the scientific basis of ergonomics (Tichauer 1978). Occupational biomechanics is defined as the physical interaction of workers with their tools, machines, and materials so as to enhance the workers' performance while minimizing the risk of musculoskeletal disorders (Chaffin and Andersson 1984).

An ergonomically designed workstation attempts to obtain an adequate balance between worker capabilities and work requirements. The objective is to optimize worker productivity and the total system and at the same time enhance worker physical and mental well being, job satisfaction, and safety. Often workstation in industry is designed in an arbitrary manner with little attention to anthropometric measurements and biomechanical considerations of the anticipated user. The situation is aggravated by the lack of usable design parameters or dimensions (Das and Grady 1983a). It is essential to understand the conditions under which biomechanics can be used for the assessment of a workstation (Marras 1997). When the magnitude of loads imposed upon the body is believed to be exceeding the tolerance of a structure, a biomechanical analysis is considered most helpful.

2. WORKSTATION DESIGN BASED ON ENGINEERING ANTHROPOMETRY

In a workstation design an attempt is made to achieve an optimum compromise between the variable anthropometry of the targeted operator population, and the physical size and the layout of the workstation components. An ergonomic analysis for a workstation design is concerned with spatial accommodation, posture, reaching abilities, clearance and interference of the body segments, field of vision, available strength of the operator, and biomechanical stress. The appropriate anthropometric data regarding body size, strength, segment masses and inertial properties from the established databases are typically used in the analysis.

Das and Grady (1983a, b) determined workstation design dimensions through the use of the existing anthropometric data, so that these dimensions can be readily employed by a designer. Workspace design dimensions were determined for industrial tasks in sitting, standing and sit–stand positions. Worker populations consisted of a combination of male and female workers and the individual male and female workers for the 5th, 50th and 95th percentiles based on existing anthropometric data. The normal and maximum reach dimensions were based on the most commonly used industrial operations, which require a grasping movement or thumb and forefinger manipulations. However, appropriate allowances were provided to adjust reach dimensions for other types of industrial operations. The normal and maximum horizontal and vertical clearances and reference points for the horizontal and vertical clearances were established to facilitate the design. The concepts developed by Farley (1955) and Squires (1956) were used to describe the workspace envelope for the individual worker. The dimensions of smaller (5th percentile) and larger (95th percentile) workers were used to determine the limits of reach and clearance requirements respectively.

2.1. Anthropometric Data Adjustment

The existing anthropometric data were derived on the basis of the measurements from nude subjects. Therefore the data were adjusted for clothing and shoe allowances. Since the data for the standing, sitting and eye heights were based on erect positions at work or rest, the data were adjusted to account for the 'slump' posture involved in the "normal" standing and sitting positions. Also, necessary adjustments were made for the reach dimensions used for performing various industrial operations (Hertzberg 1972). The corrected or adjusted anthropometric measurements to account for clothing, shoe and slump posture are presented in Table 1 (Das and Grady 1983a).

2.2. Workstation Design Parameters

To design a workstation, it is necessary to obtain relevant information or data on task performance, equipment, working posture and environment. In the case of a new workstation design, it is advantageous to obtain such information from a similar task/equipment situation. Before redesigning an existing workstation in industry, often it is desirable to conduct a worker survey through appropriate questionnaires to determine the effect of the equipment or system design on employee comfort, health and ease of use (of the equipment). However, frequently it is necessary to design a new industrial workstation. Even then, it may still be desirable to obtain feedback from the operators, who are engaged in performing a similar type of industrial task. The feedback may generate heightened awareness of workstation design problems and issues.

Table 1. Corrected anthropometric measurements to account for clothing, shoe and slump posture (Das and Grady 1983a).

Sex	Body Feature	Percentiles (cm)		
		5th	50th	95th
Male	Total height (slump)	166.2	176.1	186.3
	Body height (sitting, slump)	82.1	87.6	92.7
	Eye height (slump)	160.0	164.9	174.8
	Eye height (sitting, slump)	70.9	76.2	81.3
	Shoulder height	138.0	147.7	156.8
	Shoulder height (sitting)	54.7	59.8	64.4
	Body depth	26.7	30.2	34.0
	Elbow-to-elbow	40.0	45.1	51.7
	Thigh clearance	14.2	16.2	18.5
	Forearm length	37.1	40.4	43.7
	Arm length	68.3	75.2	82.0
	Elbow height	105.6	113.0	120.4
	Elbow height (sitting)	18.8	23.1	27.4
	Popliteal height (sitting)	42.4	45.7	48.7
Female	Total height (slump)	153.8	163.2	173.8
	Body height (sitting, slump)	78.5	82.8	87.1
	Eye height (slump)	143.6	153.5	163.4
	Eye height (sitting, slump)	68.6	72.4	76.5
	Shoulder height	127.9	137.5	146.6
	Shoulder height (sitting)	52.2	57.2	61.8
	Body depth	21.8	24.6	27.6
	Elbow-to-elbow	35.7	38.2	43.8
	Thigh clearance	12.4	14.4	16.5
	Forearm length	32.5	36.6	40.7
	Arm length	60.2	66.0	72.4
	Elbow height	99.0	105.1	111.2
	Elbow height (sitting)	18.8	23.1	27.4
	Popliteal height (sitting)	37.3	40.6	43.6

In the beginning, decisions are formalized regarding the task sequence, available space, equipment and tools. Work method needs to be established before embarking on the design of the workstation. Determination of the workstation dimensions usually proceeds according to the steps outlined in Table 2 (Das and Sengupta 1996).

The workstation design procedure commences with the collection of relevant data through direct observation, video taping and input from experienced operators and supervisors (step 1, Table 2). It is necessary to identify the appropriate user population based on such factors as ethnic origin, gender and age (step 2). The necessary anthropometric dimensions of the population are obtained or approximated from the results of the available anthropometric surveys that reasonably represent the user group. As these dimensions are taken from nude subjects in erect posture, they need to be corrected appropriately for the effect of clothing, shoe and normal slump posture during work (Das and Grady 1983a).

In developing an industrial workstation, a designer should take into account the workstation height (step 3). The work table height must be compatible with worker height, whether standing or sitting. The best working height is ~2.5 cm below the elbow. However, the working height can vary several centimeters up or down without any significant effect on performance. The nature of the work to be performed must be taken into consideration in

Table 2: A systematic approach for the determination of workstation design parameters (Das and Sengupta 1996).

1. Obtain relevant information on the task performance, equipment, working posture and environment through direct observation, video recording, and/or input from experienced personnel.

2. Identify the appropriate user population and obtain the relevant anthropometric measurements or use the available statistical data from anthropometric surveys.

3. Determine the range of work height based on the type of the work to be performed. Provide an adjustable chair and a foot rest for a seated operator and an adjustable work surface or platform for a standing operator.

4. Layout the frequently used hand tools, controls, and bins within the normal reach space. Failing that, they may be placed within the maximum reach space. Locate control or handle in the most advantageous position, if strength is required to operate it.

5. Provide adequate elbow room and clearance at waist level for free movement.

6. Locate the displays within the normal line of sight.

7. Consider the material and information flow requirements from other functional units or employees.

8. Make a scaled layout drawing of the proposed workstation to check the placement of individual components.

9. Develop a mock-up of the design and conduct trials with live subjects to ascertain operator-workstation fit. Obtain feedback from the interest groups.

10. Construct a prototype workstation based on the final design.

determining work height. For seated operators, provide adjustable chair and foot rests; for standing operators, provide an adjustable work surface or platform.

The hand tools, controls and bins that are frequently used need to be located within the normal reach spaces (step 4). The items used occasionally may be placed beyond the normal reach, but they should be placed within the maximum reach space. For locating a control that requires strength, give consideration to the human strength profile in the workspace. Extreme reach space, involving twisting of trunk, ought to be avoided at all times. Adequate lateral clearance must be provided for the large (95th percentile) operator for ease of entry and exit at the workstation and also to provide ample elbow room for ease of work (step 5).

The placement of the displays should not impose frequent head and/or eye movement. The optimum display height for the normal (slump) eye height is 15° downward gaze (step 6). Appropriate personnel from other functional units or departments should be consulted regarding material and information flow requirements (step 7). It is beneficial to consider the physical size of the individual components and make a scaled layout drawing of the proposed workstation to check the placement of the individual components within the available space (step 8). The operator–workstation fit should be evaluated with a workstation mock-up and through an appropriate user population (step 9). This will ensure that the task demand and layout do not impose an undesirable working posture. It is desirable to check the interference of body members with the workstation components. If necessary, the design should be modified. Finally, it is beneficial to construct a prototype workstation based on the final design. All the 10 steps shown in Table 2 may not be applicable in every industrial workstation design situation.

2.3. Workstation Dimensions

For the physical design of industrial workstations, the four essential design dimensions are: (1) work height, (2) normal and maximum reaches, (3) lateral clearance and (4) angle of vision and eye height.

2.3.1. Work height

Height of the working surface should maintain a definite relationship with the operator's elbow height, depending upon the type of work. The standing work heights for the 5th, 50th and 95th percentile female operators for performing different types of work for US population are presented in Table 3. The table provides guidelines especially for the design of delicate, manual and forceful work. Similar data for males can be obtained from Ayoub (1973), and Das and Grady (1983a).

2.3.2. Normal and maximum reaches

The tip of the thumb defines the normal reach while the forearm moves in a circular motion on the table surface. During this motion, the upper arm is kept in a relaxed downward position. The "maximum" reach can be considered as the boundary on the work surface in front of an operator to which s/he can reach without flexing his/her torso. For performing repetitive tasks, the hand movement should preferably be confined within the normal working area. The controls and items of occasional use may be placed beyond the normal working area. Nevertheless, they should be placed within the maximum working area.

The concept of normal and maximum working areas (Das and Grady 1983b, Das and Behara 1995) describes the working area in front of the worker in a horizontal plane at the elbow level; the areas are expressed in the form of mathematical models. The most frequently used area of the workstation preferably should be within the normal reach of the operator. The reach requirements should not exceed the maximum reach limit, to avoid leaning forward and bad posture. The maximum working area at the elbow level is determined from the data provided in Table 4. The adjusted anthropometric measurements for the arm length (K), shoulder height (E), elbow height (L), which are used to calculate arm radii (R) for the 5th, 50th and 95th percentiles for females.

Table 3. Standing work surface height for female operators in cm parameters (Das and Sengupta 1996).

Type of work	Population percentile		
	5th	50th	95th
Delicate work with close visual requirement	99-104	110-115	116-121
Manual Work	84-89	90-95	96-101
Forceful work aided by upper body weight	59-84	65-90	71-96

Table 4. Anthropometric measures for females and maximum reach in cm parameters (Das and Sengupta 1996).

Population percentile	Arm length (K)	Shoulder height (E)	Elbow height (L)	Maximum reach (R)
5th	60	128	99	53
50th	66	138	105	58
95th	72	147	111	63

Table 5. Anthropometric measurements for females for lateral clearances in cm parameters (Das and Sengupta 1996).

Population percentile	Hip breadth	Elbow to elbow	Body depth	Clearance waist level C1=W+10	Clearance elbow level C2=H+G+10
5th	40	36	22	50	68
50th	45	38	25	55	73
95th	52	44	28	62	81

2.3.3. Lateral clearance

A well known approach is to design the reach requirements of the workstation corresponding to the measurements of the 5th percentile of the representative group and the clearance corresponding to the 95th percentile measurements, so as to make the workstation compatible for both small and large persons. The minimum lateral clearances at waist level are determined by adding 5 cm on both sides or 10 cm to hip breadth (standing). Considering the elbow to elbow distance and the sweep of both the elbows within the normal horizontal working area and adding 5 cm on both sides, minimum lateral clearance at elbow level is determined. The values for lateral clearances are shown in Table 5.

2.3.4. Angle of vision and eye height

Das and Grady (1983a, b) have provided the eye height for the standing female operators: 143.6 cm for 5th percentile, 153.5 cm for 50th percentile and 163.4 cm for 95th percentile. For males, similar eye height data can be obtained from the source stated above. Using trigonometry, the angle of sight can be calculated from the horizontal distance of the display from the operator's eye position.

3. SUPERMARKET CHECKSTAND WORKSTATION: A CASE STUDY

The steps outlined in Table 2 were applied to design a supermarket checkstand workstation. The working posture and work methods of the cashiers were recorded through direct observation (step 1, Table 2). Based on the direct observation, the major shortcomings in the design of the supermarket checkstand workstation were identified.

An operator (cashier, all female) survey was conducted in 3 superstores, through questionnaires to determine the effect of (1) environmental factors, (2) general fatigue induced by the task, (3) physical demand of tasks, and (4) the postural discomfort of the operators during the course of a regular working day (step 2). The dimensional compatibility between the cashiers structural anthropometry and the workspace dimensions was evaluated using a scaled drawing of the existing checkstand which was prepared from the actual measurements taken at the site.

By employing an engineering anthropometry approach, the supermarket checkstand workstation design parameters or dimensions were determined for: (1) optimum work height, (2) normal and maximum reaches, (3) lateral clearance, and (4) angle of vision and eye height (steps 3–6). The appropriate data were used subsequently for the design of the checkstand workstation (Figure 1; Das and Sengupta 1996). The superimposition of the normal and maximum horizontal and vertical working areas on the supermarket checkstand drawing facilitated the design or

Legend:
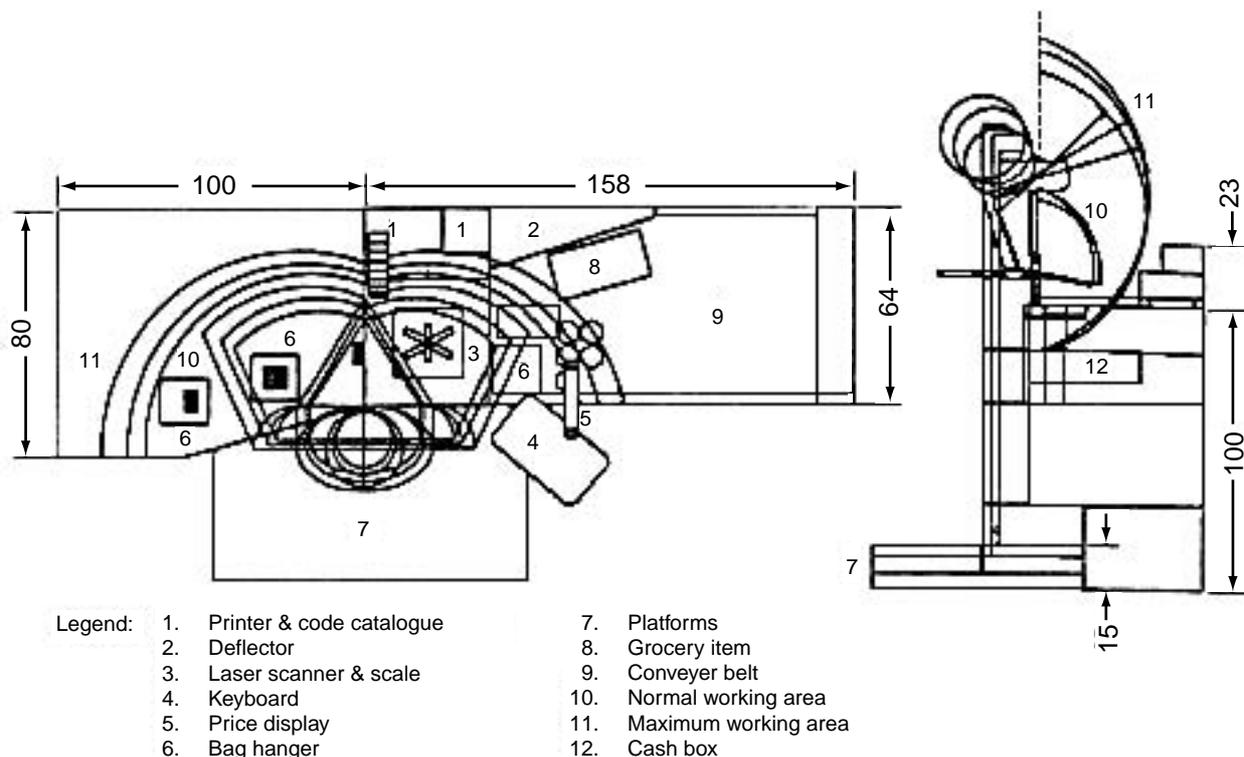|			
1.	Printer & code catalogue	7.	Platforms
2.	Deflector	8.	Grocery item
3.	Laser scanner & scale	9.	Conveyer belt
4.	Keyboard	10.	Normal working area
5.	Price display	11.	Maximum working area
6.	Bag hanger	12.	Cash box

Figure 1. Normal and maximum working areas of the 5th, 50th and 95th percentiles for the female in the horizontal and vertical planes for a supermarket checkstand workstation. All dimensions in centimeters parameters (Das and Sengupta 1996).

placement of the checkstand components. This procedure enabled placement of the components within the normal working area when possible and, failing that, within the maximum working area. Several alternative checkstand layout drawings were considered using a CAD package with human modeling capability. An adjustable and padded floor platform was provided for the cashiers. The platform height could be lowered by 6 cm at a time to accommodate taller cashiers. The padding was provided to reduce foot fatigue from prolonged standing.

The components that needed frequent operation were placed, as far as possible in the front of the cashier to reduce the twisting of the torso and neck. The width and depth of the laser scanner was reduced to accommodate the printer and code catalogue right in front of the cashiers. The reduced width scanner improved the cashiers' reach over both conveyor belt and the bag handling area. A deflector was provided on the conveyor belt to ensure that the products are within the maximum reach of the cashiers. The original product bin handling requirements at the left of the cashiers was eliminated and was replaced by plastic bag hangers at appropriate height and location. The superimposition of the normal and maximum horizontal and vertical working areas on the checkstand drawing facilitated the determination of the size and the placement of the checkstand components (step 8).

A time study of the simulated cashiers' task in the laboratory showed a 15% improvement in worker productivity. The main improvements in the proposed design were: (1) forward facing work posture and elimination of the requirement of twisting the torso (forward facing location of laser scanner, weigh scale, bag hangers, keyboard and printer/code catalogue), (2) increased area on the conveyor belt within the normal working area, (3)

placement of the visual display (item price) within the normal line of sight of the operator, and (4) adjustable height platforms to accommodate 5th, 50th and 95th percentile female operators. The proposed design would improve working posture, provide flexible work height, reduce reach requirements, improve visual display requirements and enhance productivity of the cashiers.

4. HUMAN MODELING PROGRAMS FOR WORKSTATION DESIGN

The computerized human modeling programs for an industrial workstation design provide a convenient computer interface for the user to interactively generate and manipulate true-to-scale, three-dimensional (3D) image of human and the workstation graphically on the video display terminal (VDT). Through the use of the programs, the designer can construct a large number of anthropometric combinations to represent the human. The programs give a complete control to the user over the development of the human model and provide a comprehensive package to evaluate human–machine interaction through easy to understand programmed commands. The user need not be a computer specialist to use such programs.

To illustrate the current state of development, Das and Sengupta (1995) had selected six representative programs: CYBERMAN, COMBIMAN, CREN CHIEF, JACK, SAMMIE and MANNEQIN. The individual programs were compared under four broad criteria: (1) usability in terms of hardware and software: (2) anthropometry and structure of the human model; (3) model manipulation, reach and visual analysis functions: and (4) other ergonomic evaluative functions (Das and Sengupta 1995). The programs differ considerably in terms of system requirement,

operating characteristics, applicability and the various ergonomic evaluation functions available in the human modeling programs. For the purpose of illustration two representative models are presented (Figures 2 and 3).

4.1. Recent Developments in Workstation Design Based on Engineering Anthropometry

A computerized potentiometric measurement system (CPMS) was recently developed for anthropometric measurements of the three-dimensional maximum reach envelope (MRE) (Das *et al.* 1994). The system uses four potentiometric units, a power supply unit, a computer with analogue/digital (A/D) converter to measure the position of a movable stylus in three (S, Y, and Z) dimensions (Figure 4). This system took only 15 min for reach data collection
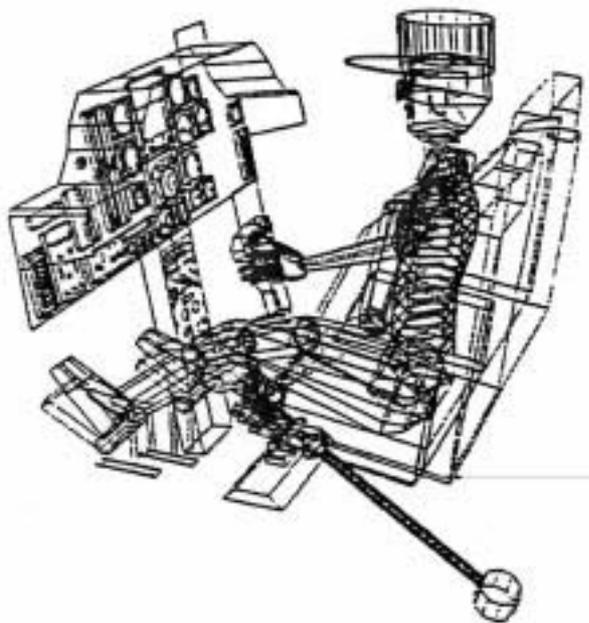
Figure 2. JACK human model seated within a workstation (Badler 1991).
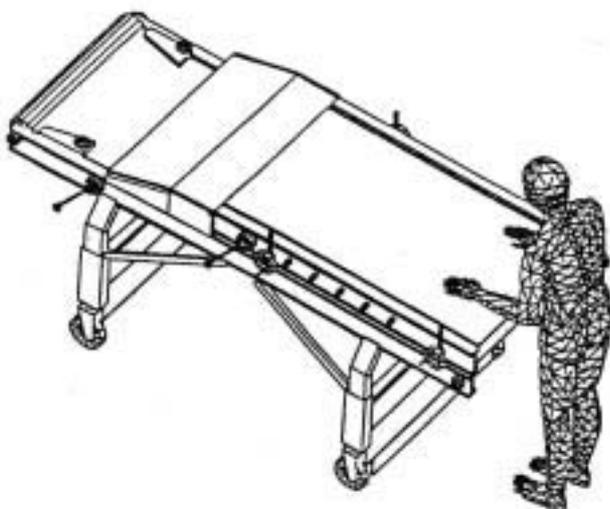
Figure 3. MANNEQUIN-generated human model in front of a meat-cutting workstation (HUMANCAD 1991).

Legend:
1. Support frame
2. Potentiometer power supply
3. Digital computer with A/D converter
4. Location of the PRUs
5. Moveable pointer
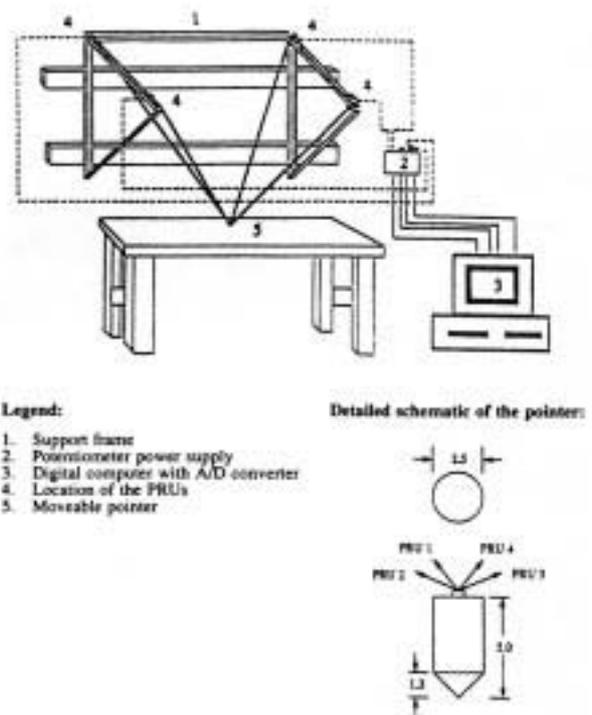
Figure 4. Schematic diagram of the computerized potentiometric system for structural and functional anthropometric measurements. All dimensions are in cm (Das *et al.* 1994).

of a participant in seated and standing positions. It recorded arm reaches while the arm was in motion, thus reflected the true dynamic nature of functional reach. The previous mechanical measurement systems recorded maximum reach envelope in terms of static arm reaches and took 3 h to collect data for the seated maximum reach envelope. The CPMS was successfully used to determine MRE for the 5th, 50th and 95th percentile males and females in seated and standing work positions (Sengupta and Das 1999).

An improved anthropometric model of three-dimensional maximum reach envelope (MRE) was recently developed (Sengupta and Das 1998). The characteristics of the model included: application of nonlinear optimization, use of relevant structural body dimensions, use of direct measurement of dynamic reach profiles and location of the spherically shaped MRE near the shoulder joint (acronium point).

4.2. Workstation Design Based on Occupational Biomechanics

A biomechanical approach to workstation design deals with the effects of exertion-related forces and awkward postures in a work situation that may cause pain or injury. For the design of a workstation it is necessary to know what a person can or cannot do. A person's physical capabilities, especially those that allow on individual to exert force or sustain external loading without causing personal injury can be determined by measuring human strengths (Mital and Das 1987). It is necessary to conduct a biomechanical evaluation of a workstation design when a worker is believed to be exposed to excessive physical stress or risk in a workstation. Through redesign of a workstation, it is possible to

minimize or eliminate work-related injuries to musculoskeletal system. The low back pain is considered to be the most costly and prevalent work related musculoskeletal disorder. According the National Council on Compensation Insurance (USA), 33% of all workers' Compensation payments are due to low back pain. The total cost estimates range from US$27 to 56 billion in the USA when indirect costs are taken into consideration (Andersson *et al.* 1991).

4.3. Determination of Human Strength for Workstation Design

Human strengths are classified as isometric or static and dynamic. The dynamic strengths are further subdivided into: (1) isokinetic, (2) isotonic and (3) iso-intertial. In the case of isometric or static muscle exertions, the body segment involved and the object held remain stationary, while in the case of dynamic muscular exertions both the body segment and object move. Because there is no effective limb–object–muscle movement in the case of isometric strengths, such strengths do not account for the inertial forces. Consequently, isometric strengths cannot be used for the determination of an individual's capability to perform dynamic tasks, such as materials handling. Thus, for the determination of a person's physical capabilities, dynamic strengths measurement is more appropriate than isometric or static strengths measurement. For designing industrial jobs and workstations, dynamic strengths should be used even though they are difficult to measure compared to static strengths. Table 6 shows isometric and isokinetic strengths of males in different postures (Mital *et al.* 1986).

Industrial workers should not generally exceed one-third of their isometric strength on a sustained basis in task performance (Putz-Anderson 1994). Overloading of muscles should be avoided to minimize fatigue. Dynamic forces should be kept < 30% of the maximum force that the muscle can exert; up to 50% is all right for up to 5 min. Static muscular load should be kept < 15% of maximum force that the muscle can exert. General guidelines suggest that hand forces should not exceed 45 Newtons. On the other hand, it is possible to handle a force of 4 kg for 10 s, 2 kg for 1 min and one-third of maximum force for 4 min.

For optimum design of a workstation, it is important to determine human strength profiles in the workspace. The ideal

Table 6. Isometric and isokinetic lifting strengths of males in different postures (Mital *et al.* 1986).

Posture	Exertion	Mean	Strength (kg) Standard deviation
Standing erect, fore arms horizontal, upper arms vertical	Isometric	40.4	6.7
Standing erect, upper arms horizontal, fore arms vertical	Isometric	52.6	8.5
Lifting floor to 81 cm height in free style	Isokinetic	110.4	30.2
Lifting from 81 cm to 152 cm in free style	Isokinetic	75.5	33.4
Pulling at 81 cm height from left to right	Isominetic	35.1	9.5

industrial workstation should be compatible with not only systems performance requirements but also with the user. The most obvious criteria are comfort and ease of use, but other equally important design criteria include work performance, safety and health (Das and Sengupta 1996). Several factors impinge upon the creation of the ideal workstation, one of which is *reach capability*, as discussed earlier under workstation design based on engineering anthropometry. Accurate reach capability data are essential to ensure that all hand-operated controls or tasks are located where they can be reached and operated efficiently. Another factor that impinges upon the creation of the ideal workstation is user *strength capability*. To ensure optimal workplace layout, it is imperative that operators strength profile be determined. The *strength profile* of a person under specified conditions is essential for the design of tools (e.g. their weight, ease of use), controls (e.g. type of grip required, spatial placement), and equipment – in other words, the workstation. Furthermore, for the selection or job placement of workers requiring strength exertion in task performance, the measurement of strength profiles of such individuals can be useful.

Studies have shown that horizontal distance and vertical height of exertion significantly affect the force exertable both in static and dynamic strength tests (Chaffin and Park 1973, Davis and Stubbs 1977). However these studies have not attempted to relate anthropometric reach space envelopes to the strength data obtained. Researchers have measured strength at varying elbow angles (Hunsicker 1955), fractions of mean reach for the population (Davis and Stubbs 1977, Kumar 1991), or fixed distances (Mital and Faard 1990). Individual functional reach regions have not determined measurement locations. For optimum workstation design a link must be established between an individual's ability to reach and exert force at functional reach regions. Recently isometric push and strength profiles were determined for the able-bodied population in the normal, maximum and extreme workspace reach envelopes (Das and Wang 1995). Also, isometric push, pull, push-up and pull-down strength profiles were determined for the paraplegics in the similar workspace (Das and Black 1999, Das and Forde 1999). Research is in progress at Dalhousie University, Canada to determine a comprehensive database for both static (isometric) and dynamic strength profiles in the workspace.

Insufficient physical capability while performing manual materials handling activities and tasks requiring hand tool usage can lead to overloading the muscle–tendon–bone–joint system and possible injury (Ayoub and Mital 1989). These two activities account for ~45% of all industrial overexertion injuries. It accounts for billions of dollars in worker's compensation cost (Waters and Putz-Anderson 1996).

4.4. Posture Analysis for Workstation Design

A common manifestation of back injury is low back pain (LBP). Epidemiological studies show a positive correlation between the exposure to mechanical overload at work and the incidence rate of LBP (Chaffin and Park 1973). However, except for traumatic and acute cases, the cause of the LBP still remains unclear (Kroemer *et al.* 1996). For the manual material handling (MMH) tasks, it is believed that the source of LBP can be traced back to the repetitive over-exertion at the lower back. This produces microtrauma at the lower spine structure over a prolonged period

of time. The microtrauma ultimately results in a permanent or temporary damage to the fibrocartilegnious disks and its surround structures and can cause LBP.

An important predictor of such structural failure is the mechanical forces acting at the lumber spine which depend on the interaction of the worker anthropometry and work characteristics. In the evaluation of workstation and work method, it is important to determine the stresses at lower back due to work performed at different trunk postures. From a biomechanical perspective, the compressive force generated at lower back (L5/S1) is believed to be the most significant factor in the development of LBP.

Safe limits of work related to heavy exertion have been well established for lifting and lowering type of manual material handling (Waters and Putz-Anderson 1996) and push–pull type of exertion (Snook and Ciriello 1991). But limits for work-related lower back stresses for postural loads are not well defined. Awkward postures are of major concern for workers who are performing repetitive jobs due to the frequency and cumulative effects of exposure. Non-neutral back postures such as flexion, lateral bending, and/or twisting increase the level of muscle fatigue and intradiscal pressures in the lumbar spine (Andersson *et al.* 1977, Chaffin 1973). Severe trunk postures can elevate the compressive force at low back, even though a load in the hands does not exist or is relatively light in weight (Chaffin and Andersson 1991).

Several whole body posture-recording schemes have been developed (Juul-Kristensen *et al.* 1997) to estimate the postural stress for various types of industrial work. Their effectiveness in quantifying the postural stress has been validated by extensive field studies (Genaidy *et al.* 1994). Primarily the posture recording schemes provide a means to estimate the level of postural stress based on the severity and duration of the work postures. However, most of them do not provide the safe limits of the postural stresses and thus, they are basically useful for comparison of postural stresses before and after modification of a work-site. The Ovako Working Posture Analysis System (OWAS) (Karhu *et al.* 1977) is a widely used postural recording scheme and has been applied to various types of industries (Pinzke 1992). A significant relationship between the back postures as defined by OWAS and prevalence in lower back pain has been established by epidemiological analysis (Heinsalmi 1986, Burdorf *et al.* 1991).

4.5. Use of Biomechanical Methods for Workstation Design

Occupation biomechanics provide a logical basis for providing data for human performance and human tissue tolerance. Biomechanical models of lower back with varied level of anatomical details and modeling capabilities have been developed to predict the compressive force in the low back due to external loads. The complex models have dealt with passive force generation and load sharing by muscles and ligaments, asymmetric postures, and inertial effects of dynamic motions. However, at the present state of development of the models, the model outputs are still not reliable enough for judging the absolute acceptability of a work situation, rather they are more suitable for comparing back loading in different work situation (Delleman *et al.* 1992). This is because of a the inherent difficulty in substantiating the validation of the model outputs and the uncertainties of the living tissue strength values against which the model predictions would be compared. Traditionally, the strength values were based on the failure strength of postmortem pine segments under axial strength values were based on the failure strength of postmortem spine segments under axial load. Lately, spine segments were tested under cyclic load, and fatigue failure data of spine segments are made available (Brinckmann *et al.* 1987). Often a computerized biomechanical model of the human musculoskeletal system is used to predict human capabilities for a particular task performance (Chaffin 1997). To predict whole-body exertion capabilities for a given population, Chaffin (1997) presents example logic for a model (Figure 5). Specific muscle group strength data and spinal vertebrae failure data are used in this model as the limiting values for the reactive movement at various body joints when a operator of particular stature and body weight attempts an exertion, such as lifts, pushes or pulls in a specific direction with one or both hands while maintaining a predicted stature. For comparing task exertions where specific localized muscle actions exist, it is possible to use electromyography (EMG) for the assessment of the active muscles and/or a subjective discomfort rating method.

Several guideline and design evaluation methods are available for the initial design of a workstation. The selection of an appropriate method will depend of the objective in terms of solving a specific problem, and whether a single exertion or multiple exertions are involved (Figure 6; Chaffin 1997). A
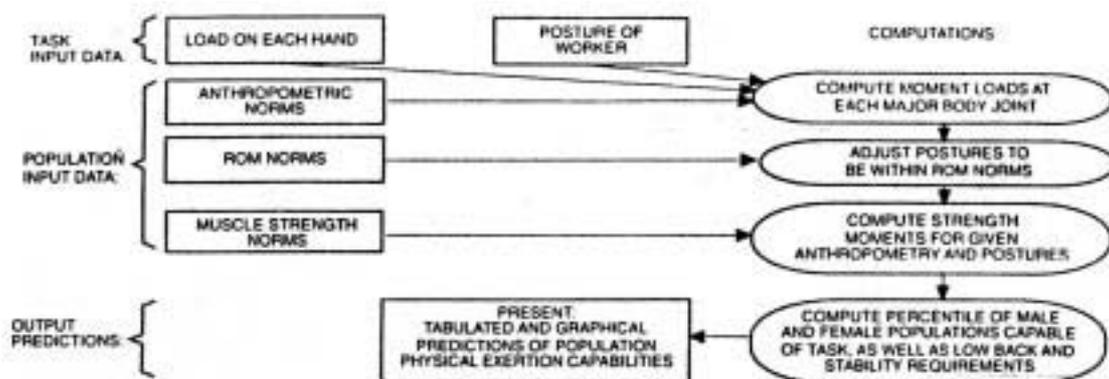

Figure 5. Biomechanical logic used to predict whole-body static exertion capabilities for given postures, hand force directions and anthropometric groups (Chaffin 1997).
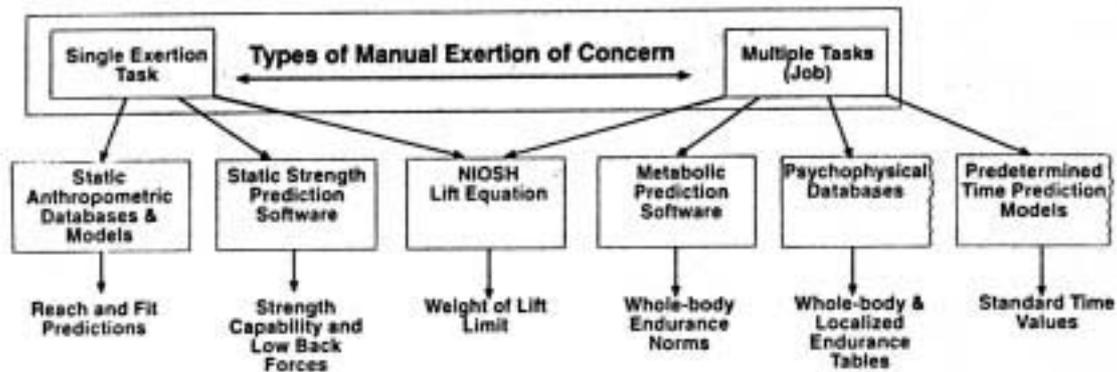
Figure 6. Models and databases available for proscriptive guidance in design of new work-places with outcome predictions listed at bottom of each model or database (Chaffin 1997).
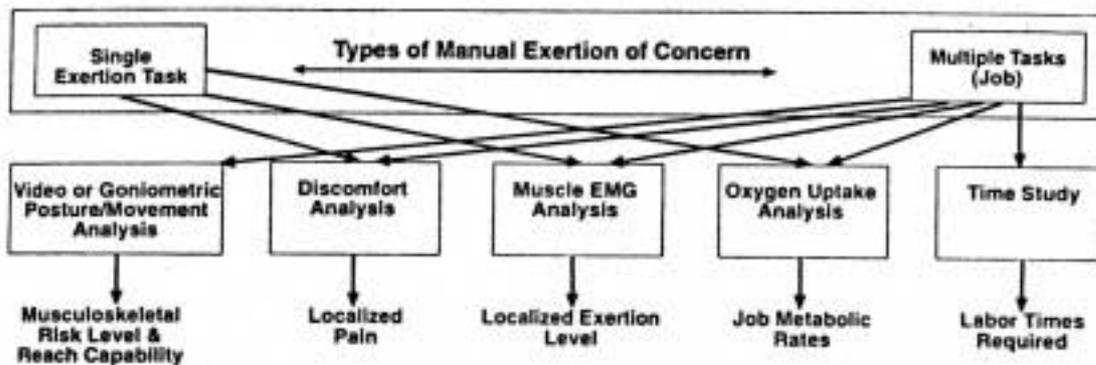
Figure 7. Methodologies for measuring human performance during prototype analyses, with outcome data listed for each (Chaffin 1997).

prototype workstation design can be evaluated by employing a suitable method or several methods presented in Figure 7; Chaffin 1997). A representative worker(s) can be used to further refine the initial design.

4.6. Workstation Design for Manual Materials Handling

From a biomechanical viewpoint, package size can be of problem, especially when the package is located on or near the floor (Chaffin 1997). If the load is of a size that cannot be easily handled between the knees at the start of the lift, then an operator must lean the torso forward. Using the 3D Static Strength Prediction Program, Chaffin (1994) had demonstrated the effect of lifting two different size boxes of the same weight. The combination of a forward torso angle and large horizontal distance between the large box and low back caused about a 30–38% increase in predicted L5/S1 disc compression force compared to small box lifting close to the body. A significant risk of low back injury would result, if the object weighed > 35 lb.

From the viewpoint of workstation design, when large packages are involved they should never be presented to an operator at a height lower than about mid-thigh (or ~30 inches). This would allow the operator to stand erect and bring the object against (or near) the torso, thus minimizing lower back bending movements and resulting spinal compression forces. When large bulky objects are involved, adjustable lift tables should be used (Chaffin 1997).

The orientation of the package when presented to the operator must be taken into consideration. If the shape of the package is not like a cube but rather one small dimension, it should be presented in a more vertical direction. This will permit the operator to list the object closer to the body by straddling the narrow dimension, or if a handle is provided on the top, by lifting close to the side of the body.

5. SUMMARY AND CONCLUSIONS

For the design of an ergonomic workstation, it is essential to give utmost consideration to engineering anthropometry and occupational biomechanics. This will ensure a reduction of many risk factors of occupational injury and contribute towards enhancing worker productivity, worker physical and mental well being and job satisfaction.

In a workstation design, an ergonomic analysis is performed to deal adequately with spatial accommodation, posture, reaching abilities, clearance and interference of the body segments, field of vision, available strength of the operator, and biomechanical stress. Typically used in the analysis are the appropriate anthropometric masses and inertial properties from the established databases. It is necessary to make anthropometric data adjustment to account for clothing, shoe and slump posture. A systematic ergonomics approach is provided for the determination of workstation design parameters. The four essential design dimensions are: (1) work height, (2) normal and maximum reaches, (3) lateral clearance, and (4) angle of vision

and eye height. A case study of a supermarket checkstand workstation is provided to illustrate the systematic manner by which design parameters can be determined. To evaluate human–machine interaction for the design of a workstation, the use of computerized human modeling programs is illustrated by presenting two representative models: JACK and MANNEQIN. The recent developments in workstation design include: (1) a computerized potentiometric system for structural and functional anthropometric measurements, and (2) an improved anthropometric model of three-dimensional maximum reach envelope.

A biomechanical evaluation of a workstation is necessary, when a worker is believed to be exposed to excessive physical stress or risk in a workstation. It is possible to minimize or eliminate work-related injuries to musculoskeletal system through redesign of a workstation. For the determination of a person's physical capabilities, dynamic strengths measurement is more appropriate than isometric or static strengths measurement. Guidelines are provided so that industrial workers do not overload muscles to avoid or minimize fatigue. It is important to determine human strength profiles in the workspace for optimum design of a workstation. Since back injury often causes low back pain, it is necessary to determine the stresses at lower back due to work performed at different trunk postures. The compressive force generated at lower back (L5/S1) is believed to be the most significant cause for low back pain. To predict human capabilities for a particular task performance, a computerized biomechanical model of the human musculoskeletal system is often used. For the initial design of a workstation, several guideline and design evaluation methods are available. The effect of lifting two different size boxes of the same weight was demonstrated by using a 3D Static Strength Prediction Program. Guidelines are provided for manual materials handling especially when large packages are involved.

REFERENCES

ANDERSSON, G., ORTENGREN, R. and HERBERT, P. 1977, Quantitative electromyographic studies of back muscle activity related to posture and loading. *Orthopedic Clinics of North America*, 8, 85–96.

ANDERSSON, G.B.J., POPE, M.H., FRYMOYER, J.W. and SNOOK, S. 1991, Epidemiology and costs. In M. Pope, G.B.J. Andersson, J. Frymore and D.B. Chaffin (eds), *Occupational Low Back Pain* (St Louis: Mosby Year Book), 95–113.

AYOUB, M.M. 1973, Workplace design and posture. *Human Factors*, 15, 265–8.

AYOUB, M.M. and MITAL, A. 1989, *Manual Materials Handling* (London: Taylor & Francis).

BADLER, N.I. 1991, Human factors simulation research at the University of Pennsylvania. Computer graphic research quarterly progress report No. 38. Department of computer and Information Science, University of Pennsylvania, Forth Quarter 1990, 1–17.

BRINCKMANN, P., JOHANNLEWELING, N., KILWEG, D. and BIGGEMANN, M. 1987, Fatigue fracture of human lumber vertebrae. *Clinical Biomechanics*, 2, 94–6.

BURDORF, A., GIVAERTM G. and DLDERS, L. 1991, Postural load and back pain of workers in manufacturing of prefabricated concrete elements. *Ergonomics*, 34, 909–18.

CHAFFIN, D.B. 1973, Localized muscle fatigue — definition and measurement. *Journal of Occupational Medicine*, 15, 346–54.

CHAFFIN, D.B. 1994, Postural considerations in lifting — or why aren't my arms five feet long? In M. Nordin, G. Andersson and M. Pope (eds), *Occupational Musculoskeletal Disorders: Assessment, Treatment and Prevention* (St Louis: Mosby Year Book).

CHAFFIN, D.B. 1997, Biomechanical aspects of workplace design. In G. Salvendy (ed.), *Handbook of Human Factors and Ergonomics* (2nd ed.) (New York: Wiley), 772–89.

CHAFFIN, D.B. and ANDERSSON, G.B.J. 1984, *Occupational Biomechanics* (New York: Wiley).

CHAFFIN, D.B. and ANDERSSON, G.B.J. 1991, *Occupational Biomechanics* (2nd ed.) (New York: Wiley).

CHAFFIN, D.B. and PARK, K.S. 1973, A longitudinal study of low-back pain as associated with occupational weight lifting factors. *American Industrial Hygiene Association Journal*, 34, 513–25.

DAS, B. and BEHARA, D.N. 1995, Determination of the normal horizontal working area: a new model and method. *Ergonomics*, 38, 734–48.

DAS, B. and BLACK, N.L. 2000, Isometric pull and push strengths of paraplegics in the workspace: 1. Strength measurement profiles. *International Journal of Occupational Safety and Ergonomics* (in press).

DAS, B. and FORDE, M. 1999, Isometric push-up and pull-down strengths of paraplegics in the workspace: 1. Strength measurement profiles. *Journal of Occupational Rehabilitation*, 9, 279–91.

DAS, B. and GRADY, R.M. 1983a, Industrial workplace layout design: an application of engineering anthropometry. *Ergonomics*, 26, 433–47.

DAS, B. and GRADY, R.M. 1983b, The normal working area in the horizontal plane: a comparative analysis between Farley's and Squires' concepts. *Ergonomics*, 26, 449–59.

DAS, B., KOZEY, J.W. and TYSON, J.N. 1994, A computerized potentiometric system for structural and functional anthropometric measurements. *Ergonomics*, 37, 1031–45.

DAS, B. and SENGUPTA, A.K. 1995, Computer-aided human modelling programs for workstation design. *Ergonomics*, 38, 1958–72.

DAS, B. and SENGUPTA, A.K. 1996, Industrial workstation design: a systematic ergonomics approach. *Applied Ergonomics*, 27, 157–63.

DAS, B. and WANG, Y. 1995, Determination of isometric pull and push strength profiles in workspace reach envelopes. In *Proceedings of the Annual International Industrial Ergonomics and Safety Conference*, Seattle, Washington, 1017–24.

DAVIS, P.R. and STUBBS, D.A. 1977, Safe levels of manual forces for young males (2). *Applied Ergonomics*, 8, 141–50.

DELLEMAN, N.J., DROST, M.R. and HUSON, A. 1992, Value of biomechanical macromodels as suitable tools for the prevention of work-related low back problems. *Clinical Biomechanics*, 7, 38–48.

FARLEY, R.R. 1955, Some principles of methods and motion study as used in development work. *General Motors Engineering Journal*, 2, 20–5.

GENAIDY, A. M., AL-SHEDI, A. and KARWOWSKI, W. 1994, Postural stress analysis in industry. *Applied Ergonomics*, 25, 77–87.

GRANDJEAN, E. 1982, *Fitting the Task to the Man: An Ergonomic Approach* (3rd ed.) (London: Taylor & Francis).

HERTZBERG, H.T.E. 1972, Engineering and anthropometry. In H.P. van Cott and R.G. Kinkade (eds), *Human Engineering Guide to Equipment Design* (New York: McGraw-Hill).

HUMANCAD, 1991, *MANNEQUIN User Guide* (New York: Melville).

HUNSICKER, P. 1955, *Arm Strength at Selected Degrees of Elbow Flexion*. WADC Technical Report 54-548. United States Air Force, Project 7214.

JUUL-KRISTENSEN, B., FALLENTIN, N. and EKDAHL, C. 1997, Criteria for classification of posture in repetitive work by observation

methods: a review. *International Journal of Industrial Ergonomics*, 19, 397–411.

HEINSALMI, P. 1986, Method to measure working posture loads at working sites (OWAS). In E.N. Corlett, J. Wilson and I. Manenica (eds), *The Ergonomics of Working Postures. Models, Methods and Cases* (London: Taylor & Francis), 100–4.

KARHU, O., KANSI, P. and KOURINKA, I. 1977, Correcting working postures in industry: a practical method for analysis. *Applied Ergonomics*, 8, 199–201.

KARWOWSKI, W. 1992, Occupational biomechanics. In G. Salvendy (ed.), *Handbook of Industrial Engineering* (2nd ed.) (New York: Wiley), 1005–46.

KROEMER, K., KROEMER, H. and KROEMER-ELBERT 1996, *Ergonomics: How to Design for Ease and Efficiency* (Englewood Cliffs: Prentice Hall), 66–7.

KUMAR, S. 1991, Arm lift strength in workspace. *Applied Ergonomics*, 22, 317–28.

MARRAS, W.S. 1997, Biomechanics of the human body. In G. Salvendy (ed.), *Handbook of Human Factors and Ergonomics* (2nd ed.) (New York: Wiley), 233–67.

MITAL, A. and DAS, B. 1987, Human strength and occupational safety. *Clinical Biomechanics*, 2, 97–106.

MITAL, A. and FAARD, H. 1990, Effects of sitting and standing, reach distance and arm orientation on isokinetic pull strengths in the horizontal plane. *International Journal of Industrial Ergonomics*, 6, 241–8.

MITAL, A., KARWOWSKI, W., MAZOUZ, A.K. and ORSARH, E. 1986, Prediction of maximum acceptable weight of lift in the horizontal and vertical planes using simulated job dynamic strengths. *American Industrial Hygiene Association Journal*, 47, 288–92.

PINZKE, S. 1992, A computerized method of observation used to demonstrate injurious work operations. In M. Mattila and W. Karwowski (eds), *Computer Applications in Ergonomics, Safety and Health* (Amsterdam: Elsevier/North-Holland).

PHEASANT, S. 1986, *Body Space: Anthropometry, Ergonomics and Design* (London: Taylor & Francis).

PUTZ-ANDERSON, V. 1994, *Cumulative Trauma Disorders: A Manual for Musculosketetal Diseases of the Upper Limbs* (London: Taylor & Francis).

SENGUPTA, A.K. and DAS, B. 1998, A model of three-dimensional maximum reach envelope based on structural anthropometric measurements. In *Proceeding of the Annual International Occupational Ergonomics and Safety Conference*, Ypsilanti, Michigan, 226–9.

SENGUPTA, A.K. and DAS, B. 1999, The determination of maximum reach envelope for industrial workstation design. In *Proceedings of the Annual International Society for Occupational Ergonomics and Safety Conference*, Orlando, Florida, 59–64.

SNOOK, S.H. and CIRIELLO, V.M. 1991, The design of manual material handling tasks: revised tables of maximum acceptable weights and forces. *Ergonomics*, 34, 1197–213.

SQUIRES, P. 1956, The Shape of the Normal Work Area. Report no. 275 (New London: Navy Department, Medical Research Laboratory, CT).

TICHAUER, E.R. 1975, *Occupational Biomechanics: The Anatomical Base of Workplace of Workplace Design*. Monograph No. 51 (New York: Institute of Rehabilitation Medicine, New York University, Medical Center Rehabilitation).

WATERS, T.R. and PUTZ-ANDERSON, V. 1996, Manual material handling. In A. Bhattacharya and J.D. McGlothlin (eds), *Occupational Ergonomics: Theory and Applications* (New York: Marcel Dekker), 329–49.

Evaluation of Work Chairs

D. Colombini, E. Occhipinti, G. Molteni and A. Grieco
Research Unit "Ergonomics of Posture and Movement" EPM – Milano, I-20145 Milan, Italy

1. INTRODUCTION

Recently, changes have been introduced in the manufacture and marketing of work chairs, leading to an widespread use of the concept of "ergonomic quality" of chairs. Nevertheless, this subject once again poses a question: what are the circumstances and requirements based on which a work chair can be truly defined as ergonomic and can these features be identified?

It is well known that technical standards and regulations, however not dissimilar, are in force in different European countries. They define the wide range of manufacturing, dimensional and performance rules of work chairs. Many rules and regulations are still being discussed by European Community Standards Authorities (CEN). Therefore, the literature as well as the regulations have produced a series of criteria, though not as yet uniformly organized. The purpose of this chapter is to illustrate an organized approach for objectively evaluating the ergonomic features of a work chair.

2. BASIC ERGONOMIC REQUIREMENTS FOR WORK CHAIRS

All aspects should be taken into account when carrying out the ergonomic evaluation of work chairs to promote the psychophysical well-being of users. On the basis of this consideration, the following are the principal requirements that a work chair should meet to be defined as ergonomic:
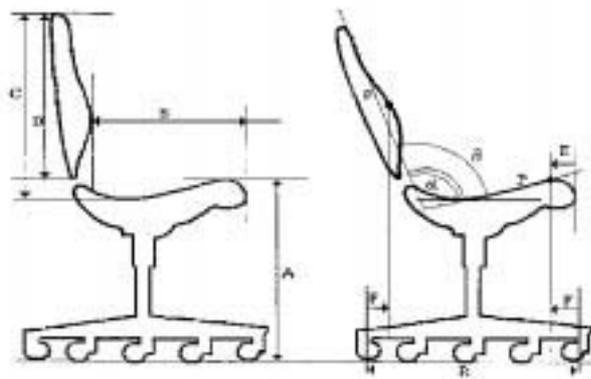
A : SEAT HEIGHT
B : SEAT DEPTH
C : BACKREST SEAT PLANE HEIGHT
D : BACKREST HEIGHT
E : DISTANCE FROM EDGE OF SEAT TO APPLICATION OF FORCE POINT (5 CM)
F : DIFFERENCE (IN PROJECTION) BETWEEN SEAT BASE AND UPPER
  SUPPORTING POINTS
P : SUPPORTING POINTS
R : SEAT BASE UNDER LESS FAVORABLE CONDITIONS
  : NET ANGLE BETWEEN BACKREST AND SEAT PLANE
  : BACKREST ANGLE

Figure 1. Lay-out of some of the measurements used for assessing safety, adjustability and comfort of the work chair (lateral view).

- Safety: the chair should never be the cause or the means of accidents.
- Adaptability: the chair and its components should have dimensions or be adjustable to meet the anthropometric needs of a wide range of users, and of at least 90% of the population.
- Comfort: the chair and its components (i.e. especially the seat and the backrest) should be upholstered, body contoured and reciprocally adjustable so as to meet the physiological needs and characteristics of many different "body shapes, curves and sizes."
- Practicality: the chair and its components should be easy to adjust by the user. The covering materials should favor hygienic conditions.
- Solidity: the chair, its components and adjustment controls should be sturdy, offering good wear and durability without changing performance over time.
- Suitability for the intended use: the chair should be appropriately designed for the intended type of and working environment: a standard chair that fits all types of jobs does not exist.

Each intrinsic factor described here requires the collection of many variables, both qualitative and quantitative, according to which specific criteria were set to determine whether or not a variable is acceptable, and if found acceptable, whether it is optimal or only sufficient.

In particular, reference was made to the criterion that 90% of the users' needs had to be met when evaluating the chair's dimensional and/or adjustment features. To this end, most appropriate, anthropometric parameters were established, and minimum and/or maximum values were set corresponding to the 5th and 95th percentile of the overall population (men and women) distribution. The anthropometric distributions referred to in this study were those of adult western populations as described by Pheasant (1986).

As far as the "shape" of the backrest is concerned, some data collected by the authors were used. These data were obtained by studying a sample of ~350 subjects (men and women) relative to the "shape" of the thoracic and lumbar spine, with the subject standing and sitting.

With regard to safety and performance, reference was made to the European Countries technical regulations in force, particularly German DIN 4551 and Italian UNI 7498.

2.1. Safety

When evaluating safety requirements the following characteristics and variables should be checked:

- Stability of the chair: the supporting base in projection should contain the surface area of the seat plane; in particular, the chair-to-ground projection of vertical force application points, placed 5 cm apart respectively should be contained within the extremity joining the two adjacent spokes. Moreover, the backrest-to-ground projection, with maximum retro-inclination, should fall within the area defined by the supporting base. While taking measurements, castors should be positioned in the most unfavorable position (Figure 1).
- Pressurized gas springs for adjusting the height of the seat plane and the slope of the backrest should be approved and tested by qualified standards authorities.

- Chair adjustment controls, especially if mechanical, should not be activated unintentionally.
- Chair components should be made of non-flammable material, especially the padding, covering materials or a combination of both.
- The chair should come in a choice of castors for different types of floors; it is better to equip chairs with anti-skid castors.
- Absence of sharp edges.
- Armrests should never pinch. In particular, they should not be open at the back.

2.2. Adaptability

The study of this requirement entails the analysis of a range of dimensional or adjustment variables of the chair components, mostly as a function of the variability of corresponding anthropometric parameters. These parameters include:

- Height of the seat plane (Figure 1): this corresponds to the anthropometric parameter known as "popliteal height." The height of the seat plane is measured from the ground to the highest point on the front edge of the chair when a weight or compression is placed on the seat plane (e.g. 50 kg). To meet fully the variability of the corresponding anthropometric parameter, this height should be adjustable between 35 cm (= 5th percentile in women) and 50 cm (= 95th percentile in men). It is surprising to note that current European regulations suggest a minimum height of 42 cm (UNI and DIN regulations): these values correspond to the mean value of the popliteal height of the general European adult population (men plus women). It is worth pointing out that part of the shorter population shall be expected to use a footrest. Therefore, bearing in mind the growth tendency of the popliteal height parameter in the population and the thickness of footwear (2–3 cm), much effort is still needed to lower the minimum height of the seat plane to at least 40 cm. The maximum height of the seat plane should be left unchanged (50 cm) or even increased to 53 cm or more (= 95th percentile in men wearing shoes).
- The usable width of the seat plane (Figure 2): this corresponds to the anthropometric parameter "hip breadth" in sitting position. In this case, aside from unrealistic adjustment mechanisms, the first step is to meet the needs of the widest subjects (= 95th percentile in women). Once this parameter is set, it becomes easier to meet the needs of other subjects. This parameter, however, becomes even more critical in the case of chairs with armrests. The suggested distance between the armrests should be equal to or greater than 49 cm.
- Depth of the seat plane (Figure 1): this corresponds to the anthropometric parameter "buttock–popliteal length." It is obtained by measuring the distance between the front edge of the seat plane and the most protruding point on the front of the backrest (i.e. lumbar support). This depth may vary in cases where the backrest is adjustable in depth or merely in slope; conversely it is fixed, with the chairs whose backrest and seat plane are joined into one body. With respect to this parameter, the most important issue is aimed at favoring the 5th percentile subjects of the buttock–popliteal length by considering measurements suitable for the proper placement of the thighs and at the same time supporting the lower back. This is particularly important in chairs without an adjustable depth, where the parameter involved should be < 41–42 cm.

- Backrest height (Figure 1): the backrest is intended to support the trunk. There are different types of backrests. Some chairs come with the backrest and seat plane all in one piece (single body). Other chairs have backrests not adjustable in height, and in other backrest adjustment in height is possible. In the best of cases, the backrest should at least support the trunk starting from the low back curvature (L5–S1) up to maximum thoracic kyphosis. Obviously, the correspondent values of the 95th percentile in men should be taken into account. An acceptable height is capable of supporting at least the entire low back segment. Generally, backrests (whether adjustable or not in height) that support only the low back are defined low. Backrests supporting the trunk up to maximum thoracic kyphosis are considered medium, while those which exceed this height are considered high. In the case of backrests with adjustable height, the backrest should be at least 32 cm high and its upper edge from the seat plane should adjusted to at least 48 cm from the seat plane. In fixed backrests the "upper edge-seat plane" distance should be at least 48 cm. Backrests with height < 32 cm (but > 22 cm) are acceptable if the measurements from the upper edge vary between 37 and 47 cm: support is only provided in the low back segment and, therefore, the chair is not advised for prolonged sitting positions.
- Backrest width (Figure 2): these measurements define minimum widths at the level of maximum lumbar lordosis and maximum thoracic kyphosis. To meet adequately the anthropometric parameters (values corresponding to the 95th percentile in men), these widths should be at least equal to 33 cm in the low back segment and 38 cm in the thoracic segment. For reference, lumbar width is measured by taking the maximum protruding point of lumbar support; thoracic width is measured ~19 cm above this point.
- Height, depth and width of the armrests: armrests are useful for supporting upper limbs while work is not being carried out, for getting into and out of the chair, especially people with functional motor disabilities, and also for keeping balance. On the contrary, armrests are not recommended for tasks involving the use of a high keyboard (especially mechanical), or when a fixed table is used under which the armrests cannot be inserted because the table is too low.

To determine the proper height of the armrest from the seat plane, it is worth recalling the anthropometric parameter "elbow–seat plane height" (obtained by measuring the distance between the elbow-to-ground height and the popliteal height, both taken in sitting position). It is better to refer to the low range of the parameter (small–medium, 5th to 50th percentile).

If the armrest is adjustable in height, its height from the seat plane should range between 15 and 23 cm (variations > 23 cm are possible); but if the armrest has a fixed height, it is better to choose a height measuring between 16 and 23 cm.

The recommended depth of the armrest is defined according to the anthropometric parameter known as "elbow-to-wrist length" in the 95th percentile of men. Since this parameter

measures ~30 cm and ideally at least two-thirds of the forearm should be properly supported, then the recommended depth for resting the arms should measure at least 20 cm or more, provided that the armrest is located in the right position for this function.

As for the width of the armrest, its dimension should be wide enough to allow for the forearms to be comfortably supported. Armrests with a width of 4 cm or more are recommended.

Table 1 (see end of article) gives a summary of the principal parameters dealt with in this study to assess adaptability. Each parameter was rated optimal, acceptable or unacceptable respectively.

2.3. Comfort

With respect to the other requirements, comfort is certainly the most difficult to assess from an analytical standpoint. The idea of comfort is easily influenced by the subjective evaluations of the users, and in literature this subject is still being widely discussed. The following reflections can shed some new light on the basic approach to the problem.

A work chair should always come equipped with a backrest to support the trunk. A backrest permits the relaxation of paravertebral muscles, and leads to less pressure overloading the intervertebral disks as compared with chairs without a backrest, even when these kinds of chairs come with a forward sloping seat plane.

A backrest also allows the trunk to alternate periods of support with periods of no support, thereby obtaining a variation in

A1 : SEAT WIDTH
A2 : ARMREST CLEARANCE
B : BACKREST WIDTH (LUMBAR LEVEL)
C : BACKREST WIDTH (THORACIC LEVEL)
d   : HORIZONTAL LUMBAR CONCAVITY (RADIUS)
e   : HORIZONTAL THORACIC CONCAVITY (RADIUS)

Figure 2. Lay-out of some of the measurements used for assessing adjustability and comfort of the work chair (frontal view).

pressure load (otherwise impossible to achieve without a backrest), which is essential to the good nourishing mechanism of intervertebral disks.

A forward sloping seat plane, which is useful only in a limited number of workplaces with specific needs (i.e. work carried out at a drawing table), may indeed contribute to maintaining the physiological curvature of the lumbar lordosis, but nonetheless it tends to overload the lower limbs, particularly the knees. In cases of extreme seat plane inclination, a knee block has compensated for this overloading, but this solution is the cause of another problem, i.e. menisci disorders.

Proper support of the lower back is achieved using a sufficiently reclinable backrest with respect of a correct profile of the lumbar lordosis. Both conditions are vital to reducing muscle and articular overloading of the spine.

Therefore, in terms of comfort, a clear choice was made in favor of chairs with backrests reclinable, autonomous from seat plane, suitably contoured with a lumbar support and ideally adjustable in height and with a slightly retro-inclined seat plane. The backrest and the seat plane should also be shaped according to the details described in this paper. It is now possible to give the analytical details of each aspect characterizing a chair's comfort.

- Inclination of the backrest (Figure 1): its adjustability is aimed to meet the needs of different people, both subjectively and task-related, bearing in mind the relationship with the overall layout of the workplace. Too much inclination tends to interfere with work tasks and it may be the cause of a man/chair imbalance with respect to the supporting base. Inclinations of < 90° with respect to the horizontal plane is ineffective for supporting the trunk. The possibility of varying the angle between the trunk and the thigh, obtained by adjusting the inclination of the backrest, is also useful during different work tasks. From this standpoint, a good synchronous balance mechanism of this opening depending on the inclination angle presents an advantage, but it is not a necessary condition. Hence, backrests equipped with an inclination ranging between 90 and 115° (with respect to the horizontal plane) are preferable (120° for high backrests). The backrest should be adjustable to the desired inclination for whatever position or range of inclinations (i.e. almost 5°). The net angle between the backrest and the seat plane should be less than these values depending on the slight retro-inclination of the seat plane or the presence of synchronous opening devices between the backrest and seat plane. As for comfort, the chairs with a backrest, albeit reclinable, that is one piece with the seat plane (single body) are less effective. Likewise, in the absence of blocking devices for desired inclinations, backrests with inclinations obtained only by putting pressure on them (even with adjustable resistance) are less advisable.
- Contours of the seat plane (Figure 3): to avoid compressing the neurovascular structures at the popliteal level, the front edge of the seat should be rounded and possibly made of non-rigid material. The correspondent curvature should have a radius contained within 4–12 cm, and the height of the rounded area should measure ~4 cm. From front to back, the seat plane should be sloped slightly backwards in order to favor the placement of the thighs and stop the buttocks
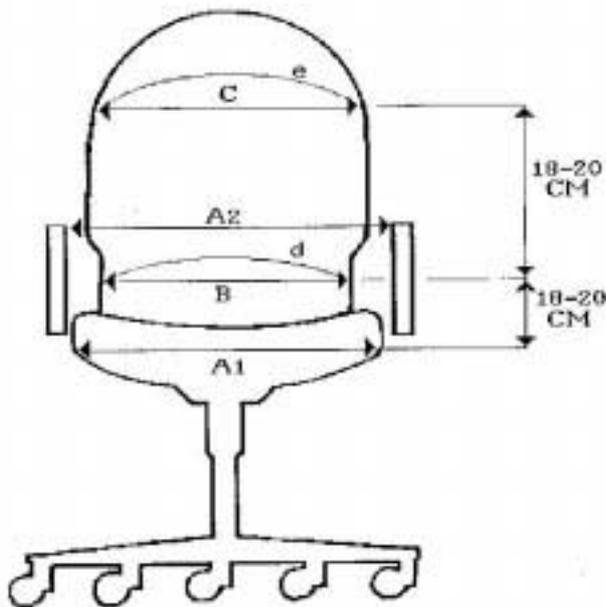
from sliding forward. A slope ranging between 3 and 10° respect to the horizontal plane with the seat in normal position is recommended. These values may be slightly increased if synchronous opening devices of the seat plane — backrest angle are present. Towards the back of the seat plane, a concavity should be designed for housing the buttocks. The center of this concavity, from back to front, should be placed within 10 cm of the most protruding point of the lumbar support with the backrest in vertical position.

- Lumbar support (Figure 3): with regard to this basic element, the height of maximum protrusion from the seat plane, the depth and the length of its extension along the backrest should be taken into consideration. The height of maximum protrusion can be defined by considering the anthropometric parameter "lumbar lordosis–seat plane height." In optimum cases, this height should be adjustable between 17 and 28 cm. In cases of chairs where the backrest is not adjustable in height, maximum protrusion is fixed between 20 and 24 cm from the seat plane. The depth of the lumbar support at maximum protrusion should be set according to data in literature and original data on the anthropometry of low back curvature. To meet anthropometric requirements and different postural choices, depth should be adjustable between 2 and 5 cm. If it is not adjustable, then the depth should be set between 2 and 4 cm. The length of lumbar support extension along the low-back padded area of the backrest is determined by using such anthropometric parameters as: "distance between 1° sacral vertebra and 1° lumbar vertebra" and "distance between 1° sacral and maximum thoracic kyphosis." Generally speaking, extension ranging between 20 and 30 cm are suggested for medium-sized backrests and 13–16 cm for low ones. As for vertical contour of the backrest padding, it is important to notice that in the case of medium or tall backrests the curvature between the 1° sacral vertebra and the maximum thoracic kyphosis is not symmetric with respect to maximum lumbar lordosis, but rather is somewhat slight on top and more accentuated at the bottom.

- Padding: the aim of a good seat plane and backrest conformation, as well as eventually the contours of the padding, is to establish the right interface between the user's body and the seat itself, avoiding the undesirable effects of compressing any protruding bones and providing an appropriate distribution of body pressure. For most office workplaces, chairs should be manufactured using semi-rigid upholstery (with a deformation due to the weight of a person < 2.0–2.5 cm), sufficiently thick (4–5 cm) and preferably applied onto a well-shaped body of the chair. In industrial environments, where chairs often come in contact with dust and liquids, it would be inappropriate to use padded and/or upholstered chairs. Therefore, other kinds of materials are preferred (i.e. wood). In this case, a good profiling of the seat plane and backrest is highly recommended.

- Covering materials: the surface that comes in contact with the user should not create any thermal undesirable sensation. In particular, these materials should allow the skin to breathe but avoid creating cold sensations. Therefore, the suggested covering materials for seats and backrests should be made of porous material (placed over the padding) or in wood,

whereas plastic, rubber and metal chair coverings are not recommended. The armrests should be covered with plastic or padded material to avoid creating undesirable thermal sensations or hitting the nerve structures of the elbow on rigid parts of the chair. The covering material, particularly of the seat plane, should be rough enough to counter the buttock's tendency to slip forward.

Table 2 (see end of article) gives a summary of the principal parameters dealt with in this study to assess comfort. Each parameter was rated optimum, acceptable or unacceptable respectively.

2.4. Practicality

Adjusting the various chair components should be made as easy as possible for the user. If adjustments have to be made using controls that are difficult to maneuver or hard to reach, ultimately they will never be used.

The adjustment controls, particularly of the seat height and the different functions of the backrest, should be easy to reach from a seated or semi-seated position, without the need for use of force or additional equipment. Any knobs or handles should not be easily removable, and, if necessary, they should be fitted with guide stops.

The covering material of the chair should also be washable in order to guarantee the minimum hygienic requirements. This aspect is particularly important in industrial or in very dirty environments, or in cases where the chair is intended for multiple users. In these cases, the chair can be covered in fabric protected with easily removable covers or made of wood.

2.5. Solidity

Resistance to wear of the chair and its components is a prerequisite not only from a marketing standpoint, but also for ensuring the durability of ergonomic performance.

The sturdiness and resistance to wear of the seat and its components can be evaluated using the specific tests and interpretation criteria developed by National Standards Authorities, with validity in each of the countries respectively. Therefore, the results of such tests are important for making an overall ergonomic evaluation of a chair. The type of warranty and customer assistance provided by the manufacturer or distribution organization is another important aspect of this evaluation.

2.6. Suitability for the Workplace and Intended Use

As mentioned earlier, every workplace has its peculiarities and special needs from an ergonomic standpoint.

The chair is a separate working instrument and, therefore, it can have different characteristics and features that make it suitable (or unsuitable) for certain job settings, work environments or potential users. In other words, on the basis of a few characteristics, it is possible to define the general areas of potential chair use.

The "intrinsic" characteristics orientating the chair towards a specific job destination were examined in this study; then a summary is presented of all the necessary characteristics of chairs intended for wide areas of utilization.

- Castors: they should be buffed according to the type of floor
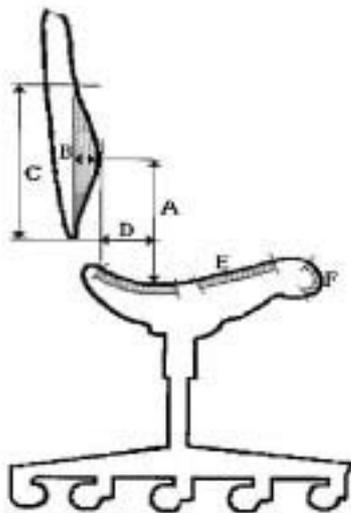
```
A : LUMBAR SUPPORT HEIGHT (FROM SEAT PLANE)
B : DEPTH  OF LUMBAR SUPPORT
C : EXTENTION OF LUMBAR SUPPORT ALONG BACKREST
D : SPACE FOR BUTTOCKS (FROM BACKREST)
E : SEAT PLANE AREA FOR MEASURING INCLINATION
       VS. HORIZONTAL
F : ROUNDED FRONT EDGE
```

Figure 3. Lay-out of some of the measurements used for assessing comfort of the work chair (lateral view).

used (smooth, carpeting, etc.). Chairs with castors for carpeting should not be used on smooth surfaces.

- Armrests: sometimes compatibility problems arise between the height (lower edge) of the worktable and the armrests.
- Covering materials and padding: in chairs intended for use in crowded areas, or areas with flames and sparks, both materials should be flameproof. In workplaces involving prolonged seated positions, a chair with semi-rigid padding is preferred. In working environments where the seat is easily dirtied, washable chair covers or upholstery should be used.
- Characteristics of the backrest: low, single body and non-adjustable backrests are not recommended for jobs involving prolonged seated positions. For these kinds of jobs, medium-to-high backrests that are separate from the seat plane are

suggested; backrest should be adjustable almost in inclination and fixable in the chosen position.

With regard to VDT workstations, it is worth recalling that EEC Directive no. 270/90 establishes that seat plane should be adjustable in height and backrests should be adjustable in inclination and in height.

For jobs involving prolonged seated positions, it is worth stressing that medium-to-high backrests with both kinds of adjustability and proper backrest profiling, especially in the low back area, are preferable. Chairs with low backrests should be restricted to jobs that do not entail prolonged sitting positions, provided they come equipped with the same two adjustments for the backrest.

REFERENCES

OCCHIPINTI, E., COLOMBINI, D., MOLTENI, G. and GRIECO, A., 1993, Criteria for the ergonomic evaluation of work chairs. *Medical Law*, 84, 274–85.

PHEASANT, S., 1986, *Bodyspace* (London: Taylor & Francis).

Table 1. Main characteristics for chair ADJUSTABILITY.

OPTIMAL (cm)	ACCEPTABLE (cm)	UNACCEPTABLE (cm)	
SEAT HEIGHT	39 - 52	FROM 42 TO MORE THAN 52	UPPER HEIGHT < 46 LOWER HEIGHT > 45
SEAT DEPTH	39 - 41 SMALL SUBJ. 42 - 55 TALL SUBJ.	39 - 42 FOR ALL SUBJECTS	FIXED < 32 OR > 45
SEAT WIDTH WITHOUT ARMRESTS	> 47	37 - 46	< 37
SEAT WIDTH + ARMRESTS	> 49	44 - 48	< 43
BACKREST HEIGHT	> 32	23 - 31 (SHORT TASKS)	< 22
ADJUSTABLEHEIGHT (LOW BACKRESTS) 48 FROM THE SEAT PLANE	MAXIMUM HEIGHT > FROM THE SEAT PLANE	MAXIMUM HEIGHT 38-47 FROM THE SEAT PLANE	MAXIMUM HEIGHT < 37
BACKREST WIDTH (LUMBAR REGION)	> 33	29 - 32	< 28
ARMREST: HEIGHT: DEPTH: WIDTH:	16 - 23 25 ±5 > 4	24 - 25 16 - 19 2.5 - 3.5	> 26 <1 5 > 31 < 15 < 2.5

Table 2. Main characteristics for chair COMFORT.

	OPTIMAL	ACCEPTABLE	UNACCEPTABLE
BUTTOCK MAX CONCAVITY (FROM BACKREST)	<10 CM	>10 CM	NOT PADDED FLAT SEAT PLANE
ROUNDED FRONT EDGE (RADIUS)	4–12 CM	< 3 CM	NONE
SEAT PLANE ANGLE (DEGREES)	3–10°	0–2° 11–14	FIXED FORWARD OR FIXED BACKWARD> 15°
LUMBAR SUPPORT *HEIGHT FROM SEAT PLANE: *DEPTH	17–28(Adj.)cm 2–5cm	20–24(fix.)cm >2 cm	NONE
BACKREST ANGLE *NET ANGLE: *ANGLE: VS HORIZ.	90–110° 90–120°	90–110° 90–105°	FIXED < 90° or > 110°
HORIZONTAL PROFILE (RADIUS)	40–80 CM	30–40 CM > 80 CM	NOT UPHOLSTERED FLAT BACKREST
PADDING	SEMI-RIGID		
COVERING MATERIALS	NON-SLIP AND THAT ALLOW THE BODY TO BREATHE		
ARMREST UPHOLSTERY	PLASTIC OR PADDED		

Facility and Workspace Layout Problems in Ergonomic Design

J. Grobelny† and W. Karwowski‡

†Institute of Industrial Engineering & Management, Technical University of Wroclaw, Wybrzeze Wyspianskiego 27, 50-370 Wroclaw, Poland

‡Center for Industrial Ergonomics, Lutz Hall, Room 445, University of Louisville, Louisville, KN 40292, USA

1. INTRODUCTION

Ergonomics design of the workspace and that of many workstations requires making decisions regarding placement (layout) of a great number of objects (machinery, devices, tools) utilized by workers. Such a layout should allow for effective functioning of the human–machine systems. The placement of objects (layout design) is, therefore, an important ergonomic design problem for simple and complex systems and structures. Ergonomic layouts should be based on optimalization of specific ergonomic criteria. For this purpose, one must formulate measurable functions, i.e. the criteria appropriate to a given design situation. Such functions must allow one to distinguish between the better and worse projects. The formal layout algorithms and optimization methods in many cases allow to find solutions which are the best or close to it.

The essence of the problem can be illustrated based on a simple example. Let us assume that in an office building there is a need to deliver different types of documents among three different types of organizational cells or company units (ABC). Daily, on average, a messenger delivering paper documents needs to travel ten times between A and B (either way), five times between B and C, and 20 times between A and C. Figure 1 illustrates a layout of the corridor linking the three rooms (offices 1–3) containing cells ABC. Assuming the first layout from Figure 1 (layout A), during an entire working day one needs to travel 165 m (10 _ 3 + 5 _ 3 + 20 _ 6). Assuming layout B, the same activity requires travel over 120 m (10 _ 3 + 5 _ 6 + 20 _ 3). It can be seen that in the second case, the messenger can perform the same task with less effort.

The above example of office work illustrates the problem of document flow between different units of the organization. In a similar way, one can analyze a single human–machine system, even though the physical layout related to the required travel is not the only way to evaluate the quality of placement of controls and displays in such a system.

The classical example of the ergonomics problem of space-related layout is discussed in the book by McCormick (1976), that refers to design problem of an aircraft cockpit. One of the tasks in this a design is placement of the controls, which must be used blindly. In such a case, a reasonable approach would be to use a layout that would guarantee maximum reliability of reaching for and grasping the correct controls. The theoretical basis for such a layout is given by Fitts (1947), who studied the accuracy of blind motions, which can be quantified, depending upon location of the target with respect to the human operator, by
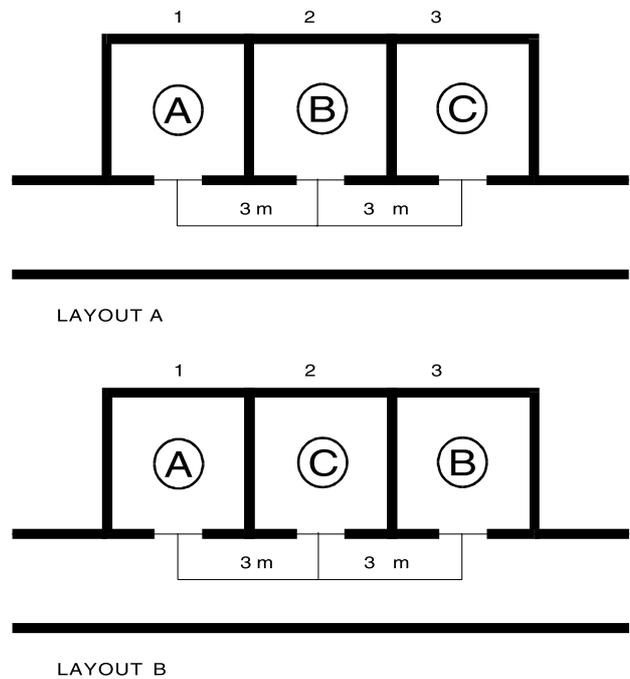
Figure 1.

considering the mean error in reaching the target. Having the information about frequency of use for the particular controls, one can calculate the overall mean error of reaching a target for a given layout. This can be done by summing over each control device the product of frequency of use and the mean error for its placement.

Traditionally, one can use many different concepts and criteria as the basis for evaluation of the layout of machines and evices in the human–work (human–machine) systems. For example, Bonney and Williams (1977), discussed the following layout criteria: (1) the type of user population, (2) comfort of use, (3) safety, (4) aesthetics, (5) closeness of devices, (6) easing their utilization, (7) appropriate distance (separation) between the devices to reduce potential error, (8) division of work between upper and lower extremities, (9) anthropometric dimensions and (10) functional relationships between devices. Proctor and van Zandt (1994), focus their attention main on the functional grouping of devices with respect to frequency and order of their use. Consideration of such issues by the designer working in a traditional way is often based on intuition, and can be improved with knowledge and experience. The attempts to formalize and use objectives methods for decision-making in this case, have also been undertaken in the past (McCormick 1976).

Owing to the very nature of many of the discussed aspects of design, it is often not possible to incorporate all of them into mathematical models, which are based on measurable quantities. In many cases, the formal models can only be used as tools that aid the design process. There are, however, in the ergonomic literature many examples where formal models are sufficiently robust and can generate final layout decisions with acceptable quality.

2. ERGONOMIC MODELS FOR LAYOUT DESIGN

The informal rules for placement of devices, tools, or machines in the workspace have been known in the ergonomics field for a

number of years. The most often cited rules recommend utilization of the following design criteria (McCormick 1976):

- Importance: the devices that are most important from the design objective point of view should be located in most important places (that is, those which are most convenient from their use) point of view.
- Frequency of use: the most frequently used elements/devices should be located in the most convenient places with respect to the ease of use.
- Order of use: the devices that are used one after another should be located close to each other.
- Functional use: the devices related to the same design function of the system should be grouped into blocks.

It should be noted that the specific design criteria could lead to contradictory solutions. For example, an important element (due to its safety connotation) can be used very infrequently. One can also imagine that there would be other set of rules applicable for some specific cases. It is difficult, however, to determine the hierarchy of different criteria, as such an attempt would lead to case-by-case (specific) classification. McCormick (1976) suggested that the design criteria A and B, discussed above, are particularly useful for layout of complex control panels (devices) in the workspace, while the criteria C and D are very useful for design of single control panels. This is because criteria A and B relate to the interrelationship between single devices and specific space locations, that is, the interrelationship between a specific point in space and the human operator. Criteria C and D are only related to mutual space relationships between different devices.

Wierwille (1980) developed formal models based on designating A and B as the criteria of the first-order (the relations measured by these criteria deal with single devices). Criteria C and D were classified as the second-order criteria, describing the relationships between pairs of devices. Given the above rules, the general problem of ergonomics placement of different elements (tools, devices, machines) at the workstation (workspace layout) can be formalized as follows: place several devices in the work space in such a way as to maximize the fulfillment of design criteria A, B, C and D.

The formal model of the above problem requires definition of an appropriate matrix that can be used to measure the degree to which criteria A–D are satisfied. This, in turn, leads to the need for defining a method for quantitative measurement for each of these criteria. Such a model is needed to distinguish between different design solutions.

A typical approach to satisfying criteria of the first-order is based on construction of the simple multipoint scales, which allow one to evaluate the location of various devices. Evaluation of the importance or frequency of use can be formulated by experts based on the knowledge about the tasks to be conducted on a given workstation. Evaluation of the quality of location for a particular workspace should be based on the ergonomics knowledge, and is often dependent upon the type of analyzed elements/system.

A given solution can be evaluated, for example, using:

$$Q_1 = \hat{A}_i W_i * I_{p(i)}, \qquad (1)$$

where W_i is importance of the device (i) in view of criterion A, or the frequency of use of the device (i) in view of criterion B;

$I_{p(i)}$ is importance (quality) for placement of the element (i) from the ergonomics point of view; and $p_{(i)}$ is an index for placement of the ith element.

If scale W is ordered in decreasing order (that is, the devices which are more important have higher values), and scale I is ordered in the increasing order, then a better design solution will have a smaller Q_1.

The criteria of second-order are most frequently being modeled using a sum of the products of indices of the length and distance:

$$Q_2 = \hat{A}_i \hat{A}_{j>i} L_{ij} * D_{p(i)p(j)}, \qquad (2)$$

where L_{ij} is the degree to which the devices (objects) i and j are linked together; and D_{pipj} is a distance between points and $p_{(i)}$ and $p_{(j)}$ in which the devices (i) and (j) have been placed.

According to criterion C, L illustrates how frequently the use of device (I) takes place before use of device (j) (or vice versa). For criterion D, the index L is the degree of functional linkage (with respect to the performed tasks) between the devices (i) and (j).

The double summation in the above equation means that in the evaluation process, one considers all pairs of devices: $j > i$, in the second sum is restricted to consideration of any given pair only one time.

L_{ij} can be put into a matrix, in which at the cross section of any row (i) and column (j) one can place the index of linkage (connections) between the devices (i) and (j). In such a case, typically one would fill out the matrix only above the main diagonal, as only these values are being considered by equation 2. For example, based on Figure 1, this can be shown:

Cell	A	B	C	
A	–	10	20	Linkage (L)
B		–	5	

In the above case, because the possible space for placement (layout) is restricted to specific rooms, one can represent the distances between such places:

Room	1	2	3	
1	–	3	6	Distance (m)
2		–	3	

Because evaluation according to equation 2 specifies, in general terms, a certain (important) distance, then it is natural that the linkage between each pair of devices will be derived in such a way so that the higher level of linkage will have the higher numerical value. In this case, the better layout solutions will have the lower overall values according to equation 2. Of course, the above evaluations are relevant for the situations that are similar to those presented in Figure 1, where equation 2 converts into simple length of travel that must be done by the messenger.

3. LINGUISTIC APPROACH

The equations 1 and 2 are not the only possible models of criteria for workspace layout design. For example, Grobelny *et al.* (1995) and Karwowski *et al.* (1999) have proposed an application of the fuzzy sets and systems theory for deriving a matrix for layout design. Such approaches have particular significance in

ergonomics analysis because they allow a consideration of the modeling process the element of imprecision, which often characterizes description of the human–artifact systems.

Linguistic problem description is more natural, especially when used to express the level of interrelationships or importance, as opposed to using the artificial scaling schemes. The layout design criteria of the first and second-order can be expressed using the language of logic in a way similar to the natural language.

For example, the design rule A can be restated:

IF Importance_level (of a given facility) is *GREAT*
THEN Location_place (of a given facility is *PROMINENT*. (3)

The rule C can be expressed as:

IF facilities pair *ij* is *USED IN SEQUENCE*   (4)
THEN *ij* are *ADJACENT* in the layout.

The underlined and capitalized words in the above equations are specific realizations of the linguistic variables, which values are words of natural language. In equation 3, such variables are the levels of importance and prominence of a location in the workplace. Equation 4 utilizes variables that can be named as the "sequential use" and "adjacency." Such equations express the idea of appropriate rules in a somewhat more formalized way, using the specific meaning. These are descriptions of the desired state for the specific model, which one can use to compare different design solutions. Utilization of equations 3 and 4 must be based on the specific evaluation mechanisms so that given rules 1 and 2, one could compare any pair of solutions, and choose the better one.

Because the above rules are logical expressions, the most natural way of evaluating then can be based on application of logical rules. The idea of linguistic patterns (Grobelny *et al.* 1995), is based on calculating the degree of satisfying such patterns (shown in equations 3 and 4), as defined by the specific solutions. This can be done by using one of generalizations for the classical implication table, which defines the truth of the sentence "IF a THEN b".

Depending on the truth (Tr) of its elements (a and b), one of such generalizations is the equation proposed by Lukasiewicz:

Tr(IF a THEN b) = min (1, 1 − Tr(a) + Tr(b)),   (5)

where (Tr) is the value of truth of the given expression (…).

It can be noted that equation 5 works in a way of a classical table of implication for truth, meaning that if Tr(a) and Tr(b) are 0 and/or 1, then the expressions are either *true* or *false*. The level of truth equals 0 (false) when a is true and b is false. In reality, the situations are more close to gradation of the value of truth. In such a case, equation 5 allows to define the level of truth of a given implication based on the level of truth for the parts a and b.

The level of truth can be defined on a scale from 0 to 1. A convenient mathematical apparatus that provides formal tools for evaluation of the level of truth is the theory of fuzzy sets (Zadeh 1968). The presented approach can also be used in a less formal way by utilizing subjective values (judgements) of truth based on the expert opinions.

For example, let us consider the situation presented in Figure

1. First one can note that in the problem of assigning specific rooms, one cannot directly apply the first-order evaluation rule because nothing is known about the value of the specific spaces for placement, or about the weight of the tasks. On the other hand, the problem of travel distance between the rooms can be evaluated by equation 4. Specifically, rule 4 can be modified slightly in order so that the expressions used are unambiguous. Given the above, one can propose the following expression:

IF the frequency of travel between the two cells, *i* and j, is BIG THEN the rooms which contain *i* and *j* are ADJACENT to each other   (6)

Cell	A	B	C	Degree of satisfaction for
A	–	0	1	the left side of rule (6)
B		–	0	

Cell	1	2	3	Degree of satisfaction for
1	–	1	0	the right side of rule (6)
2		–	1	

Let us also assume that we have the orthodox experts who only recognize a two-valued logic (true, false). If one can find data from matrices presented above, then using equation 6, and based on the knowledge and experience, one can, for example, say that frequency of travel equal to 20 is being BIG, while other frequencies are not BIG. Furthermore, the rooms that are adjacent to each other would be thought of as such rooms that have common walls between them (rooms 1 and 2; and 2 and 3). The above judgments can be set in matrices in the analogous ways as before.

Based on the above data (and Figure 1), one can perform the following evaluation for each pair of the organizational cells (units). The level of satisfying rule according to equation 5.

Cells	Layout (a)	Layout (b)
AB	min (1, 1 − 0 + 1) = 1	min (1, 1 − 0 + 1) = 1
AC	min (1, 1 − 1 + 0) = 0	min (1, 1 − 1 + 1) = 1
BC	min (1, 1 − 0 + 1) = 1	min (1, 1 − 0 + 1) = 1

As can be seen above, in the first layout solution (a), the pattern is not satisfied by a pair AC and its location. This assessment is logical also from the practical point of view, as the frequency of travel between cells A and C is BIG (according to the expert opinion), while the rooms containing cells A and C are not adjacent to each other.

In the second case, all pairs are satisfying the required pattern. Because the assessment according to the rule by Lukasiewicz expresses the level of truth, and 1.0 is the full truth, one cannot perform addition of such assessments for the specific layouts (as it is done in equation 2). Rather, the middle values of truth should be used. Therefore, such a layout satisfies the pattern according to equation 6 to the degree of 0.66, while layout B to the degree of 1 (fully).

4. CLASSIFICATION OF LAYOUT AND OPTIMIZATION ALGORITHMS

In the discussion above we have presented some methods for constructing the evaluation matrices for layout solutions. Such

matrices allow one to quantify, compare and select the best solution from a given set of solutions. A separate problem in this area of research is the rationale for selection of good solutions under different design situations. Because the above methods are very much dependent upon some characteristics of the design projects, one needs to develop a taxonomy of tasks, and then derive specific evaluation methods according to types of tasks to be solved.

The criteria for placement of objects and devices (layout) was discussed before. The first-order criteria indicate the way to place single objects in the workspace. The criteria of the second-order evaluate the mutual relationships between a pair of objects. In the example considered above, it was logical to utilize only the second-order design criteria. Such criteria are very useful in design of control panels, while the criteria of first-order are not applicable to this problem.

On the other hand, many design projects can only use the criteria of the first-order. Placement of the ready-to-use control panels in the workspace can be evaluated based on criteria on the first-order, especially when each panel serves to control different functions. There are also design situations, especially occurring in complex workstations, where one has to simultaneously seek a compromise between the rules specified by criteria of both orders. Therefore, design situations for placement of objects and devices can be classified as tasks of type 1, tasks of type 2, and the combined tasks (type 3), in which both evaluation criteria must be considered simultaneously.

The approaches to layout design problems (optimization of layout) in addition to the order of criteria also depend upon the way in which allowable places of localization (space) are restricted. An example considered above represents such restricted tasks, were space for placement of cells are predetermined. Therefore, only one cell can be placed in any given room. However, while designing a new control panel, designers may face a decision regarding placement of single elements on the panel, as well as consideration of the geometrical form and dimensions of the panel itself. Such a problem can be denoted as an "open problem".

Based on above discussion of the specific characteristics of layout design tasks, and the related design criteria, one can summarize the problem space as shown in Table 1. The above classification allows to assign to different types of tasks the specific methods and algorithms for their solutions, which can lead to either optimal or suboptimal layouts.

The restricted tasks of type 1 can be used to model common design projects, where the evaluation hierarchy for the objects to be placed, as well as an assessment of the usefulness of the predefined places are known. A typical example for this is placement of elements in the field of vision. Even though the

Table 1. Simplified classsification for layout design with respect to type of criteria and accessibility of space

	Type of design criteria		
	1st Order	2nd Order	Both simultaneously
Allowed places (pre-determined)	Restricted problems 1st Order	Restricted problems 2nd Order	Restricted problems-complex
Allowed places (any location)	Open problems 1st Order	Open problems 2nd Order	Open problems-complex

accessible places are not physically defined, one could determine the assessment values based on the knowledge of human field of vision. Therefore, the whole accessible space can be appropriately evaluated by, for example, assigning relative scores for visual acuity. Such a situation is very simple from the point of formalized optimization. The obvious solution would be to assign (consecutively) the most important objects to the most accessible places. This would lead to an optimal solution as discussed by Grobelny (1988).

McCormick (1976) also discussed the practical utilization of such an approach. Placement of control devices in an airplane cockpit was optimized by ordering devices with respect to the frequency of their use, while evaluation of the accessible space was done based on the research results of Fitts, regarding the precision of human motions under blind conditions. The presented solution called for consecutive placement of the most frequently used elements in the accessible areas, characterized by the highest precision of blind motions (based on a kinesthetic sense).

5. DESIGN ACCORDING TO CRITERIA OF THE SECOND ORDER

The placement of elements in the workstation based on criteria of the first-order are, in general, open, and can be solved in a similar way as the restricted tasks of the same type. That means that the basis for evaluation is the psychophysical characteristics of the human. If the accessible places are not determined physically, then the designer has much greater freedom to act. However, some areas with specific valuation can also be determined by objective characteristics of the human being, while the described method can give the optimal solution from the formal point of view.

Much more complex situation has to do with the tasks described by criteria of second-order. The classical optimization approaches allow seeking solutions only for tasks that are restricted (i.e. those with precisely specified placement spaces/location). These types of tasks can be defined as follows:

Place N devices in M accessible places ($M \geq N$), so that the value of the function of second-order is minimized (or the mean truth value calculated according to equation 4 is maximized).

The optimal solutions for the above can be reached by reviewing all possible layouts, and using the method of division and restriction as proposed by Gavett and Plytera (1966). Such an approach allows solving tasks with up to 20 elements. Therefore, for tasks with greater number of elements, it is necessary to use the heuristic algorithms.

Because classical algorithms require precise determination of the potential placement, the layout tasks which are open must be appropriately translated into the tasks which are restricted. Most often this is done through introduction of the modular nets, which restrict the accessible space given the specific patterns of the net. Sears (1993) proposed such an approach to optimization of the dialog windows for human–computer interface design purposes. The basis for optimization criterion was a road traveled by the cursor that is needed to accomplish basic tasks conducted using a given dialog window. The authors showed, based on empirical studies, that optimal solutions significantly improved the effectiveness of work with windows.

Drezner (1988) was first to propose a methodology for solving

the second-order layout tasks that are open on the surface, without any restrictions. A concept of the scattered plots was used for this purpose. An efficient heuristic algorithm is based on characteristics of the Eigen vectors and Eigen values of the relevant matrices, based on the following formula:

$$Q_2 = Â \, c_{ij} d_{ij}^{\,2} / Â d_{ij}^{\,2}, \quad (7)$$

where C is frequency of interactions (L) defined by equation 2; and D is distance between a given pair of objects.

The process of minimization is possible by analytical method. However, the objects must be placed on a straight line. The coordinates for such a solution (x and y) are the consecutive Eigen vector elements related to the second lowest value of matrix S, for which:

$$s_{ij} = -c_{ij} \text{ for } i \, \pi \, i, \text{ and } s_{ii} = Â_j c_{ij} \text{ for all } (i).$$

Good solutions (on the surface space) can be derived by taking as the coordinates of y, the elements of Eigen vector related to the third smallest value of an Eigen value of matrix S. This is because such coordinates represent the best solution on the orthogonal line with respect to the coordinate x.

Although theory of Eigen vectors and Eigen matrices is quite complicated, the algorithms which generate the solutions are very simple, and can be used without knowledge of such complex ideas. Given the relationship C_{ij} one needs to do the following:
- Develop matrix S.
- Calculate Eigen values and Eigen vectors.
- Select two vectors connected with the second and third smallest Eigen values, and take these as sets of consecutive for x and y solutions on the surface, respectively.

The second point of the above process must be realized using the computer program, for example, using the Mathematica.

Figure 2 illustrates the solution reached for an example where values of C were put in the form of 0,1, meaning that the elements are related to each other in the following way: the first place shows the number of the device, while the consecutive places show the number of devices related to the first device.

The scattered plot shows that, in most cases, the stars corresponding to the elements that are related to each other, are also located closely one to another. The flexible computerized
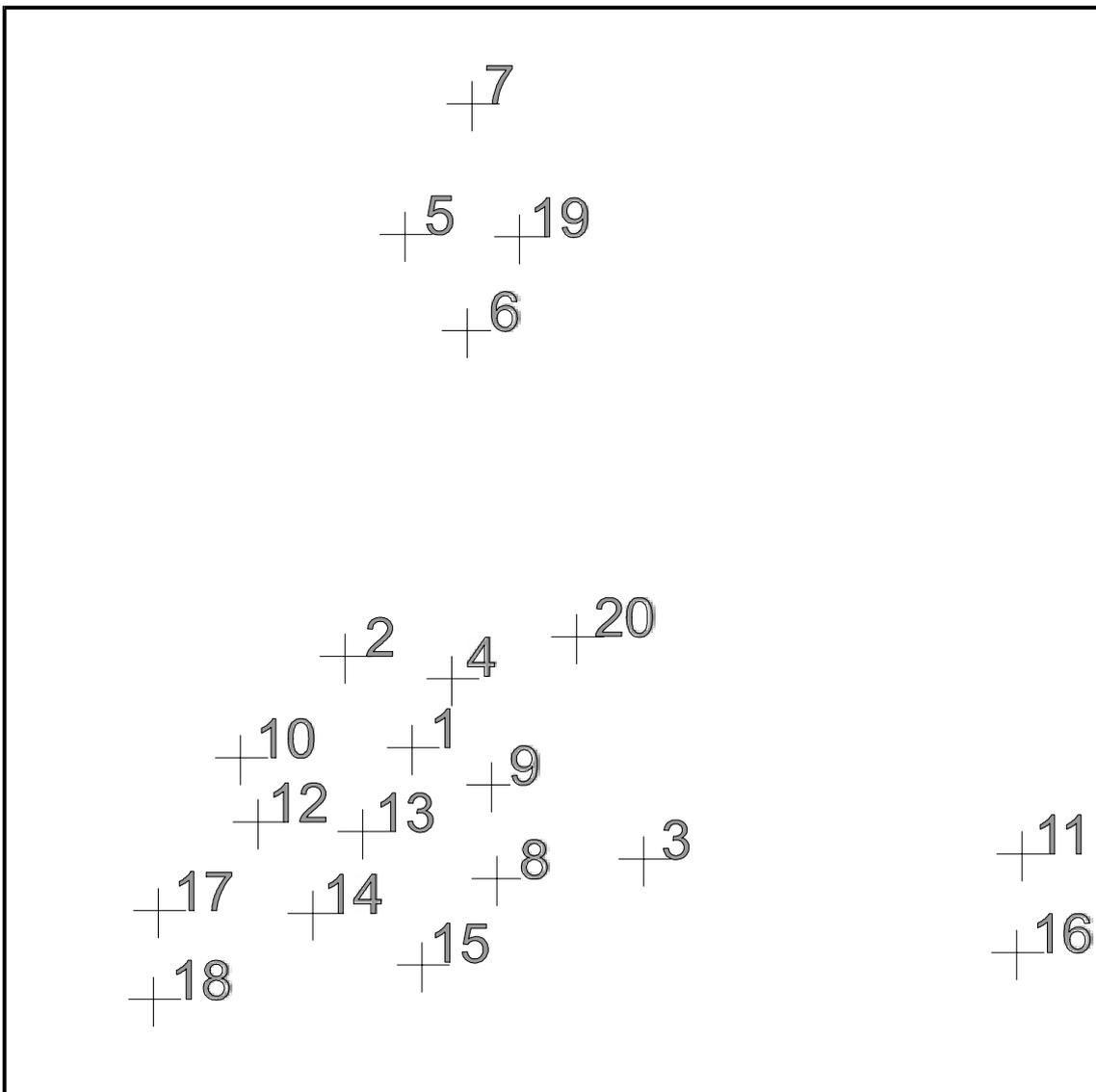

Figure 2.

method for generating the scattered plots using differential parameters was recently discussed by Grobelny (1999).

The scattered plots are the universal form for computer-aiding of layout design, especially when utilized as an aid to the designer, but one that does not replace his/her knowledge and intuition.

6. LAYOUT DESIGN FOR COMPLEX TASKS

The most complex problems in layout design are those that deal with situations requiring the use of both design criteria simultaneously, that is an assessment of design quality for complex tasks. The classical algorithm for solving these types of problems in industrial engineering is an algorithm proposed by Hillier and Connors (1966). This algorithm can only handle restricted tasks, where the number of places of localization is equal to the number of devices to be placed. The criteria of the first-order were interpreted by the authors as an assessment of the cost of placing machines/objects in given spaces. The criterion of second-order was defined as a matrix of transportation costs that are required for such placement. The algorithm progresses in the direction of minimization of the sum for both criteria.

Utilization of the above algorithm for ergonomics design purposes requires changing the task to one that is restricted, as well as developing a matrix for "installation cost", by determining the value of function 1 for each element at each place. It should be remembered that the assessment scales must be appropriately defined as the algorithm seeks a compromise that minimizes the sum of values for both criteria 1 and 2.

So far, no algorithms were proposed that would solve the combined complex tasks that are "open", in a way that is similar to the approach by Drezner, by allowing one to solve the task of second-order. Another problem that occurs in the practice of ergonomics design is the need for a consideration of the form and size of objects which are to compose the workstation. The discussion presented above simplifies classification of this problem. This is because the classical algorithmic approach described here is to a large degree independent of geometrical characteristics of the objects being placed. However, introduction of the modular nets brings the idea for defining dimensions and shapes of objects using the modular segments of the net. Such an approach was first used in industrial engineering (CORELAP). In the human–computer interaction area, the algorithm proposed by Sears (1983) utilizes modules for dividing places in a dialog window, and also for simulating functional contours of objects which are being placed in the windows.

7. CONCLUSIONS

The obvious question arises as to how such theoretical ergonomics solutions relate to practice. That is, whether the solutions developed by using methods discussed above allows one to improve the functioning of the real human–machine systems. Some of the solutions are obvious and do not require additional verification. Examples of those are related to problems with measurable, physical and economical functions of both orders. For example, the solution to the messenger travel problem considered in the beginning of this article shows that foot traffic in the office designed according to layout B will be smaller than the one designed using layout A.

Many ergonomics problems, however, do not have such uniquely defined and measurable indices, or such indices are not the only ones that determine the correctness of potential design solution. An example of such type of problems is the field of human–computer interaction. Appropriately designed dialog windows or manual structure does not necessarily improve the quality of such interface. This is because the factors that influence the subjective and objective evaluation of computer software (and the related interface) are very complex and not fully identifiable. Laboratory experiments by Sears (1993) and Grobelny (1998) indicate that optimized dialog windows (with respect to the road traveled by the cursor in the process of its use) are much more effective in real time. These are also being subjectively judged by the subjects to be better than those windows that are designed without any formal layout analysis of the objects placed on the computer screen.

REFERENCES

BONNEY, M.C. and WILIAMS, R.W., 1977, A computer program to layout controls and panels. *Ergonomics*, 20, 297–316.

DREZNER, Z., 1988, A heuristic procedure for the layout of a large number of facilities. *Management Science*, 33, 907–15.

FITTS, P.M., 1947, *Psychological Research on Equipment Design*. Report 19 (Army Air Forces Aviation Psychology Program Research Reports).

GAVETT, J.W. and PLYTER, N.V., 1966, The optimal assignment of facilities to locations by branch and bound. *Operation Research*, 14, 210–32.

GROBELNY, J., 1988, The 'linguistic pattern' method for a workstation layout analysis. *International Journal of Production Research*, 26, 1779–98.

GROBELNY, J., 1998, Index of difficulty (ID) as a quality factor for a graphical man–machine interfaces. In W. Karwowski and R. Goonetilleke (eds), *Manufacturing Agility and Hybrid Automation — II. Proceedings of the Sixth International Conference on Human Aspects of Advanced Manufacturing: Agility & Hybrid Automation* (Santa Monica: IEA Press).

GROBELNY, J., 1999, Some remarks on scatter plots generation procedures for facility layout. *International Journal of Production Research*, 37, 1119–36.

GROBELNY, J., KARWOWSKI, W. and ZURADA, J., 1995, Applications of fuzzy-based linguistic patterns for the assessment of computer screen design quality. *International Journal of Human–Computer Interaction*, 7, 193–212.

HILLIER, F.S. and CONNORS, M.M., 1966, Quadratic assignment problem algorithm and the location of indivisible facilities. *Management Science*, 13, 42–57.

KARWOWSKI, W., Lee, W. and GROBELNY, J., 1999, Applications fuzzy system methodologies in human factors and ergonomics. In H. Zimmermann (ed.), *Handbook of Fuzzy Sets and Systems* (submitted).

LEE, R.C. and MOORE, J.M., 1967, CORELAP-computerized relationship layout planning. *Journal of Industrial Engineering*, 18, 195–200.

McCORMICK, E.J., 1976, *Human Factors in Engineering and Design* (McGraw Hill).

PROCTOR, R.W. and VAN ZANDT, T., 1994, *Human Factors in Simple and Complex Systems* (Allyn & Bacon).

SEARS, A., 1993, Layout appropriateness: a metric for evaluating user interface widget layout. *IEEE Transactions on Software Engineering*, 19, 707–19.

WIERWILLE, W.W., 1980, *Statistical Techniques for Instrument Panel Arrangement* (Virginia Polytechnic Institute and State University).

Handtools

M. Braun and R. Schopp

Fraunhofer-Institute for Industrial Engineering (IAO), Stuttgart, Germany

1. INTRODUCTION

Handtools are used in all kinds of production processes, not only in the industrial production but also in the household, in medicine, arts, etc.; and particularly for the work on metals, textiles, rocks and food. The forces necessary arise from the human muscle strength. Thus, handtools are normally not used in long-term processes but they are needed for works which go beyond the human capacity (e.g. hammer drill).

In this context, *powered* handtools shall be discussed as well – although they are driven by an external source – because it is up to the human user to handle, hold and guide them differently according to the given task.

Within the ergonomic equipment design the designing of handtools plays an important role, because both the efficiency and the working process itself at the interface between user and his task are influenced. Within classification schemes handtools are distinguished from controls and machine tools.

It is the intention of this contribution to give an introduction into the handtool design process. Therefore, basic design dimensions, methodical approaches, and requirements of handedness are discussed, which are the basis of two case studies describing fundamental design dimensions by way of example.

2. DESIGN DIMENSIONS

In the ergonomic design process not only technical and formal–aesthetic aspects are of importance but also anthropometric, biomechanical, physiological criteria and aspects concerning occupational health and safety (see elsewhere in this volume).

To realize the efficient, comfortable and safe human use of handtools the design parameters are orientated to the characteristics and abilities of a specific user group (e.g. height, physical strength, freedom of movement, sensory perception and information processing). To avoid symptoms of fatigue or even occupational disease, the physical workload in its height and duration must be orientated to the user's strain limits.

Working with handtools can involve enormous risks of accidents, especially when they are not used at workstations with constant working conditions but in varying applications (e.g. at construction sites or in the food industry). Both the handtool and the object the user is working on can cause accidents. When a handtool is designed aspects of safety should be considered to protect the worker effectively from strain and risks (Hecker 1997).

The use of handtools can cause great strain on the hand because of impulsive work processes (e.g. the recoil of a hammer drill). Measures that reduce shock and vibration should already be taken when designing a handtool, e.g. concerning the minimization of production tolerances and diametrical clearances as well as of imbalances in rotating elements. Additionally, a spring-damper system is recommended to prevent shocks and vibration that affect the hand–arm system.

When handtools are used outdoors or under disadvantageous thermal conditions (e.g. in the construction industry or in forestry) a heat insulation is necessary.

Handtools can be differentiated into a hand side facing the user and a working side serving the direct work progress. The hand side as the interface to the user is of greater importance for the ergonomic design than the working side. The hand side of tools is normally represented by handles and control elements.

3. METHODICAL APPROACH

Each set of ergonomic equipment design requires the inclusion of methodical approaches, adequate instruments, and experience. In the ergonomic design process of handtools the following steps have proved successful:

- target and definition of the design task;
- analysis of the handtool to be designed;
- development of alternative design solutions;
- evaluation and selection of design alternatives; and
- realization of the selected design solution.

In the following the systematic proceeding will be described which leads to the conceptional design or corrective redesign of handtools. The aesthetic aspects like shape, size, material and surface have to be analyzed in detail before they can have an influence on the design (Bullinger and Solf 1979). Figure 1 illustrates the different steps when designing the hand side of tools. The proceeding will be described afterwards.

It is the aim of the *task analysis* to record all possible conditions of the task quantitatively. Either physical measurement or observations and interviews can provide the required data.

The next step is to examine and determine the *body's posture* of the user. Furthermore, the capable *motion range* of the hand–arm system as well as the *motion assignment* of the handtool and the hand–arm system within the task have to be taken into consideration. The best posture would be if the required movement direction corresponds with the anatomically favorable position or movement direction.

Then, the *posture of the hand* is examined, i.e. the position of the hand in relation to the forearm axis, as well as the *type of gripping*, i.e. the linkage between hand, fingers and hand side. When handtools are used static and dynamic forces have to be summoned up and transferred. Therefore, the designers of handtools have to focus on two problems: the static strain of muscles must be reduced and the transmission of great forces in the direction of the lateral and longitudinal hand axis should be prevented. For high force resistances the recommended type of gripping is steadily increasing within the types of coupling-contact (touch) to grasp.

The *type of coupling* chosen influences the strain that arises from the performed task. It would be best if the strength that is necessary to use the handtools originates from a direct, positive force transmission, for in the case of frictional transmission the strength that must be summoned up would always have to be greater. So if the optimal coupling is chosen the static physical strength can be reduced (Bullinger and Solf 1979).

In the last step the formal–aesthetic design of the hand side is considered in the *form definition*. But only when the *shape* of the handside is fundamentally determined with regard to the handedness, it can be proceeded. Then, the next steps would be to discuss the *dimensions*, the *material* and the *surface* design.
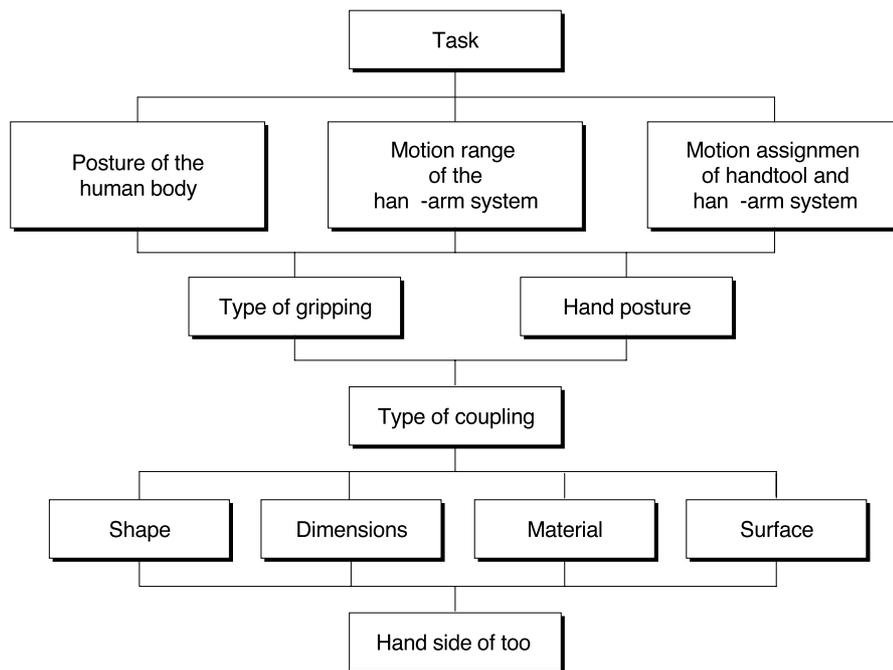
Figure 1. **Proceeding levels of hand side design of tools (Bullinger 1994).**

4. HANDEDNESS

4.1. Situation

The handedness depends on the demands of the user and the work process. As far as the user is concerned one can differentiate between left- and right-handedness. Tools can be distinguished as follows: one-handed (e.g. pliers), one-handed with additional handle (e.g. hand drill), two-handed (e.g. spade), movable (e.g. scissors) and fixed (e.g. hacksaw) handles.

Society shows a preference for the right hand side. This is confirmed when looking on most of the symmetrical handtools designed for right-handers. There is a share of 10 to 15% of left-handers in our society – and this is expected to increase because of a more tolerant, that means not right-hand-orientated, education.

Only in a few cases two symmetrical handtools – one for right- and the other for left-handers – are an optimal solution, because often the working hand is defined by the working situation and not by the handedness of the single person. Nevertheless, in many working situations advantages can be seen when viewing on handtool design adjusted to handedness.

As a result of using handtools with the unsuited hand an increased accident risk can be stated. Possibilities for adaptions on the needs of left-handed persons are often unknown in practice and therefore rarely existing (Schmauder and Solf 1992).

4.2. Design Proceeding

Figure 2 is aimed at giving an idea how the handedness of humans can be taken into consideration when handtools are newly designed or redesigned. It was conceived as an addition to the prevailing rules and is meant to optimize existing proceedings. The course diagramme is divided into task analysis, problem analysis and new design (Schmauder and Solf 1992).

Within the analysis the task is characterized by answering the four questions: "Is high exactness demanded?," "Is high quickness demanded?," "Is there high movement resistance?," and "Is continuous movement necessary?" The demands on exactness and quickness correlate with the hand-performance and are therefore of importance for the task analysis. For tasks with low and both-handed equivalent demands on ability and aptitude of the hand–arm system it is not necessary to include handedness explicitly into handtool design.

If the task corresponding to the four questions is characterized as "demanding" or when the task must be performed one-handedly, it has to be proved in the problem analysis whether the user has free choice of hand-assignment; this is e.g. favorable in spatial hindrance. In case the performing hand is defined by the work process or by the existing design solution, the handedness must be considered, for the user has no choice of hand-assignment. Even when hand-assignment is not clearly defined it is principally possible to couple the preferred hand – which need not be necessarily the favorable hand – to the tool. The question on less- or non-optimal results is relevant when the user has free choice of the working hand and when the linking to the tool is not determined. Here it should be examined whether this determination causes non-optimal results (with consideration
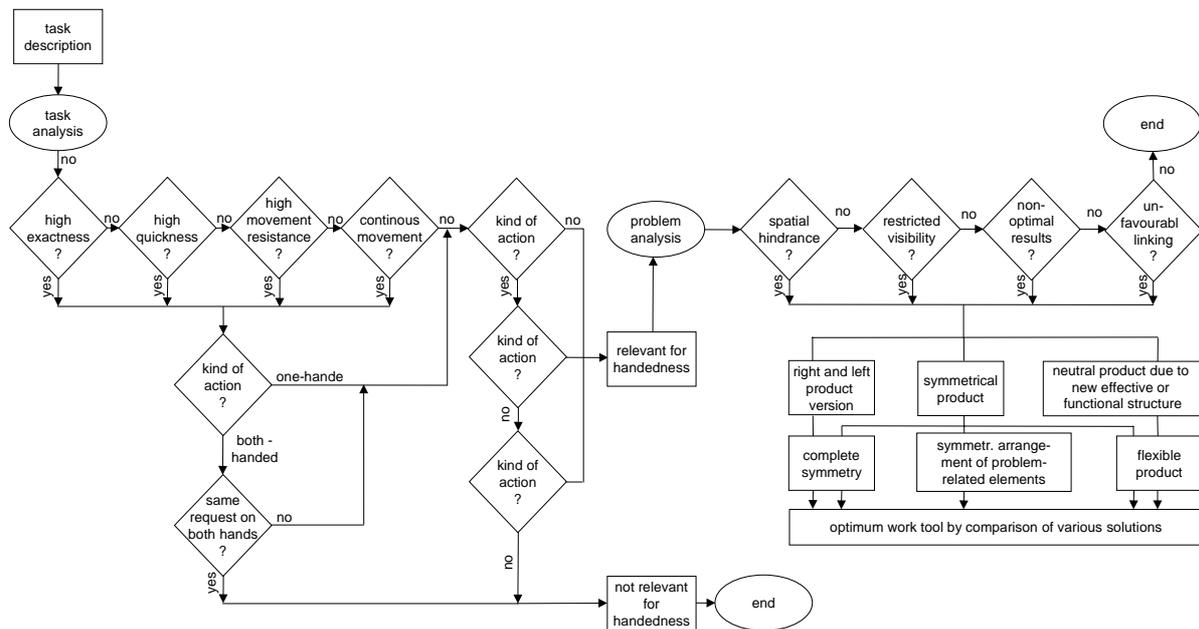
IEEHF

Figure 2. From task analysis to problem analysis and new design.

on different shaping of performance). In a definite case there might be free choice of hand but a user-oriented handtool design makes it impossible to choose the unsuited one. Nevertheless, a non-optimal result might be achieved when different aptitudes are demanded, e.g. quickness and strength, which do not have their maximum constantly either in the right or left hand–arm system.

Handtools with high sensorimotor demands on the user should be designed so that they can be used with both the right and the left hand. This might be advantageous not only for left-handers but also for right-handers in case working with the left hand would be more favorable. This possibility of free hand choice should be realized in the handtool design – a left and right hand version should be an exception. Often it is sufficient to design the problem-relevant elements symmetrically or flexibly. Additionally the following design aspects should be taken into account:

- The visibility towards the working center should be good in any working position.
- Rotating handtools should – if possible – be designed for both clockwise-rotation and reverse action.
- Feed lines should be fixed symmetrically to the handtool, so that they cannot hinder the work in one direction.

5. DESIGN CASES

The two design cases described in the following are to be used as a basis for demonstrating the demands and design dimensions

in handtool design process by way of example. In doing so, it becomes obvious that a design process – considering the complex problems inherent in human and work – is dependent on the most varied factors.

5.1. Kitchen Knives

The knife is one of the most original tools of human. It is used in nearly all areas of everyday life when performing different work tasks in which material is to be cut or separated such as pricking, carving, splitting, peeling and scraping. In the course of automation the simple knife was replaced by new technologies; nevertheless, the hand-held knife cannot be replaced by machines or special-purpose tools in many areas still today. Above all in the area of the preparation of dishes and meals the knife has been the most frequently used tool now as before.

Professional users working in the kitchen frequently complain about pain in the shoulder/neck region. These complaints can above all be attributed to the cramped posture during cutting operations. On the other hand, pains in the hand–arm system and in the hand–finger system have their origins in the insufficient handle design of the used professional cutting tools.

The conventional styling of professional kitchen knives neglects ergonomic requirements that result from the handling of the tool by users with hands of different sizes and by application in different cutting tasks. A considerable ergonomic improvement of the kitchen knife can be achieved by a hand side (handle)

adapted to the working side (blade). The objective of designing a set of kitchen knives for professional use is to permit the optimal performance of all cutting tasks encountered with as few different types of knives as possible.

In an investigation of the human–knife–material system six typical cutting movements and six typical manners of gripping could be classified within the analyzed cutting operations. The cutting movements are dependent on the cutting task, i.e. from the work result achieved and from the spatial position of the cutting plane. The cutting tasks could be divided into the categories "cutting on a tabletop" and "free cutting in the room." The cutting movements are directly influenced by the possible movements of the hand–finger system and the hand–arm system. They can be regarded in dependence on the manners of gripping which are decisively influenced by the force to be exerted in performing the cutting task (Mangol and Eckert 1995).

There is a direct interrelation between cutting task, form of movement, manner of gripping and blade type. Based on this fact, seven different types of knives were defined, whose design requirements can be derived from a specific cutting task as well as from cutting movement and manner of gripping associated with it. The design requirements were worked out separately for each of the seven types of knives. In doing so, the requirements made on the blade geometry and on the handle geometry were detailed. This fact permitted seven different blade types and three different handle shape types to be defined (Mangol and Eckert 1995).

A comparison of the requirements regarding the handle geometry specific to the seven types of knives reveals correspondence for specific types. Therefore, all requirements made on the handle design can be satisfied with only three different forms of handles. These three forms of handles are not only different in an adaptation of the proportions of the handle to the blade length, but they have basic geometries that deviate from each other. The three forms of handles are each assigned to one of the seven blade types. Consequently, this knife set consists of seven different types of knives altogether, as shown in figure 3. Each of these knives has been optimized for a specific range of applications. All cutting tasks encountered in the professional kitchen will be covered by the application of the seven types of knives.

5.2. Powered Hand Drill

The second example deals with the design of a powered hand drill. Here, it is above all endeavored to improve the general handling of the tool for holding and guiding. A reduction in the total weight and a reduction of the tilting moment by displacing the tool's center of gravity near the handle or by arranging a supporting handle make a marked improvement in the handling conditions.

The handle of a powered hand drill should be arranged and shaped in such a way that a manual exertion of high feed forces is possible, at the same time as damping developing reaction forces and torques. In doing so, the feed force should be favorably generated by positive power transmission. The dimensional
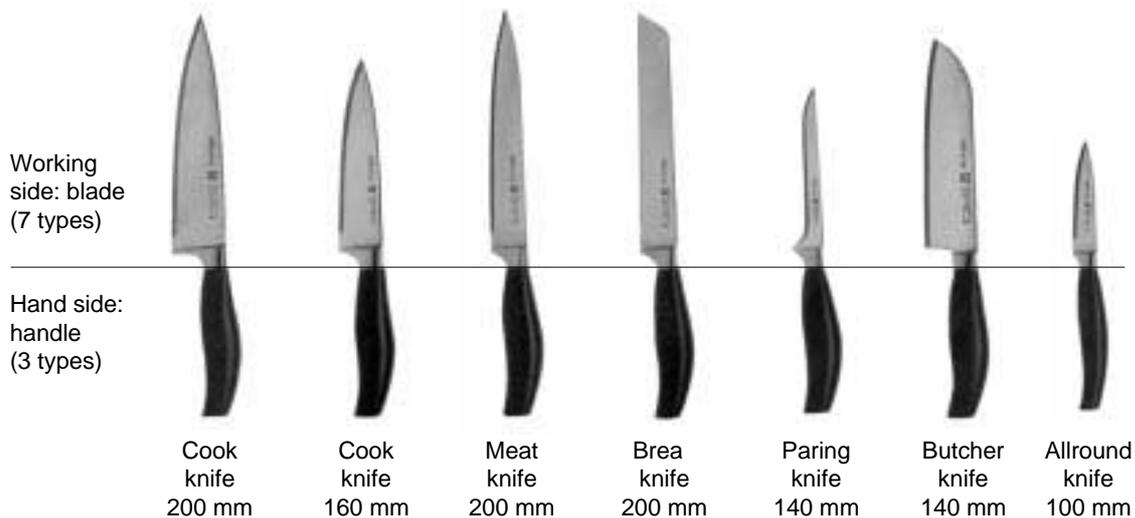
Figure 3. Set of ergonomic designed kitchen knives for professional use.
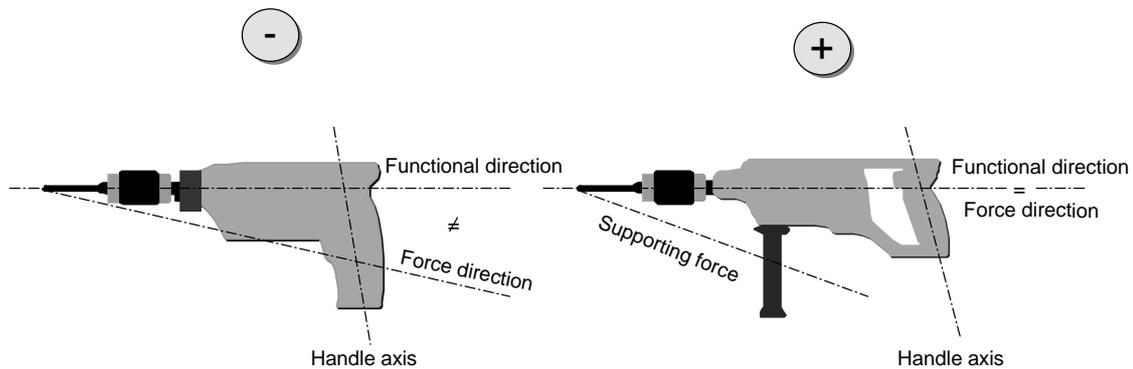
Figure 4. Design alternatives of a powered hand drill: suitable and unsuitable motion assignment.

design of the handle and the switches is oriented at the anthropometric and physiological properties of the hand–arm system.

It is the aim of the motion assignment that an agreement, as suited as possible, is achieved between the functional direction of the tool and the anatomically favorable movements of the user. The kinematics of anatomically suited modes of movement favors the exertion of force and prevents forced motions and postures. Figure 4 shows both the unsuitable and the suitable assignment of motion in a powered hand drill. By displacing the handle to the rear motor region, a direction of force will be obtained in the hand drill designed according to ergonomic criteria, with this force being in alignment with the functional direction resulting from the drilling axis.

REFERENCES

BULLINGER, H.-J., 1994, *Ergonomie – Produkt- und Arbeitsplatzgestaltung* (Stuttgart: Teubner).

BULLINGER, H.-J. and SOLF, J. J., 1979, *Ergonomische Arbeitsmittelgestaltung I–III*, Schriftenreihe der Bundesanstalt für Arbeitsschutz (Bremerhaven: Wirtschaftsverlag NW).

HECKER, R., 1997, Handwerkzeuge und von Hand angetriebene Geräte. In Luczak, H. and Volpert, G. (eds), *Handbuch Arbeitswissenschaft* (Stuttgart: Schäffer-Poeschel), pp. 594–8.

MANGOL, P. and ECKERT, R., 1995, Analysis of the use of professional kitchen knives. In Bittner, A. C. and Champney, P. C. (eds) *Advances in Industrial Ergonomics and Safety VII*, pp. 519–23.

SCHMAUDER, M. and SOLF, J. J., 1992, *Einfluß der Händigkeit von Arbeitsmitteln*, Schriftenreihe der Bundesanstalt für Arbeitsschutz, Fb 661 (Bremerhaven: Wirtschaftsverlag NW).

Instruments and Design

P. Rabardel† and P. Béguin‡

†Psychologie, Université Paris 8, CNRS, 2 rue de la liberté, 93526 Saint Denis cedex 02, France

‡Laboratoire d'ergonomie, Cnam, 41 rue Gay Lussac, 75011 Paris, France

Today, the necessity of designing technical systems, products, machines and tools so that they are good instruments for users is more and more evident. This chapter formulates some theoretical proposals to contribute to anthropo-centered design, that is design organized around and for human beings, whether they be users of a product or professionals who use tools or machines. Proposed is a conceptualization of the instrument that design processes can be based on. We then develop the idea that design continues in use, and as a result, design processes need to consider this extension from the outset.

1. WHAT IS AN INSTRUMENT TO THE PERSON USING IT?

The literature proposes several definitions that vary depending on the disciplines and their aims. Norman (1991) distinguishes two perspectives: the "system view" and the "personal view":

- The system view is when the outside observer looks at how the man–artifact whole accomplishes a task. This is the traditional view of the man–machine system where the task is considered as unchanging and the instrument is seen as an amplifier of users' skill.
- The personal view is based on the analysis of modifications brought about by an artifact's use: what is transformed in the task, what must be learned, the processes that must be dropped.

The instrumental approach works within a personal view that we extend to activity. We think it necessary to:

- on one hand, analyze the influence of the artifact on the reconstruction of tasks, the appearance of new tasks and the disappearance of old ones.
- on the other hand, analyze the complete reconstruction of activity in the view opened up by Vygotsky (1931–78) with the concept of the instrumental act. This view is developed under various forms, particularly in current research on human–computer interaction and socially shared cognition (Bødker 1989, Caroll 1991, John-Steiner 1995, Kuutti 1996).

An essential precept of this work and that of the French-speaking school is that the instrument is a medium ground, half way between the subject and what needs to be transformed (matter or information) that we will call the "object". Between the three poles, subject, object and instrument, there are several interactions (direct subject–object interactions (dS-O) or interactions mediated by the instrument (mS-O); subject–instrument (S-I) and instrument–object interactions (I-O); Figure 1) where each type of interaction represents an analysis plan.

It is important to underline the fact that in this conceptualization, the instrument is distinguished from the artifact. Artifacts are built by humans with certain purposes in mind. Nonetheless, users do not always use artifacts as designers intended. This empirical fact has constituted, since the initial
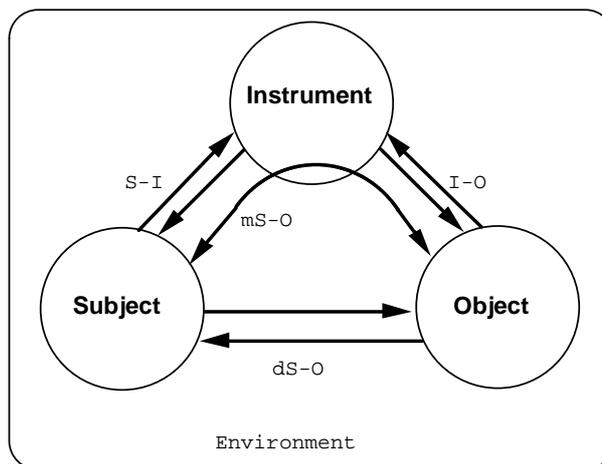
Figure 1. I.A.S. model: the triad characteristics of Instrumentalized Activity Situations (Vérillon & Rabardel 1995).

work of Ombredane and Faverge (1955), one of the bases of French language ergonomics in which operator inventiveness is considered as an essential feature of work intelligence. In situations, operators explore, interpret, use and transform their technical, social and cultural environment.

It is within this activity that the artifact, whether material or symbolic, is instituted as an instrument by the user (Rabardel 1991, 1995). We propose that the instrument mobilized by the user in his/her activity be considered as a mixed entity, coming from both the subject and the object (in the philosophical sense). The mixed entity instrument includes both:

- a material or symbolic artifact produced by the subject or by others; and
- one or more associated utilization schemes resulting from a construction particular to the subject, or an appropriation of pre-existing social utilization schemes.

Schemes organize subjects' actions. They correspond to stabilized aspects of actions for known situations. It is because there are stable aspects in situations that actions also include stable aspects. An artifact is in itself a stable aspect within the situation in which it is used. This is why users construct utilization schemes which form the instrument once they are associated with the artifact. An example is the association between writing schemes and an artifact such as a pen which constitutes the instrument that allows the text to be written.

Utilization schemes have a private dimension particular to each individual. They also have a social dimension: they are common to all, or several members of a social group, a collectivity, a country, etc. This is why we speak of social utilization schemes.

Schemes assimilate. They can apply to several types of artifacts. The hammering scheme, usually associated with a hammer, can temporarily be associated with a spanner. This property, combined with the social nature of utilization schemes, is relevant to design because the designer is able to use a utilization scheme that exists within the collectivity in question to invent an artifact that can be easily assimilated into pre-existing utilization schemes. Schemes are also accommodating. They can change and develop with a changing situation. This characteristic is also

relevant to design because the designer can use a scheme that is merely a neighboring utilization scheme to the new artifact.

2. DESIGN CONTINUES IN USE ...

It has been seen that users establish the artifacts elaborated by designers as instruments. This activity can be almost instantaneous (like the transformation of the spanner into a hammer), or take place over long periods (like the constitution of writing schemes associated with the pen). Often there is an *instrumental genesis* (Rabardel 1991) that concerns both the artifact and the utilization schemes through two processes:

- The *instrumentalization* process concerns the emergence and evolution of the artifact and instrument components. It is based on the characteristics and properties specific to the artifact and gives them a status depending on the function of the action taking place and on the situation (in the example of the spanner replacing the hammer, the important properties are bulk, hardness and prehensibility). Beyond the action taking place, these properties can become, for the subject, a permanent characteristic of the artifact component of his/her instrument. Instrumentalization can go as far as transforming the artifact's behavior and in some cases, its material structure.
- The *instrumentation* process is relative to the emergence and evolution of utilization schemes: their constitution, their functioning and their evolution as well as the assimilation of new artifacts into already constituted schemes, etc.

These two dimensions of instrumental genesis are related to the subject and are different in their orientation. In the instrumentation process, the activity is oriented towards the subject himself and him managing his own activity. In the instrumentalization process it is oriented towards the instrument's artifact component and mastering the object of the action. The two processes work together to contribute to the constitution and evolution of instruments, even if, depending on the situations, one of them can be more developed, dominant or even the only one exercised (Rabardel 1995).

According to the traditional scheme, which temporally distinguishes design and use, the real use phase begins only after preparation, and possibly installation. The real use phase is supposedly nothing more than the functional use of the artifact. Our approach demonstrates that the design process does not stop with an instrument's use. It continues throughout its use in instrumental geneses. We think that users, as agents of these processes, are also agents of change in design as a whole. Of course they do this in a different manner to "institutional" designers. Users design for themselves. This self-designing can also result from work collectives (Béguin 1997).

3. TOWARDS DESIGN PRACTICES FROM AN INSTRUMENTAL VIEW

What are the consequences for design of this approach to instruments? Operator creativity and inventiveness constitutes an ontological characteristic of the design process and is not only an indicator of a deficit: needs change and use evolves following the changing dynamic of situations. Design continues in use and forms of exchange must be organized between the activity of one or more designers and the activity of one or more users. This is true before, after and throughout the design process (Béguin 1997, Rabardel 1997).

First we will look at before. We propose to undertake instrumental analyses preceding design with two objectives: to draw inspiration from artifacts born of previous instrumental geneses and to design around schemes (Rabardel 1997).

- on one hand, the analysis of instrumentalized artifacts by users reveals both perceived needs, accepted or rejected artifact functions and new created functions; the structural and behavioral solutions imagined by users;
- on the other hand, we have underlined the ability of utilization schemes to accommodate and assimilate. The analysis of utilization schemes appears as necessary given that the artifact's characteristics can lead to the mobilization of badly adapted schemes, or on the contrary, make difficult the application of schemes essential to reach the action's objectives.

Nonetheless, in our approach, the user's contribution is not limited to enriching data and ensuring the increased validity of technical choices. The analysis of instrumental geneses shows that while reducing effort and improving efficiency are definite motives, they can also come from the subject's sense of action (Minguy and Rabardel 1993).

Next we look at the post-design process in the designing of instrumentalizable artifacts. It is not only a question of designing artifacts that tolerate a range of uses. While a certain anticipation of uses and outcomes can be grounds for an artifact's design to an extent, only a small number of uses can be anticipated due to the contingencies that contribute to their appearance.

It should also not be imagined that artifacts adapt themselves to the user, even if this perspective seems interesting in certain cases. It is more a question of providing the user with artifacts that facilitate the process of instrumental genesis by bringing appropriate assistance to the elaboration of uses by novices, and by allowing users to act on the behavior and structure characteristics of the artifact. An artifact designed in this way must not only allow instrumentalization. It must also help the user to carry out an action and ensure that a coherent, collective use is maintained by making possible a shared construction between individuals. Shared is used to mean both "placed at everyone's disposal" and "true of everyone". As we can see, the instrumental approach opens up new perspectives to solve many questions. At the same time, it opens up new problems and we must be careful not to weaken users by leaving it up to them to find solutions to problems that are too complex given their resources and means, particularly when the solutions require them to go against directions.

REFERENCES

BÉGUIN, P., 1997 Le schème impossible ou l'histoire d'une conception malheureuse. *International Journal of Design and Innovation Research*, no. 10, 21–40.

BØDKER, S., 1989, A human activity approach to user interfaces. *Human computer Interaction*, 4, 171–95.

CAROLL, J., 1991, The little Kittle little house. In J. Caroll (ed.) *Designing Interaction: Psychology at the Human Computer Interface* (Cambridge: Cambridge University Press), 37–54.

JOHN-STEINER, V., 1995, Cognitive pluralism: a socio-cultural approach. *Mind, Culture and Activity*, 2, 2–11.

KUUTTI, K., 1996, Activity theory as a potential framework for human–computer interaction research. In B. Nardi (ed.) *Context and Consciousness: Activity Theory and Human Computer Interaction* (Cambridge, MA: MIT Press).

MINGUY, J.L. and RABARDEL, P., 1993, Control of a Fisching trawl: a multi-instrument process control situation. In H.G. Stassen (ed.) *Analysis, Design and Evaluation of Man–Machine Systems* (Oxford: Pergamon), 157–62.

NORMAN, D.A., 1991, Cognitive artefact. In J.M. Caroll (ed.) *Designing Interaction Psychology of Human Computer Interface* (Cambridge: Cambridge University Press).

OMBREDANE A. and FAVERGE J.M., 1955, *L'analyse du travail* (Paris: PUF).

RABARDEL, P., 1991, Activity with a training robot and formation of knowledge. *Journal of Artificial Intelligence in Education.*

RABARDEL, P., 1995, *Les hommes et les technologies, une approche cognitive des instruments contemporains* (Paris: Armand Colin).

RABARDEL, P., 1997, Men's instruments, from project to use. *Ergonomia*, 9, 21–7.

VYGOTSKY, L.S., 1931–78, Mind in society, the development of higher psychological processes. In M. Cole *et al.* (eds) (Cambridge, MA: Harvard University Press).

Laptop Computer Use

L.M. Straker

Department of Physiotherapy, Curtin University of Technology, GPO Box U1987, Perth WA 6001, Australia

1. INTRODUCTION

Use of laptop computers has increased rapidly in the 1990s. More people are using laptop computers and each person is using them for more hours per day. This means that human-machine interaction is becoming more critical.

The rapid spread of laptops has outstripped the scientific community's research on the possible implications of their use. A few research reports have recently been published; together with the extensive research on desktops, they will be used to inform this article. What literature is available suggests laptops may be a problem for users; according to Harris and Straker (1999), 60% of laptop users report associated musculoskeletal discomfort.

This article considers some key issues and makes recommendations. The specifics are limited to the prevailing technologies of the late 1990s, but the principles should be valid for future technologies. However, this is one of the fastest areas of development, so the key features of the interaction could change quite rapidly. For example, the use of efficient speech recognition input and high quality displays may alleviate many of the postural and visual concerns.

Here are some brief definitions of the terms "laptop" and "desktop" as used in this article. And figure 1 shows a photograph of the different computers. Desktop computers are characterized by a screen (typically a cathode ray tube monitor but increasingly tubeless flat screens) that is separate from the keyboard and other input devices. It is often separate from the central processing unit (CPU) and disk drives, though the iMac is a notable exception. Desktop monitors and CPUs are typically rather heavy and are usually permanently located on a desk or other stationary work surface, though sometimes on a trolley. In the 1990s anyone asked to imagine a computer would typically imagine a desktop.

Laptop computers, sometimes called notebook computers, are smaller than desktop computers. Laptops were initially introduced to provide a portable computer, with 1990s laptops typically weighing about 4 kg (depending on battery size) and

Figure 1. Desktop, laptop and palmtop computers

being about the size of an A4 piece of paper and around 50 mm thick. With their processing power now equaling desktops, they are increasingly being used as replacements for desktops. Laptops can be characterized by a flat screen attached to a restricted keyboard which usually sits over the CPU and disk drives.

A third group of computers is also currently available. Palmtop computers are smaller still, around the size of an adult hand. They typically have liquid crystal display screens and small keys and/or stylus input. Palmtops currently have considerably less power than laptops, though still enough to run simple word processing, spreadsheet and database software. Their most common use is as "personal organizers" based on real-time diaries. However, their power can be expected to increase over the next decade, increasing their potential to be used as desktop replacements. Although not mentioned again here, palmtops raise similar issues to laptops.

2. POTENTIAL PROBLEMS WITH LAPTOP USE

Laptop computers are increasingly being used as replacements for desktop computers, and they are widely marketed as such. However, in the late 1990s they differ from desktop computers in at least four ways which have critical implications on the human-machine interaction: display device, input device, connection of screen and keyboard, and environs of use.

2.1 Display Device

Laptops use flat screen technology instead of cathode ray tubes. Originally they were simple liquid crystal displays but increasingly they are active matrix or thin film transistor displays. Some users report poorer screen clarity with laptops, even in ideal viewing conditions, whereas others find the flat screens more comfortable than cathode ray tubes. As newer technologies improve the clarity of flat screens, this may be less of an issue in ideal viewing conditions.

However, laptop viewing conditions are often far from ideal (section 2.4) and the possibly poorer screen clarity is compounded because flat screen technology is more susceptible to problems with viewing angle, lack of adjustability and reflections from the environs.

The result is common complaints of difficulty viewing laptop screens. The likely impact of this is a greater chance of visual discomfort and possibly visual disorders. Given the close linkage of visual and musculoskeletal systems, it is also likely that users will compromise their musculoskeletal system in an attempt to reduce visual discomfort. There is therefore also likely to be a contribution to increased risk of musculoskeletal discomfort and disorder.

2.2 Input Device

Current laptops typically use a smaller keyboard and a pointing device like a trackpad, trackball or mini-joystick (Batra *et al.* 1998). Keyboards usually have a reduced number of keys (number pads typical of desktop extended keyboards are usually missing) and often the size and spacing between keys is reduced. Digit and wrist postures are therefore likely to be more constrained and, especially for larger hands, more awkward.

Pointing devices are usually smaller than those associated with a desktop. The smaller and finer control movements required

may put greater strain on intrinsic hand muscles for movement control and forearm muscles for hand stabilization (Batra *et al.* 1998, Fernstrom and Ericson 1997).

Much of the common software of the late 1990s is written expecting a computer mouse as the main pointing device. Although laptop users sometimes attach a computer mouse, they often use the other built-in pointing devices. This is partially due to the environs of use; laptops are often used where a mousing surface is not available. Although there have been reports of musculoskeletal problems associated with mouse use (Cooper and Straker 1998), the finer movements associated with many laptop pointing devices may cause even more problems.

Different input devices have also been reported to affect performance speed and error in a variety of pointing and dragging tasks. Although the trackpad, trackball and mouse have usually provided better performance than other pointing devices, the assessments usually depend on the task undertaken.

The likely result of the more restrictive input devices normally used with laptops is a greater chance of musculoskeletal discomfort and disorder, with a potential effect on productivity.

2.3 Connection of Display and Input Devices

Nearly all laptops have their screen attached to their keyboard/pointing device. This connection limits the options available to the user for placement of screen and keyboard/pointing device.

Both the visual system and the hand movement system have zones of comfortable reach. Where the display and the control interfaces are separate, it is possible to try to optimize the position of each for the user. Thus two people working with desktop computers who have different visual accommodation and convergence and different arm lengths can vary the locations of screen and keyboard to suit their individual characteristics. But a laptop does not offer separate adjustability, so the user who prefers a larger eye-screen distance has to work with extended arms or compromise the screen distance to gain a comfortable arm posture.

There is considerable debate about the optimal positioning of computer interfaces, in particular the computer screen (see under VDU positioning). Recent research supports a keyboard position forward from the user, allowing nearly full forearm support on the desk surface. There is also considerable evidence that a low screen position results in less visual strain, although there is some concern that this comes at increased neck strain. When using a laptop on a desk, users typically adopt the flexed neck posture similar to a low desktop screen position (Harbison and Forrester 1995, Straker *et al.* 1997). The position for laptop use may therefore not be as problematic as first thought, though this is the subject of considerable debate. However, both sides would probably agree that it is desirable to have separate adjustability of screen and keyboard. Price and Dowell (1998) trialed "docking" a laptop with a desktop monitor and keyboard, but their results were inconclusive.

Straker *et al.* (1997) found no significant performance difference between desktop and laptop computers over short working periods, but there was a trend for faster word processing with the laptop. Several subjects who were not touch typists noted how it was quicker to move from viewing screen to keyboard with the laptop. The more restricted laptop posture probably has little short-term impact on productivity.

The likely result of reduced adjustability of screen and keyboard is a greater chance of musculoskeletal discomfort and disorder.

2.4 Environs of Use

Desktop computers are normally used at a single location, often dedicated to computer use. The desk, chair, foot support and associated documentation supports are often adjustable to obtain a comfortable posture. The lighting is often specifically designed for work with VDUs

In contrast, laptops are used in a wide variety of situations: in a car, on the floor in a bedroom, between machinery on a factory floor, and in an airport terminal. Laptops are often used without furniture designed for computer use and in bright daylight or other unsuitable lighting conditions. A critical feature of the environs of laptop use is that the furniture and lighting designs are often uncontrolled by the user. This may be because it is a
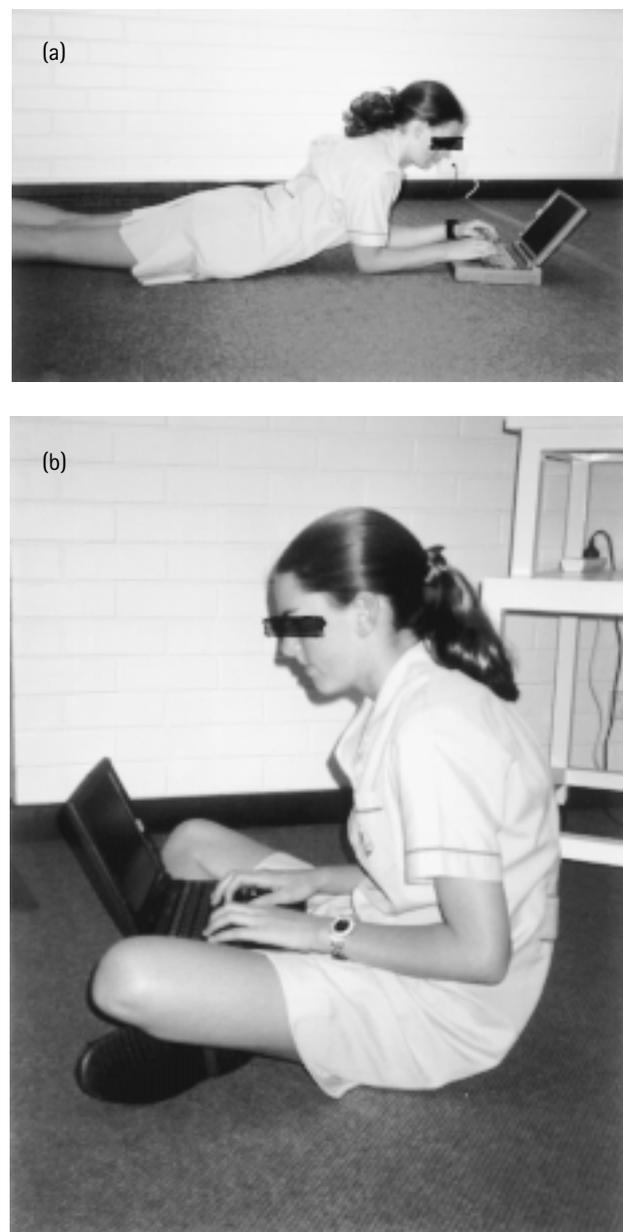
Figure 2: Postures adopted for laptop computer use

client's home (e.g., for home sales) or because there are other design constraints (e.g., when using a laptop in a car).

Together with furniture constraints, the desire to minimize reflections and optimize viewing angle leads to a wide variety of postures being assumed. Although varying posture is an aim of physical ergonomics, the recommended variation is between "good" postures. Often postures adopted during laptop use are "poor"; that is, they involve extremes of range and sustained static muscle contraction (figure 2).

Even when a reasonable posture is used, e.g., upright sitting, there is usually limited opportunity for forearm support — important for reducing trapezius muscle loading (Aaras *et al.* 1998).

Laptops are also transported by users, and this creates a potential carrying problem which rarely occurs with desktops. Although current laptops usually weigh only about 4 kg, they are often carried with a power transformer, portable printer, mobile phone and paperwork. A survey of schoolchildren using laptop computers found 60% reported discomfort associated with carrying their laptops. Laptops may therefore add a critical mass to the load carried.

Again the likely result of the lack of control of environs of use and more frequent carrying is a greater chance of musculoskeletal discomfort and disorder.

2.5 Problem Summary

The design and use characteristics of laptops may result in increased potential for visual and musculoskeletal disorders compared with desktop computers. The small available literature on laptops supports these concerns, as do the implications from earlier research on desktops. The main problems are likely to be visual and musculoskeletal (mainly neck and upper limb) as a result of the poorer display clarity, more restricted input devices, lack of independent adjustability for screen and keyboard, and lack of control over environs of use.

3. RECOMMENDATIONS

It seems prudent to be cautious in the use of laptops. Pending further research, here are some tentative recommendations for acquisition and use:

- Preferably do not use laptops simply as replacements for desktops.

- Where laptops are used in office locations, use them in conjunction with a separate high clarity display and normal-sized keyboard — use docking.
- Buy one with the best quality screen (considering image clarity, angle of view, reflections).
- Buy one with full-sized keys.
- Use it for shorter periods of time between breaks (the poorer the posture, the shorter the work period); be guided by visual and musculoskeletal discomfort.
- Use a variety of supported postures in the middle of your joint range.
- Buy the least heavy laptop.
- Carry it in a comfortable backpack with two shoulder straps or use a wheeled "aircraft" case.

REFERENCES

AARAS, A., HORGEN, G., BJORSET, H.-H., RO, O., and THORESEN, M., 1998, Musculoskeletal, visual and psychosocial stress in VDU operators before and after multidisciplinary ergonomic interventions. *Applied Ergonomics*, 29, 335–54.

BATRA, S., DYKSTRA, D., HSU, P., RADLE, K., and WEIDENBECK, S., 1998, Pointing device performance for laptop computers. *Proceedings of the Human Factors and Ergonomics Society 42nd Annual Meeting*, pp. 536–40.

COOPER, A. and STRAKER, L., 1998, Mouse versus keyboard use: a comparison of shoulder muscle load. *International Journal of Industrial Ergonomics*, 22, 351–57.

FERNSTROM, E. and ERICSON, M.O., 1997, Computer mouse or Trackpointæ effects on muscular load and operator experience. *Applied Ergonomics*, 28, 347–54.

HARBISON, S. and FORRESTER, C., 1995, The ergonomics of notebook computers: problems or just progress? *Journal of Occupational Health*, 11, 481–87.

HARRIS, C. and STRAKER, L., 2000, Survey of physical ergonomics issues associated with school children's use of laptop computers. *International Journal of Industrial Ergonomics* (in press).

PRICE, J.M. and DOWELL, W.R., 1998, Laptop configurations in offices: effects on posture and discomfort. *Proceedings of the Human Factors and Ergonomics Society 42nd Annual Meeting*, pp. 629–33.

STRAKER, L.M., JONES, K.J. and MILLER, J., 1997, A comparison of the postures assumed when using laptop computers and desktop computers. *Applied Ergonomics*, 28, 263–68.

Performance Prosthetic Hands

I. Gilad

Faculty of Industrial Engineering and Management, Technion (Israel Institute of Technology), Haifa, Israel

1. INTRODUCTION

A primary goal in the design and provision of body-powered hand prostheses is to make them efficient and simple in operating. A hand prosthetic device should be designed along ergonomical principles to provide a good substitute for the missing organ. Obviously, the higher the efficiency the less energy is required to use or operate in daily life situations. It should be remembered that any prosthesis would always remain a secondary, artificial substitute hand and will be an inefficient device when compared to the real organ. The main contributor to the inefficiency is the prosthetic prehensor.

Arm amputees have a basic choice of a hook or a hand for the prosthetic prehensor. There are compromises between these two options. The hook is more functional, and the hand is more cosmetic. Some amputees solve the dilemma by having one of each and interchanging them as work and social situations dictate. Arm amputees prefer to have one acceptable prehensor and they want one which is functional and attractive, but it does not necessarily have to look like a hand. The study reports the findings about five mechanical prosthetic designs and one mechanical hand. These were examined in object-geometry handling and ADL (activities of daily living) tasks.

2. PREHENSOR DESIGN

The mechanical devices present four non-conventional designs for prosthetic prehensors, shown and compared to the traditional Hosmer/Dorrance hook and to a hand-like prosthesis. The various prehensor designs include two that incorporate different kinds of grasp and combine voluntary opening and closing features (LeBlanc 1988).

2.1 The Hosmer/Dorrance hook

The first record of an artificial hand is an iron hand fitted to a Roman general who had lost his own hand in battle about 200 BC. In the later years of knights and armor, various prosthetic hands were fashioned out of metal, and at some point Captain Hook and other "pirates" were using hooks. In the 1890s, D.W. Dorrance developed the split hook. The Hosmer/Dorrance 5X hook (figure 1), which is made of stainless steel or aluminum, is the most common and is considered the "standard" for comparison with other prehensors. While many upper-limb amputees get used to the appearance of the hook, it is not what most people would consider an attractive looking device. For that reason, some amputees prefer the use of a prosthetic hand, although it may be less functional. In addition, some amputees try to have the best of both by having a hook and a hand and interchanging their use. Even then, sometimes the hand is relegated to the closet because of the inconvenience of switching prehensors and the superior function of the hook.
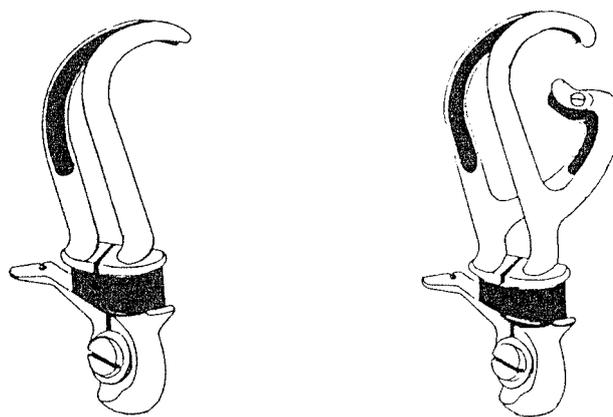
Figure 1. The Hosmer/Dorrance 5X hook (left) and the Gilad design (right)

2.2 The Gilad Design

This design (figure 1) is based on the lyre-shaped hook finger type "SS-555 Dorrance"; it features two additional "thumb fingers" which enhance the grasp performance. This device was introduced in Gilad (1985) after a time and motion study which advocated a design with greater ability to grasp various geometrical objects, but which still kept the simplicity of the popular two-finger design.

2.3 The Parker Design

This prehensor (Figure 2) is derived from study of the anatomy of the human hand. It is voluntary closing (VC), uses primarily three-jaw-chuck grasp, has curvatures for multi-point grasp of cylindrical objects, and has a cutout in the dorsum for use of pencils and utensils.

The operation of the Parker design is conceptualized as normally closed, thumb opening completely with a slight pull on the cable, and thumb closing with further pull on the cable. One unique feature being considered is cable take-up with thumb opening, which would allow a normally looser harness that would tighten for operation as the thumb opened for grasp.

2.4 The LeBlanc Design

This prehensor (Figure 2) is based mainly on functional considerations. It is voluntary opening (VO) in the normally closed position and is voluntary closing against the opposite finger. It has fingertip grasp in the VO position and palmar grasp in the VC position. The fingertips handle objects up to 1.5 inches in diameter, and the proximal area handles 1.5–3 inches in diameter. Prior work has indicated that 90% of activities can be handled with 1.5 inches opening or less.

2.5 The Nelson Design

This prehensor (Figure 2) is based mainly on esthetics. It is unique in that it has a rotary thumb which is normally closed in the VO position for finger tip grasp and then can be rotated around for palmar grasp in the VC position.

The LeBlanc and Nelson designs are unique in that they offer different grasping surfaces and the option of using VO or VC prehension. That is, one could hold objects in VO position with fixed prehension force and no harness pull, or hold objects in VC position with variable prehension proportional to harness pull.

Figure 2. The Parker design (left), the LeBlanc design (middle) and the Nelson design (right)

■ Geometrical Performance □ ADL Performance
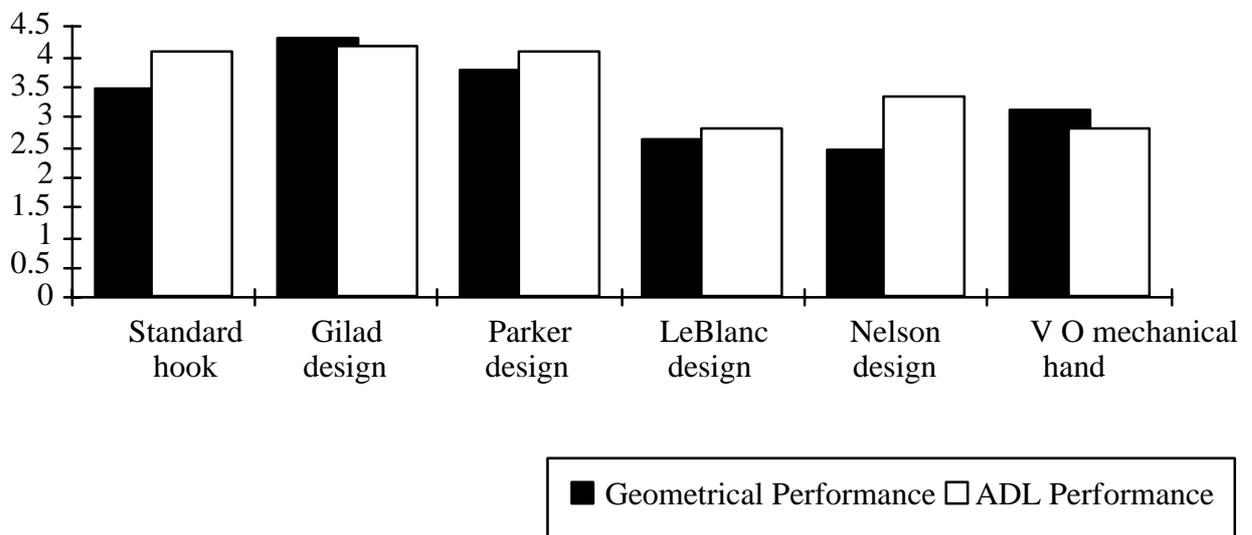
Figure 3. Performance scores for five hook designs and the VO mechanical hand

Note: The challenge for the manufacturer is to design a prehensor which is functional as a hook and esthetically acceptable. The upper-limb amputee will prefer the functional design that provides an attractive-looking device. The ability to perform a set of 15 activities common to daily living was also tested. The scores from this performance analysis indicated that the Standard Hosmer/Dorrance hook, the Gilad design, and the Parker design were preferred. Of these, the two designs referred to as the Gilad and the Parker meet the challenge for good performance devices. Although the Gilad design got better scores in both geometrical and ADL tests, it is believed that the Parker design will be more suitable for users who seek the better esthetic appearance.

2.6 The Mechanical Hand

When appearance and body image are important to amputees, the cosmetic non-functional or the mechanical hand is preferred. The mechanical hand, like the Otto Bock which has a voluntary opening and spring closing mechanism, is then popular. These were not considered in this study.

3. COMPARATIVE STUDY

Subjects were asked to rank their preference after examining the use of the different prosthetic devices in two tests. The first test was the ability to grasp, maintain hold, transport and position four sets of geometrical objects differing in shape and size using time criterion (Gilad 1986). The second test was the ability to perform a set of 15 activities common to daily living.

From this performance check, it is clear that the Standard Hosmer/Dorrance hook, the Gilad design and the Parker design were preferred. The scores were presented in a 1–5 scale: 1 given when unable to perform the given task, through to 5 for excellent performance. Findings are presented in figure 3. From this performance check, it is clear that the improved hook of the Gilad and Parker designs were preferred.

4. DISCUSSION

The use of various types of prehensors varies significantly around the world. European countries have much the same technology available as in the USA but apply it differently. Cultural and psychological factors play a big role on the way technology is used in clinical practice.

REFERENCES

GILAD, I., 1985, Motion pattern analysis for evaluation and design of a prosthetic hook. *Archives of Physical Medicine and Rehabilitation*, 66(6), 398–407.

GILAD, I., 1986, Objective performance analysis of hand amputees, toward the improvements of function, *Ergonomics*, 29(4), 553–61.

LeBLANC, M., 1988, Use of prosthetic prehensors, *Prosthetics and Orthotics International*, 12, 152–4.

Principles of Handtool Design

M. Aptel, L. Claudon and J. Marsot

Laboratoire d'Aide à la Conception d'Outils à Main Ergonomiques, Service de Physiologie, Environnementale, INRS Avenue de Bourgogne Bp 27, 54501 Vandoeuvre Cedex, France

The development of ergonomic tools responds, first, to health protection needs on the part of users. This is particularly the case in so far as MSD risks are concerned. Second, the development of ergonomic tools responds to the needs of the companies who use these tools and to their desire for better and more rapidly produced products. Only an ergonomic design process can enable tool manufacturers to meet these requirements. Three factors are involved: integration of ergonomics into the design process, definition of the different ergonomic stages involved in such processes and knowledge of the different factors involved in handtool design. The objective of this paper is to examine these three elements in more detail.

1. INTRODUCTION

For some years now, ergonomic handtool design has stimulated renewed interest among users, manufacturers and researchers. In the past, in fact since the Industrial Revolution at the beginning of the 19th century, emphasis was placed on handtool function to improve efficiency and to allow for standardization. The tool was required satisfactorily to fulfil the task for which it had been designed, to respond to the needs of the greatest possible number of users and to be as cheap as possible. Consequently, a given tool was designed to be used by all potential users. However, in recent years, approaches have changed and new notions of increased comfort and reduced biomechanical solicitation with regard to users' functional capacities have been introduced into tool design. There are several reasons for this development.

The rise of musculoskeletal disorders of the upper limbs (MSD) is the most important. These work-related disorders are found throughout the industrialized countries (USA, Sweden, France and Australia and are particularly widespread in industries that make use of handtools). As a result, the food-processing industry and car and household goods manufactures are production sectors in which the risk of MSD is particularly high. In a report published by the National Institute for Occupational Safety and Health (NIOSH) (Radwin and Smith 1993) it was estimated that MSD represents ~24% of all handtools injuries.

Second, the development of new technologies (artificial intelligence, robotics) and new forms of production process organization (just-in-time, ISO 9000 quality certification) have also had an impact. Such factors require a greater range of skills, greater know-how and deeper implication in the work process on the part of employees. New tool requirements have emerged from this background.

Finally competition between handtool manufacturers has led to widening of the skills and know-how required of manufacturers, including ergonomics, if they are to respond to market forces.

In practice, tool manufacturers must take three new types of need into account in the manufacture of handtools:

- Integration of ergonomics into the design process.
- Definition of the different ergonomic stages involved in the design process.
- Knowledge of the different factors involved in the design of handtools.

It is worth stressing, at this point, the relation between ergonomic design and the risk factors associated with MSD. Research into ways of improving ergonomic tool design dates from well before the systematic linking of tool design to MSD in the workplace.

The objectives that ergonomic study into handtools aims to achieve are constantly changing with technical progress, changes in the organization of work and the expectations of operators. Safety objectives, comfort and even considerations of style have been added to considerations of improved efficiency in tool design. To meet these new, interrelated requirements, researchers have resorted to a variety of methodological approaches derived from different disciplines (mechanics, physiology, psychology, sociology).

Hence, awareness that ergonomic design represents a means of preventing MSD leads to the addition of another objective to design specifications: the reduction of biomechanical and vibration risks. An "ergonomic tool" must be safe and efficient, and must also make fewer demands on upper limbs if risk of MSD is to be reduced.

The objective of this paper is to analyze these three requirements, first, to help manufacturers master ergonomic handtool design processes and, second, to help users develop specifications that will adequately express their particular needs in this field.

2. INTEGRATING ERGONOMICS INTO THE DESIGN PROCESS

The design process can be defined as a transformation of information allowing for movement from an initial concept to the final result. This process should be structured methodically. Indeed, it is generally accepted today that 75% of the total cost involved in the development and industrialization of a product is determined at the very outset of the design process.

Method in design process (Figure 1) involves an approach in which different "models" (language, technical experience, know-how. etc.), on the one hand, and different "tools", on the other, are brought together. By "tools", it is understood to be all the techniques involved in the analysis of the functioning of a part or the whole of the design process.

General design procedure involves a group of project participants (marketing, design, manufacturing) and a number of phases (definition of needs, specifications, general and detailed design). The approach adopted must achieve the highest possible integration of these different elements if the project is to run smoothly.

Traditional design process techniques have envisaged project participants and phases in a sequential manner. This sort of approach is characterized by:

- the erroneous belief that needs and specifications can be fixed once and for all at the outset of the design process;
- a difficulty in integrating project participants other than those directly involved in the different phases (ergonomics, etc.); and
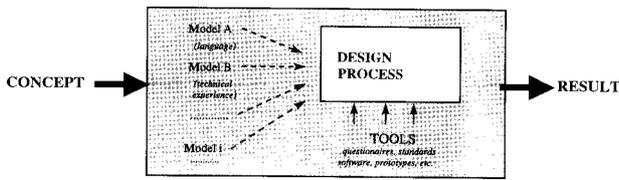
Figure 1. Method Structure.

• a difficulty in, or even absence of, communication between the different participants involved in non-consecutive project phases. (Lack of communication of this sort is aggravated by the fact that protagonists in the process have difficulty in understanding all the intermediary objects involved: technical drawings, nomenclatures, technical sheets, specifications, etc. Lack of communication of this sort is one of the main reasons for the non-inclusion of ergonomic recommendations in the design process.)

The result of such an approach is that considerable difficulties are to be had in introducing ergonomics into the design process.

To overcome such difficulties, two approaches are to be recommended. On the one hand, iterative models, among which are to be found the spiral model for phase organization, should be adopted. On the other hand, concurrent engineering for the management of the different protagonists in the process is advisable.

The notion of "concurrent engineering" appeared at the beginning of the 1980s. Without going into detail, this concept aims to provide for the simultaneous integration of process management and the different phases in product development. It is essentially an organizational device allowing for improved communication between the different project participants (including ergonomics). Concurrent engineering has been defined by several authors (Evbuomwam *et al.* 1994), one of which is "A concurrent development of project design functions, with open and interactive communication existing among all team members for the purpose of reducing lead time from concept to production launch".

The spiral model (Figure 2) makes use of both functional analysis and prototyping techniques. It allows for the integration of all project participants before completion of each design phase. In addition, prototypes — which are intermediary objects that
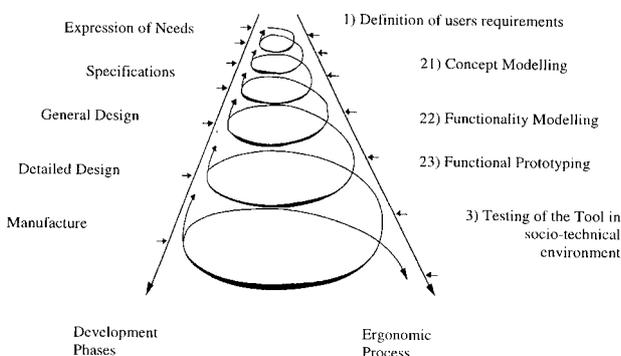
each project participant readily understands — re-center attention on the product and improve communication in general.

3. DEFINITION OF THE DIFFERENT ERGONOMIC STAGES IN THE DESIGN PROCESS

Increasing awareness that (1) the user, (2) the tool and (3) the workplace, environment and task are inextricably linked has led to reconsideration of the content and meaning of the design process. Indeed, each of the three elements emphasized above interacts with the two others. Figure 3 gives this interaction.

Conventional study of tool function should be completed by inclusion of task and user characteristics (Figure 3). Only by using ergonomics can this objective be achieved. Three basic stages can be outlined:

• Stage 1. Definition of user requirements (Figure 2) and expectations after detailed observation of the work process and work context. Employee characteristics (training, anthropometric measurements, etc.) are also defined during this phase. Tool specifications finally emerge from the study of user needs and work processes.

• Stage 2. Design of a new tool prototype based on tool specifications and laboratory simulation and study of the biomechanical solicitation produced by the new tool by comparison with the tool previously used. This stage includes all other phases — that means, concept modeling, functionality modeling, functional prototyping (Figure 2) — relating to tool function, styling, etc.

• Stage 3. After formal completion of the second stage, test of prototypes by a large sample of users in real workplace situations. The trial should be conducted over a sufficiently long period (several weeks) and feedback on the user's perception of the new tool should be regularly sought according to a procedure similar to that used in the first phase. If satisfactory results are obtained, the tool can then be considered duly certified for those situations in which it has been tried and tested. Users must be trained and encouraged to use the prototypes over a sufficiently long period before final judgement is given.

Such an approach illustrates the extent to which final users are at the center of design preoccupations as they are the ones who do the work. Nevertheless, it is also likely that manufacturers will be keen to refer to the positive commercial advantages that ergonomic handtool design provides. Certain manufacturers already do this.

Figure 2. Spiral model for an ergonomic development process. (See Section 2 for the three stages of the ergonomic process.)

Figure 3. Relations between users/tool/workplace.

4. CRITERIA INVOLVED IN THE DESIGN OF HAND TOOLS

This section deals with a number, but not all, of the different elements making up Figure 3.

4.1. Tool Design

- Tool mass — tool mass is directly linked to the muscular effort expended by the user.
- Center of gravity — if the center of gravity is not situated on the grasp axe with regard to handle grasp, greater muscular effort on the part of the user is required.
- Handle form and dimensions — it is important that the form and dimensions of the tool handle are adapted to the form of the user's grip and allow for a satisfactory distribution of pressure in the user's hand.
- Handle length — the handle must be longer than the size of the user's hand to avoid points of excessive pressure on the palm.
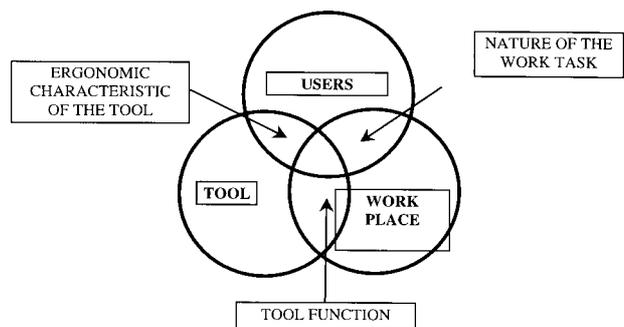- Handle material and texture — pressure distribution across the hand, shock and vibration absorption, electrical and thermal insulation, impermeability to certain chemical products and the possibility of allergic reactions to certain materials (chrome, nickel) should all be taken into account in choosing material. Handle texture determines the friction coefficient between the hand and the handle. This should be high to guarantee a firm grip when necessary and low when functions requiring a less firm grip are to be undertaken.
- Trigger — according to the type of work task involved and tool use required, the trigger on a power tool might be activated by the thumb, by one or several fingers or by the palm of the hand. Generally, surface contact for the trigger must be large enough and the thickness of the trigger fine enough to avoid points of concentrated pressure.
- Guards — the presence of a guard on certain tool handles (such as knifes) prevents the hand from sliding towards the functional part of the tool and thus reduces the risk.
- Inclination of the tool handle in relation to the functional part of the tool — according to the type of work task involved, the inclination of the handle in relation to the machine's functional part prevents the bad postures of upper limb.
- Vibration and reaction torque — exposure to vibration involves both the length of time the tool is used and the intensity of the vibrations transmitted to the upper limbs. Reduction of this vibration is achieved by vibration attenuating handle design. Similarly, it is important to reduce the reaction torque generated by certain rotating tools by introducing, for example, reaction torque bars.
- Other factors — other factors should be taken into consideration, such as noise, handle temperature for tools in which compressed air passes through the handle, the positioning and direction of compressed air exhaust pipes, etc.

4.2. User Comfort

- Anthropometric considerations — the use of anthropometric data relating to, for example, hand length, width and circumference is recommended in determining handle dimensions.
- Age — during tool testing, users of different ages should be chosen to eliminate the risk that the reduced physical capacities of older users are unduly taxed.
- Sex — handtool design should take differences between male and female hand size and physical force into account.
- Right- and left-handed users — tool design should cater for both right-handed and left-handed users.
- Experience and technique — instruction and training should be given, on a regular, systematic basis, to users with regard to tool use, strain reduction and performance optimization.

4.3. Environmental Factors, Work Tasks and Work Stations

- The right tool for the job — certain tools are adapted to particular work tasks. Different screwdrivers, for instance, are required for different work tasks (pistol right angle, in-line).
- Posture — user posture may influence the ability to exert an effort. Generally, the greater the distance between the user and the workplace, the greater the likelihood that an awkward or uncomfortable posture will be adopted. This results in diminished capacity to exert an effort.
- Physical conditions — the environment (thermal, lightening, noise) in which a tool is used should be taken into account during the design phase so that the tool does not become uncomfortable or even dangerous for the user.
- Gloves — according to the material used in their making, gloves can increase the external dimensions of the hand, reduce hand flexibility and sensitivity to touch or modify the friction coefficient. These all influence tool grasp factors.
- Tool supports and reaction torque bars — the use of certain accessories such as tool holders, etc. and reaction torque bars allows for diminution in the effort expended by the user.
- Tool maintenance — a well-maintained tool reduces biomechanical solicitation and improves performance and work quality.

All these factors are described in detail in the literature.

5. CONCLUSIONS

Recent scientific work illustrates the importance of adopting a participatory ergonomic approach to the study of tool design in authentic work situations involving habitual tool users. Ergonomic field observation techniques should be applied to handtools to provide specifications that will guide choices made by designers. It is generally admitted that no tool can in be considered "ergonomic". It can only be considered as such if it is properly adapted to all the different uses to which a specific user puts it. Given that work situations are by nature diverse, it is difficult to design a universal tool that is adapted to all possible work situations.

Moreover, in certain work situations, it may well be more profitable and efficient, as much in terms of production quality as in terms of reduced risk of MSD, to modify certain parameters of the work situation (the object on which the tool is used or an element in the production process, for instance) than to modify

the tool. Consequently, tool design includes not only the tool, as previously defined, but also advice, training and above all the ability to study the specific details of a given work process.

REFERENCES

EVBUOMWAM, N. F. O., SIVALOGANATHAN, S. and JEBB, A., 1994, A State of the art report on concurrent engineering. In *Proceedings of the conference Concurrent Engineering: Research and Applications*, 35–44.

FRASER, T. M., 1980, *Ergonomic Principles in the Design of Hand Tools*. Occupational Safety and Health Series, 9 (Geneva: International Labour Office).

GJESSING, C. C., SCHOENBORN, T. F. and COHEN, A. (Eds), *Participatory Ergonomics Interventions in Meatpacking Plants* (US Department of Health and Human Services, Public Health Service, CDC/NIOSH), pp. 94–124.

LAPRIE, J. C., 1996, *Guide de la sûreté de fonctionnement* (Toulouse: CEPADUES).

MITAL, A. and KILBOM, A., 1992, Design, selection and use of hand tools to alleviate trauma of the upper extremities: Part I — Guidelines for the practitioner. *International Journal of Industrial Ergonomics*, 10, 1–5.

POMIAN, J. L., PRADERE, T. and GAILLARD, I., 1997, *Ingénierie et ergonomie* (Toulouse: CEPADUES).

RADWIN, R. G. and SMITH, S. S. (Eds), 1993, A Strategy for Industrial Power Hand Tool Ergonomics Research: design, selection installation and use (Cincinnati, OH: NIOSH).

RADWIN, R. G. and HANEY, J. T., 1995, *An Ergonomics Guide to Hand Tools* (Fairfax: American Industrial Hygiene Association).

Product Development Approach

M. (Matti) Ameli

Toyota Technical center (LA, California) phone: (310) 787-5581
Fax: (310) 787-5555 e-mail: mameli@ttc-usa.com

1. INTRODUCTION

The main goal of human factors specialists, as part of the design team, is to help develop products and systems that are compatible with the users' logic, so they can use them easily and efficiently. As time goes on, the market for the majority of consumer products will become more competitive and therefore hard to penetrate because consumers have very high expectations: they want to choose the "perfect" product. To better comply with the expectations of the market, it is important for human factors specialists to go beyond the criteria of usability by taking into account other dimensions such as aesthetics (i.e., product appearance, style) and innovation. This requires a more active collaboration with all design partners and also a better preparation of students in human factors programs to face the realities of the industrial world.

2. PROBLEMS RELATED TO THE "TRADITIONAL" HUMAN FACTORS APPROACH

Human factors specialists are mainly faced with two types of problems that are connected:
- lack of consensus about important methodological parameters such as usability metrics and context specificity (is it possible to generalize results observed in one design context to other design contexts?) ;
- lack of credibility in the industry.

Because our discipline has not defined standard methods to measure usability, it is very hard for human factors specialists to argue with the quantitative-based logic used by design engineers. Also, the ambiguity in the context specificity limits the generalization of our results and recommendations. These methodological weaknesses convey a "fuzzy" image of human factors and therefore could lead to lack of credibility in the product development teams.

The lack of credibility is also sometimes due to the "conservative" attitude adopted by human factors specialists: engineers propose and initiate (active role), human factors specialists execute and follow (passive role). The usual approach of human factors specialists is to "wait" for the engineers to ask them to evaluate a given prototype or product. Such an attitude is perpetuated by opinions such as Perrow (1983) who suggests that upper management should make an effort to introduce human factors specialists to design engineers and publicize for them. However, I disagree with this premise. Why can't human factors specialists initiate discussion with engineers and other design partners to present and explain the usefulness of their methods in the product development process?

If we want to have credibility, not only should human factors specialists get to know directly all relevant design team members but we should also dare to become leaders and propose creative product ideas and concepts. Innovation is not the reserved domain of engineers. As Norman (1998) points out, human factors specialists are very good at criticizing products but are not very good at proposing good ones.

3. MARKET REQUIREMENTS FOR A "SUCCESSFUL" PRODUCT

The first attribute that attracts a buyer is the appearance of a product, its aesthetics, its style. The product should look "good". The other important criteria is usefulness: buyers look for products that are necessary or valuable for their everyday life. Of course, all users expect the product to be easy to use and reliable. These four aspects, aesthetics, usefulness, usability and reliability, in concert are strong determining factors in a purchase decision.

In the current market, many consumer products become more similar because previous ideas that worked are "copied" by everybody. Therefore, the market tends to become more "uniform". For example, electronic products such as cellular phones, televisions or VCRs have almost all the same functions and their reliability is comparable. Similarly, looking at compact or even medium size cars, even though their reliability has considerably improved in the few past years, they all offer basically the same features. Nowadays, in order to attract the customer, it is not enough to design a nice, useful, easy to use and reliable product; it becomes necessary to offer products that are different from others, products that are innovative. Innovation is an important key for a successful product in the market.

One of the best examples of successful and lasting innovation is the post-it, by 3M. It is a very simple concept that is flexible (different sizes) and that can be used by everybody and everywhere. Everybody in the product design team, including human factors specialists, can think of such concepts, it's just a matter of imagination!

4. MOVING TOWARD A MORE GLOBAL AND AGGRESSIVE APPROACH

Several steps are necessary for human factors specialists to become more efficient and persuasive in the product development process:

4.1. Being Known in their Companies

Human factors specialists should take the step to explain their purpose, methodology and approach to all the persons involved in the product development process by focusing on the positive impact that their input can have on the product. They should also be able to attract the upper management by making clear and convincing presentations with concrete examples of successful "ergonomic" products. Training could be set up to make everybody aware of the importance of human factors in the early stage of the product development.

4.2. Being Accepted by the Design Team

As Norman (1998) suggests, human factors specialists should be less self centered and more open to other disciplines. They should make more effort to understand engineers' "techno-centered" (focused on the technical performance of a product) approach. To deal more effectively with engineers, it is necessary to present quantitative and objective data whenever possible, and understand the value in an interaction between quantitative and qualitative approaches.

Human factors specialists should also develop very close relationships with the Marketing department. This can be achieved by learning the "business" language (Norman, 1998), and also by developing collaboratively tools such as questionnaires and focus groups to apprehend the customer preferences: a good product is one that is useful for the market and easy to use. Lund (1998) proposes to "measure" usability with three scales: ease of use, usability and satisfaction. Those evaluation scales are common to human factors and marketing. According to Caplan (1990), it is also useful for human factors specialists to borrow the focus group methodology from marketing and add quantitative techniques to it because their recommendations will be more likely to be understood and accepted by management. Interaction and collaboration with Marketing is critical to gain credibility within the company.

Aesthetics is a critical dimension of an attractive product. Human factors specialists should also be more sensitive to the stylists' approach and try to find a balanced compromise between usability and aesthetic dimensions.

To summarize, in order to be really accepted and appreciated by the product team, human factors specialists should take into account all design dimensions in their approach: their recommendations should go beyond the criteria of usability and reflect a balance between technical, aesthetic, marketing and usability parameters.

4.3. Being Innovative

Innovation is critical for a successful product in the current market. Therefore, Human factors specialists should include creativity in their approach. We should think more in terms of "unique" products and original ways to present a product to customers. In fact, we should combine expertise and imagination to propose innovative products to engineers (during brainstorming sessions, for example) and show them that we could also be pioneers and leaders in product development.

4.4. Being more efficient

Efficiency means doing the best job possible in a short period of time. In order for human factors to be perceived as efficient in the product team, it is necessary to:
- use shorter and less complex evaluation methodologies. Approximate methods (short and less detailed or rigorous evaluation methodologies) should be preferred to long and "precise" methods: "good enough is good enough" (Norman, 1998).
- Develop specific guidelines for different product domains (computers, cellular phones, cars) in order to avoid long user evaluations.
- Develop sharable databases of different methodologies, results and recommendations published in various human factors articles, for different products.

Building credibility in industry requires four factors:
1. being known by all relevant persons in the company (including upper management),
2. being accepted by the product team,
3. being innovative,
4. being efficient.

Formal education is the best place to prepare those interested in applied human factors to enter successfully into industry and deal adequately with product design partners.

5. TIME FOR A CHANGE IN TEACHING HUMAN FACTORS

Currently, most psychology oriented human factors programs prepare students to become good researchers and average practitioners. For example, students are required to know in details all the theories of information processing and master long and heavy methods of task analysis. Their "theoretical" preparation is excellent. However, after graduating, human factors students know very little about the realities that they have to deal with in the product team: the engineers' "techno-centered" approach, short product development timelines, management expectations, the importance of innovation and aesthetics etc.

It is important to balance theory courses with concrete and realistic examples of the role that a human factors specialist should adopt in the product team. It becomes crucial to "build" students' character in college in order for them to be able to defend themselves in the tough industrial world.

The following points would help students be integrated easier in the product team:
- the techno-centered approach should be explained with concrete examples of the "appropriate" ways to deal with it (e.g., correlate objective and subjective data);
- the "business" language (Norman 1998) should be taught with the help of marketing experts;
- product stylists should be invited to present their approach and the influence human factors has on it;
- experienced human factors specialists involved in product development should be invited to explain the short and "approximate" evaluation methods that they use to deal efficiently with the design partners;
- students should be put in situations where they have to propose an innovative product and defend it using marketing, aesthetics and usability criteria;
- students should be taught to "sell" their discipline to upper management by preparing clear and attractive slides demonstrating the importance of human factors in the early stages of product development;
- focus should be on proposing creative and original concepts and ideas rather than criticizing existing products;
- students should have "competitive" group exercises where they have to argue, justify and defend their recommendations for a given product against each other.

6. CONCLUSION

Unlike engineering or marketing, human factors is not considered as primordial for product development. In fact, in many companies, human factors is still considered as a "luxury". In order for human factors specialists to "survive" in product teams, the first criterion is to earn credibility by being proactive, innovative, efficient and persuasive. The survival of our discipline does not only depend on our methodological strength or the efficiency of our recommendations; it depends also on our character, on our ability to compromise, argument, justify and convince.

It is about time to "revolutionize" our mentalities and become leaders in product development. It is about time to prepare ourselves and future human factors experts to enter the twenty first century with more confidence, an aggressive approach and a broader knowledge of other disciplines involved in product development.

REFERENCES

CAPLAN, S., 1990, Using focus group methodology for ergonomic design. *Ergonomics*, 33, 527–33.

LUND, A.M., 1998, The need for a standardized set of usability metrics. Proceedings of the human factors and Ergonomics Society 42nd annual meeting (pp.688–90).

NORMAN, D. A., 1998, *The Invisible Computer* (London: the MIT Press).

PERROW, C., 1983, The organizational context of human factors Engineering. *Administrative Science Quarterly*, 28, 521–41.

REECE, W.J., and BLACKMAN., H.S., 1998, Cross-industry performance modeling: toward cooperative analysis. Proceedings of the human factors and Ergonomics Society 42nd annual meeting (pp.697–91).

Safety in Public Offices in Italy

A. Baldacconi

INAIL (Italian State-Owned Workers' Compensation System), University of Urbino, Italy

1. INTRODUCTION

A novel approach to international organization and coordination has emerged with the Community Framework Directives 89/391/EEC, 89/655EEC, 89/656/EEC, 90/269/EEC, 90/270/EEC, 90/394/EEC, and 90/679/EEC aimed to extend health, safety, protection and prevention in working environments.

The emerging theme is one that applies not only to occupational health and safety but also to other areas of social regulation where command/control instruments have been used extensively to environmental health and safety regulation.

In Italy such a directives were accepted by two Acts of the Italian government: Legislative Decree of 19 September 1994, no. 626 and Legislative Decree of 18 March 1996, n. 242.

Even if the former laws had been applied in the past, the above EEC directives are going to give a "New Approach" as "New Frontiers in Occupational health and safety" dealing with a "Management System Approach according to the ISO Model".

The new approach represents a bridge between the past and the present to encourage active debate on the issues raised and to avoid the inefficiencies and the limited applicability by tying environmental health and safety performance to market participation and procurement contract language.

2. PUBLIC OFFICES

2.1. Development of Requirements

In public offices usually the importance of a high-quality workplace has never been more evident than in today's climate of rising real estate costs and increasing business competition to give comfort and safety.

To attract and retain high-caliber employees, an organization must be able to offer a safe, comfortable, and stimulating work environment. To support those employees in executing their responsibilities and to maximize their contribution to the organization, the office must be efficient, functional, and flexible. Meeting these objectives requires that an organization plans strategically and designs facilities to meet the needs of the organization both now and in the future.

It is necessary to identify the major technical and design issues related to office environment, and to communicate those issues in terms that can readily be understood by the manager who does not have training in the interior architectural professional but is responsible for the planning and design of organizational facilities.

The safety management system has to guide through:

- office planning end design: *the strategic facilities planning process to understand in detail how to lay the proper groundwork for a specific interiors project by identifying clear objectives and selecting a qualified team;*

- reference to specific issues: *the various planning components of the office, building support system, and elements of interior design to yield information about planning and design consideration such as general office areas, support areas, government requirements, mechanical and electrical engineering systems, furniture, ceiling systems, lighting, color, acoustics, and finishes.*

This is the way to promote better understanding of the factors that constitute a high-quality workplace and the elements that combine to ensure successful implementation in health and safety. The objective in the development of requirements is to determine the total occupiable space required for the organization currently and in the future.

To develop the organization's requirements, the strategy facilities planner collects both quantitative and qualitative needs through management and personnel interviews, questionnaires, and facility surveys. For example, programming during a design project would identify a specific number and type of file cabinets, the information necessary to plan and design the space, the characteristic that would be affected by the building architecture or building support system (mechanical and electrical systems). In this way the strategic facilities planner evaluates existing planning standards to determine their adequacy for use in the study of planning standards and furniture inventory — workstation planning standards identify configuration for administrative through executive levels, equipment standards identify space to accommodate equipment and stand space for operating the equipment/support area such as conference room. Once the strategic facilities planner has compiled the quantitative and qualitative requirements for the organization, the data are analyzed in the context of trends, sensitivities, and potential risks. The purpose of conducting this analysis is to test the assumptions made during the study and the validity of the information developed, and to investigate the potential risk associated with each alternative solution. Ultimately, the sensitivity and trend analysis will assist the strategic facilities planner in designing an appropriate amount of flexibility in to the plan to accommodate the potential risk.

The first step in the trend and sensitivity analysis is to review all organizational and industry trends to determine if the projected requirements for the organization are consistent with those trends. If any inconsistencies arise, the strategic facilities planner must identify the causes and their potential impact on future solutions.

The second step in conducting the trend and sensitivity analysis is to determine the sensitivity of the study's result. To accomplish this, the strategic facilities planner must identify all factors affecting the validity of information and decisions made during the strategic plan's overviewing.

The third step is to assign potential risk to each factor. For instance, the strategic facility planner has assumed that the program information obtained during the development of requirements is accurate. In an office the interaction between man-machinery-environment must take into proper account the workstation standards.

In fact, a workstation is the space allocated to house a person to execute a job or task on an ongoing basis. The term workstation refers both to a private office with full height partitions and a door, and to an open-plan "cubile" configured from systems furniture or low-height partitions. The term standard indicates

that the workstation size and configuration are assigned on a standard basis for all personnel performing a similar function or at a comparable organizational level or job classification. Ideally, the number or workstation standards are kept to a minimum to simplify planning, future inventory, and facility maintenance. Workstation standards are generated based on:

- functional requirements of each position,
- organizational culture,
- organizational status,
- industrial or professional standards.

To ensure that workstation standards are responsive to user needs, it is necessary to survey and interview representatives from each job classification or standards category to solicit detailed information that will affect workstation size, components, and configuration of those components. The workstation standard size is the space necessary to accommodate:

- tasks performed,
- surface area requirements,
- technology/equipment requirements,
- storage requirements,
- conference/meeting requirements.

To develop the workstation standards it has to be satisfied of the following requirements:

1. Job classification function: *workstation user and job function: executive, manager, supervisor, professional/ technical, clerical.*

2. Worksurface area: *number and sizes of worksurfaces — primary, secondary, tertiary.*

3. Machine use: *amount, types, and sizes of electronic equipment — VDT, PC, printer, typewriter, other.*

4. Workstation area: *amount of space to be allocated for individual or task.*

5. Workstation dimensions: *length and width that will comprise the area.*

6. Conference requirements: *number of guest chairs.*

7. Storage requirements: *amount and type or unit size of the material to be stored and storage locations: letters or legal files, computer printout, binder, bulk.*

8. Configuration: *configuration of the worksurfaces, primary orientation, and opening for the workstation.*

9. Wire management: *type and location of wire management related components: baseline wireway; beltline wireway; grommet locations; wire management clips or trays.*

10. Lighting: *quantity and location of any task or ambient lighting components.*

11. Accessories: *type and number of accessories: task surfaces; pencil drawers; other.*

Once the decision-making process is in place, the team can develop a project definition for use in assembling a project team of professionals and in developing a strategy for resolution of the requirement, including specifically the two Acts of Italian government. A comprehensive project definition addresses the following issues: (a) project need, (b)reason for the need, (c) goals for the project, (d) parameters for a solution, (e) predisposition to a solution. According to the above issues many public offices must be subjected to interior redesign because they were set up in historical sites before and after the Second World War.

3. ACTS NO. 626/94 AND NO. 242/96

3.1 Objectives and Scope of the Acts

The objectives of these Acts are:

- To secure a working environment which affords the employees full safety against harmful physical and mental influences and which has safety, occupational health, and welfare standards that correspond to the level of technological and social development of the society at large at any time;

- To provide a basis whereby the enterprises themselves or the public administrations can solve their working environment problems in cooperation with the organizations of employers and employees and under the supervision and guidance of the public authorities;

- To provide in any working environment a unit so named "Prevention and Protection Services" (working environment committee). The aim of this unit is to edit a paper on all sources of risk. According to Webster's Unabridged New International Dictionary, risk is ''hazard; danger; peril; exposure to loss, injury, disadvantage, or destruction''. It distinguishes risk from hazard by suggesting that a risk is more often voluntary, a hazard the product of chance. These definitions contain the essentials of risk, combining the idea of loss with that of chance or probability because we need to know the probability, the likely consequences, and the best ways to reduce the probability and the consequences. The probabilities come from probabilistic risk assessments (more later) with inevitable uncertainties, and the estimates are apt to generate disagreement and confusion. Multiplying the probability of the event by the amount of potential loss it comes out what is called the product of expectation of loss, used as a final measure of risk. The working environment committee shall work to establish a fully satisfactory working environment and shall participate in planning safety and environmental work and shall follow up developments closely in questions relating to safety, health, and welfare of employees.

- To organize and plan health and safety according to the above paper, especially regarding the Matrix of Risk, with the aim to control, reduce or to leave out the expectation of loss and the potential hazards.

- To organize and plan information, instruction, assistance, consultation, and training.

The main scope of these Acts are:

- Provided that it does not expressly state otherwise, these Acts apply to all enterprises that engages employees;

- For the purpose of these Acts "employee" shall mean any person who performs work in the service of another. If an enterprise is conducted by two or more persons jointly for their own account, only one of those persons shall be deemed an employer under these Acts, the others being regarded as employees.

- For the purpose of these Acts ''employer'' shall mean any person, physically or legally representative, who has engaged employee(s) to perform work in his service. The provisions of these Acts relating to the employer shall apply correspondingly to the person conducting the enterprise in the employer's stead.

- The working environment in the enterprise or in the public administrations shall be fully satisfactory when the factors in the working environment that may influence the mental and physical health and welfare of the workers are judged separately and collectively (matrix of risk).
- The workplace shall be arranged so that the working environment is fully satisfactory as regards the safety, health, and welfare of employees. In particular, it shall be ensured that: (a) workrooms, passageways, stairways, etc. are suitably dimensioned and equipped for the activities being conducted; (b) good lighting is provided, if possible with daylight and a view; (c) climatic conditions are fully satisfactory as regards volume of air, ventilation, humidity, draughts, temperature, etc.; (d) pollution in the form of dust, smoke, gas, vapours, unpleasant odors, and radiation is avoided, unless it is known that the pollution cannot lead to undesirable effects upon employees; (e) noise and vibration is avoided or reduced to prevent undesirable effects upon employees; (f) the necessary precautions are taken to prevent injury to employees from falls and falling or sliding objects or masses; (g) precautions are taken to prevent fire and explosions, and to provide adequate means of escape in the event of fire, explosions, and other emergencies; (h) sanitary installation and welfare rooms are satisfactory in size and design; (i) workrooms, sanitary, and welfare rooms, etc. are kept in good repair, clean, and tidy; (l) first-aid equipment is readily accessible.
- Technical apparatus, machinery, and equipment shall be designed and provided with safety devices so as to protect employees from injury and disease; when technical apparatus, machinery, and equipment are being installed and used, care shall be taken to ensure that the employees are not exposed to undesirable effects from noise, vibration, uncomfortable working position, etc.; technical apparatus, machinery, and equipment should be designed and installed so that it can be operated by or be adapted for use by employees of varying physique; technical apparatus, machinery, and equipment shall always be maintained and attended.
- Toxic and other noxious substances must not involve a health hazard and so working processes and other work shall be fully satisfactory in order that employees are protected against accidents, injury to health, and excessive discomfort. The employer shall keep a record of such substances showing the name of the substance, its composition, physical and chemical properties, as well as information concerning possible poisonous effects (toxicological data), elements of associated risk, preventive measure, and first aid treatment.
- Conditions shall be arranged so that employees are afforded reasonable opportunity for professional and personal development through their work; technology, organization of the work, working hours and wage systems shall be set up so that employees are not exposed to undesirable physical or mental strain and so that their possibilities of displaying caution and observing safety measures are not impaired.
- The employees and their elected representatives shall be kept informed about the systems employed for planning and effecting the work, and about planned changes in such systems. They shall be given the training necessary to enable them to learn these systems, and they shall take part in planning.
- If the work involves particular hazard to life and health, a special directive shall be issued prescribing how the work is to be done and the safety precautions to be observed, including any particular instruction and supervision.
- Employees shall take part in the creation of a sound, safe, working environment by carrying out the prescribed measures and participating in the organized safety and environmental work of the enterprise; employees shall perform their work in conformity with orders and instructions from superiors and shall use the prescribed protective equipment, display caution, and otherwise cooperate to prevent accidents and injury to health.

4. OCCUPATIONAL HEALTH IN AN OFFICE ENVIRONMENT

It has been estimated that in Italy more than 12 million people are now working in offices or in settings similar to offices. Like the industrial environment, there are also some potential health and safety hazards in offices, which, if unattended, may cause illness, injury, and discomfort. For this reason, INAIL — Italian State-Owned Workers' Compensation System — has published a guide entitled "Working in safety and comfort in an office environment" in accordance with the above Acts.

The office environment is usually considered comfortable and healthy, but there are various factors which could affect or influence the health of the staff. In mild cases, complaints usually involve tiredness, nose and throat discomfort, back pain, eyestrain, headaches, tiredness, repetitive strain injury; more rarely health factors may include respiratory diseases (extrinsic allergic alveolitis, humidifier fever, asthma, allergic rhinitis), sick-building syndrome, indoor air pollution (nitrogen oxides, mineral fibers, radon, formaldehyde, solvents, environmental tobacco smoke, contributions from outdoor pollutants — traffic and industrial pollution, etc.).

An office workstation is the specific area that includes desks, chairs, footrests, VDT, plan files, printer tables, copying machines, storage cabinets, lighting, air conditioning, noise, and other environmental risk factors. The design of the workstation must be related to the ergonomic principles that include layout, seating, clearances, and adjustments to accommodate individual differences in size and strength. Failure to do so may result in errors, reduced efficiency, discomfort, and sickness or even injury

The design of a comfortable working posture is important because most office work is sedentary and bad postures can cause musculoskeletal injury, discomfort, pain and/or strain of the back, neck, shoulders, upper arms, knees, and feet.

Lighting is another important factor because if it is not well-designed, workers are unable to see comfortably; good lighting also helps to reduce accidents, eye strain, and discomfort.

Efficient ventilation systems are necessary to control airborne pollutants (for example, gases, dust, fungal spores, viruses, bacteria, and mites).

REFERENCES

EASTMAN KODAK COMPANY, 1983, Ergonomic Design for People at Work (New York: Van Nostrand Reinhold).

ITALIAN STATE-OWNED WORKERS' COMPENSATION SYSTEM (INAIL), 1996, Commentario alla Sicurezza del Lavoro:d.lgs 626/94 e 242/96 (Milano: Il Sole 24ore Pirola).

KROEMER, K.H.E., 1994, Ergonomics: How to Design for Ease and Efficiency (New Jersey: Prentice-Hall).

KROEMER, K.H.E., 1997, Fitting the Task to the Human: A Textbook of Occupational Ergonomics, 5th ed., (London: Taylor & Francis).

LESLIE, G.B. and LUNAU, F.W., 1994, Indoor Air Pollution: Problem and Priorities (Cambridge: Cambridge University Press).

MATTALUCCI, L., 1992, L'Analisi del Lavoro d'Ufficio (Milano: Franco Angeli).

Ships and Maritime Systems: Design Process

T. B. Malone, Ph.D. and C. C. Baker

Carlow International Incorporated

3141 Fairview Park Drive Suite 575

Falls Church VA 22042

1. INTRODUCTION

One of the more important tools for the human factors analyst to design systems in terms of the capabilities and limitations of humans is a standardized, formalized, and comprehensive human factors design process. This process is comprised of the activities to be conducted by human factors specialists in the determination of requirements and concepts for (a) improving human performance and safety while reducing error potential, workloads and manning levels, and (b) designing and evaluating human-machine interfaces.

A generalized human factors and engineering design process is presented in Figure 1. The phases of the process correlate with the phases of ship design described in Malone and Baker (1996). This process is focused on specific human factors outputs for each process phase, as indicated in the figure.

The following sections address each phase of system design.

2. HUMAN FACTORS CONCEPTUAL DESIGN PROCESS

Human factors objectives in this phase are to develop concepts for the roles of humans vs automation in conducting system functions, and to assess concepts in tradeoffs and modeling and simulation. In the conceptual design of ships and maritime system design, the major activities are those involved in a top down requirements analysis (Bost *et al*, 1998, Bost *et al*, 1996, Anderson *et al*, 1998, Anderson *et al*, 1997, Malone *et al*, 1996), as depicted in Figure 2.

The activities shown in Figure 2 are described below.
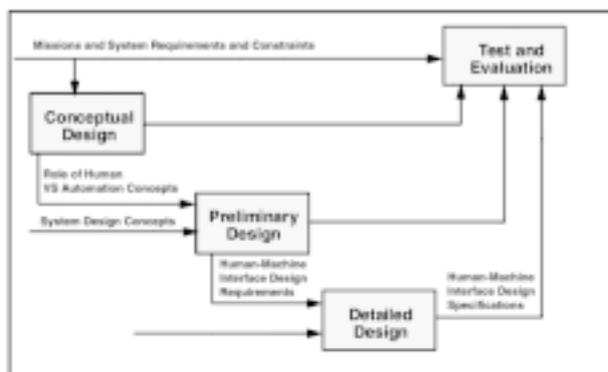
Figure 1. Ship/maritime system development HFE design process.

2.1. Identify and Analyze Missions and Functions

The initial step is to identify missions, mission requirements, and mission scenarios. Scenarios will include the ships and maritime system requirements, system top level functions by mission, and conditions under which functions will be performed.

System functions constitute the major activities to be performed by the system. The ships and maritime system function analysis should, for each scenario, identify functions to be performed and provide a functional flow depicting the sequence of functions for the system completing a mission scenario. The functions will be decomposed to successively greater levels of detail in an iterative manner based on requirements associated with each function.

2.2. Conduct Function Allocations and Identify Roles of Humans

Through a reverse engineering process, functions and tasks in existing systems which impose heavy workloads on humans or which have performance or safety problems associated with them can be identified, and requirements for alternative allocations to humans or automation can be specified. The rationale for allocation decisions in the existing ships and maritime systems can be made explicit and opportunities for alternate allocations can be explored. Alternative role of the human concepts involve alternate approaches to automation, providing decision aiding to reduce human workload, leading to reduction in manning, and improved design of human–machine interfaces to simplify tasks and reduce workloads.

Figure 2. Human factors activities in the conceptual design phase.

2.3. Identify Human Factors Alternate Concepts and Technologies

The human factors approach to ships and maritime system design is focused on design requirements associated with automation, consolidation, simplification, and elimination of functions. Feasible concepts will be selected based on the results of simulation exercises assessing the workloads and performance problems associated with alternate human factors conceptual strategies. For these concepts, assessments will be made of how to reduce the manning required at functional duty stations, and the impact of manning reductions on readiness, performance effectiveness, and safety.

2.4. Conduct Simulations to Assess Workloads and Human Performance

The next step is to identify workloads and manning requirements associated with alternate function allocation schemes. The assigned roles of human vs automation for each function and task will be assessed in terms of impact on workloads and human performance through use of the task network simulation. The simulation identifies potential performance problems and quantifies the workload of operators for a simulated mission under the candidate function allocation strategies.

The net result of the application of the simulation is a first approximation of which roles of the human are feasible, what workloads are associated with these roles, what problems are to be expected in specific role of human models, and what human performance characteristics should be further investigated.

2.5. Assess Design and Readiness Requirements

Feasible manning reduction concepts will be selected based on the results of simulation exercises assessing the workloads and performance problems associated with alternate manning strategies. For these concepts, assessments will be made of how to reduce the manning required at functional duty stations, and how to enhance human performance, workload and safety.

3. HUMAN FACTORS PRELIMINARY DESIGN PROCESS

In the Preliminary Design Phase the details are defined for individual systems and subsystems necessary to meet system functions identified in the conceptual design phase. The major human factors activities in the Preliminary Design Phase are depicted in Figure 3 and are described below.

3.1. Conduct Task Analysis

A task analysis represents a model of human task sequences in the conduct of a series of functions, and identifies requirements associated with performance of each task. Task analysis provides the bases for making design decisions; e.g., determining, to the extent practicable and before hardware fabrication, whether system performance requirements can be met by combinations of anticipated equipment, software, and personnel, and assuring that human performance requirements do not exceed human capabilities. This analysis shall also be used as basic information for developing preliminary manning levels, equipment procedures, skill, training and communication requirements, and as logistic support analysis inputs, as applicable.

Figure 3. Preliminary design phase human factors activities.

3.2. Conduct of Human Factors Studies

Human factors studies are conducted in support of the identification of human-machine interfaces and requirements. Studies include additional analyses, (such as cognitive analysis, information handling analysis, timeline analysis, and workload analysis) and empirical (laboratory or simulation) investigations of human performance capabilities and limitations for specific interfaces.

3.3. Identify Human–Machine Interface Requirements

3.3.1. Analyze functional interfaces

The elements of functional interfaces include (a) the roles of humans versus automation in system operation, control, maintenance and management; (b) human functions and tasks; and (c) roles of system personnel in automated processes (e.g., monitoring, management, supervision, intervention, etc.).

3.3.2. Analyze informational interfaces

These interfaces constitute the information needed by a human to complete a function or task, required characteristics of the information (source, accuracy, currency, quantity), and protocols and dialogues for information access, entry, update, verification, dissemination and storage.

3.3.3. Analyze environmental interfaces

This class of interface is concerned with the system's physical environment (illumination, noise, temperature, vibration, ship motion, weather effects, etc.), workspace arrangement, facility layout and arrangement, and environmental controls. This class of human interfaces will be optimized by determining requirements for environments which are within performance, comfort and safety limits, designed in terms of task requirements with consideration for long term as well as short term exposures.

3.3.4. Analyze operational interfaces

Operational interfaces include operating, maintenance, and emergency procedures; workloads; personnel skill requirements; personnel manning levels; and system response time constraints.

The major impacts of operational interfaces are on human error probability, and safety. Design criteria for procedures address the extent to which required levels of human performance can be assured given time constraints. Human factors improves the accessibility, content, and organization of procedures by ensuring that the procedure is complete, correct, clear, concise, current, consistent, and compatible with the reading/language/skill levels of the users.

3.3.5. Analyze organizational interfaces

Organizational interfaces include the factors impacting the organization of system management functions, policies and practices, personnel jobs, and data. Criteria for optimization of organizational interfaces include determining that position descriptions are based on functions allocated to the position and include duties, jobs, responsibilities, levels of authority, tasks, and decisions appropriate for each position; that assignment of duties and tasks to each position is realistic; that duties and jobs are consistent with those found in existing systems; and that data required to perform functions and tasks are available, current, and identifiable.

3.3.6. Analyze cooperational interfaces

These interfaces are primarily concerned with communication, collaboration, and team performance. Human factors objectives in optimizing communications are directed at improving both the media and the message. Specific requirements for media design include speech intelligibility and communications device operability. human factors concerns for the message include message standardization, use of constrained language, controlled syntax, and restricted vocabulary, methods of coding message priority, and human error potential in message transmission.

3.3.7. Analyze cognitive interfaces

Components of cognitive human interfaces include decision rules, information integration, problem solving, instructional materials and systems, short term memory aids, cognitive maps, and situational awareness. Design requirements for cognitive interfaces focus on design for usability, and conceptual fidelity. A major cause for human error is the fact that the human is operating on the bases of erroneous cognitive expectancies concerning what the problem is, what the system is doing, and how it will respond. In attempting to diagnose a problem event, an operator relies on expectancies. These expectancies are developed based on information presented to the operator, his or her procedures and training, past experience, design conventions, and, when all else fails, intuition. Expectancies will support the diagnosis when the cognitive model that the operator has of the system is in close agreement with what is actually happening, i.e. has high conceptual fidelity.

3.3.8. Analyze physical interfaces

Physical interfaces include the physical, structural, and workstation elements with which the human interacts in performing tasks. Interfaces include: workstations, control panels and consoles, displays and display elements (screens, windows, icons, graphics), controls and data input and manipulation devices (keyboards, action buttons, switches, hand controllers), labels and markings, structural components (doors, ladders, hand holds, etc.), and maintenance design features.

3.4. Integration of Human–Machine Interface Requirements

Human-machine interface requirements are integrated through conduct of human factors studies and through requirements assessment. The assessment of interface requirements will focus on the extent to which interfaces will address human error potential of specific functions and tasks, and the extent to which interfaces will support the design of the system to be error-tolerant.

4. HUMAN FACTORS DETAILED DESIGN PROCESS

In the Detailed Design Phase of a ship or maritime system design and development, human factors issues of workspace design, control and display layouts, and environmental factors are addressed. In this phase the actual design features of human-machine interfaces are defined and developed. The steps to be achieved in this phase are presented in Figure 4. The activities to be conducted in this phase are described below.

4.1. Conduct of Human Factors Studies

Human factors studies are conducted in support of the development of design concepts and criteria. Studies include: Human-Error Likelihood Analysis; Tradeoff Analysis; and Modeling and Simulation.

4.1.1. Human-error likelihood analysis

The human-error likelihood analysis identifies tasks and task sequences which are critical from a systems effectiveness, and human and public safety point of view. For each task an identification will be made of the types of errors which have occurred in existing systems, or which could be postulated on the bases of human performance requirements.

4.1.2. Tradeoff analysis

Tradeoff studies will address the comparison of alternate design concepts, leading to a selection of an optimal concept.

4.1.3. Modeling and simulation

Modeling and simulation will enable conceptual or actual performance of human operators and maintainers with aspects of alternate design concepts. In this manner, the potential for human error can be assessed through observation of task performance.

4.2. Define Design Concepts and Criteria for the Design for Operability

System elements which impact human-machine interface design for operability include workstations, I/O hardware, software, data bases, networks, computation systems, peripheral devices, communications systems, and software engineering environments. Human–machine interfaces include displays, displayed information, display characteristics, display formats, integration of displays, labels, instructions, alarms, symbology and graphics, decision aids, decision support systems, input devices, data designation and manipulation devices, controls and controllers, control systems, control and display arrangements, communications, workspace layout, workspace environment, help features, embedded training, intelligent tutoring systems, and procedures.
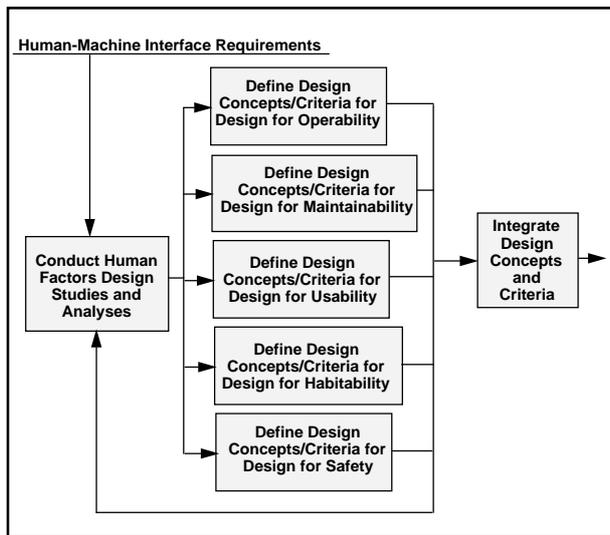
Figure 4. Human factors activities in the detailed design phase.

4.3. Define Design Concepts and Criteria for the Design for Maintainability

Maintainability design requirements include information requirements, design for accessibility, equipment arrangement to facilitate maintenance, procedures, troubleshooting diagnostics and decisions, skill levels and maintenance training, equipment design for maintainability, allocation of maintenance responsibility to man or machine, equipment installation requirements, requirements for special tools and support equipment, job aid requirements, communication requirements, facility design requirements, and safety design requirements.

4.4. Define Design Concepts and Criteria for the Design for Usability

The development of detailed HMI design for usability will focus on prototyping HMI concepts to assess and reduce the risks associated with integrating available and emerging human factors technologies into a system design approach to satisfy a validated mission need. Test and evaluation of prototypes will confirm the feasibility of specific design approaches relative to its ability to satisfy the mission need and achieve minimum acceptable operational performance requirements within affordability constraints. Prototyping will be used to assess cost and performance tradeoffs.

4.5. Define Design Concepts and Criteria for the Design for Habitability

Habitability design involves specifying workspace free volume, environmental effects, traffic patterns, workspace layout, facility compartmentalization, quality of life, and adequacy of the design for habitability. The human factors concepts to be developed will address the major user-machine and user-facility interface issues. Human factors concepts will either be developed or will reflect an assessment of architectural/engineering design concepts from a human factors point of view.

4.6. Define Design Concepts and Criteria for the Design for Safety

The development of human–machine interface design concepts and criteria as they relate to safety will be concerned with identifying, evaluating, and providing safety considerations or tradeoff studies to identify concepts for: guarding the hazard; labeling the hazard; alarming the hazard; training/procedures for avoiding the hazard; or designing out the hazard

The effort will entail the review of appropriate engineering documentation (drawings, specifications, etc.) to make sure safety considerations have been incorporated. These activities will extend to reviewing logistic support publications for adequate safety considerations, and ensuring the inclusion of applicable USCG, EPA, and OSHA requirements; verifying the adequacy of safety and warning devices, life support equipment, and personal protective equipment; and identifying the need for safety training.

4.7. Integrate Human Factors Design Concepts and Criteria

Design concepts and criteria will be integrated through modeling and simulation efforts which will produce prototypes of the interfaces for selected scenarios.

5. HUMAN FACTORS TEST AND EVALUATION PROCESS

The initial activity in human factors test and evaluation is to identify ship and marine equipment, systems and operations which are expected to be high risk from a human error point of view. High risk situations are those for which human error likelihood is relatively high and those for which human errors, whatever their likelihood, would produce results catastrophic to human or environmental safety.

Requirements and constraints for human factors T&E will then be specified. This will begin with an identification of constraints, including time limitations, legal barriers to evaluations, and availability of data. Requirements for evaluations include functional requirements, information requirements, performance requirements, decision requirements, support requirements, and interface requirements.

The next step will entail identification of human factors evaluation scenarios. When the requirements for evaluation of high risk equipment, systems, and operations have been identified, evaluation scenarios will be described. These scenarios include tasks and test conditions to be included in the evaluation.

The next step will be to identify human factors evaluation measures, criteria, and data requirements including the actual data required from the evaluation, and factors influencing the quality of these data.

The final steps in test and evaluation is to integrate the T&E requirements into a human factors test plan, conduct the tests and evaluations according to the plan, and analyze and interpret T&E data.

REFERENCES

ANDERSON, D.E., MALONE, T.B. and BAKER, C.C. (1998) "Recapitalizing the Navy through optimized manning and improved reliability". Paper submitted to the Naval Sea Systems Command's Association of Scientists and Engineers for the AS&E 1998 Symposium.

ANDERSON, David E., OBERMAN, FREDERICK R., MALONE, THOMAS B., and BAKER, CLIFFORD C. (1997), (6):67–76. "Influence of human engineering on manning levels and human performance on ships". *Naval Engineers Journal*, Vol. 109, (6): 67–76.

BOST, J.R., BAKER, C.C., HEASLY, C.C. and MALONE, T.B., (1996) "Technology and people: striking a balance". *Marine Log Conference on Marine Technology*.

BOST, J.R., ANDERSON, D.E., Malone, T.B., Baker, C.C., Heasly, C.C. and Williams, C.D. (1998) "Human-centered automation in total ship engineering". Paper accepted for presentation at the ASNE Engineering the Total Ship Symposium 1998.

MALONE, THOMAS B., BAKER, C.C. and ANDERSON, David E. (1996). "Advanced manning designs and tools for marine structures". Society of Naval Architects and Marine Engineers.

Ships and Maritime Systems: Requirements and Issues

C. C. Baker and T. B. Malone

Carlow International Incorporated, 3141 Fairview Park Drive, Suite 575, Falls Church, Virginia 22042, USA

1. INTRODUCTION

This article presents an overview of the performance requirements placed on humans in maritime systems and on many of the personnel issues associated with humans working at sea. Manned maritime systems are small floating and mobile societies that are, for varying lengths of time, entirely self supportive and autonomous. Nearly all human needs of land based society must be fulfilled by at-sea systems, including providing; sustenance, recreation, transportation, communication, waste disposal, socialization, shelter, warmth, sleep, hygiene, and medical needs.

The missions of the maritime platforms themselves are varied, but generally involve one of the following: transportation of goods (merchant ships for carrying varying cargo); projection of power (military and maritime law enforcement); or mining (fishing industry, oil and natural gas recovery). To meet it's mission, the marine system must: generate power, navigate, manage waste, provide environmental and self protection, and accommodate the human inhabitants. All of these requirements must be met under the extremely diverse and often harsh conditions of the sea (Anderson, et.al. 1998).

2. HUMAN FACTORS, SYSTEM MISSIONS AND FUNCTION REQUIREMENTS

2.1 Mission–Function Analysis

Missions of maritime systems are the goals, or objectives, of what is intended to be done or accomplished. A typical mission for a merchant container ship, for example, would be to load containers beginning at point A, transport them to Point B, and from point B, to pick up additional containers and transport them to other points. All of this cargo hauling would also happen within the constraints of shipping schedules, fuel costs, and so on. A mission then is what the system is expected to do within a set of constraints. Functions are mechanisms by which mission requirements are performed. For example, moving cargo from point A to point B suggests a need for motive power, and therefore, a system propulsion function exists. Mission-Function analysis is a technique for identifying and analyzing system functions that must be provided to meet system objectives. Functions are analyzed hierarchically, such that high level functions spawn lower level functions. For example, providing motive power can be broken down to sub-functions, such as: generation of mechanical power, providing cooling to mechanical processes, providing for equipment monitoring and control, and providing means to diagnose malfunctions. Table 1 shows a highly generalized top-level function breakdown for a notional military ship.

In summary, (1) mission statements identify *what* is to be accomplished by the system, and (2), functional analyses and

Table 1. Typical Total Ship Functional Breakdown (top-level)

Mission Operations (vary by mission)	Special Operations
Boat Handling/Boarding	Underway Replenishment
Area Surveillance	Electrical Failure
Communications – Internal	Extreme Weather
Communications – External	Fire – Large and Small
Weapon and /Combat Information	Flooding/Ballast control
Helicopter or other air operations	Collision/ Grounding/ Stranding
NOAA/Weather services	Internal Security
War fighting	Loss of Propulsion
Cargo Handling	Tug and Pilot interfaces
Oceanographic	Boarding/Law Enforcement
Search and Rescue	Anti-Terrorism
Deck Operations	Man Overboard/Rafters
Anchoring	Fuel Spills/Environment Hazards
Docking Undocking	Lifeboat
Helo Operations	**General Operations**
Boat operations	Bridge - Housekeeping
Line Handling/Mooring	Bunkering (Refueling)
Anchor	Deck Equip Maintenance
Towing Operations	Direct Shore Gangs
Underwater Lighting	Docking/Undocking
Navigation/Bridge	Line/wire Maintenance
Approach Berth	Medical
Berthing	Stores Breakdown
Collision Avoidance	Steering Gear Maintenance
Depart Berth	Stores Handling/Breakdown
Hull Performance/Station Keeping	Structure Maintenance
Maintenance	Wash down - Deck
Maneuver	Wash down - Engine
Lookouts	**Administration**
Signals	Reporting
Position Fixing	Health Care
Record/chart Keeping	LAN Maintenance
Track Keeping	Finance/Payroll
Voyage Planning	Ships Meetings
Weather Monitoring	Mission Planning
Hotel/Unit Support	Shore Operations
Catering/Messing	Stores loading
Laundry	Port Logs/Records
Provisioning	**Maintenance – Ship area**
Space Cleaning	Vital Systems Test
Waste Disposal	Equipment Surveillance
Maintenance – Overview	Computers
Maintenance Philosophy	Electric System
Depot/Shoreside	Evaporators
Intermediate	Fuel Oil
Hull	Fuel Transfer
Unit	Generators
Training and Personnel Support	Inert Gas
Off ship and Schools	Propulsion
OJT	Pumps/Valves
Cross training	Tools/Test Equipment
Engine and Auxiliaries Operations	Communications
Record Keeping	Boat Maintenance
Routine Operations	HVAC
Watch Standing	Housekeeping/Cleaning

statements begin to identify *how* elements of the mission are to be accomplished. HF requirements begin to emerge when decisions are made as to who or what (humans, machines, or a combination) is responsible to provide specific system functionality (Baker 1997).

2.2 Specifying the Role of the Human

A central HF issue in any system design is the allocation of

functions to man or machine, thereby establishing and defining the role of man in the system and allocating responsibility to maximize overall system performance. Function allocation is based on an assessment of the differential capabilities and limitations of men and machines in terms of the requirements of a specific function. Machines excel at generation of vast amounts of power, performance of highly repetitive tasks, and performance of tasks in noxious environments. Machines are also highly reliable, capable of performing their functions for months on end. Humans are resourceful, have marvelous sensory capabilities, can render decisions and generate plans with varying amounts of information, and can apply control and manipulate objects in moderately precise, but in a highly variable and adaptive, fashion. Humans are also much more variable, unpredictable, and error prone compared to machines. The general allocation of function that determines the relative roles of humans and machines stem from these relative differences in capability.

There is an increasing need for interactive dialogue between humans and computers in automated systems. It underlines a requirement to consider the interactions between man and machine because few operations are either purely manual or totally automated; most are "semi-automatic". The role of man in automated operations is as activator, monitor, manager, and under certain circumstances, as the intervening decision maker, taking over control from the automated process. The role of man is situationally dependent.

Highly generalized contributions of humans to maritime systems include:

- Sensory and perceptual functions - notably vision and hearing as sensory inputs, and highly adaptive and learned filtering and interpretation of sensory data as perceptual input.
- Decision making and planning - including surveillance and supervision of machine function, and intervention of machine function where deemed necessary.
- Object manipulation - generally those of very low force requirement, but that are of an unpredictable and variable nature. For example, manipulating and repairing electronics equipment.

2.3 The Problem of Human Error

Human error refers to any situation where an observer fails to perceive a stimulus, is incapable of discriminating among several stimuli, misinterprets the meaning of a stimulus, makes an incorrect decision, fails to select a correct response, or performs a response in an incorrect manner. A definition of human error states that it is an action that violates some tolerance limit of a system. Human errors have been classified as:

- errors of omission (tasks that are skipped)
- errors of commission (tasks performed incorrectly)
- sequential errors (tasks performed out of sequence)
- temporal errors (tasks performed too early, too late, or not within the required time).

While human error will always be with us, there are characteristics of people which have an influence on the frequency of errors. These include such factors as fatigue, disorientation, training, motivation, forgetting, confusion, incorrect expectancy or set, excessive stress, boredom, inadequate skills and knowledge, and

inadequate or impaired perceptual or cognitive ability. Such factors can certainly contribute to the occurrence of errors, and in some cases even cause errors (Malone, et.al., 1997).

It is also well established that factors external to the individual can influence the potential for human error. Elements of the job or task, design of equipment, operating procedures and training can affect the potential for error. These external factors can be classified as situational factors and design factors. Situational factors include those aspects of the operational setting, other than design, which influence human error incidence. These include: task difficulty, time constraints, interfering activities, poor communications, and excessive workloads. Design factors include aspects of the system hardware, software, procedures, environment and training which affect human error likelihood. Design factors encompass such aspects of the system as: man-machine interface design features; information characteristics (availability, access, readability, currency, accuracy and meaningfulness); workspace arrangement; procedures; environments; and training.

Human error has been cited as a causal or contributing factor in an increasing number of incidents and accidents. The GAO report quoted above reported that at least 50% of the failures of major weapon systems are due to human error. The GAO listed five "types" of human errors which cause the most failures. These include: 1) failure to follow procedures, 2) incorrect diagnosis, 3) miscommunications, 4) inadequate support, tools, equipment and environment, and 5) insufficient attention or caution.

A number of other sources have estimated the incidence of human error as a causal factor in a variety of accidents and incidents. These include the following (Malone, et. al.)

- an NTSB investigation of 82 major marine accidents in the 1970-80 timeframe revealed that 33 were collisions and that human error was the predominant cause of ship collisions;
- An American Nuclear Society (ANS) Report in 1986 stated that 90% of facility emergencies involve human error;
- A study of the role of human error in chemical plant safety in 1987 found that most accidents were due to human errors;
- The Office of Technology Assessment (OTA) estimated in 1986 that 62% of hazardous material spills were due to human error;
- A Boeing study attributed 65% of all airliner accidents from 1959 to 1986 to human error;
- The Navy Safety Center cited human error as the cause of 85% of ship accidents;
- The Nuclear Regulatory Commission attributed 50% of nuclear power plant accidents to human error;
- According to the US Department of Transportation, 90% of all automobile accidents involve human error.

3. HF ISSUES ASSOCIATED WITH HUMANS AT SEA

From the above discussion it is clear that many factors influence the occurrence of error. The questions that are presented from this are (1) what are the factors and issues associated with error and (2) what control can be exercised over these factors and associated issues. The following present samples of HF issues related to human performance, safety, and quality of life of people at sea.

3.1 Physiological Issues

These deal with factors influencing human performance such as work-rest cycles, environment, and fatigue. Life at sea is inherently fatiguing. Just living on a rolling, pitching platform increases the physical energy needed to walk, stand, and even sit. Further, few people aboard ship experience predictable sleep-wake cycles, and work-sleep cycles shift on a daily basis. Sleep deprivation is a chronic condition at sea.

The physical characteristics of ship design also tend to induce fatigue. Steep ladders, narrow passages, and cramped quarters increase the energy expenditures of crew in all aspects of living such as simple walking from space to space, accessing equipment, or carrying equipment and goods to workstations.

Other factors of the environment also challenge the human at sea, including extremes in weather ranging from arctic cold to tropical heat, job performance in the dead of night and in intense sunlight. Visual function is required under extremely diverse conditions, from display reading in darkened rooms (radar areas, for example) and under intense sunlight. Ships are also noisy, and the characteristics of ambient background noise (low to middle frequencies) tend to mask audible signals and speech.

3.2 Maintenance Design and Practices

This includes design and procedural issues related to conduct of maintenance activities. It also includes issues related to maintenance philosophy and distribution of functions to ship vs. shore-based support. Marine systems exist in harsh environments that adversely effect equipment reliability, further, constraints (such as ship size) impose limitations on the extent of redundancy that be provided for vital systems. Design of systems for ease of maintainability therefore become extremely important. The same constraints that limit the level of redundancy also tend to limit the physical space available to support corrective and preventive maintenance. Design for physical access to equipment is often sacrificed due to these space limitations.

3.3 Design and Arrangement Issues

This group of issues addresses the physical design of workspaces, operational and maintenance, to enhance human performance. Issues address physical access, safety, and communications supporting distributed task performance. Much of any marine systems arrangement is driven by it's mission and resulting functionality. Humans are integrated into this arrangement and often must contend with designs that are optimized to facilitate the physical functioning of the hardware. This can distribute job performance over large areas of a marine system (for example, logistics areas and machine shops that are distant from the hardware that they support). This increases the work requirements of the crew, and tends to induce fatigue and human error.

3.4 Training

The functional diversity of marine systems, in concert with limited numbers of people available to participate in the performance of functions, results in the need for people to provide many functions and to have diverse skills. Crew must be crossed trained to cover all the job requirements associated with human functions. Review of Table 1 shows over 100 functions at the very top level. If a ship with a functionality of Table 1 were to be manned at a level of 20 crew (as some ships are), then it becomes apparent how much cross training and specialization each crew member must posses in order to meet all the functional requirements of the system and it's mission.

3.5 Environmental and Safety Issues

These include issues such as damage control in optimally manned ships, effects on crew workload as a function of crew member disability, and manning implications of off-normal operations (such as engine failure, electronics failure, etc.). Special skills are required at sea to meet situations such as fire and flooding.

4. SUMMARY AND RECOMMENDATIONS

Marine systems place heavy demands on the performance of humans who operate and maintain them. Maritime systems (1) have a diverse set of mission requirements and systems functions, (2) operate in a often difficult environment, (3) must be self supportive and autonomous to succeed in their missions, and (4) often have a limited crew of humans to operate and maintain the hardware. In order to meet mission requirements of the systems while also meeting the needs of the human operator/maintainer, the following are recommended:

i. Develop and apply a process that guides HF analysis, design, and test and evaluation. The article in this document entitled 'Human Factors in Ships and Maritime Systems: Design Process' by Thomas B. Malone provides a foundation for an HF process.

ii. Apply Human Engineering (HE) standards and guidance to the design of major marine systems interfaces. Two good sources of HE design guidance are:

1. American Bureau of shipping (1998). Guidance Notes on the Application of Ergonomics to Marine Systems. ABS, Two World Trade Center, New York, New York.

2. American Society of Testing and Materials (1995). Standard Practice for Human Engineering Design for Marine Systems, Equipment and Facilities. Standard F 1166-95a. West Conshohocken, Pa.

REFERENCES

AMERICAN BUREAU OF SHIPPING (1998). Guidance Notes on the Application of Ergonomics to Marine Systems. ABS, Two World Trade Center, New York, New York.

AMERICAN SOCIETY OF TESTING AND MATERIALS (1995). Standard Practice for Human Engineering Design for Marine Systems, Equipment and Facilities. Standard F 1166-95a. West Conshohocken, Penna.

ANDERSON, D. E., OBERMAN, F. R., MALONE, T. B. and BAKER, C. C. (1997). Influence of human engineering on manning levels and human performance on ships. *Naval Engineers Journal*, 109(6), 67–76.

ANDERSON, D.E., MALONE, T.B., and BAKER, C.C. (1998). Recapitalizing The Navy through Optimized Manning and Improved Reliability. Paper presented at the Association of Scientists and Engineers (ASE), 35th Annual Technical Symposium 13 April, Washington, DC.

BAKER, C.C. (1997). Advanced Manning/Human Systems Integration (HSI) for Sealift Ship Technology Development Program (SSTDP). Technical report prepared for John J. McMullen Associates, Inc and the Naval Sea Systems Command.

BOST, J. R., BAKER, C. C., HEASLY, C. C. and MALONE, T. B. (1996). Technology and people: Striking a balance. Marine Log Conference on Marine Technology.

BOST, J. R., ANDERSON, D. E., MALONE, T. B., BAKER, C. C., HEASLY, C. C. and WILLIAMS, C. D. (1998). Human centred automation in total ship engineering. ASNE Engineering the Total Ship Symposium, Washington, DC.

MALONE, T. B., BAKER, C. C. and ANDERSON, D. E. (1996). Advanced manning designs and tools for marine structures. Society of Naval Architects and Marine Engineers, 20 November, Washington, DC.

MALONE, T.B., BAKER, C.C., ANDERSON, D.E., and WILLIAMS, C.D. (1997). Requirements to Ensure Crew Safety in Reduced Manning Ships. NAVSEA Association of Scientists and Engineers (ASE) Technical Symposium.

Tactical Cockpit Technology

M. W. Haas[†], A. Leger[†] and G. Rood[#]

†Human Effectiveness Directorate, Air Force Research Laboratory, Wright Patterson Air Force Base, OH, USA

‡Sextant Avionique, Saint-Medard-en-Jalles, France

§DERA, Farnborough, UK

1. INTRODUCTION

The interaction between machines and man is, and will continue to be, a pervasive technological challenge engaged by defense-related science and technology communities internationally. Optimal effectiveness of current and future weapon systems will only be achieved if the interactions between the operator, or group of operators, is efficient and intuitive. According to Leger *et al.* (1997), the tactical cockpit of the future will be influenced by both the future combat environment and technology affordances. This chapter provides a vision of the future combat environment and the human–system interface technology for tactical cockpits operating within that environment.

2. THE FUTURE COMBAT ENVIRONMENT

Aircraft present three unique and powerful advantages to the combat environment. They operate within an expansive three-dimensional volume, operate at higher speeds than do other combat vehicles and their operations are extremely flexible and unpredictable. Because of this, over the next 20 years it is unlikely that the basic operational roles that aircraft will be called on to complete will be significantly different from those laid down in the current air power doctrine. It is clear that in the future the highly dynamic and time-critical nature of the combat environment that currently exists for tactical aircraft will become more dynamic and require an increased pace of decision by the pilot.

There are four technology innovations that will dramatically affect the look and feel of future tactical cockpits. They are un-inhabited vehicles, directed-energy optical weapons, stealth and affordability. These technologies will affect the perceptual and cognitive workload of the pilot by manipulating the content and character of information availability. The portrayal of information to the pilot and the ability to control the acquisition and subsequent transmission of prioritized and synthesized information from the pilot are central themes of future tactical cockpit design.

The increased use of uninhabited vehicles for both reconnaissance and combat is gaining more acceptance within the military community. This will translate into concurrent operation of inhabited and uninhabited vehicles in the same airspace. In the future, the pilot of a tactical aircraft may be physically located inside the air vehicle or may be remotely located, either on the ground, at sea or in another air vehicle.

Directed optical energy weapons such as lasers will affect the look and feel of future tactical cockpits. Laser weapons have the ability temporarily or permanently to damage the human visual system. The speed of a laser weapon, calculated from when it is activated to when it reaches its target, is orders of magnitude greater than any projectile weapon, such as a gun or missile. In addition, the frequency of the laser can be modulated within the visible spectrum in ways that make it difficult, if not impossible, to counter without completely removing the pilot's view of the outside world.

Stealth technology renders an aircraft, or other vehicle, almost invisible to sensors used in military aircraft such as radar or infrared detectors. This invisibility, more precisely described as a reduced signature, allows the stealth user more freedom of movement within the combat environment at the expense of not operating active sensors. The use of stealth technology by an adversary makes targeting more difficult and the possibility of surprise by the adversary increases dramatically, potentially turning an offensive situation into a defensive situation in a matter of seconds.

A critical driver in the design of future tactical cockpits will be affordability. In fact, affordability will have a pervasive impact on every device and concept within the future tactical cockpit.

2.1. The Pilot's Job

In simple terms, the job of a flight crew is to aviate, navigate and communicate. The three functions are typically described as flying the aircraft. For the military flight crew, they must fly *and* fight. More specifically, the aircrew of a combat aircraft must fly the machine to a location, often via a predetermined routes, where the target, in the air or on the ground, can be found and attacked before returning to a friendly base.

Different aircraft, roles, locations, weather conditions, targets, defenses, equipment failures and interactions with other forces generate a plethora of specific demands and concerns that the crew must deal with to complete the mission satisfactorily. However, it is reasonable to describe the job as being composed of the following separable high-level tasks: aircraft control, navigating, managing the mission, managing aircraft systems, communicating, targeting, managing weapons, and countering threats.

Pilot's of current, as well as future, tactical aircraft must maintain situational awareness, which has been described by tactical pilots as knowing the following within a volume of space and time over which they can exert control:

- Knowing where the threats are and what they are doing.
- Knowing where the friendlies are and what they are doing.
- Knowing what my flight knows and options for offense and defense.
- Knowing what other flights know – to a lesser extent.
- Knowing what I do not know directly.

3. FUTURE COCKPIT TECHNOLOGY (NEXT 3–7 YEARS)

The cockpit technology of current tactical aircraft includes 6-inch monochrome and color head-in displays, 10–20° monochrome heads-up displays (HUD), analog intercom combined with auditory warning tones, and the control concept of hands-on stick and throttle (HOTAS). However, move advanced interface devices and concepts, such as helmet-mounted displays (HMD), digitized and localized auditory displays, wider field-of-view head-up displays, collimated head-level displays, large area liquid crystal displays, haptic displays, direct voice interaction, and touch screens/touch pads may be applied in future tactical

Figure 1. Within 7 years, the tactical cockpit will include a HUD, HMD, HLD, localized audio, DVI and haptic displays.

cockpits. Figure 1 depicts a tactical cockpit that could exist within the next 3–7 years.

3.1. Hands-on Stick and Throttle

Most modern aircraft have now adopted the HOTAS concept, initially implemented on the F-15 but generalized with the F-18 for sensor and weapons management. This concept is closely linked to the notion of *real-time command*, where the pilot can immediately operate a control, without the need to release the stick or throttle to reach a switch on a dedicated control panel. It is now well recognized that this concept greatly enhances pilot effectiveness during combat missions.

Although it was initially intended to control the weapon system in the final phase of the mission, the HOTAS concept has been progressively extended to other mission phases. The intent of this extension was to reduce pilot workload, such as the control of the aircraft system configuration during ingress. As a consequence of this extension, the number of real time commands available from HOTAS has sometimes increased considerably, heavily soliciting the pilot working memory. The problem is basically linked to the saturation of working memory, inducing the need for additional training hours, without totally suppressing the risk of high error rates and poor results.

3.2. Large Area, In-cockpit Displays

The notion of a large image was a key element in the origin of the *little picture–big picture* concept proposed by Adams (1995). Both head-in and head-out displays are likely to co-exist in the cockpit of the future. Besides redundancy, these display concepts bring different advantages in regard of building good situational awareness.

Integrating large field-of-view head-in displays in combat aircraft is technically challenging due to the cockpit environment and the lack of space. Currently, the largest image displayed in a tactical aircraft is found in the Rafale, with the color head-level display (HLD) concept. Using LCD valves and collimation optics the HLD offers a square field of view of 20 x 20°. At the normal distance of vision for the HLD, achieving an identical image size would require a 12 x 12-inch display.

While the HLD is intended to display complex tactical situation, the Rafale's HLD is also used for map presentation, more demanding in regard of display resolution. Recent studies indicate that numerous situations would require a simultaneous management of short-, mid- and long-term information, requiring a frequent swapping between different map scales. In contrast, and in an operational context, it may be of great interest to allow different information to be displayed contiguously with a variable window size. For the future, such considerations are leading to concepts of large, re-configurable and interactive displays. The size of such a display could be, as an example, in the range of 20 x 15-inch. Some of the expected benefits would be:

- Increased flexibility, as the displays windows size or location could rapidly be reconfigured following mission type, phase or even user experience or cognitive style. On request, most adapted size for a given situation could be rapidly obtained, from full screen to iconic.
- Capacity to display complex tactical situations with an always appropriate size and resolution.
- Allow in-flight mission planning and rehearsal.

3.3. Head-up and Helmet-mounted Displays

A HUD is a cockpit-mounted virtual visual display in which the pilot views symbology or imagery overlaid on the real-world scenery as viewed through the display. A HMD is similar in concept except that it is mounted on the pilot's head, is much smaller and lighter, and takes advantage of a sensor that measures head position and attitude to spatially stabilized symbology and/or imagery.

It is quite likely that the HUD, if they remain in the cockpit, will see very little evolution from their current status for military aircraft application. Progress has continued on color HUD development. In fact, a color HUD has been tested on a Mirage 2000 test-bed aircraft at the French Flight Test Center at Brétigny sur Orge. Nevertheless, with cost effectiveness consideration becoming a crucial factor, further development of such equipment is very unlikely. It can also be expected, based upon progress in HMD technology that helmet mounted equipment could supersede HUD in the cockpit.

The first generation of helmet-mounted equipment (sight only, using a head-tracker) is already in service on fighter aircraft as the Mig-29. Second-generation HMD (symbology only) or even third-generation (symbology and sensor imagery) are now quite close to application on fighter aircraft. Interest in third generation HMD for helicopters, as an important part of the aircraft system during night operations, is now well recognized and clearly established for both mono- or binocular designs. Testing of second generation HMD has convinced tactical pilots, who had the opportunity to use such equipment in flight, about the great potential it has for close in air-to-air combat superiority. Advantages, in terms of radar or seeker lock on, reverse cueing and off-boresight missile launch have been reported in many places. Several human factors and system integration issues need to be rapidly clarified to accelerate the introduction of the helmet-mounted technology.

3.4. Auditory Displays

After the visual channel, the audio channel has always been an important display in aeronautics and essentially based on voice

communication. More recently, attempts have been made to investigate the use of acoustic orientation cues. What is new with the current trend is the expansion of *audio symbology* far beyond the current status of auditory warnings.

Synthesized voice messages are now a classical feature of commercial and military cockpit. the advantages and disadvantages of such communication mode have been widely reviewed. Recently, the use of iconic and metaphoric representations associated with comprehensive understanding of sound signal characteristics were used to build a set of auditory threat warning symbology in the UK.

Over the past 10 years laboratory work and some flight-testing has been devoted to the investigation of 3D localized sound. Results obtained in-flight, and within flight simulation, have confirmed the interest of localized auditory information, especially when coupled to HMD. Applications include following areas:

- Threat or waypoint spatial localization.
- Spatial separation of audio communication.
- Warnings and alarm.

3.5. Haptic Displays

The use of tactile interface to convey spatial orientation information has been suggested. Application of such techniques could also be advocated in the context of a synthetic environment. The first application of haptic display is certainly to restore feedback cues for aircraft control, which were lost with the introduction of fly-by-wire controls. Such techniques may also prove to be useful in synthetic environments.

3.6. Direct Voice Interaction (DVI)

Numerous studies and test flights have been devoted to voice control in the first half of the 1980s. Results obtained at that time were quite mitigated and the general feeling was that substantial progress had to be made in this area before direct voice interaction (DVI) can really be used in a cockpit. Nevertheless, DVI was included very early in the EFA program. It is currently considered as an option on the Rafale. Significant progress has been made in the early 1990s, which now makes DVI quite attractive for complex cockpit environments.

The UK and France have independently flight-tested voice recognition systems currently under development in their respective countries. Despite the fact that different test-bed aircraft and methodologies were used, there is a striking convergence of results obtained in UK and France. Recognition rates reach, in both cases, 98% for short sentences in stable flight conditions and remain between 85 and 89% under G-load. The effects of stress, acceleration, noise reduction and modeled dialogue are current subjects of research.

3.7. Touch Screen–Touch Pads

Strictly speaking, touch screens and touch pads belong to the manual control category. The use of touch screens is now well established in civil applications such as in terminals for banking operations. Touch pads can replace joysticks, to move a cursor as an example, and allow input validation. Both technologies have been flying for quite a long time on the Rafale. They are now used routinely by pilots, with very positive results so far. They may rapidly become a common feature for other cockpits.
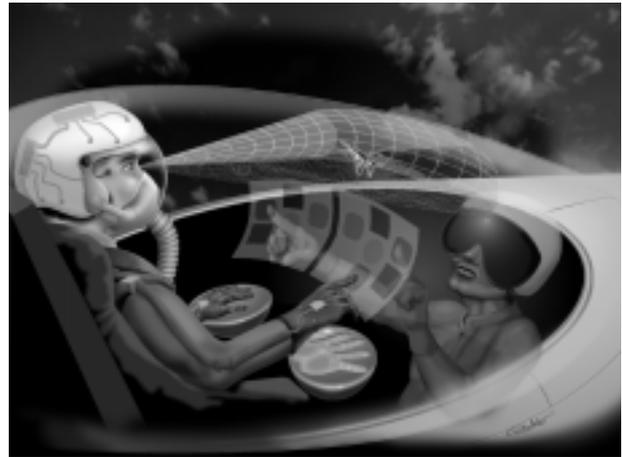

Figure 2. Advances in pilot state measurement and modeling, combined with advanced control and display technology, will revolutionize the tactical cockpit within the next 15 years.

4. FUTURE COCKPIT TECHNOLOGY (NEXT 7–15 YEARS)

These interface concepts include control using eye line-of-sight and gestures, interfaces that measure pilot state and adapt in real-time, and potentially totally enclosing the cockpit. Figure 2 depicts a tactical cockpit that could exist within the next 15 years.

Eye tracking, or gaze control, measures the eye line-of-sight and uses it to select and activate discrete or analog controls, or to orient symbology. Gesture-based control recognizes static postures of the hands and fingers as well as dynamic movements of the hands and fingers as a control or communications device.

4.1. Windowless Cockpit

The windowless cockpit, which is probably the greatest change in cockpit configuration to future tactical aircraft, will come about from the need to protect aircrew against optical/directed energy weapons. A reliable counter to the optical threat is essential, and this will be in the form of protection for the pilot's eyes. Both operational requirements and peacetime considerations suggest that a close-able solution, protection deployed as required under aircrew control, is an acceptable solution when current protection, the use of helmet visors with single or multi- become ineffective against variable-frequency lasers. Reactive dyes or other elements may well provide a longer-term solution against agile lasers, but the density of the visor is liable to leave a pilot, in a crucial part of the mission, with a partial or fully restricted vision of the world outside of the cockpit. Full closure may be achieved either on the helmet visor, or the cockpit canopy, but, in either case, the direct view of the outside world would be obstructed, and the information needed would have to be presented either from the range of sensors, or synthetically from aircraft data bases, or both.

4.2. Adaptive Interfaces

Adaptive interfaces will utilize knowledge of the operator's state, both measured and modeled, to enable automated, real-time decisions regarding information management and display characteristics. These characteristics include information modality, spatial arrangement, temporal organization and control

utilization. The adaptation of the interface to operator state can occur across two different domains, the first being dynamic function allocation and the second being information portrayal characteristics. The adaptation is enabled by the availability of three technologies: (1) highly flexible display and control devices, (2) computational models of situation awareness, workload and operator performance, and (3) direct real-time physiologic and behavioral measurement of the operator.

REFERENCES

ADAM, E.C., 1995, Flat Panel Imperatives. In *AGARD CP-575, Symposium on Situation Awareness: Limitations and Enhancement in the Aviation Environment*, 9, 1–7.

LEGER, A., ROOD, G. and HAAS, M., 1997, Human–Machine Interaction*, AGARD-CP-600, NATO AGARD/RTA's Future Aerospace Technology in the Service of the Alliance*, 1, 3-1–16.

Virtual Workplace Design

G. Andreoni and A. Pedotti

Dipartimento di Bioingegneria, Politecnico di Milano, via C. Golgi, 39, I-20133 Milan, Italy

1. INTRODUCTION

In society, with the increasing role of machines in everyday life, the topics concerning the man–machine interface are becoming a major subject in designing home and job environments, production systems, transportation, and education tools.

In particular, ergonomic design based on a deep knowledge of the human psycho-physiological capabilities and limitations in relation to the requirements of machines and environments can be very incisive and effective. But it is very difficult to make ergonomics participate also in the earliest phases of the design. In general, the work and the role of the ergonomist begins when the physical mock-up of the product (tool, environment, workplace, etc.) is built and ready to be tested. In that way they can intervene only in a second time, even if we have to remember the general ergonomic guidelines that should be followed by designers.

Nowadays the virtual design (also called *virtual prototyping*) of workplaces and the simulation of the man–machine interface is going to have more and more importance to save time, money and risks during the design phase of manufacturing lines and cockpits (Haslegrave and Holmes 1994, Miller 1997). Therefore, to make a real intervention of ergonomics possible in this phase, one of the arising problems in this field is the realization of a virtual manikin able to move in a human like way inside the virtual environment. This will allow for testing the man–machine interactions before the mock-up is built cooperating directly in the early CAD design (Wilson 1997).

For this reason, the manikin has to represent a fundamental part of the whole system for design, since the virtual environment has usually a very good CAD description and is thus well identified. Nowadays the manikins can simulate a lot of possible movements but it can be seen how this movement can be far from the movement humans are likely to perform.

2. ANALYSIS OF THE POSTURE IN THE CAR

2.1 Car Seat

In general manikins are controlled by kinematics or inverse kinematics or a mixture (Badler *et al.* 1993, Das and Sengupta 1995). These methods produce reliable solutions to the motion generation problem, but observing the moving manikin it is easy to detect an "artificial" motor task. So we can say that they are reliable but not realistic approaches.

In fact, we can say that there are two ways to generate the movement of the manikin. One is the use of *deterministic mechanical models*. For these models some kind of criterion of optimization (energy, power, trajectory's length) must be introduced in order to deal with the overdetermination of the inverse kinematics problem (Uno *et al.* 1989). The other is to use *neural networks* to drive the model starting from examples acquired on humans (Massone and Bizzi 1989).
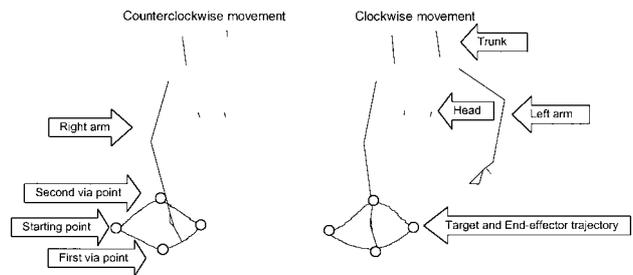

Figure 1. CAD representation of the environmental condition (left) where the subject performs reaching movements: the stick diagram of the moving subject (right) shows the body segments considered in the biomechanical model.

Some examples of this second solution regard the studies on the analysis and the simulation of working tasks in a seated position. A mock-up represents a working bench on which a seated operator assembles some parts or he has to perform reaching and grasping movements like pushing buttons or handling car commands: this representation of the experimental setup is shown in Figure 1. This CAD design is also a virtual working environment where the movements are to be performed. In this case the aim of the research is to investigate the motor strategies and the coordination of the movements of the upper limbs with the head–trunk complex. A 29 degrees of freedom model was defined for the upper trunk including both arms and the head. The ELITE system was used to collect kinematic data. This automatic motion analysis system uses small passive retro-reflective markers to highlight the body landmarks. The designed model was implemented by using 24 markers. The computer vision techniques for obtaining coordinate data on human motion have many potential advantages because they use optical (non-contacting) principles and can be automated to work rapidly and accurately without human intervention.

The kinematic patterns of the variables are computed from this markers distribution applying to the data set the biomechanical model; they are also needed for the *neural network training*.

In fact the data to be provided to the neural network are the angles of the various joints acting in the movement, and the trajectory of the end-effector (the index fingertip since the hand was modeled as a single rigid body). A campaign of acquisitions was conducted with different modalities on three population percentiles (5% female, 50% male and 95% male). Reaching movements, point to point and including via-points, were performed both on the bench and on the vertical direction. In the case of trajectories with via-points, the typical bird-eye pattern was observed (see in Figure 2 the stick diagram of the model as well as the trajectory of the end-effector during a clockwise and counter clockwise movement including via points). This behavior reinforces the need of the use of real data, in this case processed through neural networks, instead of deterministic models that can not account for these strongly human features.

Neural networks need many training examples, and the number of examples is directly proportional to the level of complexity of the motor task to be simulated. Therefore, for very
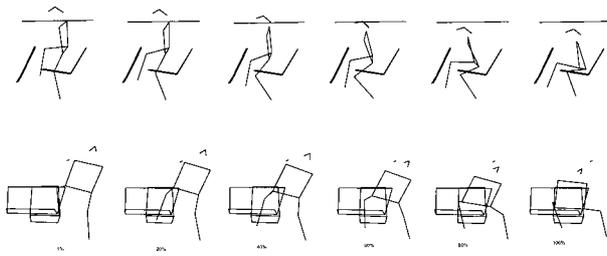
Figure 2. Typical non-linear human behavior: reaching target and return to the starting point passing through two via points with the right upper limb: counter-clockwise (left) and clockwise (right) movements.

complex task, involving the whole body (this means to consider a biomechanical model with a high number of degrees of freedom) and eventually performed in a complex environment (i.e. an environment with many objects inside it to be considered in the collision avoidance) both the number of examples to be given and the experimental set-up to be adopted could represent a severe problem. For this reason another approach for the simulation (that is also based on actual data) was developed and applied in studies of the movements of man–machine interface in car related tasks.

The car is the most common transportation mean and, therefore, the most lived environment for millions of humans. Therefore, the study of man–machine interface is very relevant both for the evaluation of the comfort related to a given car set-up and in relation to muscular fatigue and load distribution associated to long-term tasks involving car driving and usage.

Among these, tasks accessibility movements, driving movements can be identified at a high level of priority.

For accessibility movements we mean the set of motor tasks performed to enter and to take the seated position inside the car and to get out from the car considering the various situations: anterior seat, driver seat, posterior seat (a typical example is shown in Figure 3).
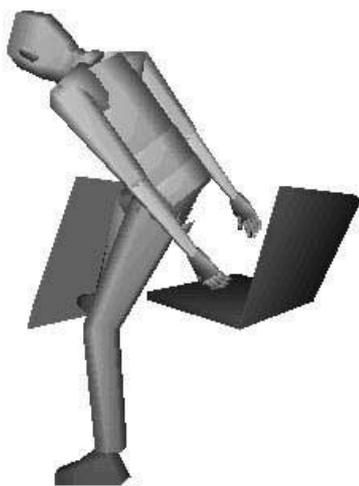

Figure 3. Stick diagram sequence of the entry movement in a car cockpit for one subject adopting the most common motor strategy (the lateral sliding) in the sagittal (up) and frontal (down) plane.

The performance depends on the anthropometric characteristics of the subject and on the shape, structure and dimensions of the car and can be of great discomfort.

Driving movements refer to the motor performance requiring coordination between upper limbs and trunk to reach and manipulate various commands: steering wheel, lights, switches, air conditioning commands, shift on the floor, seat-belt, etc. They refer also to the lower limbs and foot movements performed while acting on the brake, acceleration and clutch pedals. Therefore, the analysis of driving movements and the optimal design in order to facilitate their execution is strictly related to the general optimization of driving improving also safety.

Optimal design of cars should optimize accessibility movements and driving movements execution to improve comfort, safety and to prevent the long-term pathologies related to unsuitable loads and movements.

Up to now these problems have been faced by using simulation programs on computer or an approach based on dummies. Moreover the traditional procedure followed in developing a new vehicle begins with the exterior styling: in a second time man is "fitted" inside this volume. That is what is called an "outside-in" approach. All these methods and computer tools are very useful for designing an acceptable structure in terms of dimension and shape. However, they are based on the concept of the human being as a *passive load* and are not considering the real execution of movements in term of control strategies adopted: kinematics and coordination among the various body segments, dynamics, muscle forces and loads on the major joints. Therefore, this simulation approach leads to the definition of a stereotyped man–machine interface in which the concept of comfort is related to several parameters in a deterministic way.

Future design of MMI could greatly take advantage from these observations by considering the "real" behavior of a statistically significant population performing the required motor task. This means to reverse the idea into an "inside-out" approach to the new vehicle design (Porter and Porter 1997). The development of new technologies makes possible the analysis in three dimensions with the required level of precision of complex movements performed in car usage. It becomes, therefore, possible to investigate the different motor strategies involved in such interactions and to identify in a quantitative way several factors that are important for the most ergonomic solution in the design of the machine (Andreoni *et al.* 1997). In this way it is absolutely necessary to begin from the real human behavior properly to model all the different modalities that can be used to carry out a specific movement. This means to have a set of movements described in their kinematic variables that represent the *grammars of the movement*. If the complete movement can be divided in different phases it is possible also to extract specific and elementary tasks.

The experimented application is oriented to the design of cars and takes into account primarily the complex interaction between the user and the vehicle in terms of movement required for accessing and driving the car. This approach is based on the direct and dynamic evaluation of such interaction, on the identification of the motor strategies in relation to the structure of the car, on the relative constraints and the specific features of the human neuromotor system. Also in this case the three-dimensional measurement of human movement is reduced to

the detection of the trajectories of several points that identify the positions of the body segments in space. In the first stage the protocols, the models and experimental setups were fully implemented and applied on a small sample population.

The first outcome was the identification of the main motor strategies (Pedotti and Crenna 1990) used by different subjects in entering the car. Despite the great complexity of the movement involved in this task, each considered subject showed a high level of repeatability (invariant individual motor strategy which represents the grammar of the movement) as presented in Figure 3 for the most common and used motor strategy. At the same time we observed a constant energetic characteristic among such different strategies: the vertical center of gravity displacement, i.e. the potential energy is the most correlated parameter with the lowering of the door height. The flexion of the left hip demonstrated to be another correlated parameter: in fact the values increased during the lowering of the doorway. This is very important from an ergonomic point of view because this angle is related to the limb entirely supporting the body weight while the subject is entering and we could assume it as an indicator of the postural status, or also of the effort spent in the task (the movement corresponds to a muscular work done). This result points out a quantitative relationship between the dimension of the door and some kinematic variables independently form the motor strategy used by the subjects. The outstanding consistency and sensitivity of these data validate the method and confirm its reliability as a feasible and innovative tool to evaluate and classify car set-ups in terms of energetic accessibility cost (measure of comfort).

The approach developed for the simulation of the movement consists in the modulation of the real movement in response to a varying parameter of the set-up; it means to modify through a mathematical function the non-linear component of the temporal numerical series of the degrees of freedom of the movement beginning from a grammar of movement extracted from an observed movement. To extract a grammar of movement means to calculate the temporal numerical series of the kinematic variables defined with respect to the biomechanical model.

The biomechanical model has an identical general structure for all the subjects (same body segments of the kinematic chain and constant number of degrees of freedom) but it is customized on the single subject through a procedure of matching (minimizing the sum of the square distances of the points of the biomechanical model from the real anatomical points of the subject acquired in the standing posture).

Starting from the actual data the movement is modulated to adapt it to the new environmental conditions. The real kinematic data corresponded to the mathematical model of the motor strategy referred to a particular set-up. The first step of the procedure was the separation of the movement in two components: a linear function which was obtained by connecting the starting point and the ending points defined by the first and last frames by a line for each kinematic parameter (or degree of freedom of the biomechanical model), and a non-linear component which represented the modulation of the movement. We then reconstructed the virtual kinematic data of the simulated movement by an algorithm which minimized a target function. This target function was designed to:

- minimize the joint movement with respect to the standing posture;
- respect a collision avoidance function;
- ideally reach the same modulation for the non-linear component of each anatomical angle (e.g. to have a same coefficient of the non-linear component for each degree of freedom: it means to distribute the modification along the body districts);
- respect particular constraints (such as keep in touch ground and left foot until the subject is not seated); and
- respect joint constraints (such as physiological limits or joint ranges of motion).

The initial value of the target function for each joint angle and for each frame of the movement was given by the linear function value at that time.

The algorithm computed the new joint angles in the virtual movement and then re-calculated the 3D coordinates of the
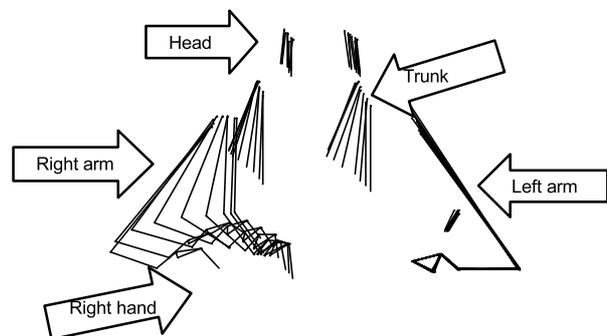

Figure 4. Simulated frame of the entire entry to the left posterior seat in VRML. This allows for CAD importing and testing different environmental conditions (e.g. the height of the doorway from the floor or the distance between the seats).

anatomical marker on the basis of the biomechanical model previously defined.

In this way the kinematic data of a simulated or synthetic or virtual movement are computed: they correspond to a subject whose anthropometric dimensions were known and adopting a precise motor strategy to enter a vehicle of given (virtual or real) dimensions. A frame of the simulated movement is presented in Figure 4 in VRML environment.

The comparison between the simulated and the real kinematics in a test movement demonstrated the reliability of the method with statistical significance.

3. CONCLUSIONS

In conclusion, we said that a good simulation tool is fundamental for correct the design of workplaces. In this way also specific and proper algorithms for the ergonomic evaluation can be included to compute an assessment of the ergonomic condition and give the information of how to improve the working conditions.

Many common methods for the ergonomic evaluation are used in these programs: for example static and dynamic biomechanical calculations, reaching zones, RULA index, OWAS, NIOSH lifting equation. Also databanks with information about comfort joint range and zones, field of vision and the computation of strength under different conditions can be used as reference.

So it is possible to integrate ergonomics and engineering in the technical design with all the benefits of a more effective collaboration between ergonomists and designers for a better product design and manufacturing organization (Haslegrave and Holmes 1994). This is the realization of the "inside-out" approach for ergonomic design, with a human-centered philosophy applied to the design itself.

REFERENCES

ANDREONI, G., RABUFFETTI, M. and PEDOTTI, A., 1997, Biomechanical method for the ergonomic evaluation of the entry–exit movements to a car or to a virtual prototype. In *4th International Conference and Exhibition on Comfort in the Automotive Industry: Recent Developments and Achievements*, 6–7 October, Bologna, Italy, 309–17.

BADLER, N.I., PHILLIPS, C.B. and WEBBER, B.L., 1993, *Simulating Humans. Computer Graphics Animation and Control* (New York: Oxford University Press).

DAS, B. and SENGUPTA, A., 1995, Computer-aided human modelling programs for workstation design. *Ergonomics*, 38, 1958–72.

HASLEGRAVE, C.M. and HOLMES, K., 1994, Integrating ergonomics and engineering in the technical design process. *Applied Ergonomics*, 25, 211–20.

MASSONE, L. and BIZZI, E., 1989, A neural network model for limb trajectory formation. *Biological Cybernetics*, 61, 417–25.

MILLER, J.S., 1997, 3D simulation: a virtual environment for Proactive Ergonomics. In *13th Triennial Congress of the International Ergonomics Association*, 29 June–4 July, Tampere, Finland, 2, 379–81.

PEDOTTI, A. and CRENNA, P., 1990, Individual strategies of muscle recruitment in complex natural movements. In Winters and Woo Savio (eds), *Multiple Muscle Systems, Biomechanics and Movement Organization* (New York: Spinger), 542–9.

PORTER, C.S. and PORTER, J.M., 1997, An "inside-out" approach to automotive design. In *13th Triennial Congress of the International Ergonomics Association*, 29 June–4 July, Tampere, Finland, 2, 90–2.

UNO, Y., Kawato M. and SUZUKI, R., 1989, Formation and control of optimal trajectory in human multi-joint arm movement: minimum torque-change model. *Biological Cybernetics*, 61, 89–101.

WILSON, J.R., 1997, Virtual environments and ergonomics: needs and opportunities. *Ergonomics*, 40, 1057–77.

Visual Display Units: Positioning for Human Use

L.M. Straker

Department of Physiotherapy, Curtin University of Technology, Perth WA 6001, Australia

1. INTRODUCTION

As an increasing proportion of the workforce spends a greater proportion of their work (and leisure) time interacting with computers, the object of the visual interaction, the visual display unit (VDU), becomes a critical component in the interaction.

This chapter will review the options for the optimal placement of a VDU in terms of vertical height, lateral location, eye–screen distance and VDU tilt. The discussion is focussed on a single computer screen office workstation, although the principles may be applicable to other workstations.

2. VERTICAL HEIGHT

2.1. Vertical Placement Options

VDU for office work are usually placed in one of three vertical locations, as described below.

2.1.1. "High" — top of screen level with sitting eye height

The traditional recommendation for VDU placement has been to locate the top of the screen at sitting eye height (Figure 1). With a typical 13-inch monitor at 750 mm from the user's eye this results in viewing angles (angle between horizontal and a line from the eye to the visual target) from 0 to 15°.

2.1.2. "Mid-" — bottom of screen level with desk surface

Recently, more researchers have advocated a lower VDU position, with the bottom of the screen at desk height (Figure 1). For a user of average height sitting upright at a desk just above elbow height and viewing a 13-inch monitor at 750 mm, this would give viewing angles from 15 to 30°.

2.1.3. "Low" — bottom of screen just visible over keyboard

The third option for VDU vertical placement is as low as possible given cathode ray tube monitor and keyboard input device (Figure 1). Using current desktop computer and keyboard technology, the lower limit is probably formed by the viewing angle to the far edge of the keyboard. For the same situation as described in Section 2.1.2, where the user has full forearm support on the desk with the keyboard furthest edge ~500 mm from the user's abdomen, the lowest possible viewing angle is ~45°. Thus, with a 13-inch screen the viewing angle range could be 30–45°.

2.1.4. "Ultra-high" and "ultra-low" positions

"Ultra-high" positions are above sitting eye height. For example, de Wall *et al.* (1992) recommended a VDU position that would result in viewing angles of 5–25° above the horizontal.

"Ultra-low" positions would also be possible with flat-screen technology by placing the screen on the desk just in front of the user's abdomen. This would be comparable with reading a book
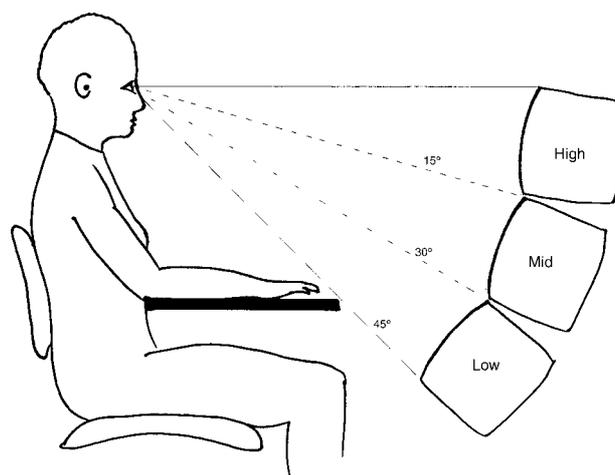

Figure 1. Vertical VDU placement options: (a) "high," (b) "mid-," (c) "low."

placed flat on the desk and result in viewing angles from 45 to > 65°.

2.2. Evidence for a "High" VDU Position

The main argument for the "high" position is to reduce musculoskeletal strain on the neck by keeping the head in the "normal" upright standing position.

A "high" VDU position is likely to result in less neck flexion, and to a lesser extent less thoracic flexion. Based on modeling of neck muscle forces de Wall *et al.* (1992) suggest neck extensor muscle forces are minimized by a head position with 30° *extension* relative to "normal" standing head posture. At the other extreme, excessive neck flexion results in increased neck muscle loads and discomfort. There is general agreement that extremes of neck extension and flexion should be avoided, however the optimal neck position is debated. "High" VDU placement will result in less flexion than lower positions, but may result in undesirable neck extension.

Research in the author's laboratory has found reduced cervical and thoracic erector spinae muscle activity in the "high" position compared with the "mid-position", with a similar trend in upper trapezius (Mekhora 1996). Turville *et al.* (1998) also found a reduction in cervical and thoracic erector spinae activity in the higher position, although they found an increase in trapezius activity on one side. Aaras *et al.* (1997) found no difference in trapezius activity between a higher and lower VDU position when there was no forearm support. As neck flexion becomes near end-of-range, one would expect extensor muscle activity to reduce due to increased passive tissue tension. Whether or not this is desirable is not known.

In summary, there is reasonable evidence that when users work in an upright sitting posture, a "high" VDU position is associated with reduced neck flexion and reduced cervical and thoracic erector spinae activity. Most researchers consider the trapezius to be a critical muscle (due to its prevalence in reports of work related symptoms), but the current evidence for whether one VDU height is preferable for trapezius is conflicting. Further, it is not known whether the increased neck flexion with lower VDU placement is a problem.

2.3. Evidence for a "Mid-" to "Low" VDU Position

The two main arguments for the lower positions are to reduce visual strain by locating the VDU where it is easier to accommodate and converge for close visual work (Lie *et al*. 1997) and to reduce musculoskeletal strain.

Visual accommodation is the change in the shape of the eye lens capsule to achieve focus. Insufficient accommodation results in blurred images. Convergence is inward turning of the eyes to keep a single binocular visual image. Inaccurate convergence results in double images. Muscles control both accommodation and convergence. The eyes have natural resting points for accommodation and convergence, and as a visual target moves in closer from these points more eye muscle activity is required. Thus, just as there is a convenient reach zone for touching, there is a similar zone for visual "touching." However, unlike the tactile reach zone, an arc around the shoulder does not define the visual reach zone. Rather, the visual "reach" zone extends out from the resting points and comes closer to the user with increasing downward viewing angle. The resting point of accommodation averages ~800 mm with young people and shifts farther with increasing age (Krueger 1984). The resting point of vergence averages 1120 mm with a horizontal viewing angle and comes in to 900 mm with a 30° downward viewing angle (Heuer and Owens 1989). A lower monitor position will allow for closer viewing distances without an increase in visual strain.

Aside from less strain on eye muscles, with lower gaze angles the eyelid covers more of the eyeball that helps reduce the likelihood of dry eyes (Villanueva *et al*. 1996).

The support for a lower VDU position being less straining visually comes from the general knowledge of human visual performance but also from laboratory and field studies. There is little doubt that from a visual system point of view, lower VDU placement is preferable (assuming suitable non-reflective environmental lighting).

The second argument for the lower monitor positions is to reduce neck musculoskeletal strain by reducing head–neck extension (Ankrum 1997), lengthening neck muscles (Burgess-Limerick *et al*. 1998) and allowing a greater variety of appropriate neck postures (Ankrum and Nemeth 1995).

Proponents of lower VDU positions argue that users typically flex their lower cervical spine and extend their upper cervical spine (to give the chin jutting forward posture) when working with "high" VDU monitors. Upper cervical spine and head–neck extension is maintained by static muscle activity, probably by small, deep neck muscles in particular. The sustained activity may result in discomfort and disorder.

Fatigue in the muscles supporting head–neck extension is more likely as the muscles would be working in shortened positions. (There is a relationship between the length of a muscle — compared with its resting length — and the tension or force it can generate such that the more shortened a muscle is the less force it is able to generate. In essence a shorter muscle equals a weaker muscle.)

Ankrum and Nemeth (1995) suggest that lower VDU positions allow a range of neck flexion postures while keeping appropriate gaze angles. Turville *et al*. (1998) attempted to determine whether users do vary their neck posture more when working with a lower VDU, however their measure was not adequate.

While the preference for lower VDU positions from a visual point of view is in little doubt, the preference from a neck musculoskeletal point of view is still contentious.

2.4. Issues Complicating Recommendations for VDU Height

Several limitations in the available research literature make weighing the evidence for various VDU height options difficult. Different definitions of postural and visual angles make comparison of research difficult. This is further complicated by inadequate descriptions of working postures, and by different postures being used (upright with unsupported forearms and forward tilt with supported forearms) and different equipment (some studies have used what are now considered old monitors with poor clarity; this may have affected vision and therefore posture).

Current technology in office furniture and lighting design often restricts the option of "low" VDU height as a separate support surface below desk height is required and office luminaries may not be designed to reduce reflections on "low" VDU. A further example of possibly important interactions between VDU height and posture is that the "low" position may restrict the use of the backward tilting sitting posture. This posture has some research support and is probably a useful alternative sitting posture. In contrast, the "low" VDU position may encourage a forward tilting posture, which is also probably a useful sitting posture.

Technological changes are tending to merge the options as large screen sizes usually mean an overlap of positions. For example, with a 21-inch screen (vertical screen dimension of 300 mm, compared with 240 mm for a 13-inch screen), the viewing angle range at 750 mm is nearly 25°. Thus the "high" position would be 0–25° and the "mid-" position would be 5–30° and the "low" position 20–45°. Further, flat screen technology may provide the option of using a VDU like a book, at an even greater viewing angle (the "ultra-low" option described above).

Individual variation may also mean one position is not suitable for all users. For example, a user who has played a piano regularly since childhood may be musculoskeletally adapted to an upright sitting posture and "high" to "low" visual target.

3. LATERAL LOCATION

The lateral location of a VDU is less contentious than the vertical placement. After only ~7° (100 mm at 750 mm) of eye lateral movement the head responds by postural change (Fisher 1925, cited in von Noorden 1985).

Sustained head and neck rotation is likely to result in musculoskeletal discomfort. Therefore, if the visual task is largely screen based the screen should be directly in front of the user. If the visual task entails a mix of screen and paper, or multiple screens, then a compromise position should be selected while trying to avoid sustained neck rotation to one side.

4. EYE–SCREEN DISTANCE PLACEMENT

Appropriate eye–screen distance interacts with height, so that as the VDU is placed lower, it can be placed closer to the user. Lower gaze angles bring the resting point of vergence closer, and

sustained viewing of visual targets closer than the resting point of vergence has been found to increase eyestrain (Owens and Wolf-Kelly 1987). Therefore, at any one VDU height, users will tend to find viewing more fatiguing at closer eye–screen distances. Optimal eye–screen distances are also dependent on the user's visual capacity (which is age related) and the clarity and size of the visual image. While recommendations have varied from less than 500 mm to near 1000 mm, when the VDU is in the "mid-" vertical position somewhere ~750 mm is likely to be comfortable for many users. Greater eye–screen distances are probably also suitable, as long as screen font sizes are increased. However office space and increased costs for large screens may constrain more distant VDU placement.

5. VDU TILT

It is generally recommended that VDU should be tilted to maintain an approximate 90° between line of sight to center of screen and screen surface.

Ankrum *et al.* (1995) found vertical monitor tilt influenced postural and visual discomfort. Tipping the monitor forward so that the top of the monitor was closer to the eyes increased neck, upper back, and visual discomfort when compared with a monitor tipped backward at the same gaze angle. They attributed the results to a visual phenomena called the vertical horopter (the locus of points in space that stimulate binocular singularity in peripheral vision) The vertical horopter tips away from the viewer at much the same angle that a person would read a magazine.

Optimal tilt may need to be compromised in work places where luminaries are poorly designed and thus cause screen reflections.

6. RECOMMENDATIONS

VDU workstations should allow for considerable adjustability in VDU position to allow for individual visual and musculoskeletal system differences, task differences and environment differences (especially lighting).

VDU workstations should preferably be adjustable enough to allow the VDU to be:
1. directly in front of the user;
2. at 600–1000 mm distance;
3. tilted slightly back from a right angle to the line of sight;
4. with the top of the screen to be as high as a 5° viewing angle; and
5. with the bottom of the screen to be as low as 50° viewing angle.

A single height desk surface can not provide this adjustability.

When allowing a user to determine a comfortable VDU location the starting point could be with the VDU directly in front, at 1000 mm, tilted slightly back from at right angles to the line of sight and with a viewing angle of 30° to the bottom of the screen.

REFERENCES

AARAS, A., FOSTERVOLD, K. I., RO, O., THORESEN, M. and LARSEN, S., 1997, Postural load during VDU work: a comparison between various work postures. *Ergonomics*, 40, 1255–68.

ANKRUM, D.R., 1997, Integrating neck posture and vision at VDT workstations. In *Proceedings of Work With Display Units 97* (Tokyo: Waseda University), 63–4.

ANKRUM, D.R., HANSEN, E.E. and NEMETH, K.J., 1995, The vertical horopter and the angle of view. In A. Grieco, G. Molteni, B. Piccoli and E. Occhipinti (eds) *Work with Display Units "94* (Elsevier: Amsterdam), 131–6.

ANKRUM, D. and NEMETH, K. J., 1995, Posture, comfort, and monitor placement. *Ergonomics in Design*, 7–9.

BURGESS-LIMERICK, R., PLOOY, A. and ANKRUM, D. R., 1998, The effect of imposed and self-selected computer monitor height on posture and gaze angle. *Clinical Biomechanics*.

DE WALL, M., VAN RIEL, M. P., AGHINA, J. C., BURDORF, A. and SNIJDERS, C. J., 1992, Improving the sitting posture of CAD/CAM workers by increasing VDU monitor height. *Ergonomics*, 35, 427–36.

HEUER, H. and OWENS, D., 1989, Vertical gaze direction and the resting posture of the eyes. *Perception*, 18, 363–77.

KRUEGER, H., 1984. Visual functions in office-including VDU [Introductory paper]. In *Ergonomics and Health in Modern Offices* (London: Taylor & Francis).

LIE, I., FOSTERVOLD, K. I., AARAS, A. and LARSEN, S. 1997, Gaze inclination and health in VDU operators. In *From Experience to Innovation: IEA "97 Proceedings of the 13th Triennial Congress of the International Ergonomics Association*, 5 (Tampere: International Ergonomics Association), 50–2.

MEKHORA, K., 1996 An evaluation of visual display unit placement by electromyography, posture, discomfort and preference. Masters thesis, Curtin University of Technology, Perth.

OWENS, D.A. and WOLF-KELLY, K. 1987. Near work, visual fatigue, and variations of the oculomotor tonus. *Investigative Ophthalmology and Visual Science*, 28, 743–9.

TURVILLE, K. L., PSIHOGIOS, J. P., ULMER, T. R. and MIRKA, G. A., 1998, The effects of video display terminal height on the operator: a comparison of the 15° and 40° recommendations. *Applied Ergonomics*, 29, 239–46.

VILLANUEVA, M. B. G., SOTOYAMA, M., JONAI, H., TAKEUCHI, Y. and SAITO, S., 1996, Adjustments of posture and viewing parameters of the eye to changes in screen height of the visual display terminal. *Ergonomics*, 39, 933–45.

VON NOORDEN, G.K., 1985, *Binocular Vision and Ocular Motility* (St Louis: Mosby).

Wearable Computers

W. Barfield and E. B. Nash

Virtual Environment Laboratory, Department of Industrial and Systems
Engineering, Virginia Tech, Blacksburg, VA 24060, USA

1. INTRODUCTION

Over the past few years, developments in a new information technology known as wearable computers have been rapidly taking shape. Wearable computers are fully functional, self-powered, self-contained computers that allow users access to information anywhere and at any time. A wearable computer is designed for hands-free mobile computing and is a battery-powered system worn on the user's body. Wearable computers may be configured in many different forms from see-through head-mounted displays (HMD) with speech input, to wearable computers with input devices worn on the wrist or body. This chapter reviews the components of wearable computers, addresses developments and design issues, and summarizes application areas.

2. COMPONENTS OF WEARABLE COMPUTERS

One of the main goals of wearable computers is to allow users to access computational resources whenever and wherever they are in the environment. Compared to virtual and augmented reality display systems, in which users are confined to the range of motion associated with the length of the cable connecting the display hardware and rendering platform, wearable computers allow complete freedom of motion within the real world. This mobility generally requires three standard system components, each worn on the user's body. These are the computer unit, the power supply, and the input devices. The computer unit functions to store, retrieve, and process information. Typically belt or wrist worn, it houses the CPU, motherboard, hard drive, and memory.

In terms of computational resources, the processing power and storage capability of PC-based wearable computers are generally comparable to those of current laptop technology. The power supplies usually consist of lithium-ion-based battery packs that are worn on the body and connected to the computer. They can generally last for 2-4 hours before recharging is necessary. But other developments are also being investigated, such as lithium ion polymer batteries (which may allow 11 hours of usage before recharging), and harnessing human motion as a power source (Starner 1997). Output may be viewed using either a flat panel display (FPD), which is typically worn on the wrist, or a head-mounted display (HMD) connected to the computer unit.

Head-mounted displays for wearable computers may be characterized as opaque or see-through. Each type may be viewed in monocular or binocular mode. Monocular displays require the user to integrate the display viewed using one eye with real-world scenes viewed using the uncovered eye (Figure 1).

When worn over both eyes, opaque head-mounted displays still allow real-world scenes to be perceived, in this case by using

Figure 1. Picture of a user with a wirelessly networked "Mobile Assistant" wearable computer accessing a map of Virginia Tech through the World Wide Web (Picture courtesy of Virtual Environment Laboratory, Virginia Tech).

video captured via a camera mounted on the HMD (or some other location). In contrast to opaque HMDs, see-through HMDs admit light from the outside environment, allowing users to directly view the real world with the naked eye. Half-silvered mirror systems fixed in front of the user's eyes reflect light from computer-generated graphic images. The resulting image is an optically combined view of the real world superimposed with graphical images.

Because the user is often using one or both hands to perform a task in the real-world environment, input devices used with wearable computers need to be designed to allow the user efficient manipulation and interaction with both the displayed information and objects in the real world. The input devices that have evolved for use with wearable computers are diverse and ever changing to accommodate user needs. For data entry or text input tasks where hands-free operation is not needed, body-mounted keyboards, wrist keyboards, and hand-held keyboards are often used. Cursor movements used to select options or manipulate data formerly using mice are being substituted by other devices such as track pads. When real-world tasks require the use of both hands, hands-free operation can be provided through speech recognition along with gesture- based, EMG-based, and EEG-based input device technology.

3. WEARABLE COMPUTERS AND AUGMENTED REALITY

One important application for wearable computers is use with augmented reality display technologies. Augmented reality in the context of display design refers to the ability to project or merge graphics or text with real-world imagery (Barfield *et al.* 1995). Feiner *et al.* (1993) refer to this ability as "knowledge enhancement" of the world.

There are different types of wearable display technologies that can be used to combine real-world objects with computer-generated imagery to form an augmented scene. Optical-based wearable computers allow the observer to view the real world directly with computer graphics or text overlaid using a head-mounted display (HMD) which is worn like glasses. Video-based wearable computers use cameras mounted near the user's eyes to present live video with overlaid computer graphics or text. Both types of display can present stereoscopic images if two slightly offset cameras are used.

4. EXTENSIONS OF WEARABLE COMPUTERS

4.1. Network Applications

Due to the advent of wireless technology, wearable computers are now permitted to be wirelessly connected to local area networks (LANs) or wide area networks (WANs). This has important applications for users of wearable computers. Namely, users may now access databases from shared networks such as the World Wide Web at any time or place.

4.2. Integration of Sensor Technology

Another interesting extension is the integration of sensor technology with wearable computers for the creation of "smart spaces." We define smart spaces as environments in which objects have associated with them (1) a CPU allowing some minimal level of intelligence to exist in the object or environment, and (2) sensors that detect the presence of other objects within the environment

or the state of the object itself. In order for any digital system to react to events in its environment, it must be able to sense the state of the environment. This can be accomplished by incorporating sensors, or arrays of various sensors into systems. Sensors are devices that are able to take an analog signal or stimulus from the environment and convert it into electrical signals that can be interpreted by a digital device with a microprocessor. The stimulus can be a wide variety of energy types but most generally it is any quantity, property, or condition that can be sensed and converted into an electrical signal (Fraden 1996).

In general, there are two kinds of sensors: active and passive. The characteristics of these two types of sensors will affect their potential use as components of a wearable computer system. Active sensors require an external power source or excitation signal in order to generate their own signal to operate. The excitation signal is then modified by the sensor to produce the output signal. Therefore, active sensors consist of both a transmitting and receiving system. A thermistor is an example of an active sensor. By itself it does not generate a signal, but by passing an electric current through it, temperature can be measured by the variation in the amount of resistance (Fraden 1996). In contrast, passive sensors directly convert stimulus energy from the environment into an electrical output signal without an external power source. Passive sensors consist only of a receiver. An infrared motion detector is a passive sensor; it uses radiant heat energy from the objects as its source of detection energy.

When building and employing sensors into intelligent environments (including wearable computers), designers should consider the application of the intelligent environment and select the sensor technology which best complements it. Among these considerations are the sensor's accuracy, range, calibration error, power consumption, physical dimension, saturation, reliability, sample rate, and noise filtering.

5. WEARABLE COMPUTERS AS ACCESSORIES TO THE HUMAN

Wearable computers can also be thought of as personal information devices (Starner *et al.* 1997). With experience, the user can personalize the system to ensure appropriate responses to everyday tasks. As a result, the user's wearable computer system becomes a mediator for other computers and interfaces in the environment, providing a familiar, dependable interface.

Effective provision of this type of "intelligent assistant" depends on several considerations. For one, it is easy to cross the boundary between information that is actually useful and clutter that is overwhelming. Thus, in order to assist the user more effectively, a wearable computer might model its user's actions, goals, and even emotions. Systems have been developed to track the user's position, visual field, and current interests as revealed by what is being typed (Starner *et. al.* 1997). This type of interaction is well suited to wearable computers because continual close contact with their user allows a unique opportunity for sensing the user's actions. In practice, such a system using bio-sensors to measure physiological states might allow a wearable computer to track the state and actions of its user and react accordingly.

A more personal and striking interface may be possible if the user's emotional affect can be sensed as well (Picard 1995). It is well known that emotional affect plays a large part in everyday

life (Damasio 1994). Picard (1995) discusses ways in which computers might recognize affect as well as a number of potential applications of affect-sensitive wearable computers. Specifically, by combining a wearable computer with biosensors that can sample a user's temperature, blood pressure, galvanic skin response, and electromyogram, it is possible to create an "affect model" of the user. Using these sensor data and the user model, a wearable computer can track the state and actions of its user, and react accordingly.

Finally, user models implemented within a wearable computer system should be able to predict future actions. In turn, this information can be used to allocate resources preemptively. For example, if a user enjoys hard rock music when working on a late night project to keep alert, but prefers classical music during the day to lower stress, the wearable computer can predict what the user may want to listen to next and can download potential selections over a wireless network.

Human interaction with the real world can provide the dimensions by which wearable computers can be evaluated. Two of these dimensions include the level of mobility provided by the computing system and the level of scene fidelity. The level of scene fidelity refers to the quality of the image provided by either the virtual reality display, the real world, or the augmented world. Not surprisingly, real-world experiences result in high levels of mobility and scene fidelity. Essentially, this is because humans have developed the capacity to carry their sensors with them as they move around the environment, and they experience the world with the resolution provided by these biological sensors.

Wearable computers also allow a high degree of mobility, but current limitations in the technology do not permit as wide a range relative to our human bodies. For example, currently we cannot swim with wearable computers although our biological sensors easily allow for this. And mobility is further inhibited by the extra weight that must be tolerated. On the other hand, wearable computers using see-through optics or live video for output allow a high degree of scene fidelity. Wearable computers using computer-generated images may suffer in the area of scene fidelity because currently they do not have the rendering capa-

bility of workstations that are often used (e.g., SGIs) to render high resolution graphics for augmented reality environments.

6. CONCLUSIONS

Present advances in microelectronics and wireless networking are making truly ubiquitous computing a reality. However, as the number and complexity of wearable computers continue to grow, there will be increasing needs for systems that are faster, lighter, and have higher resolution displays. In addition, wearable computer systems must become unobtrusive and socially acceptable. Thus, components of the wearable computer (CPU housing unit, input devices, and output devices) will have to look far more like clothing or clothing accessories than the commercial wearable computer systems available now. Lastly, with wearable computers and with appropriate sensors integrated into objects within the environment, it is possible to create "smart spaces", in which users will interact with each other and with networked objects within the environment to gain information, solve problems, and "think together".

REFERENCES

BARFIELD, W., ROSENBERG, C., and LOTENS, W., 1995, Augmented-reality displays. In Barfield and Furness (eds) *Virtual Environments and Advanced Interface Design* (New York: Oxford University Press), pp. 542–75.

DAMASIO, A., 1994, *Descartes' Error* (New York: G.P. Putnam).

FEINER, S., MACINTYRE, B. and SELIGMANN, D., 1993, Knowledge based augmented reality. *Communications of the ACM*, 36 (7), 5362.

FRADEN, J., 1996, *Handbook of Modern Sensors: Physics, Design, and Application* (2nd ed.) (New York: AIP Press).

PICARD, R., 1995, Affective computing. Technical Report 321. MIT Media Lab, Perceptual Computing Group.

STARNER, T., 1997, Human powered wearable computers. *IBM Systems Journal*, 35(3/4), 1996.

STARNER, T., MANN, S., RHODES, B., LEVINE, J., HEALEY, J., KIRSCH, D., PICARD, R. and PENTLAND, A., 1997, Augmented reality through wearable computing. *Presence: Teleoperators and Virtual Environments*, 6(4), 384-98.

Wheelchairs

I. Gilad

Faculty of Industrial Engineering and Management, Technion–Israel Institute of Technology, Technion City, Haifa 32000, Israel

1. INTRODUCTION

Mobility aids are the basic social needs for individual that suffer from limitations in gait abilities. Henceforth mobility must be viewed as being essential to the outcome of rehabilitation for wheelchair-dependent persons and to their successful integration into society and active life. In the case of the disabled, a wheelchair is more than simply a mobile aid to transport an individual from one place to another. For the young disabled a wheelchair is essential to socialization, education, and employment; for the older disabled it is an around-the-house moving device. Young and old use wheelchairs in very different ways, but both groups need them to satisfy a set of subjective physical comfort requirements, fit individual anthropometry, relate to their age group, and comply with any other special needs.

Today's wheelchairs propose a variety of mechanical bodies and supporting equipment to stabilize the user's body. To best fulfill technical requirements and address the needs of different users to match the appropriate wheelchair, a systematic approach should be introduced. The match between the appropriate wheelchair and the needs of an individual is usually provided by professional social and medical systems or agencies, acting as local or federal authorities. But these systems use different procedures and policies when providing the right wheelchair. To learn about selection of wheelchairs, two surveys were conducted in Israel; they indicated that much work remains to be done in optimizing the selection procedures for mobility aids.

The number of actual wheelchair users is difficult to determine. An estimated number for Europe is 2.5 million wheelchair-dependent individuals (Van der Woude 1999). For the USA two numbers are quoted in the literature: Gaal *et al.* (1996) gives a figure of 1.4 million wheelchair riders, and nearly half of them live outside of institutions. Kopardekar and Mital (1994) estimate 0.7 million regular users of wheelchairs; an additional 1.0 million Americans are using wheelchairs, while suffering from cardiovascular, nervous system, or related diseases until they become mobile. A study conducted in Israel by Gilad (1997) found that about 1% of the population use wheelchairs permanently or in a temporary mode of mobility.

These groups are large enough to allow research on effectiveness of use and effectiveness of rehabilitation programs. The number of wheelchair users is high in the three estimates cited here and in many other estimates. It is therefore costly to provide mobility aids, and in the light of this expenditure some people are now concerned about the low level of satisfaction expressed by the average wheelchair user.

Wheelchairs are generally used by three sectors: young people, adult people, and elderly people. Advances in medical treatment have helped children and the younger generation (ages 3 to 18) to survive diseases of the nervous system and the musculoskeletal system (e.g., multiple sclerosis and cerebral palsy). The children and young handicapped use manual and semipowered wheelchairs fitted with special seating supports. The adult sector (ages 18 to 64) is three times as big as the young sector; it is divided into three groups:

- People handicapped from birth or through childhood diseases of the nervous system or the musculoskeletal system.
- Individuals in middle age who survive a traumatic accident related mainly to the spine or cervical regions. Their handicaps result from work, car crash, military, sports, and other physical activities.
- Individuals suffering from progressive physical malfunctions, due to diabetes or cardiovascular problems. This sector is the most active: 55% use motorized or powered wheelchairs, 45% use manual wheelchairs.

The number of people in the elderly sector (ages 65+) is increasing at a rate of about 15% annually. In my survey, 46% of them are stationed in adult homes and geriatric facilities; this population uses basic manual wheelchairs.

A cross-handicapped population survey was conducted to analyze the mechanical fit and personal satisfaction between wheelchair users and their actual wheelchair. Another survey was conducted in parallel to examine the procedures used by the different agencies while assigning wheelchairs to the handicapped. Two kinds of questionnaire were designed: (1) a questionnaire for wheelchair users, and (2) a questionnaire for professionals who choose the wheelchairs and allocate them to users; among the professionals are occupational and physical therapists, nurses, administrative personnel, and sales executives.

The user questionnaire was so designed that every wheelchair user can easily answer it. It was distributed to war-disabled, car and work accident victims, musculoskeletal injury patients, and elderly users, people who use wheelchairs in their own home or in a nursing home. Its questions cover physical fitting, level of mobile activity, purpose of use, chair features, ease of use, and safe use, limitations imposed by the wheelchair, moving ability in a range of given activities, ease of ingress and egress, and personal experience of the social and medical services.

The professional questionnaire was distributed to rehabilitation agencies. Its aim was to analyze procedural topics and to learn about the fit of the wheelchair and common needs of the users. The answers provided information on wheelchair design, such as seat height, length, width, slope, shape, backrest, clearance for feet and calves, armrest, required force to operate, and cushions. It contained questions on safety, because about two-thirds of wheelchair injuries require medical attention.

The overall figures on wheelchair users are 74% male and 26% female; these figures differ in each category. About 60% of the users reported that they were appropriately fitted for seat height, width, length, and slope measures. About 50% of the wheelchairs were adjustable for adaptability and comfort. Some 70% of users were satisfied with their current wheelchair. Many safety incidents and injuries are reported among wheelchair riders. A recent study reported in the USA showed that incidents among wheelchair users break down like this: 42% due to tip and fall, 33% due to component failures, and 25% due to other events. Statistics derived from the Israeli survey show similar figures. In the 77% that are not the result of component failure,

many of the underlying causes arise because the wheelchair does not fit the user's needs.

2. SELECTION ALGORITHMS

It was found that the various financing authorities pursue very different policies when assigning a wheelchair to an individual. The majority of the individuals who participated in the study have criticized this. Findings indicate that, when design is considered, there is low correlation between the human needs, the mechanical supporting aid, and the maneuvering solution. In many cases the device is uncomfortable and therefore is not used in the proper way. The users were not informed about their optional needs and the available solutions. In providing an inappropriate tool to the needed person, the policy makers unfortunately failed to optimize the efforts and the heavy costs involved.

Neither of the above policies has a computerized information base from which a wheelchair user can find out how to choose a wheelchair, or what is available from manufacturers. Most of the information is transformed via verbal communication, in nonformalized venues. No formal guide is used on educational issues or selection of the proper solution for nonstandard personal requirements. Older people require escorts and help in folding their wheelchair; these aspects should receive adequate attention. These needs were reported by 36% of the Health Ministry population and 25% of the Ministry of Defence population. When these needs are reported, it is important to assess the capabilities of the handicapped person and their spouse. Very little attention is given to the type of wheelchair or the technical solution to be used.

The financial aspects for purchasing a new wheelchair differ among the financing agencies. It is obvious that financing agencies aim to lower their expenses, and this conflicts with the decision on which components and which design of wheelchair will be recommended to an individual. This causes arguments between the authorities and the wheelchair user. The budget involved in supporting the wheelchair population is enormous. I could not get access to the real figures, but my estimates are high compared to what I used when accumulating the cost per wheelchair and the cost of solving any problems. The handicapped popula-
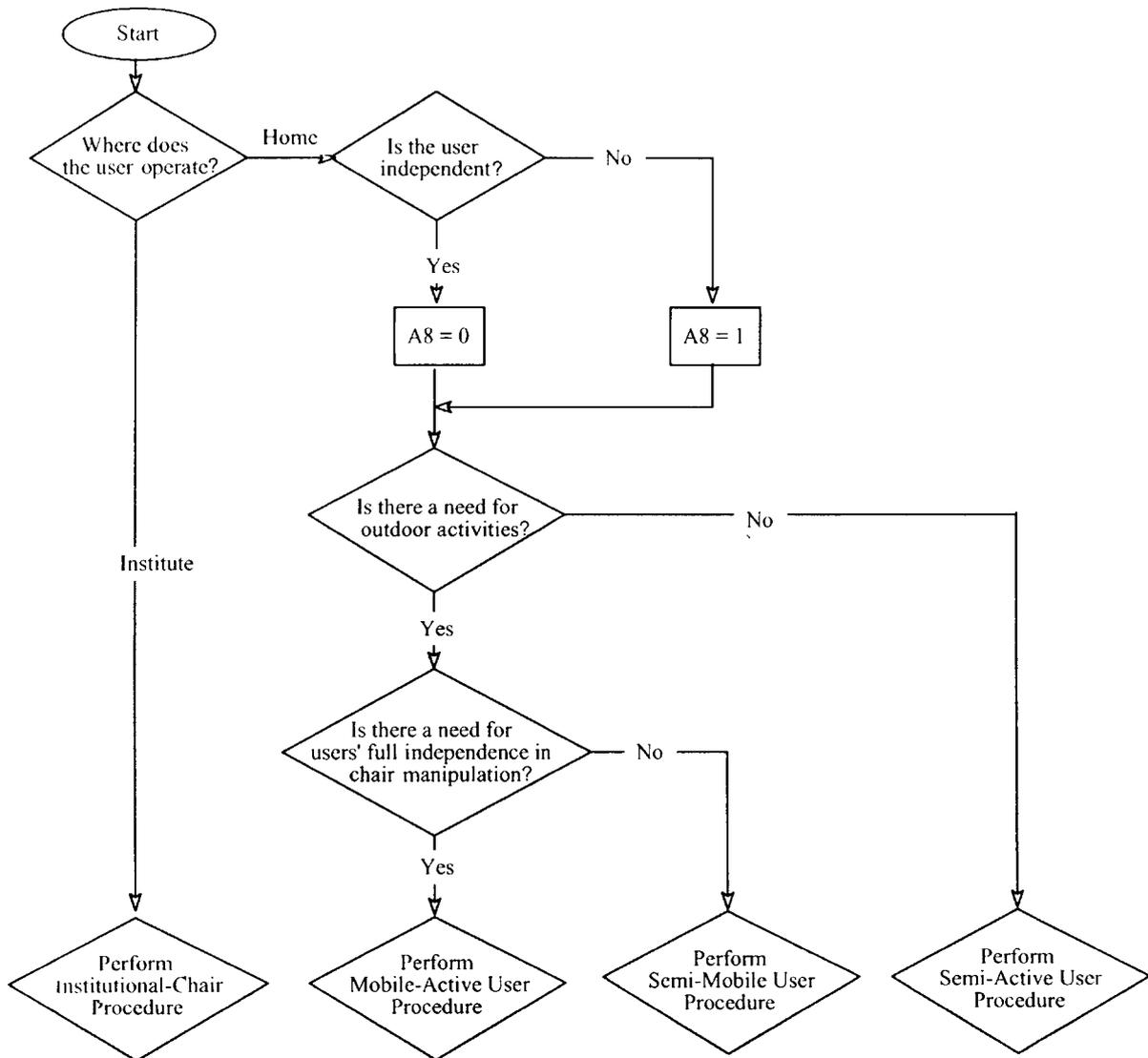

Figure 1. First-step algorithm for wheelchair selection and definition of user's need
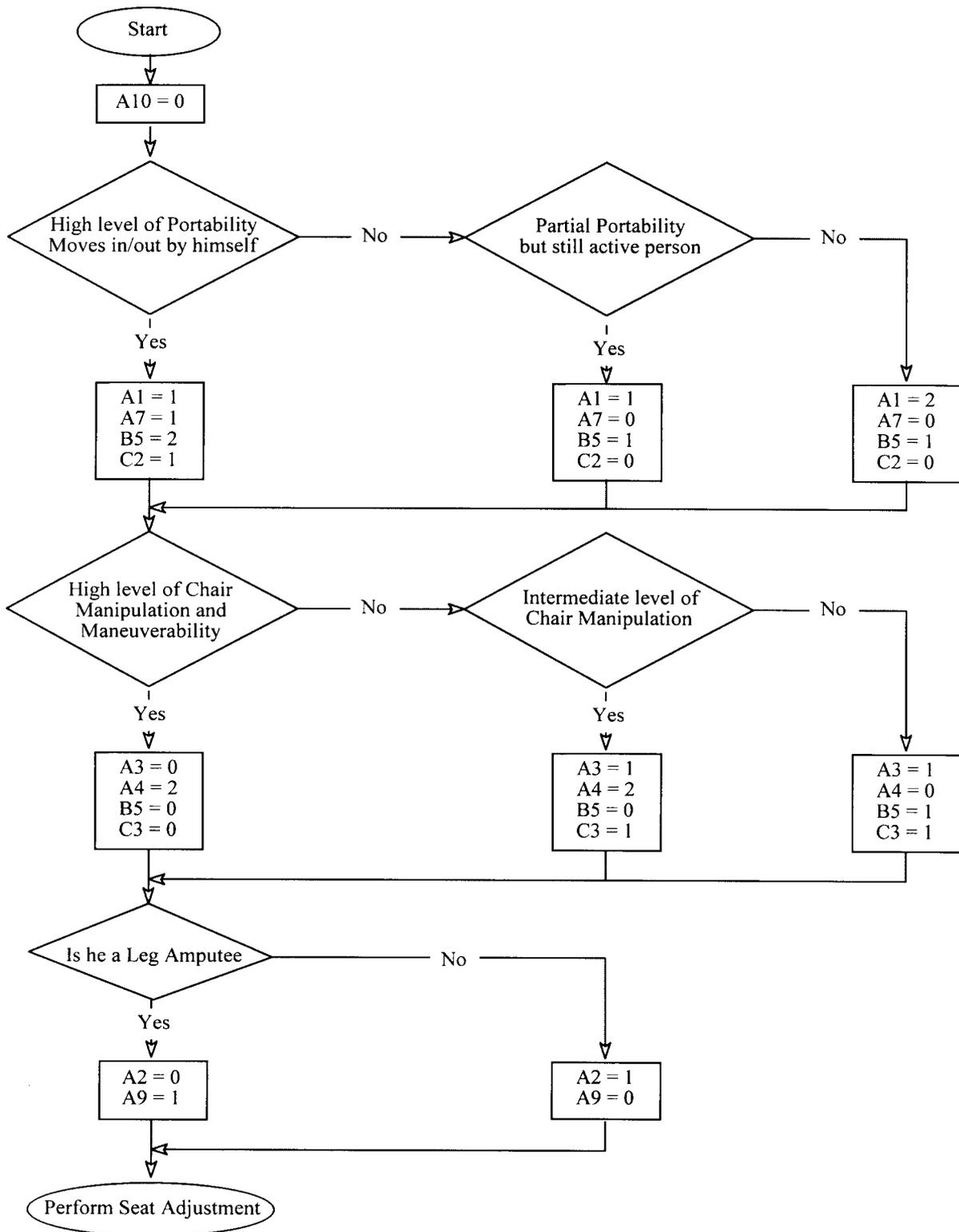
Figure 2. Algorithm to select a wheelchair for an active user

tion is very diverse and there is no way to organize its sectors to improve their current status, as a healthy society will do. Trying to work out a better system, I have suggested an algorithm that provides a good policy for the social and medical authorities. It is divided the into three categories: chair frame, chair seat, accessories and special aids.

Each category is then divided into components or adjusting features that have an impact on the fitting procedure. Figure 1 shows how to define a user's need as the first-step algorithm for wheelchair selection; the categories used in this algorithm are given in Table 1. Wheelchair selection is then performed in four steps:

1. Define the type of chair user, according to the actual user's needs: institutional chair, mobile active user, semimobile user, and semiactive user.
2. Perform chair selection for each type of user, observing the definitions and required components for each user's specific needs.
3. Perform the seat adjustments.
4. Perform the accessory adjustments.

Figure 2 shows the selection algorithm for an active user who is a self-operator, commonly known as a free person; the categories used in this algorithm are given in Table 2.

REFERENCES

GAAL, R.P., REBHOLZ, N., HOTCHKISS, R.D., and PFAELZER, P.F., 1996, Wheelchair rider injuries: causes and consequences for wheelchair design and selection. *Journal of Rehabilitation Research and Development*, 34(1), 58–71.

GILAD, I., 1997, A study on wheelchairs and users in rehabilitation. In Seppala, Luopajarvi, Nygard, and Mattila (eds) *Musculoskeletal Disorders Rehabilitation*, Vol. 4 (Helsinki: Finnish IOH), pp. 500–502.

KOPARDEKAR, P. and MITAL, A., 1994, Preliminary results from a wheelchair usability survey. *Proceedings of the 12th Triennial Congress of the International Ergonomics Association*, Toronto, Canada, Vol. 4, pp. 393–96.

VAN DER WOUDE, 1999, *Biomedical Aspects of Manual Wheelchair Propulsion* (Amsterdam: IOS Press).

Table 1. Categories and values for selecting user's need

Category	Function	Value for category	
A1	Ability to fold the frame	A1 = 0 A1 = 1 A1 = 2	no fold at all self-fold fold by third party
A2	Footrest	A2 = 0	no footplate
A3	Distance between front and rear wheels	A3 = 0 A3 = 2 A3 = 1	short distance standard distance long distance
A8	Assistant handle	A8 = 0 A8 = 1	no handle needed with assistant handle
A10	Wheelchair weight	A10 = 1 A10 = 0	light frame regular frame

Table 2. Categories in three groups: components and meanings

Category	Adjustment/feature	Category	Adjustment/feature
	A IS FOR CHAIR FRAME		
A1	foldable frame	A2	footrest
A3	distance between front/rear wheels	A4	rear wheel width
A5	caster wheel size	A6	caster wheel fork size
A7	detachable wheel	A8	assistant handle
A9	centre of gravity	A10	wheelchair weight
	B IS FOR CHAIR SEAT		
B1	backrest height	B2	seat' depth
B3	seating angle	B4	armrest height
B5	detachable arm rest	B6	seat cushion
	C IS FOR ACCESSORIES AND SPECIAL AIDS		
C1	foreleg harness	C2	rear safety wheel
C3	tipping lever	C4	support accessories
C5	head and neck supports	C6	assisting table
C7	spasm leg belts	C8	upper body harness
C9	hip supporting belts	C10	seat add-on
C11	hand rim		

Workstations Organization

E. Górska

Institute for Organization of Production Systems, Warsaw University of Technology, ul. Narbutta 85, 02-524 Warsaw, Poland

1. INTRODUCTION

An ensemble of enterprises is best visible on workstations. First of all the defects, which are the consequence of the organization of production process and the work of sections directing the process, are seen.

Organizing the workstations is a very complicated problem. It is a result of workstations functioning in production process, which focuses all technical, organizational and ergonomical elements relative to the needs and wants in the work process.

Organization of the workstation should connect all these components and create a harmonic entirety. It will provide optimal realization of the planned goal. The growth of the workstation organization level can have an important influence on production effects. Thanks to rationalizations in the organization range, it is possible to have an increase of works' efficiency with relatively little investment in circulation.

2. DIFFERENCE BETWEEN WORKPOST AND – STATION

In organization theory the notion of "a workpost" and of "a workstation" is usually used. In general, a workpost denotes a room or a part of a room that is at the same time a spatial unity and a separate place, and it serves for some purposes to a specific working team consisting of a few people who do the same job and use the same set of tools.

The notion of a workpost can be understood as *a place occupied by an employee in some system or organization*. In this case a specific partial activity done by a given employee and resulting from the work division is taken into consideration.

This term also suggests a basic organization unit where human work is carried out. Its main characteristics are: static system of components, dynamics of human work, operation of machines and influence of environmental factors.

A workstation is generally understood as the smallest production unit.

Thus, "a workpost" refers to every post where work is carried out (including the work of a manager of a big company or the work of a usual factory worker). The starting point of this analysis does not constitute a production process or a part of production space but the functions to be fulfilled and which have been assigned to an employee. Therefore, the notion of "a post" also includes an employee. According to the definition, both types — a workpost and a workstation — have the same logical structure of their notions. It enables one to analyze the organization of these two types using the same methods. It also enables one to transform some rules of the organization of workposts to workstations. The complexity of this division (workposts and -stations) and the problems connected with clear logical naming result in the mixing and exchanging of these two types in the literature of the subject.

Furthermore, it must be stated that every workstation is necessarily a workpost but not every workpost is a workstation.

3. WORKSTATION

The notion of a workstation both in the literature and in practice is defined in many different ways. In general it can be said that a workstation is the smallest production unit in which a man acts in a deliberate way, using machines and necessary tools, working with material and carrying out some stage of the production process. Putting it in a more detailed way, "a workstation" can be defined as the smallest production unit whose main characteristics are: necessary instrumentation for the given work, the personnel employed and the relevant work standards and safety regulations.

Thus, it is the smallest basic unit capable of production in which there occurs, in the precisely determined space, an organic unification of the three main factors of this process: producers (people), the means of work and the work objects.

A workstation can be a stationary or a movable means of production (in buildings, at machines or work facilities, in the means of transport, vehicles) that is operated by an employee to carry out all the preparatory and production activities or to control and monitor the production (in the case of the high-tech automatic machines).

A workstation is therefore an object that fulfills strictly determined basic production tasks with the help of human and technological means constituting a part of production process. This kind of object is functionally connected with other objects of the same kind. In other words, there is a workpost when there is a man performing his assigned production duties. The conclusion is obvious: one workstation can be equipped with a few machines or instruments, with one machine or with a part of a machine. Apart from a machine (or its part) and instruments being the basic technical component of a workpost, a workpost includes also some indispensable space necessary for man to move at his work and to store the material before and after the treatment.

Many specialists analyze a workstation from the functional point of view and consider it "the smallest subsystem of the company, its smallest so-called work system." It is an often heard definition that a workstation constitutes "a set of relevant components, factors of production process interconnected and functioning as a homogenous production element which performs some specific tasks."

Taking system theory as a starting point, a workstation can be described as "the smallest system (subsystem) of work, separated from a superior system and integrally connected with it by fulfilling the partial target (task) which necessarily refers to the main target of the superior system."

If one treats a workstation as a system, it should involve not only a working space and instruments, but also a human and his relations with the environment.

In the descriptions of workstations, a technical and organizational point of view usually prevails, which is connected with the work division and specialization of a station. Unfortunately the problems about worker's needs, social status in the structure of the company, rights and prestige are left behind. Describing the fragmentary structure of a workstation involves the loss of information about the worker's motivation, attitude to duties and responsibility.

The desirable behavior of a worker can be achieved only when

the process of planning the work system takes into consideration psychosocial determinations of a worker's behavior and the technical and organizational requirements of a workstation.

While analyzing the scope of commonly accepted definitions, one often comes across the statement that a workstation is "a part of a production or industrial space," "a part of a machine and production space," "a place of work, an employee, an instrument and machines," "space, equipment and energy," "the smallest production units, "space, machines, object of work and performers," etc.

Some indispensable components of a workstation necessary for the production process to take place include:

- Production tasks.
- An employee (group, team).
- Equipment (machines, instruments, subsidiary equipment, tables, chairs, containers).
- A product or a service which is the object of the work.
- Size and shape of the production space occupied.

Designing a modern workstation involves not only tasks connected with the performance of the production process, but also adjusting a workstation to the human needs and determinations. It is the issue that concentrates the efforts and interests of many specialists from different fields: machines designers, technologists, economists, and the representatives of the humanities and other sciences like anthropologists, psychologists, physiologists, etc. It not only demonstrates the contribution of sciences dealing with technology, economics and organization, but also it emphasizes the significance of medical and social sciences.

This is a good example of tendencies which opt for a greater humanization of work and adjusting it to the psychophysical characteristics of humans. In regard to this problem, a workstation at every stage of its design, analysis, evaluation or improvement should be recognized as a coherent system integrating three main elements: the human–machine–material environment.

The relations between these elements change with the modernization and mechanization of a workstation and show the evolution of links, which become more and more complex as the level of mechanization and automatization increases. In this case the significance of human duties decrease and other functions become of greater significance, which is in turn conditional on a machine's work.

The human–machine system is shown as a separate system to make the interpretation more clear. A workstation as a human–machine system not only must meet the modern requirements of production, but also above all it must fulfill the needs of human.

Thus, one must aim at such technical and ergonomic designs and solutions that would guarantee the best harmonization of the two elements of human–machine system or human–technical instrument system.

The human–technical instrument system, defined and described in a proper manner, and reflecting a given workstation, is a basis of an analysis and evaluation of the phenomena which occur within this system.

The human–technical instrument system has a dynamic character: it changes in time, it interacts with physical, chemical, biological and social environment and this environment also influences it. During its operation it can interact with different systems, which can even result in modifying its own previously arranged plan of action.

The name of a basic human–technical instrument system has been adopted for a system consisting of a man and a determined kind and number of technical means which are used to carry out a planned task in the specific time and in the specific way. The elementary character of the system — as a feature of its structure — consists in the fact that depriving the system of whichever element violates its ability to carry out the planned tasks. The basic human–technical instrument systems can differ greatly from each other in many following ways:

- The degree of complexity of tasks.
- The intensity of environmental influence (above all physical, chemical and biological factors of the environment).
- The strength of relations between other basic systems.
- Significance for the subsystem operation.

There are three types of the basic human–technical instrument systems that depend on technology and the level of modernization:

- Systems consisting of a man and a machine (instrument), e.g. a turner and his turning lathe, a blacksmith and his hammer, a driver in his vehicle.
- Systems consisting of a man and a fragment of a larger technical instrument, e.g. a worker working at a streamline or at a lathe.
- Systems consisting of a man and a larger number of single machines that work at the same time, e.g. a worker inspecting the work of lathe machines.

4. ORGANIZATION OF WORKSTATIONS

The organization of a workstation aims at such collection of means and their use (and the application of these means and methods) which makes it possible for a performer to carry out his task in the proper manner.

The proper organization of workstations and production processes aim at achieving most favorable economic results by increasing the productivity and better work quality — using the least effort on the part of the workers — and providing safe and comfortable work conditions.

The organization of a workstation depends on the organization system of the group of workstations consisting of the individual stations. Some claim that the organization of a workstation is a system made of properly applied elements, which interact to achieve a given target. The aim of the accepted organization of a workstation is to provide an effective fulfillment of tasks in the situation of some inner disturbances and mutable influence of the environment.

Nowadays the development of sciences concerning work like the organization of work, the analysis of work and, particularly, ergonomics has changed in the significant way the attitude towards the organization of work. The organizers leave behind the traditional methods and become more and more interested in the system solutions. They treat a workstation as a system separated from the environment — a subject to changes due to the influence of the environment and a subject interacting with the environment. The system approach in the organization practice is very effective.

The main components of the organization system at a workstation are:

- Methods of work.
- Organization of work.
- The spatial organization of a workstation.
- Service.
- Material environment of work.

The organization of a workstation can be analyzed from the three points of view, i.e. space, time and relations to other workstations.

While adopting the spatial point of view, some elements taken from the entire analysis of the three main factors of a workstation find their reflection in the specialized fields of scientific research, e.g.

- Human adaptation to work (the physiology of work, psychotechnology, etc.); it involves the analysis of the performer's features — a man in his working space — in a static manner, separated from the other main factors of a workstation.
- Work adaptation to human (ergonomics); it involves the analysis of the means and conditions of work in the working space. This set of problems is most often the object of research that results from the common understanding of a workstation where the equipment is the richest and the most important source of changes and improvements.
- Human adaptation to human (sociology and psychology of work); it involves the analysis of a human–performer in his relations with other performers working at workstations and managers.

Analyzing the main factors of a workstation (performer, means and objects of work) in time, i.e. the harmonization of the work time of the individual factors, which is usually analyzed separately and then analyzed in their mutual relations.

The main object of analysis of this approach is human (performer), which is also an object of interest of many specialized institutions with a whole spectrum of methods, techniques and means.

Analyzing the work time of a worker and looking for the ways of increasing his efficiency is not always a right and effective approach. While operating unique and high-tech machines, the main problem is to provide work continuity of these machines and not the problem of work pauses for the workers, engineers or technicians who operate the machines (e.g. operating electronic code machines, energy systems, etc.)

When no main factor of a workstation is engaged in the production process, there is a pause in the work of a workstation.

While analyzing the organization of workstations, some special attention should be paid to the individual factors during the work pauses. These pauses are described as technological breaks.

The analysis of workstations is from the third above-mentioned point of view: in their relation to other workstations — apart from the analyzed problems of the psychology and sociology of work — it brings into view some work means indispensable to a workstation like transport machines, communication and signaling means, etc.; and particularly work objects in their production course starting from a stock (in the form of the initial material) through all the production and inspection stations (in the form of semiproducts) to the store and market (in the form of ready products). The time of the material course in the production process together with

expectations determines the length of the production cycle, which in many cases can be a measure of the level of the organization of production.

With the higher types of production (of an average and large series mass production) the relations between workstations are more and more straight.

While analyzing the organization of a workstation, it cannot be identified exclusively with the work organization and the organization of production. The humanocentric approach must be taken into consideration, that is ergonomic aspects, and also a classical human–machine system must be connected with the environmental factors and the technological problems whose degree of modernization determines the organization of a workstation.

Concluding the organization of a workstation must involve elements of work organization, production, technology and ergonomics.

Taking for granted that a workstation is a human–technical object system, one can define the organization of a workstation as a purposely designed system of organizational, technical and ergonomic elements that are included in the system of interaction with the environment and which will allow one achieve this goal/target. While designing as well as analyzing the organization of a workstation, the role of the environment must be considered and its significance must be revealed, i.e. the influence on the human–technical object must be analyzed if its disturbing, stimulating or neutral/indifferent impact is to be proved.

The organization of a workstation therefore must be a result of three factors from the field of the organization of work and production (O), technology (T), science about human being, his psychophysical abilities, the motivation of his actions (E) and his behavior in the working team:

$$O = f\{(T_1 \ldots T_n), (O_1 \ldots O_n), (E_1 \ldots E_n)\},$$

where T is technological factors — technology and properly chosen machine determines a type and form of a workstation; O is organizational factors (of production and work) — a workstation is under a constant influence of the organization of material supplies and receipt/collection of the product (the pace of work, the means of transport, the means of mutual synchronization and spatial relations of instruments and the objects of work during the following operations, methods of work, etc.); and E is ergonomic factors — adaptation of a workstation to the psychophysical features of a working subject/human and the shape of the occupational environment.

Some of these factors are interconnected, which means that the change of one of them involves disturbances in the system of the others and requires their change. Ergonomic factors are connected with both other groups of factors.

The changes in the first two groups of factors often influence the form of a workstation when one considers its adaptation to a human being. Technology and the organization of production very often become invariable determinants to which the entire body of ergonomic designs must be adjusted (so-called corrective ergonomics). When analyzing and evaluating a workstation one must bear in mind a most favorable model, which means that one must know in advance what the aims are and if they can be achieved.

REFERENCES

GÓRSKA, E., 1998, *Ergonomic Diagnosis of Workstations* (Warsaw: Oficyna Wydawnicza PW).

GÓRSKA, E. and JUCHELKO, H., 1994, *Ergonomics and Organization of Workstations* (Warsaw: WPW).

Part 7

Environment

Environmental Ergonomics

A. Hedge

Department of Design and Environmental Analysis, NYS College of Human Ecology, Cornell University, Ithaca, NY 14853-4401, USA

1. INTRODUCTION

Environmental ergonomics studies our physiological and behavioral reactions to the ambient environment, and the design of effective barriers that allow us to survive in otherwise inhospitable settings. Research on physiological and behavioral reactions includes studying topics such as responses to cold and heat stress, our capacity to perform work in extreme climates, and acclimatization processes. Research also focuses on requirements for clothing and enclosures that allow us to live in extreme terrestrial climates that range from deserts to Polar regions, or that allow us to venture into extremely harsh environments, such as the ocean floor or outer space. Effective barrier design encompasses the following:

- Clothing — the primary barrier that functions as our second skin and provides for thermal comfort and protection from adverse conditions (personal protective equipment), such as extremes of cold, heat, wind, wetness, air pressure, chemical contaminants, etc., as well as providing adornment and modesty values.
- Enclosures — two types of secondary barrier function as our third skin to provide safe and comfortable conditions:

1. Vehicles — most types of vehicles function as a protective barrier against climate conditions. Everyday transportation vehicles (cars, buses, trains, etc.) provide internal conditions within the vehicle that are regulated for human comfort. Specialized transportation vehicles (planes, submarines, space shuttle, etc.) provide a comfortable climate that allows human survival in otherwise fatal conditions.
2. Structures — climate conditions within buildings are designed to provide us with appropriate ambient environmental conditions (thermal comfort, ventilation and indoor air quality, lighting, acoustics, etc.) in even the most inhospitable terrestrial locations. Buildings regulate daily and seasonal variations in climate and this allows people to populate a much wider range of environmental niches, well beyond those in which the clothed body alone is capable of surviving.

2. AMBIENT ENVIRONMENT CONDITIONS

2.1. Thermal Conditions

Being warm-blooded creatures, we strive to maintain a thermal balance with our environment within relatively narrow limits ($37 \pm 1°C$). For survival, acceptable limits for the deep body core temperature are between 35.5 and 39.5°C. The skin can tolerate a much wider temperature range with limits between –0.6°C (skin freezes) and 45°C (skin begins to burn). Thermal balance occurs whenever net body heat production matches the rate of heat gain or loss via physiological processes. The body generates heat through metabolism and muscle activity not related to external work, and it exchanges heat with the environment via several processes. We can lose or gain heat by:

- *Radiation* between skin or clothing surface and surrounding surfaces, e.g. windows, sun.
- *Conductance* to the surfaces by direct contact with skin or clothing, e.g. sitting on a warm or cold surface.
- *Convection* — a form of heat conductance where the surrounding air close to body absorbs heat. Natural convection occurs as air warmed by body heat rises. Forced convection occurs when air is moved past the body, such as by a fan, or by loose clothing (bellows effect).
- *Evaporation* of water through the outer layers of skin (insensible perspiration) or from the skin surface when this is wetted by sweat (perspiration) or some other external agency affects heat exchange. At maximum sweating, evaporation accounts for 15% of heat loss. Heat is also lost through warming and wetting of air that is inhaled and then exhaled and by excretion of feces and urine.
- *Storage* of heat in the body. Ideally the rate of heat storage should equal 0 when the body is in heat balance i.e. heat production = heat loss with no storage.

In practice the body rarely attains or maintains heat balance and environmental ergonomists study the many factors influence the effectiveness of the heat exchange processes.

2.2. Heat Exchange Mechanisms

Three physiological mechanisms for heat exchange allow us to adapt to variations in thermal conditions. These are:

- *Vasomotor* — the skin needs a minimal blood supply to keep it alive. For thermoregulation this blood flow can be increased many times. Increasing skin blood flow raises skin temperature, which increases heat transfer to the environment, and cools the body core. Decreasing skin blood flow cools the skin, reduces heat transfer to the environment and warms the body core. Changes in skin blood flow are most marked at the extremities of limbs (hands and feet) and less marked in the trunk and head.
- *Sweating* — transferring heat from the core to the skin becomes increasingly difficult as skin temperature approaches the body core temperature. In hot environments sweat evaporation cools the skin surface allowing heat transfer from the core. Sweat rate varies from a minimum of ~650 ml/day, to a daily maximum continued total sweat rate of ~1 liter/h. Over shorter periods of up to 6 h the maximum short-term sweat rate may be as much as 2.5 liter/h. Sweat is secreted by two types gland:

1. *apocrine* — forehead, back, palms of hands, armpits — produce protein-containing sweat; and
2. *eccrine* — produce dilute salty and watery sweat from all other skin areas.

- *Shivering* — when skin blood flow is minimal there may be excessive heat loss from the body core by conduction through the skin. Maintenance of body core temperature requires an increase in heat production and shivering is disorganized muscular activity that has this effect, and increases heat production 300–400%.

2.3. Effects of Heat and Cold on the Body

Our ability successfully to thermoregulate depends on individual factors, such as age, alcohol consumption, general health and acclimatization to thermal conditions. For a normal adult, when

the body core becomes too warm, blood flows from the body core to the skin, and sensible perspiration starts, which dissipates additional body heat by sweat evaporation. *Hyperthermia* occurs if heating of the body continues beyond its ability to lose heat through perspiration (i.e. the upper critical temperature). Hyperthermia risks are greater in hot industries, such as steel or glass making factories, and in hot climates. Hyperthermia can be fatal, with between 240 and 1700 deaths per year in the USA depending on heatwave conditions.

When cold, skin vasoconstriction occurs to shunt blood to the body core, *piloerection* (i.e. the hairs of the skin stand up) helps trap an insulating layer of air over the skin, and shivering starts. *Hypothermia* occurs if body cooling continues beyond our ability to generate and conserve heat (i.e. the lower critical temperature). Hypothermia risks can occur in cold industries, such as refrigerated cold storage, arctic or antarctic settings, mountain climbing, deep sea fishing or diving, outer space, etc., or in cold climates. Hypothermia can be fatal with 600–1000 deaths per year in the USA, depending on the severity of the winter.

2.4. Thermal Comfort

Thermal comfort is a psychological concept that is affected by at least six variables:

- *Air temperature* — temperature (°C) of the ambient air.
- *Mean radiant temperature* — average temperature (°C) of radiant sources, such as a fire. Radiant asymmetries (warm front, cold back; warm head, cold feet) can produce thermal discomfort.
- *Air speed* in meters per second past the body. Air moving over the skin cools this.
- *Humidity* — percent relative humidity. (Percent relative humidity is the ratio of the moisture content of the air divided by the maximum moisture content of that air at the same air temperature and barometric pressure, _ 100.) Sweat evaporation is less effective as humidity rises.
- *Activity* — muscular activity generates heat.
- *Clothing insulation* — different clothing designs and materials have different thermal insulation properties (clo units; 1 clo = 58 W/m²). Summer attire is ~0.6 clo and a winter ensemble ~1.0 clo. A well-padded high-back office chair insulates the body at ~0.3 clo. Clothing insulation is reduced by wind or body movements that disrupt the air layer between the skin and the clothing, or by wetting the clothing.

2.5. Thermal Comfort Standards

For most indoor spaces we attempt to construct settings that provide for thermal comfort. Two thermal comfort standards apply to indoor environments.

2.5.1. Thermal comfort zone (ASHRAE [American Society for Heating, Refrigerating, and Air-conditioning Engineers] 55-1981)

Thermal comfort conditions are defined as those for which > 80% of the people will not report thermal dissatisfaction. These conditions are specified as a zone on a psychrometric chart representing various combinations of air temperature and relative humidity for particular air velocities, clothing ensembles, and seasons (winter and summer). Air temperature and mean radiant

temperature are assumed to be equivalent, and activity levels are assumed to be typical for indoor settings.

2.5.2. Fanger's comfort equation (ISO 7730)

Thermal comfort is defined by the physical state of the body rather than environmental conditions (Fangor 1970, ISO 1984). A comfort equation is computed that predicts comfort temperature (°C), the average thermal comfort votes for any group of people (predicted mean vote, PMV), and the percentage of people who will be dissatisfied with the thermal environment (predicted percentage dissatisfied, PPD). Several experiments show that for sedentary work and light clothing, the preferred temperature is close to the 25.6°C predicted by Fanger's equation. Preferred temperatures appear to be uninfluenced by age, race, or individual differences (Fangor 1970). Studies of sex differences in thermal comfort have produced conflicting results. Recently, computer programs and portable thermal comfort instruments based on ISO 7730 have been developed. However, in some laboratory and field research the PMV does not always give good predictions of thermal comfort ratings (McIntyre 1980, Bensel and Santee 1997, Konz 1997).

2.6. Temperature and Performance

Many studies show how changes in thermal conditions can affect human performance (Konz 1997, Parson 1998). Extremes of heat or cold can decrease work performance. The effects of heat on task performance may depend upon the cognitive complexity of the task and the level of heat exposure that occurs. Changes in thermal conditions can affect worker performance in office buildings, for example, typists with no incentives or knowledge of results produce more work at 20°C than at 24°C. Students' *t*-test scores generally are better with 22–23°C than > 26°C. Cold conditions can impair manual dexterity when hand skin temperatures are between 13 and 18°C. Cold conditions can also affect psychomotor task performance.

2.7. Ventilation and Indoor Air Quality

Indoor air quality focuses on the nature and concentrations of air pollutants (particles, fibers, gases or bio-aerosols) inside of enclosures that can adversely affect comfort, performance and health. Ever since the energy crisis of the early 1970s, which resulted in reductions in the design standard for mechanical ventilation rates, there has been a growing awareness of the impact of indoor air quality on comfort, health and performance.

2.7.1. ASHRAE Standard 62-1989

In the USA the ASHRAE Standard 62-1989 sets minimum ventilation rates for occupant comfort. Acceptable indoor air quality as "air in which there are no known contaminants at harmful concentrations as determined by cognizant authorities and with which a substantial majority (80% or more) of the people exposed do not express dissatisfaction." Occupants of modern, sealed, air conditioned buildings, may experience symptoms of the sick building syndrome that may be caused by poor indoor air quality (O'Reilly *et al*. 1998). (Sick building syndrome describes complaints of non-specific symptoms among building occupants, such as headaches, sore and irritated eyes, sore throats, and flu-like symptoms.) Indoor air pollutants can affect cognitive performance, cause eye, nose, throat, and lower respiratory tract

irritation, elicit acute and chronic health symptoms, provoke allergic reactions and fatal illnesses. Many factors affect indoor air quality, including outdoor air quality, ventilation system design and performance, and emissions from building materials, technology, workers, and their activities. Indoor air quality studies typically investigate a wide array of contaminants that include:

- *Combustion gases* — exhaust emissions of combustion gases can include:

1. *carbon monoxide* — a colorless, odorless gas. It affects psychomotor task performance and is toxic at high concentrations;

2. *carbon dioxide* — a colorless, odorless gas that we exhale. High carbon dioxide levels promote drowsiness, and ultimately asphyxiation;

3. *nitrogen oxides* — several gases that can cause mucus membrane irritation and provoke bronchospasms; and

4. *sulfur oxides* — several gases that can cause mucus membrane irritation and affect lung function.

- *Ozone* — a colorless, pungent gas that causes mucus membrane irritation, provokes asthmatic reactions, and has toxic effects. A major component of photochemical smog that occurs in may cities.

- *Volatile organic chemicals (VOC)* — a wide variety of carbon chain or ring compounds. Indoors they may be emitted from construction materials, paints, adhesives, furniture, office technology, cleaning agents, smoking, etc. Studies of building products emissions report ~80% of indoor VOC may cause eye irritation, and ~30% are suspected carcinogens. Exposure to low level mixtures of VOC adversely affects health and performance. Chamber studies show that exposure to a mixture of 22 VOC, typical of indoor settings, produces detectable neurotoxic symptoms within 2 h of exposure to a concentration of 25 mg/m^3 (O'Reilly *et al*. 1998).

- *Formaldehyde* — a colorless gas with a pungent odor. Toxicity occurs through contact with skin and mucous membranes (eyes, nose, throat). Some people can become sensitized to formaldehyde. Exposure can induce fatigue, memory lapse, headache, difficulty sleeping, respiratory problems (including asthma), and eventually chronic respiratory disease and possibly nasal and respiratory tract cancers.

- *Radon* — a colorless, odorless radioactive gas that emits radiation as it decays into its progeny through a decay chain. Radon is thought to cause some 20 000 deaths per year among non-smokers.

- *Bioaerosols* — In the USA, respiratory tract infections (mainly viral illnesses) account for 75 million visits to physicians, 1.25 million hospitalizations, $15 billion direct medical costs, 150 million lost workdays and $59 billion in indirect costs (lost income, absenteeism). Three groups of microorganisms, viruses, bacteria and fungi, are responsible for many of the bio-aerosol illnesses associated with indoor exposures. Bio-aerosol exposure can be fatal, such as those deaths from Legionnaire's disease. The relationship between the risk of respiratory tract infections and the spatial design of buildings and ventilation system performance is only poorly understood. Research shows increased risk of infection in shared offices and in air-conditioned spaces (O'Reilly *et al*. 1998).

2.8. Lighting

Visible light is electromagnetic radiation that is emitted by a source. Since ancient times people have used artificial light sources, such as oil lamps or candles, to supplement or replace daylight. Today, electric lighting provides a wide array of options for creating artificial lighting conditions 24 h a day that support diverse kinds of visual work, irrespective of natural light levels.

Environmental ergonomists studying lighting pay particular attention to six factors that determine appropriate lighting for visual tasks (Boyce 1997):

- *Contrast* — the difference between the luminance (brightness) of an object and the luminance of the background. It is easier to see high contrast targets than low contrast targets. Contrast is affected by several factors, including the type and amount of light sources, by surface reflectance, by viewing angle and by target size.

- *Size* — the larger an object, the easier it is to see. The size of the image on the retina is important, not the size of the object, and this is affected by viewing distance. As we bring an object into closer focus we can see finer details.

- *Time* — seeing takes time. The retinal biochemical processes take time to occur, visual processing and recognition also takes time, especially for unfamiliar objects. Consequently viewing time is important. We need bright light if objects are to be briefly seen.

- *Luminance* — the quantity of incident light arriving from a source. This determines *illuminance* (i.e. the amount of light falling over a unit area) which affects the luminance of the task.

- *Color* — this affects both to contrast and luminance factors because the eye is not equally sensitive to all wavelengths, and ~8% of men and 0.5% of women have color vision deficiency (Boyce 1981).

- *Visual acuity* — the resolving power of the eye is affected by anatomical factors. The shape of the cornea produces ~70% of its refracting (light bending) capabilities, with the crystalline lens providing ~30%. Changes in lens curvature (accommodation) alter refraction and focus light onto the photoreceptive retina.

2.9. Visual Processes

Light is detected by photoreceptors in the retina of each eye. Each retina has ~130 million rods (monochromatic vision), and ~7 million cones (color vision) concentrated in the fovea, the region of greatest sensitivity and acuity. Plentiful light allows high visual acuity and color vision (photopic vision). Low light levels reduce visual acuity and impair our ability to see colors (scotopic vision). The eyes adapt slowly when moving between light and dark spaces, and visual recovery can take up to 1 h. The eyes adapt more rapidly when moving between dark and light spaces, but immediate visual function can be impaired for a short while. Partial adaptation can occur if visual field contains adjacent dark and bright areas. This is why ergonomists do not recommend placing a VDU screen either facing or backing against a bright background, such as a window.

2.9.1. Aging on visual performance

With age, especially > 40 years of age, the lens loses flexibility and presbyopia occurs. The lens also becomes increasing opaque

with age, which reduces light transmission and can result in a cataract. Fluid filled compartments of the eye, the aqueous humor and vitreous humor, function to provide nutrients to the non-vascular structures within the eye and maintain the shape of the eye. With age, more light scattering occurs, especially at the shorter wavelengths. Retinal sensitivity decreases with age, especially at low luminance levels. Sensitivity to glare increase with age (Anshel 1990).

2.9.2. Lighting, computer work and eyestrain

Workers, especially computer users, may experience glare in their workplace. *Disability glare* is direct glare from a bright light source, such as bright sunlight through a window. *Discomfort glare* is indirect glare from uncomfortably bright reflections, such as ceiling light reflections in a computer screen. Screen reflections can be distracting, and cause visual discomfort and eyestrain. Surveys of computer users find that inadequate lighting is frequently reported, and eyestrain afflicts almost half of all workers. Field studies testing different lighting have found that computer workers report fewer complaints of eyestrain and eye focusing problems, greater satisfaction, and express strong preferences for lensed-indirect uplighting compared with parabolic lensed downlighting (Hedge *et al.* 1995).

2.10. Acoustic Environment

Sounds are an integral part of human communication for most people. Sound is the result of pressure changes in the air, and its intensity is measured in decibels (dB). A very quiet room would have a sound level at ~40 dB, conversations are mostly between 55 and 65 dB, and a jet taking off may emit sound > 130 dB. Sounds > 120 dB usually feel painful to the ears. Brief exposure to high intensity sounds can produce a transient impairment in hearing, called a temporary threshold shift. Prolonged exposure to loud noise can cause a permanent threshold shift and permanent hearing damage.

When sounds are intense, annoying or otherwise unwanted they become noise and that can adversely affect our performance and health. The noisiness of a sound often results from the subjective impression of how annoying it is. Noise can be unwanted sound that carries unpleasant information about the sound source (e.g. a crying baby), or it can be loud, annoying unwanted sound (e.g. a jackhammer outside one's window late at night). The ear is least sensitive to lower sound frequencies, and most sensitive to frequencies in the speech range, especially ~3 kHz. Lower frequency sounds generally are less annoying than higher frequency sounds of the same subjective loudness (Kryter 1985).

2.10.1. Occupational noise

Sources of occupational noise include the vibrations of structures, machines, or their components, and also aerodynamic turbulence (e.g. high-pressure air jet). Occupational noise exposure assessed by quantifying a permissible noise dose that is based on daily exposures. Permissible noise dose is regulated by OSHA (Sanders and McCormick 1993).

Industrial plants typically transmit noise to the external environment via certain sources or routes (e.g. open windows, roof ventilators, steam injectors, compressors, diesel engines).

Planning regulations have reduced the impact of noise generating industries on many residential areas. However, indoors in manufacturing industries 80% of noise levels > 80 dB and 20% > 95 dB. Exposure to noise this loud can cause hearing loss (i.e. physiological deterioration of hearing due to destruction of hair cells in the organ of Corti, and decreased number of associated nerve fibers). Exposure to uncontrollable and unpredictable noises can also have stressful aftereffects on behavior. Studies of occupational noise show that this is associated with annoyance, health problems, accidents and reduced efficiency of performance (Crocker 1997).

2.10.2. Speech privacy

In many modern open-plan offices worker performance is impaired by inadequate speech privacy and distracting noise from computers, telephones, printers, photocopiers, air conditioning and conversations. Speech privacy increases as the intelligibility of encroaching sounds decrease. Solutions to office noise problems include increasing the sound absorptance of room surfaces, especially the ceiling, and reducing source strength (e.g. choosing quieter equipment, encouraging workers to speak more quietly). Sound masking systems can also be effective because they add background "white" noise to the office that lowers the signal-to-noise ratio.

2.11. Electromagnetic Fields

Electric and magnetic fields (EMF) emanate from power lines and from office equipment, such as cathode ray tube computer screens and system units with power transformers. There are concerns about prolonged EMF exposures and health effects, such as nervous system cancers and leukemia. Research findings are conflicting. Most modern computer monitors are low-emission designs and with older computers EMF issues can be minimize by putting the computer monitor and system unit at arms length on a non-conductive surface.

2.12. Vibration and Motion

Oscillating motions can cause vibrations that can adversely affect the body and performance in several ways (Griffen 1997). Whole-body vibration occurs when the body is supported by an oscillating surface, such as a vibrating truck seat or a boat moving up and down on water. Very low frequency whole-body oscillations, usually < 1 Hz can cause *motion sickness*. Whole-body vibration also reduces efficiency for manual tasks such as tracking, typing or writing. Segmental vibration occurs when only a part of the body is in contact with an oscillating surface. Sometimes this is useful, for example, a vibrating pager silently signals a call. Often the segmental vibration is associated with hand-transmission from a tool, such as a chain saw, weed trimmer or jackhammer. In extreme cases hand-transmitted vibration can lead to a disease caused *vibration white finger*.

2.13. Gravity

Riding a roller coaster subjects the body to high rates of motion brief fluctuations in gravity. For some people this is a thrill, for others a fright, and for some a cause of motion sickness. These transient changes are magnified for people in other high-speed settings, for example, a pilot flying a high-G turn, an astronaut during lift-off. High-G situations affect the body in many ways,

increasing loads on the musculoskeletal system and modifying blood distribution, which sometimes causes symptoms ranging from elbow pain to loss of consciousness. At the other gravitational extreme our continued exploration of long-term effects of space requires knowledge about the effects of microgravity on the body and on performance (Albery and Woolford 1997).

3. CONCLUSIONS: WORK–ENVIRONMENT DESIGN — A PROCESS OF INTEGRATION

Environmental ergonomics is making an increasingly important contribution to improving the design of environments. Poorly designed workplaces can create many problems for occupants that affect their attitudes and motivation (e.g. satisfaction with work), their health and stress level, and reduce their efficiency. Well-designed work environments provide good ambient conditions that satisfy the physiological and personal requirements of workers. Environmental ergonomists typically use a systems approach to analyze the interactions between users, technology, the organization, and the ambient environment. They also work as members of a design team that includes other professionals, such as architects, interior designers, and engineers (ventilation, illumination, acoustics). In many respects environmental ergonomics is still in its infancy, yet as human influence spreads into remote and hazardous places on and beyond the earth, environmental ergonomics has an increasingly important role to play.

REFERENCES

ALBERY, W.B. and WOOLFORD, B., 1997, Design for microgravity and microgravity environments. In G. Salvendy (ed.), *Handbook of Human Factors and Ergonomics* (2nd ed.) (New York: John Wiley), 935–63.

ANSHEL, J., 1990, *Healthy Eyes Better Vision* (Los Angeles: Body).

ASHRAE, 1981, *Standard 55-1981: Thermal Environmental Conditions for Human Occupancy* (Atlanta: American Society of Heating, Air Conditioning, and Refrigerating Engineers).

ASHRAE, 1989, *Standard 62-1989: Standard for Ventilation for Acceptable Indoor Air Quality* (Atlanta: American Society of Heating, Air Conditioning, and Refrigerating Engineers).

BENSEL, C.K. and SANTEE, W.R., 1997, Climate and clothing. In G. Salvendy (ed.), *Handbook of Human Factors and Ergonomics* (2nd ed.) (New York: John Wiley), 909–34.

BOYCE, P.R., 1981, *Human Factors in Lighting* (London: Applied Science).

BOYCE, P.R., 1997, Illumination. In G. Salvendy (ed.), *Handbook of Human Factors and Ergonomics* (2nd ed.) (New York: John Wiley), 858–90.

CROCKER, M.J., 1997, Noise. In G. Salvendy (ed.), *Handbook of Human Factors and Ergonomics* (2nd ed.) (New York: John Wiley), 790–827.

FANGER, P.O., 1970, *Thermal Comfort* (New York: McGraw-Hill).

GRIFFEN, M., 1997, Vibration and motion. In G. Salvendy (ed.), *Handbook of Human Factors and Ergonomics* (2nd ed.) (New York: John Wiley), 828–57.

HEDGE, A., SIMS, W.R. and BECKER, F.D., 1995, The effects of lensed-indirect uplighting and parabolic downlighting on the satisfaction and visual health of office workers. *Ergonomics*, 38, 260–80.

INTERNATIONAL STANDARDS ORGANIZATION, 1984, *International Standard 7730. Moderate Thermal Environments — Determination of PMV and PPD Indices and Specification of the Conditions for Thermal Comfort* (Geneva: ISO).

KONZ, S., 1997, Toxicology and thermal comfort. In G. Salvendy (ed.), *Handbook of Human Factors and Ergonomics* (2nd ed.) (New York: John Wiley), 891–908.

KRYTER, K.D., 1985, *The Effects of Noise on Man* (New York: Academic Press).

McINTYRE, D.A., 1980, *Indoor Climate* (London: Applied Science).

O'REILLY, J.T., HAGAN, P., GOTS, R. and HEDGE, A., 1998, *Keeping Buildings Healthy: How to Monitor and Prevent Indoor Environmental Problems* (New York: John Wiley).

PARSON, K., 1998, *Human Thermal Environments* (London: Taylor & Francis).

SANDERS, M.S. and McCORMICK, E.J., 1993, *Human Factors in Engineering and Design* (7th ed.) (New York: McGraw-Hill), 589–681.

Flight Motion Effects on Human Performance

F. H. Previc

Flight Motion Effects Branch, Air Force Research Laboratory, Brooks Air Force Base, Texas, USA

1. INTRODUCTION

The aerial environment can be very dangerous and unforgiving to humans because of its abnormal accelerations, excessive motions and G-forces, and frequently distorted visual cues. Several "sensory" syndromes occur during flight are known to severely impact aircraft control and mission performance, the most prominent of these being spatial disorientation (SD) and motion sickness (MS). These syndromes and other physiological stressors such as high G-forces may compromise perceptual judgments (for example the perceived direction of gravity), cognitive skills (for example temporal estimation), motor function (for example inability to control or initiate eye or arm movements), and overall psychological functions (for example motivation). The remainder of this chapter will describe in greater detail the problems associated with SD and, to a lesser extent, MS and high G-forces. Finally, implications for human factors research and design of information displays in the aerial environment will be discussed.

2. FLIGHT MOTION SYNDROMES

2.1. Spatial Disorientation

A commonly accepted definition of SD is "an erroneous sense of one's position and motion relative to the plane of the earth's surface" (Gillingham 1992). This definition encompasses all erroneous perceptions of aircraft attitude (pitch and bank) and altitude above ground. In its mildest form, SD is almost universally experienced by pilots and other aircrew during their flight careers, but it can also lead to serious safety risks and even mishaps in its severe manifestations. For example, it is estimated that 30% of all fighter pilots have experienced at least one serious SD episode (Kuipers *et al.* 1990), with an actual SD mishap occurring once every 300,000 flying hours. During the 1980s, the United States Air Force suffered 81 Class A mishaps due to SD at an annual cost of over $100M (Gillingham 1992), while a similar percentage of mishaps (~15% of the total) was reported by the United States Navy during this period (Bellenkes, Bason and Yacavone 1992). Although the percentage of general aviation accidents attributable to SD is less than 5%, the percentage of fatal mishaps in this flying community is also estimated at about 15% (Kirkham *et al.* 1978).

Spatial disorientation is typically categorised into three major types: Type I (unrecognised), Type II (recognised), and Type III (incapacitating) (Gillingham 1992). Type I involves a situation in which the pilot feels the aircraft is properly oriented when it is not; information from the primary flight instrument displays that could contradict the pilot's erroneous percept may not be registered if he or she suffers from distraction, task-saturation, or merely negligence in his or her instrument cross-check. Type I SD contributes to over 50% of all SD mishaps, often in the form

of "controlled flight into terrain". Type II SD involves a recognized conflict between the pilot's "natural" orientational percept and his or her "synthetic" orientational percept derived from the instrument readings. The pilot does not always recognise that an SD episode is occurring is such an instance, but may instead attribute the discrepancy to instrument malfunction (which almost never is the case). In Type III SD, the pilot may recognise the SD episode but may be too incapacitated to make the appropriate control response due to vestibulo-ocular disorganisation (for example uncontrollable ocular nystagmus in a spin) or to an inability to establish voluntary motor control (for example the "giant hand" phenomenon). Although Types II and III are less likely to result in an actual SD mishap than is Type I, Type II SD is extremely common and even Type III SD has been reported in about 10–15% of fighter pilots.

The causes of SD are multifold and involve several sensory and perceptual systems as well as cognitive and motor factors. As will be discusssed later in this article, erroneous visual and vestibular inputs are the source of most SD illusions. Cognitive impairments, mainly in the form of loss of situation awareness, can precipitate an SD episode, whereas motor-control problems in the form of a conflict between the voluntary cortical (lateral) and primitive reflexive (ventromedial) motor systems can exacerbate an SD episode to the point of Type III.

2.1.1 Visual aspects of SD

Visual interactions in 3-D space

Various visual illusions and other aspects of SD have traditionally been interpreted using the distinction between focal and ambient vision (Gillingham and Previc 1993), but this dichotomy has recently been incorporated into a more general model of 3-D space that includes four major realms of visual processing (Previc 1998). The four realms, shown in Figure 1, are the 1. *peripersonal*, used for reaching and manipulation of objects, 2. *focal extrapersonal*, in which visual search and object recognition occur, 3. *action extrapersonal*, devoted mainly to navigation and orientation within topographically defined space, and 4. *ambient extrapersonal*, in which postural control and orientation within earth-fixed space occurs. The ambient extrapersonal system, which extends over the entire visual field and represents the most radially distant portion of our visual world, provides the framework that allows us to maintain our spatial orientation on earth preconsciously. Through a combination of perceptual, oculomotor, and postural mechanisms, ambient vision helps to maintain a stable, vertically oriented world despite the constant movement of the visual field during various types of self-motion. By contrast, the focal extrapersonal system maintains a high-resolution but attentionally demanding visual processing realm limited to the central 30° or so of the visual field. From the standpoint of SD, the fundamental visual problem is the loss or distortion during degraded visual conditions of the "natural" ambient visual reference frame with its wide field-of-view (FOV), which must be supplanted by a mental representation of the aircraft's orientation in space derived from aircraft instrument displays using spatially limited focal-extrapersonal vision. Such a "synthetic" orientational percept usually consumes considerable attentional resources and is frequently unable to achieve "visual dominance" over nonvisual (for example vestibular) orientational inputs.
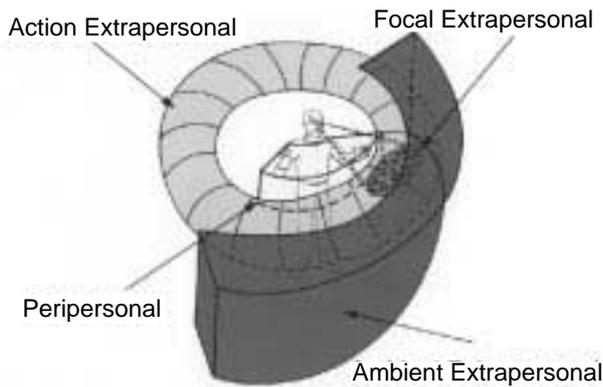
Figure 1. The four regions of interactions in 3-D space: peripersonal, focal extrapersonal, action extrapersonal, and ambient extrapersonal. From Previc (1998, Fig. 3).

Visual SD illusions

The most prominent types of visual illusions are those derived from distorted or absent ambient vision, especially involving terrain or horizon information (Gillingham and Previc 1993). False horizons and/or surface planes can be especially dangerous, leading to illusions of pitch, bank, and altitude above ground. Some of the most dangerous ambient visual distortions in bank are sloping cloud decks (as when a moving weather-front expands the cloud deck on one side), sloping terrain, and tilted sky formations such as the Northern Lights. A well-known example of a false horizon in pitch occurs when flying over a shoreline at night, which recedes beneath the aircraft during level flight in the same manner as an actual horizon does while the aircraft pitches upward and can, therefore, lead to an illusion of pitching-up if the lights of the shoreline are misperceived as the horizon. The sloping walls of canyons, if sufficiently large, can also lead to a misperceived slope of the canyon floor and a misjudgment of terrain elevation. Finally, distorted illumination gradients can induce powerful inversion and other orientational illusions, if the normal tendency for the lighter regions of the visual field to be on top is violated; two of the best-known illumination illusions involve 1. flying over water at low sun-angles toward an incoming cloud-front that darkens the sky relative to the water, and 2. flying in clouds when the sun angle is low, which can severely tilt the ambient illumination gradient.

An absent ambient visual reference frame can also be very dangerous to pilots. In daylight conditions blowing sand and snow can create situations known as "brownout" and "whiteout" respectively. A sparse and irregular terrain can destroy the beneficial effects of many monocular depth cues such as texture gradients and perspectival splay in judging altitude and distance, especially in the case of foreground ridges obscured by direct sun angles. Night-time conditions can also destroy the ambient reference frame by blending ground lights and starlights or removing the horizon and terrain features altogether. This can lead to SD during aerial refueling, formation, deployment of flares, and many other situations containing only limited visual references that move independently of the horizon and ground planes. Night-time landings are especially difficult, giving rise to the "black-hole" illusion in which an isolated runway appears be more distant than it is, leading to a short approach on landing. The "black-hole" illusion is especially dangerous when combined

with a narrow or upsloping runway, upsloping terrain beyond the runway, or weather conditions such as rain or fog, all of which create the illusion of a runway being further away than is actually the case.

Generally, focal extrapersonal processing is used to help us orient only when ambient vision fails us, as in the above situations. However, sometimes focal-mode vision can dominate our orientational percepts even when an ambient reference frame is available, as in the case of "known size". Because of size-constancy mechanisms, ground features become smaller as our distance to them increases, but narrow runways or forests composed of atypically small trees can contradict this information.

2.1.2 Vestibular aspects of SD

Vestibular processing and its limitations

The vestibular system contains two major sets of organs: the six semicircular canals (one pair for each of three orthogonal planes of motion) and the four otolith organs (the utricle and saccule on each side). The cardinal principle of vestibular processing is that 1) the semicircular canals act as angular accelerometers in sensing, by means of the hair cells attached to a gelatinous structure known as the cupula, the inertial lag of the endolymph inside them as the head moves, and 2) the otolith organs, by virtue of the relatively dense calcium-carbonate crystals located on their membranes, respond to the shearing forces produced by linear acceleration or head movements relative to gravity. The vestibular system is a diffusely projecting system that is closely linked to the ambient extrapersonal visual system in its cortical projections; hence, it complements the earth-referenced ambient system in providing an gravitational frame for our perceptual-motor activity (Previc 1998).

The vestibular system is ideally designed to detect transient motions on earth, such as horizontal head movements and walking (which typically occur in a frequency range above 1 Hz). Unlike the ambient visual system, it is not designed to detect *sustained* head rotations or linear accelerations. For example, during an angular rotation lasting only a second or so (see left portion of Fig. 2), the canals effectively integrate the angular acceleration signal to accurately sense head-velocity in space; because the inertial lag of the endolymph dissipates after ~5–10 sec, however, the canals may signal a turning in the opposite direction if a slowing of a sustained constant-rate turn occurs (see right portion of Figure 2b). Similarly, the otoliths correctly sense the velocity of the head in space if the linear-acceleration signal is less than a second or two, whereas the movement of the otolithic membrane during a more sustained acceleration is interpreted as a shift of the head relative to gravity (which produces the same hair-cell shearing motion in the otoliths, as shown in Figure 3). The difficulties in interpreting vestibular inputs during the sustained angular and linear accelerations found in the aerial environment constitute the major source of vestibularly generated SD illusions in flight.

Vestibular SD illusions

There are two basic types of vestibular SD illusions: those caused by angular motion and those caused by linear motion (Gillingham and Previc 1993). In many situations, both angular and linear motion illusions are present, often in conjunction with visual SD illusions as well.

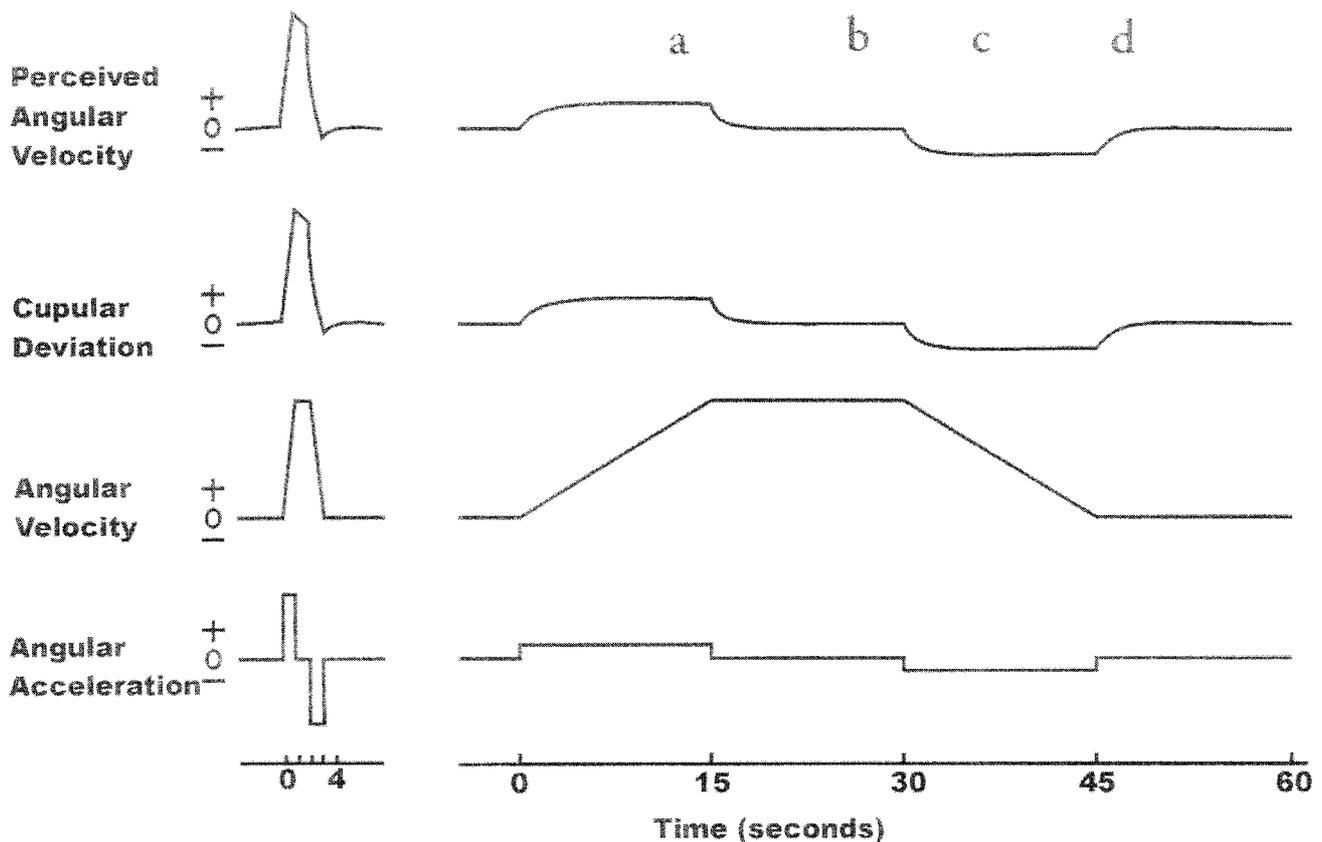One of the historically most dangerous of the angular (or

Figure 2. The different response of the semicircular canals to a transient acceleration and deceleration (left panel) versus a sustained acceleration and deceleration (right panel). As shown at right, the output of the cupula (as well as the subject's perceived turning velocity) is underestimated after the first second or so of the constant-acceleration, increasing-velocity rotation (part a). No cupular response or perceived velocity occurs after the acceleratory stimulus ceases and the rotation attains a constant velocity (part b). The cupula responds to the deceleration in part c and the subject's perception is of rotating in the direction opposite to the original turn, although the actual rotation remains in the same direction as before but with decreasing velocity. From Gillingham and Previc (1993, Fig. 25).

"somatogyral") illusions is the "graveyard spin", whose underlying mechanism is the inability of the canals to relay correct velocity information during a sustained turn. For example, a pilot may attempt to use his right rudder in recovering from a prolonged left flat spin, but even though this effort may successfully stop the spin momentarily it gives rise to a perceived spin to the right because the horizontal canals now sense the rightward deceleration. Another angular illusion is the Coriolis or "cross-coupled" illusion, in which a head movement during a continuous turn stops the motion in a canal that is no longer in the plane of motion and produces an illusion of motion in an orthogonal plane (for example a 90-deg pitch of the head while yawing can lead to a rolling percept because the horizontal canals, now in the roll plane, receive a deceleratory stimulus once they are removed from the plane of yaw motion. The angular SD illusions still contribute to a large number of mishaps in general aviation, but they are considered less of a danger in modern high-performance military aircraft because of the generally smaller turning rates in these aircraft.

The most prominent linear SD illusions are caused by 1. a deviation of the resultant gravitoinertial force vector away from the true vertical or 2. a change in the magnitude of the gravitoinertial force vector. These changes can be produced by sustained accelerations (as during takeoff) or by centrifugal forces that emanate from the inside of turns or level-offs. One of the most dangerous linear SD illusions is the "somatogravic" illusion, which gives rise to a pitch-up illusion during sustained forward acceleration at takeoff and a pitch-down illusion during deceleration on approach (see Figure 3). If a pilot is climbing during the former situation, a lowering of the nose of the aircraft may create a level-off whose centrifugal force is directed through the bottom of the aircraft and leads to a feeling of inversion. The "graveyard spiral" is another dangerous SD illusion caused by a resultant gravitoinertial force that deviates substantially from the gravitational vertical. This illusion, which is manifested in an erroneous perception of being level while banked in a sustained turn, is caused by two main factors: 1. a failure to sense a continued turning after being in a coordinated turn for several seconds, as occurs during several types of maneuvers; and 2. the imposition upon the force of gravity of a centrifugal force emanating from the inside of the turn, which leads to a resultant gravitoinertial vector that is directed through the top of the aircraft. As the pilot rolls out of the turn to level flight, a misperception of banking in the opposite direction (an example of the "leans") will occur. If the pilot follows his or her "natural" orientational percept rather than that indicated by the cockpit instruments, then an unintended return to the original bank may ensue. Finally, a head movement in an excessive-G environment can result in
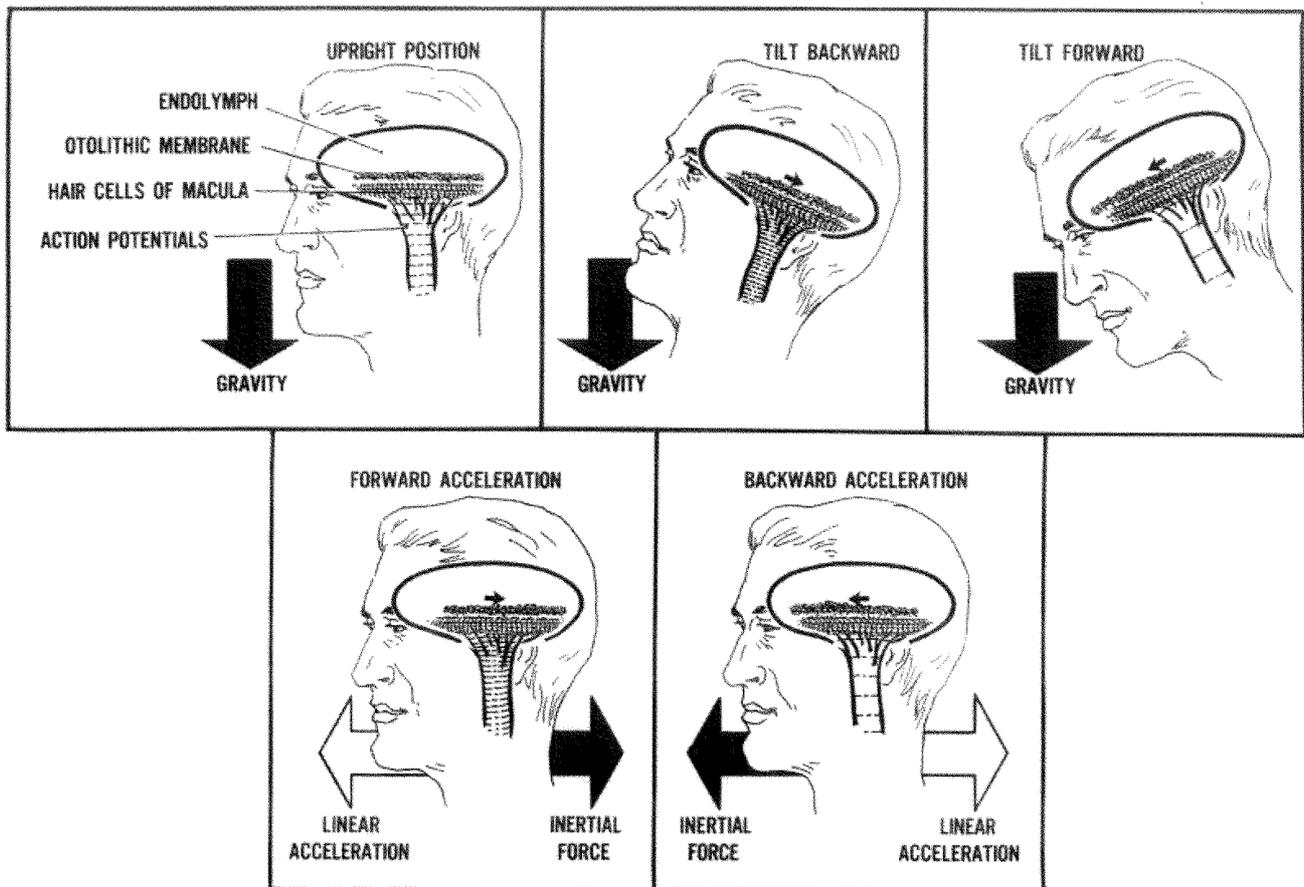
Figure 3. The ambiguity between the otolithic response to a sustained linear acceleration (top part) and a head tilt relative to gravity (bottom panels). From Gillingham and Previc (1993, Fig. 11).

an illusion of aircraft motion in pitch or roll, because greater shearing motion of the otolithic membrane occurs at higher G-levels for the same angle of actual head tilt. This "G-excess" illusion is believed to especially dangerous when a pilot looks upward from the inside of a turn, because the excessive otolithic shearing may be interpreted as a rolling out of the bank by the aircraft; hence, the pilot may inadvertently overbank the aircraft to maintain the coordinated turn percept.

Although a good ambient visual reference such as the horizon can usually overcome illusions caused by erroneous vestibular and other nonvisual orientational percepts, this may be less true of more impoversihed visual stimuli that are perceived as lying with the frame of motion of the aicraft. Under the latter conditions, the visual world may tend to move in concert with the vestibular SD illusions (as in the "oculogyral" and "oculogravic" illusions) rather than break them.

2.2. Motion Sickness

Motion sickness refers to a general autonomic syndrome produced by certain types of motions or conflicts between different sensory inputs during simulated or actual motion. The most prominent symptoms of MS are pallor, cold sweating, nausea, and vomiting, but a host of other cardiovascular, gastrointestinal, respiratory, thermal, ocular, and mental (for example drowsiness) symptoms are all associated with this syndrome (Money 1970). Both parasympathetic and sympathetic activity are altered in MS, although the most dramatic manifestation of MS—vomiting—is

believed to be mainly a parasympathetic response. A combination of drugs that block parasympathetic activity (for example scopolamine) or increase sympathetic (for example amphetamine) activity appears to be the most successful method of preventing MS (Money 1970).

The etiology of MS is not entirely understood, but evidence from a number of sources suggest that certain types or combinations of motion are extremely provocative, as are certain types of sensory conflicts. The types of motion that appear to be most likely to lead to MS include low-frequency vertical oscillations and cross-coupled (Coriolis) angular motions created by head movements in a rotating vehicle. Exposure to 0-G environments also generally leads to MS, as in the space adaptation syndrome experienced by about 75% of astronauts; head movements relative to gravity (for example pitch and roll movements) appear to be the most provocative stimulus for such "space sickness". The role of sensory conflict in MS is well-documented while wearing prismatic lenses that distort or invert the visual scene; a similar mechanism may underlie what has been termed "simulator sickness", which is experienced while flying wide-FOV simulators whose actual motion (or lack thereof) does not always match that expected from the motion of the visual scene. Although the relationships of the various types of MS to each other are not entirely clear, it has been confirmed that an intact vestibular system is necessary to experience MS—i.e., persons with deficient labyrinths are not susceptible to any form of MS (Money, 1970). Even when distorted or asynchronous wide-FOV visual inputs

contribute to MS, the conflicts they engender may ultimately rely on interactions occurring at the level of the vestibular nuclei or the vestibularly responsive portion of the cerebellum (Henn, Young and Finley 1974), especially given that distorted visual inputs do not create MS in labyrinth-defective subjects (Money 1970).

Motion sickness is not considered a serious problem in commercial aviation, as the incidence of "airsickness" among commercial airline passengers is only about 0.5%. The incidence of MS among military pilots is much greater, presumably due to the more stressful motions encountered in military aviation. About 10–15% of military pilots suffer from MS during the early stages of training, but only 1% of all pilots are ultimately eliminated because of a failure to adapt to MS (Money 1970). However, even in those cases where frank MS is no longer a problem, more subtle signs of MS such as drowsiness may be manifested at times.

2.3. High–G Stress

Although high G-stress exerts its most profound effects on human performance by diminishing the oxygen supply to the brain and restricting the motions of skeletal motor systems with large inertial masses such as the head and limbs (see Albery, this volume), there are also specific visual and vestibular effects that can affect performance. For example, retinal vascular insufficiency can result in shrinkage of the peripheral visual field and changes in color perception, while vestibular effects include abnormal vestibulo-ocular reflexes (for example upbeating nystagmus) and the previously described G-excess SD illusion. The precise consequences of these G-related visual and vestibular effects are not known, although it can be assumed that they pose a much more serious problem for pilots of high-performance aircraft than for general and commercial aviation pilots.

3. IMPLICATIONS FOR HUMAN FACTORS

The sensory problems created by the abnormal acceleratory environment of flight can, if not properly understood and combated, seriously degrade pilot performance. The best defense against such threats are effective information displays that are vigilantly monitored by pilots who are highly trained in instrument procedures. However, human factors engineers must design displays, switches, and other human-interface elements for worst-case scenarios, rather than ideal ones. They must take into account the fact that, while disoriented or in some other form of visual–vestibular conflict, pilots may suffer from a disorganised cognitive state in which attention resources are reduced, temporal perception is distorted, voluntary motor

commands are difficult to execute, etc. Whereas, for example, recovery from an unusual attitude in a laboratory may require approximately 1 s on average, pilots during actual flight may require up to 5-10 s to clearly determine the attitude of their aircraft using their HUD or other current display. Even highly overtrained responses may fail to be executed under such circumstances, as the pilot may behave more like a novice than an experienced aviator. It is, therefore, important to present primary flight and other control information in a highly effective and intuitive manner, consistent with how the brain naturally processes such information (Previc in press). Greater use should be made of nonvisual (for example tactile and auditory) sensory systems and motor systems that are less affected by G-forces and visual–vestibular conflict (for example the fingers).

It is ultimately through improved design of aircraft control stations that the consequences of SD and related "sensory"-induced flight syndromes can be most effectively counteracted.

REFERENCES

BELLENKES, A., BASON, R., and YACAVONE, D. O., 1992, Spatial disorientation in naval aviation mishaps: A review of Class A incidents from 1980 through 1989. *Aviation, Space and Environmental Medicine*, 63, 128–131.

GILLINGHAM, K.K., 1992, The spatial disorientation problem in the United States Air Force. *Journal of Vestibular Research*, 2, 297–306.

GILLINGHAM, K.K., and PREVIC, F.H., 1993, Spatial orientation in flight, AL-TR-1993-0022, Armstrong Laboratory, Brooks Air Force Base, Texas.

HENN, V., YOUNG, L. R., & FINLEY, C., 1974, Vestibular nuclei in alert monkeys are also influenced by moving visual fields. *Brain Research*, 71, 144–149.

KIRKHAM, W. R., COLLINS, W. E., GRAPE, P. M., SIMPSON, J. M., and WALLACE, T. F., 1978, Spatial disorientation in general aviation accidents. *Aviation, Space and Environmental Medicine*, 49, 1080–1086.

KUIPERS, A., KAPPERS, A., VAN HOLTEN, C.R., VAN BERGEN, J.H.W., and OOSTERVELD, W.J., 1990, Spatial disorientation incidents in the R.N.L.A.F. F16 and F5 aircraft and suggestions for prevention, AGARD-CP-478, AGARD, Neuilly-sur-Seine, France.

MONEY, K. E., Motion sickness. *Physiological Reviews*, 50, 1–39.

PREVIC, F.H., 1998, The neuropsychology of 3-D space. *Psychological Bulletin*, 124, 123–164.

PREVIC, F.H., Neuropsychological guidelines for aircraft control stations. *IEEE Engineering in Medicine and Biology*, in press.

Free Flight

R. Parasuraman
The Catholic University of America, Washington DC, USA

B. Hilburn and J. Hoekstra
National Aerospace Laboratory NLR, Amsterdam, The Netherlands

1. INTRODUCTION

After almost 100 years of service, commercial air travel can look back with some satisfaction on an excellent safety record. Nonetheless, rapid growth in air traffic is threatening the ability of air traffic control (ATC) systems worldwide to handle traffic efficiently and safely. Moreover, even if the airline accident rate remains at its current low value, it is estimated that on average one major commercial airline accident will occur somewhere in the world *every week* simply because of the increased volume of traffic. The recent call for a fivefold improvement in safety by the Gore Commission in the US was clearly motivated by the need to avoid the public outcry that would result from such an outcome.

1.1. What is Free Flight?

To meet this challenge, several proposals to change the ATC system have been put forward. In the US there has been much interest in free flight (FF), which has been promoted by the Radio and Technical Commission for Aeronautics (RTCA) advisory group. FF represents a radical departure from the current method of controlling air traffic, which is based on radar surveillance and ground control of aircraft movements along well-specified paths in the sky. Under FF, aircraft could fly preferred routes and have greater flexibility in maneuvering, including "self separation" from other aircraft. As stated by the RTCA in 1995, the goal of FF is to allow aircraft under instrument flight rules (IFR) the ability to choose, in real time, optimum routes, speeds, and altitudes, in a manner similar to the flexibility now given only to aircraft operating under visual flight rules (VFR). In theory, FF will result in the utilization of more fuel-efficient routes and a reduction in the delays imposed by ATC. To achieve this goal, traditional strategic (i.e. flight-path based) separation will be replaced by tactical separation based on flight position and speed. The responsibility for separation will gradually shift from the current ground-based system to a more co-operative mix between the air and ground.

An alternative approach to FF is to extend the current system of ground-based ATC. Responsibility for separation would remain firmly on the ground, but there would be greater use of automation to support air traffic controllers (ATCos) in the management of an increasingly dense airspace (Wickens *et al.* 1998).

The US Federal Aviation Administration (FAA) Free flight Phase I Plan includes elements of both these proposals. Current views of FF are less ambitious than those initially proposed by the RTCA in 1995, and do not include removal of separation authority from ATCos. Moreover, European proposals for future ATC differ from those in the US, but also include some elements of FF.

1.2. Human Factors in Free Flight

Developing future ATC systems that allow for increased capacity while improving safety levels will pose considerable technical and human factors challenges. For convenience we divide the human factors issues in FF into three categories: (1) air (pilots); (2) ground (controllers and dispatchers); and (3) air—ground (pilot—controller—dispatcher communications, and collaborative decision-making). It should be noted, however, that all three are highly interrelated. In a systems view they should be considered simultaneously, along with the automated systems and technologies (such as the Global Positioning System (GPS) and Automatic Dependent Surveillance-Broadcast (ADS-B)) that will support FF.

2. AIRCRAFT AND PILOT FACTORS

According to a vision of mature FF as espoused by the RTCA, aircraft would generally be free to fly user-preferred routes, modify their trajectories en route, and self-separate with minimal intervention by ATC. By shifting much of the responsibility for these tasks to the flight deck, it is thought that airlines can derive significant economic benefits through increased use of direct routes. FF could also lead to a dramatic increase in airspace capacity, especially over oceanic areas in which only a limited number of flight tracks are currently available.

Although the advent of FF assumes certain enabling technologies, it would represent as much an operational (or procedural), as a technological, evolution. From a human factors perspective, there are many critical flight-deck issues that must be addressed. Two are briefly addressed here:

- Will pilot workload be increased with the use of a cockpit display of traffic information (CDTI)?
- Will pilots be able to use CDTI and other systems to achieve self-separation from other aircraft safely and efficiently?

2.1. Cockpit Display of Traffic Information (CDTI)

Work is beginning on examining these issues in a systematic manner. The most research has been done on CDTI display design, because the concept of presenting traffic information in the cockpit is an old one that was extensively examined by NASA during the 1980s. A primary area of concern is whether a CDTI will negatively impact pilot visual workload, by increasing heads-down time in the cockpit. The early studies established the feasibility of the CDTI concept, but also found some evidence for an increase in pilot workload associated with use of the CDTI. More recent studies have shown that when predictive information is added to the CDTI display (e.g. a threat vector indicating the approach of an intruder aircraft), pilot workload is reduced, although the heads-down problem remains (Morphew and Wickens 1998).

2.2. Airborne Self-separation

Can pilots use CDTI and other cockpit systems to maintain self-separation from other aircraft, with minimal ATC intervention? This is an ultimate goal of FF, and a recent study indicates that it is feasible, at least in a simulated environment (Van Gent *et al.* 1998). These investigators modified an algorithm developed at the MIT Lincoln Laboratory for airborne conflict resolution that would not involve any communication between pilots. The algorithm assumes that aircraft and the destinations to which they fly to can be modeled as positive and negative electrical

charges, respectively. Simplistically speaking, aircraft therefore "repel" one another and are attracted to their destination. Combining these forces creates an avoidance vector for maintaining minimum separation standards (e.g. 5 nm and 1000 ft) between two aircraft. Van Gent *et al.* (1998) tested eight flight crews in an advanced cockpit simulator under varying levels of traffic density. The highest level of traffic represented three times the current normal density of traffic in Europe. All eight crews flew successfully, with no separation losses, and participants rated the system quite favorably. Van Gent *et al.* (1998) concluded that a FF concept of automated airborne separation with no ATC involvement "could not be refuted."

2.3. Other Airborne Human Factors Issues

Several other human factors issues need additional work or are still to be addressed. These include:

- How to display a complex four-dimensional problem when a conflict is predicted?
- How much of the conflict detection and resolution process can be automated?
- How to present a resolution advisory: as a flight plan change or as a more short-term solution such as a heading or altitude change?
- Will the required cooperative behavior still occur in a competitive situation, or will competing airlines abuse the system for economic gains?
- How robust and safe can such a human–machine system be?

3. GROUND FACTORS: AIR TRAFFIC CONTROLLERS AND DISPATCHERS UNDER FREE FLIGHT

The work described so far suggests that pilots could use CDTI to maneuver to avoid conflicts with other aircraft, particularly if predictive information is provided. Airborne self-separation has also been shown to be feasible in simulation studies. Before FF can be implemented safely, many other human factors problems on the airborne side will have to be resolved. However, their resolution alone will not be sufficient: ground and air–ground issues will also have to be addressed.

What are the most important ground issues? To begin with, note that the ground side includes not only ATCos but also airline dispatchers. The role of dispatchers will expand dramatically under FF (Smith *et al.* 1997). Currently, airline dispatchers possess the best information on global weather patterns that can be passed on to pilots seeking to fly fuel-efficient routes. How will competition between airlines impact the role of dispatchers under FF? Will the best route suggested by the dispatcher for one airline also be selected by other airlines? This might lead to congestion and delay, and possible cancellation of FF by controllers due to impending loss of separation, thereby negating any potential benefits of FF. Recent studies of the National Route Program (NRP) suggest that this might occur. The NRP allows aircraft to file direct routes above specified altitudes and at certain distances from major airports. The idea of the program is to lessen cruise flight time en route between airports. However, a study by Smith *et al.* (1997) revealed that many aircraft participating in the NRP requested the same routing, creating heavy traffic on a single preferred route that resulted in delays and in some cases re-routing

back to the original non-preferred route. Such unanticipated effects may also occur under FF and may not be predictable prior to operational tests.

How will FF impact the ATCo? First, it must be recognized that an advanced FF system that implemented airborne self-separation would nevertheless require ATCo intervention under four general cases identified by the RTCA: (1) to ensure separation, (2) to preclude exceeding airport capacity, (3) to prohibit unauthorized flight through special use airspace, and (4) to ensure safety. Two of these exceptions specifically involve cases in which the ATCo may be called upon to resolve a conflict (e.g. if there is an outage in a GPS satellite, or a failure in the conflict detection system). The question thus arises, can the ATCo "step in" and do this effectively if he or she has ceded all or some separation responsibility to the air side?

3.1. The Changed Role of the Controller

FF, if implemented to its fullest extent, will dramatically change the role of the controller. Strategic air traffic management will change to tactical control, and active management to management by exception. The airspace will become relatively less structured, thereby introducing a degree of uncertainty. Responsibility for separation and maneuvering, now firmly under ground control, will move to the air, and surveillance will be shared between the air and the ground. Finally, under advanced FF, the controller will become more of a passive monitor of the airspace rather than an active participant in control.
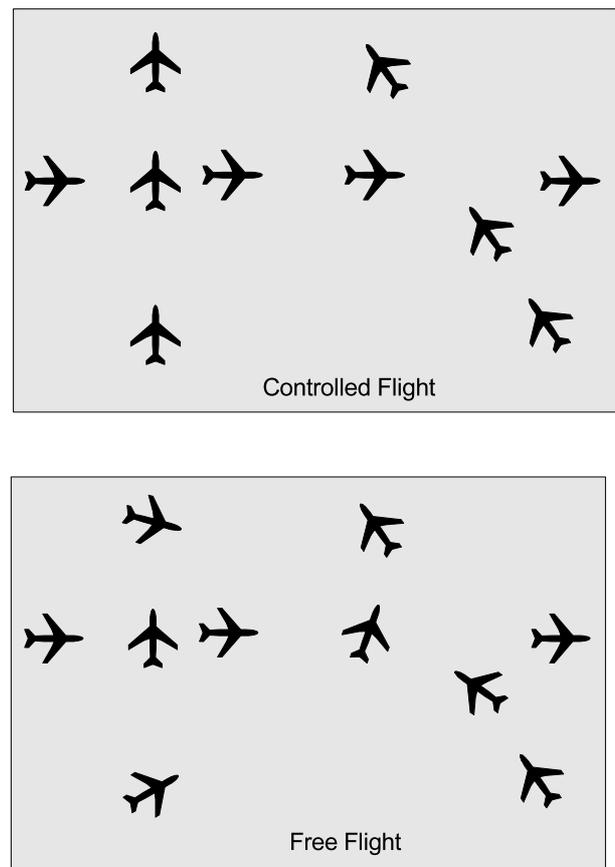

Figure 1. Representation of aircraft movements under controlled flight (top) and free flight (bottom). Potential conflicts are more difficult to predict under FF.

To understand how profoundly the change to FF could influence controller workload and monitoring, consider the simple diagram of figure 1, which depicts the principles of controlled flight and FF in the en-route phase. The diagrams are identical, except that the angle of four of the ten aircraft has been changed under FF. Notice how much more difficult this makes the task of anticipating traffic conflicts. Under controlled flight (top), there are a limited number of areas at which conflicts are likely to occur. Indeed, the historical reasons behind the current-day fixed route structure have to do more with human limitations than with technical or procedural concerns. Under FF (bottom), on the other hand, assuring separation of the same number of aircraft is now a daunting task for the controller.

3.2. Monitoring, Situation Awareness, and Failure Recovery

While relinquishing active responsibility for separation, the ATCo will become more of a monitor as FF moves towards a mature level. A recent report by a National Research Council panel on future ATC raised some concerns regarding this development (Wickens *et al.* 1998). One concern is how effective a controller can be in a purely monitoring role. Human monitoring of automated systems can be poor, especially if the operator has little active control over the automated process and is engaged in other tasks (Parasuraman and Riley 1997).

Recent simulator studies have examined how FF may influence controller performance, situation awareness, and workload (Endsley, *et al.* 1997; Hilburn *et al.* 1997). The studies disagree on the effect of FF on controller workload, with one finding an increase and the other a decrease. The discrepancy may be related to the use of civilian ATCos in one study and military ATCos in the other (cf. Hilburn and Parasuraman 1997), as well as other methodological differences. However, there is agreement that in FF conditions in which the controller is given no information about the aircraft's intent, controller workload is increased. Thus, the provision of intent information will likely be a key issue in FF implementation.

These studies suggest some caution in implementing high levels of FF in which authority for separation is increasingly given to the air side. In the ultimate extension of this trend, *all* authority is ceded to the aircraft. This is the so-called *mature* stage of FF. Such a stage will have been reached when the technologies and procedures are in place for full aircraft self-separation, with ATC in a monitoring role. The end result will be an airspace that is considerably more dense and complex than it is today. As noted previously, Van Gent *et al.* (1998) found that pilots could self separate without ATC involvement in traffic with three times current density in Europe. Will the controller of the future still be able to monitor such a saturated airspace and intervene effectively if, for example, the pilots cannot resolve a conflict, there is a GPS outage, bad weather encroaches into the FF airspace, or there is some other system failure? To examine this issue, a recent study investigated how well controllers could detect conflicts and aircraft self-separating events when all separation decision-making authority was removed from them, as it would be under advanced FF (Galster, *et al.* 1999). Controllers missed a very high proportion of the conflicts, especially under high traffic. Self-separating events went unreported even more

frequently. When conflicts and self-separations *were* identified, they were reported at very short times prior to their occurrence.

These studies suggest that although FF may or not increase ATCo workload, situation awareness and monitoring performance may be adversely affected. This might make ATCo-based recovery from failure in a timely manner nearly impossible (Wickens *et al.* 1998). Requiring ATCos to "step in" to assure separation might therefore put them in a very difficult position. Furthermore, if controllers are to be held legally liable in such a situation, they may legitimately demand a greater controlling role than envisaged by FF advocates to be commensurate with this responsibility.

4. AIR–GROUND COMMUNICATION AND INTEGRATION

Both pilots and controllers will be affected by the technological and procedural changes brought by FF. In addition, as mentioned earlier, airline dispatchers represent another group of agents involved in the airspace system whose role will be changed by FF (Smith *et al.* 1997). Over and above the domain-specific human factors issues that arise for pilots in the cockpit and for controllers and dispatchers on the ground, effective air–ground communication and integration will be necessary for FF to be implemented safely. In commenting on the change from the current system of centralized control and clear-cut division of responsibilities to a distributed system under FF with sharing of responsibilities, Smith *et al.* (1997) have questioned whether effective cooperative decision-making is possible when the different agents have different information sources and differing local goals.

Cockpit studies of FF have focused on pilots' use of CDTI for navigation and for self separation. Ground ATC studies of FF have examined the effects of different FF scenarios on controllers' ability to monitor the airspace and to resolve conflicts if called upon to do so. Since all versions of FF envision a role for controllers (though that role will be different), it would therefore be appropriate to include both pilots and controllers in a single study of FF. This was done in a recent study in which a flight crew flew a 747-400 simulator in a FF airspace that was also monitored by air traffic controllers (Mackintosh *et al.* 1998). As in the study by Van Gent *et al.* (1998), pilots were able to use CDTI to detect and resolve conflicts with intruder aircraft. Under such self-separation conditions, there was greater air–ground communication, usually by pilots informing controllers of the maneuvers they had taken. However, there was a much greater rate of air–air communication between pilots, which has consequences for such issues as pilot workload and frequency congestion.

5. DISCUSSION AND CONCLUSIONS

Human factors studies of FF have just begun. There are several system-level issues that must also be addressed, in addition to the more local human factors problems that we have discussed. These include reconciling the different and sometimes competing goals (e.g. vis-à-vis efficiency and safety) of airlines, dispatchers, pilots, controllers, and regulatory agencies. Predicting the implications of various FF concepts for safety and efficiency also raises significant challenges. Moreover, several research methodologies may need to be used to evaluate the safety impact of FF, in addition to the traditional human factors methods

(Wickens *et al.* 1998). Among these are large-scale simulations and modeling studies that make valid assumptions about the human agents involved in the airspace system; and extensive human-in-the-loop studies of the type that have already been conducted, but with the participation of *all* human and machine agents that would play a role in FF. Such studies may be beyond the capability of any one group or research center, thus necessitating cooperative projects between centers. They may be also quite expensive and time-consuming to conduct, but if the goal of *increased* safety (rather than level safety) is to be reached, the funds for such evaluation efforts should be procured, even as much larger sums of money are being devoted to develop the technologies needed for FF.

Human factors studies of FF are also of broader interest from the perspective of human–automation interaction in complex systems. This is because FF will require the development of several automation tools to aid both pilots and controllers (Parasuraman, Duley, and Smoker 1998). Over the past decade, human factors researchers have documented in some detail the system and human performance benefits as well as costs associated with certain forms of automation (Parasuraman and Riley 1997). Some of these automation tools will be designed to have relatively high levels of autonomy, which will raise significant issues of whether and how humans and "intelligent" agents can share decision-making responsibility in a distributed system (Smith *et al.* 1997).

Another reason why the study of FF should be of broad interest to human factors and ergonomics is that FF represents an emerging set of technologies and procedural changes that has not yet been fully implemented. There is recognition of the need for human factors evaluations of the concept and for such evaluations to continue through the stage of design, operational deployment, and field experience (Wickens *et al.* 1998). Human factors is often a reactive discipline, being considered well after the design of a system is fixed and at a time when only minor changes can be made. In the case of FF, there is some hope that human factors can play a more proactive role by providing results that can be easily translated into design guidelines. Whether this promise is realized, however, will depend not only upon the quality of the human factors input that is provided to the designers, manufacturers, and regulatory agencies, but also on political and economic factors that may fall outside the influence of the discipline

6. REFERENCES

ENDSLEY, M. R., MOGFORD, R., and STEIN, E., 1997, *Effect of Free Flight on Controller Performance, Workload, and Situation Awareness* (Technical Report) (Atlantic City, NJ: FAA Technical Center0.

GALSTER, S. M., DULEY, J. A., MASALONIS, A. J., and PARASURAMAN, R., 1999, Effects of aircraft self separation on controller conflict detection performance and workload in mature free flight. In M.W. Scerbo and M. Mouloua (eds) *Automation Technology and Human Performance: Current Research and Trends* (Mahwah, NJ: Erlbaum), pp. 98–101.

HILBURN, B.G. and PARASURAMAN, R., 1997, Military controllers as an appropriate population for evaluation of advanced ATM concepts. In *Proceedings of the 10th International CEAS Conference on Free Flight*, Amsterdam.

HILBURN, B.G., BAKKER, M. W. P., PEKELA, W.D., and PARASURAMAN, R., 1997, The effect of free flight on air traffic controller mental workload, monitoring and system performance. In *Proceedings of the 10th International CEAS Conference on Free Flight*, Amsterdam.

MACKINTOSH, M., DUNBAR, M., LOZITO, S., CASHION, P., McGANN, A., and DULCHINOS, V., 1998, Self-separation from the air and ground perspective. In *Proceedings of the 2nd USA/Europe Air Traffic Management R&D Seminar,* Orlando, FL.

MORPHEW, M. E. and WICKENS, C. D., 1998, Pilot performance and workload using traffic displays to support Free Flight. In *Proceedings of the Human Factors and Ergonomics Society 42nd Annual Meeting* (Santa Monica: Human Factors and Ergonomics Society), pp. 52–6.

PARASURAMAN, R. and RILEY, V.A., 1997, Humans and automation: Use, misuse, disuse, abuse. *Human Factors,* 39, 230–53.

PARASURAMAN, R., DULEY, J. A., and SMOKER, A., 1998, Automation tools for controllers in future air traffic control. *The Controller: Journal of Air Traffic Control,* 37, 7–13.

SMITH. P., WOODS, D., McCOY, E., BILLINGS, C., SARTER, N., DENNING, R., and DEKKER, S., 1997, Can automation enable a co-operative future ATM system? In *Proceedings of the 9th International Symposium on Aviation Psychology* (Columbus, OH: Ohio State University Research Foundation).

VAN GENT, R., HOEKSTRA, J. M., and RUIGROK, R. C. J., 1998, Free flight with airborne separation assurance. In *Proceedings of the International Conference on Human Computer Interaction in Aeronautics* (Montreal, Canada: Editions de l'Ecole Polytechnic de Montreal), (pp. 63–9).

WICKENS, C. D., MAVOR, A., PARASURAMAN, R., and McGEE, J., 1998, *The Future of Air Traffic Control: Human Operators and Automation* (Washington DC: National Academy Press).

Human Exposure to Vibration

D. E. Wasserman

Human Vibration & Biomedical Engineering Consultant, Cincinnati, OH 45242, USA

1. INTRODUCTION

Worldwide there are millions of people who go to work each day who are exposed to mechanically induced vibration. These include truck, bus, heavy equipment, fork lift, farm vehicle operators, train engineers and conductors, miners, helicopter and fixed-wing aircraft, ship personnel and fixed plant sites. These people are exposed to head-to-toe whole-body vibration (WBV). Users of hand-held power tools such as chain saws, various electrical and air-driven pneumatic tools such as chippers, grinders, sanders, mining tools, drills, road rippers, etc. are exposed to segmental or hand–arm vibration (HAV).

2. VIBRATORY MOTION

Mechanical vibration can be oscillatory, impact or random and each is characterized as a vector quantity and thus consists of both an intensity or magnitude and a direction. Vibration measured at any one point consists of six vectors: three linear mutually perpendicular vectors and three rotational vectors called pitch, yaw and roll. Generally in human vibration work only linear measurements are obtained. Root-mean squared (RMS) acceleration is usually the intensity measurement quantity of choice and is expressed in either $m\ s^{-1}\ s^{-1}$ or gravitational g's, where $1\ g = 9.81\ m\ s^{-1}\ s^{-1}$. There are specific vibration frequencies (Hz) where WBV is measured (1–80 Hz) and where HAV is measured (5–1500 Hz). Finally, there is a undesirable phenomenon known as resonance or natural frequency where humans are optimally tuned to impinging vibration resulting in involuntary amplification and exacerbation its effects at these resonant frequencies; WBV resonance occurs at 4–8 Hz for vertical and 1–2 Hz for the side-to-side or front-to-back directions. For HAV resonance is thought to occur at 100–250 Hz (Wasserman 1987).

3. HUMAN EFFECTS OF VIBRATION EXPOSURE

The major human effects generally can be subdivided by either WBV or HAV exposure.

3.1. Major WBV Effects

At WBV frequencies < 1 Hz, sea sickness or kinetosis can occur which usually ceases when exposure stops. At frequencies > 1 Hz both acute and chronic WBV effects can occur. Generally, if acute exposure occurs at the 4–8 Hz resonance band there can be a potential safety issue, because, for example, a vehicle operator might not be able safely to control the steering wheel; there also can be discomfort due to WBV exposure (Griffin 1990).

The major medical effects of chronic WBV exposure are principally, but not limited to: (1) severe low back pain; (2) lumbar spine disc :degeneration, loss of moisture, buckling, slipping/prolapse, herniation, tearing, distortion, excessive pressure (Dupuis and Zerlett 1986, Wilder *et al*. 1994); and (3) spon-taneous abortions and related gynecological problems have been reported in female vehicle drivers chronically exposed to WBV (Abrams 1990).

3.2. Major HAV Effects

Chronic HAV exposure has been causally linked for > 80 years to a generally irreversible condition of the fingers and hands known as Hand–Arm Vibration Syndrome (HAVS); previously called Vibration White Fingers, and Raynaud's Phenomenon of Occupational Origin. HAVS (Pelmear and Wasserman 1998) is initially characterized by attacks of paresthesia (tingling and/or numbness in the fingers and hands). As HAV exposure continues, finger blanching attacks in one or more fingers, lasting 5–15 min, occur particularly in the presence of cold. With continued HAV exposure, this condition is cumulative and progressive, involving multiple fingers; in its terminal stage (rarely seen) gangrene of the fingers can occur requiring digit amputation. Medical treatment involves the use of so-called "calcium channel blocker" medications which are not curative, but are palliative (i.e. easing pain and suffering without curing). Thus both early detection of HAVS and the cessation of HAV exposure are essential for these hand power tool operators.

Typically HAVS prevalences as high as 50% have been found in pneumatic vibrating hand-tool workers in the USA. Finally, it is important to note that a physician's early and careful differential diagnosis is essential to distinguish HAVS from other similar conditions such as idiopathic Primary Raynaud's Disease and (non-vibration) Carpal Tunnel Syndrome.

4. VIBRATION HEALTH AND SAFETY STANDARDS

Numerous nations worldwide have participated in the development of consensus health and safety standards/guides in an effort to protect workers against the hazards of both WBV and HAV exposure. Initial standards efforts were begun by the International Standards Organization (ISO) in the late 1960s and continue to this day in cooperation with national standards setting organizations such as the American National Standards Institute in New York, British Standards Institute in London, and others including the European Union.

Examples of WBV standards/guides include ISO 2631, ANSI S3.18, BSI 6841. Examples of HAV standards/guides include ISO 5349, ANSI S3.34, BSI 6842. Worldwide virtually all WBV standards employ the same measurement coordinate system; the same is true for HAV standards. This permits easy comparison of vibration data made by various groups.

5. CONTROLLING VIBRATION EXPOSURE

The major engineering methods of vibration reduction include isolation (the intentional alteration of the pathway between the vibration source and the human receiver) and damping (vibration impinging upon a viscoelastic material, causes the material to deform thus converting mechanical energy into a small amount of heat).

5.1. Reducing WBV Exposure

Air-ride seats are commonly used in trucks, buses, farm and construction equipment, ships, fixed plants, etc. to isolate the operator's upper torso/spine from WBV, especially in the 4–8 Hz region. Also in use are isolated truck cabs and special vehi-

cle suspensions. In fixed plant locations, isolating machinery and the remote placing of controls all help to reduce WBV exposure. Administrative controls including regular rest breaks and not lifting cargo immediately after emerging from driving a vehicle.

5.2. Reducing HAV Exposure

First, workers are advised to use tools that are both anti-vibration (A/V) designed and ergonomically designed since an ergonomically designed tool may only represent the optimum interface between the worker and the tool (i.e. handle design, weight, etc.) but these tools do not necessarily have reduced vibration characteristics within the tool *per se*.

Second, do not as a rule use "tool wraps" that use a wade of vibration damping material merely wrapped around a conventional tool handle, because of both the increased handle diameter of the handle and the limited ability of the wrap to damp HAV from certain tools.

Third, use only full-finger, good fitting, A/V gloves that reduce vibration and keep the fingers and hands warm and dry. Do not use "exposed fingers" gloves that only protect the palm since HAVS nearly always begins at the fingertips.

Fourth, good work practices include: keeping the entire body warm and dry especially the hands; no smoking while using vibrating hand-tools since nicotine, cold and vibration all constrict blood flow. Let the tool do the work grasping it as lightly as possible consistent with safe work practices. If signs and symptoms of HAVS appear see a physician immediately. Keep the tool and implements in good working order; replace worn-out tools promptly.

REFERENCES

ABRAMS, R., 1990. Sound and vibration in pregnancy. In R. Creasy and J. Warshaw (eds) *Seminars in Perinatology, Part II* (Philadelphia: W.B. Saunders),pp. 273–334.

DUPUIS, H. and ZERLETT, G., 1986, *The Effects of Whole-body Vibration* (Berlin: Springer).

GRIFFIN, M., 1990. *Handbook of Human Vibration* (London: Academic Press).

PELMEAR, P. and WASSERMAN, D., 1998, *Hand–Arm Vibration: A Comprehensive Guide for Occupational Health Professionals* (2nd ed.) (Beverly Farms: OEM Press).

WASSERMAN, D.E., 1987, *Human Aspects of Occupational Vibration* (Amsterdam: Elsevier).

WILDER, D., WASSERMAN, D., POPE, M., PELMEAR, P. and TAYLOR, W., 1994, Vibration. In P. Wald and G. Stave (eds) *Physical and Biological Hazards of the Workplace* (New York: Van Nostrand Reinhold), pp. 64–83.

Human Aspects of Lighting in Working Interiors

A. Wolska

Safety Engineering Department, Central Institute for Labour Protection,
ul. Czerniakowska 16, 00-701 Warsaw, Poland

1. INTRODUCTION

Lighting is one of the main environmental factors in the work area. Over 80% of information on the work stands is collected (received) by the eyes. We used to talk about lighting the workplace, but in fact we should talking about lighting for the people in the workplace. This is why, lighting design should provide an environment in which people, through the sense of vision, can function safety, effectively and comfortably (Rea 1993). Lighting of an interior should fulfil three functions (ISO 8995, 1989, prEN 12464 1996):

- Ensure maximum safety (making any hazards visible).
- Maintain an appropriate level of visual performance (workers should be able to perform their visual tasks, even under difficult circumstances and during extended periods).
- Provide acceptable visual comfort (interior appearance and good frame of mind).

Safety should be ensured in all circumstances but different emphasis is given to visual performance and the appearance of the interior depending on the nature of the interior. For example, lighting for a tool room in a factory should put much more emphasis on lighting the task than on visual comfort and on the pleasant appearance of the room, but in a restaurant or a hotel room the priorities should be reversed. In all cases, energy consumption and cost effectiveness of a lighting system should be considered.

Lighting of an interior by daylight and by electric lighting should provide optimum conditions for performing a task and an appropriate environment.

2. LIGHTING TERMINOLOGY (PARAMETERS)

2.1. Light (visible radiation)

The Illuminating Engineering Society of North America (IESNA) has defined as radiant energy that is capable of exciting the human retina and creating a visual sensation (Rea 1993). Light is visible radiation. The visible range (spectrum) is between 380 and 760 mm on the electromagnetic spectrum (some sources quote slightly different ranges like 380–780 or 380–770 mm). The visible spectrum can be divided into eight subranges, which correspond to the relevant color of light: violet, blue, blue-green, green, yellow-green, yellow, orange and red (Figure 1).
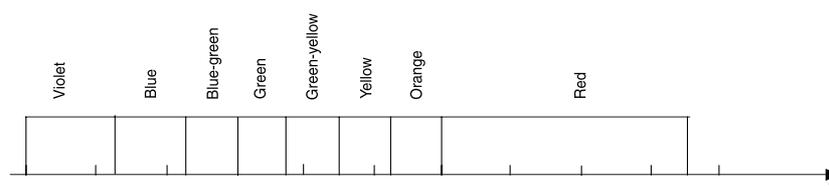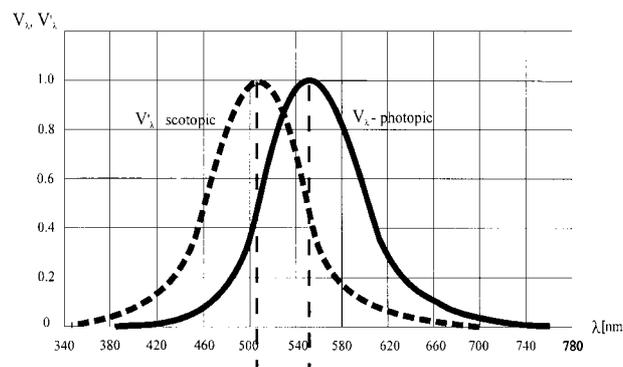
Figure 2. Luminous efficiency curves for photopic and scotopic vision

2.2. Spectral Luminous Efficiency Curves

Physical quantities describing the radiation (like radiant flux, irradiance, radiant intensity) are not sufficient to define (designate) efficiency of radiation to create a visual sensation. Sensitivity of human eye is different for particular wavelengths of visible radiation. It means that eye is more sensitive to some wavelengths than others. There are individual variations, and not everyone has the same sensitivity. In 1924 International Commission on Illumination (CIE) adopted and agreed on a standard response called the CIE standard observer and established luminous efficiency curves for photopic and scotopic vision (North 1993). As seen in Figure 2 the photopic function V_λ describes the spectral luminous efficiency function for photopic (cone) vision with maximum sensitivity at 555 nm (green-yellow) and scotopic function V describes the spectral luminous efficiency function for scotopic (rod) vision with maximum sensitivity at 507 nm (blue-green).

2.3. Photometric Quantities and Units

Photometric units (quantities) are the physical quantities, which are assessed on the basis of the visual sensation they create.

Luminous flux – Φ (unit 1m, lumen) – quantity of light emitted from a light source per time unit. The lumen is a unit relating radiant flux (in watts, W) visually to effective light for a CIE standard observer.

Luminous intensity – I (unit cd, candela) – angular density of the luminous flux in a given direction. It is luminous flux emitted in a very narrow (unit) solid angle containing the given direction. One candela is equal to one lumen per steradian (1 cd = 1 lmsr^{-1}).

Illuminance – E (unit lx, lux) – area density of a luminous flux incident at a point on a surface (luminous flux per unit area - lx = 1 mm^{-2}).

Figure 1. Subranges of visible radiation

Luminance – L (unit cdm^{-2}, candelas per m^2) – physical measure of brightness. It represents the luminous intensity emitted (or reflected) in a given direction per projected area of a luminous (or reflecting) surface.

Contrast – C, luminance difference of two objects, surfaces or parts of the objects in the visual field, which are adjacent to each other (viewed simultaneously) or observed one after the other (successively). Contrast is defined in several ways:

Luminance ratio – ratio between the luminances of any two areas in the visual field (usually observed successively):

$$C = \frac{L_1}{L_2} \qquad (2.1)$$

or:

Luminance contrast – relationship between the luminances of an object and its immediate background (for surfaces viewed simultaneously):

$$C = \frac{L_1 - L_2}{L_1} \qquad (2.2)$$

where $L_1 > L_2$.

or when the areas of different luminances are comparable in size, the following formula may be used (ISO 8995, 1989):

$$C = \frac{L_2 - L_1}{0.5(L_1 + L_2)} \qquad (2.3)$$

where
L_1 = luminance of background or the largest part of the visual field
L_2 = luminance of the object.

Reflectance – ρ, physical property of surface. It is the ratio of the luminous flux reflected from a given surface to the luminous flux incident on it. Reflectance values that approximately equal 1 correspond to bright (white) surfaces and values that almost equal 0 correspond to black surfaces. Depending on the type of texture of surface three main kinds of reflection can be distinguished: specular (mirror or polished surfaces), diffuse (matt surfaces) and mixed combination of polished and matt surfaces, i.e. polished scratch on matt surface). For matt surfaces the following relationship can be used:

$$L = \frac{\rho E}{\pi} \qquad (2.4)$$

where
L = luminance (cd/m2)
ρ = reflectance of the surface considered
E = illuminance (lx)
π = 3.14.

2.4. Visual Performance and Lighting

There are many different types of work stands in industry, offices and in other working interiors, which require specific visual conditions to perform visual tasks, often very complex. The goal of lighting design is to minimize strain of the visual system and to achieve efficient visual performance. Prolonged visual work in poor lighting conditions induces visual fatigue, which can manifest itself in a variety of asthenopic symptoms like headache, redness, tender, itchy, burning, dry, throbbing eyes, focusing problems and blurred or double vision. Visual fatigue symptoms are temporary and after a rest period those symptoms should disappear. On the other hand the term "visual performance" is used to indicate quantitatively how workers "perform" in terms of speed, accuracy and probability of detection in their visual field (ISO 8995, 1989). The optimum values of lighting parameters, from the human point of view, are determined on the basis of visual performance assessment. Thus effectiveness of the visual system is measured in terms of visual performance and it depends on visual and non-visual factors like visual capability of the individual, visibility of the task, and psychological and physiological variables (North 1993). Lighting engineering can influence only on the one group of factors, which are connected with the visibility of the task. The visibility of task depends on two groups of factors:
- Task attributes (size, distance, luminance, contrast, color, surface properties).
- Illumination (illuminance level, illuminance uniformity, luminance distribution, spectral content of light sources, glare).

Knowledge of the task attributes is essential that lighting designer should prepare an appropriate project of illumination from the point of view of visual performance.

3. LIGHTING CHARACTERISTICS

Interior lighting conditions are described by the following lighting parameters, which determine a satisfactory luminous environment:
- Illuminance level.
- Illuminance uniformity.
- Luminance distribution.
- Glare.
- Spectral characteristic of light.
- Flicker.

3.1. Illuminance Level

Illuminance and illuminance uniformity are determined and measured on the work plane. The work plane is a plane at which work is usually performed and it is usually a task surface (on a real height above the floor). When there is no information about the height of the task surface, a horizontal reference plane is assumed. Its height depends on particular national requirements. For example, in US standards it is 0.76 m above the floor; in Polish ones 0.85 m. For corridors, stairs, bathrooms, communication areas, etc. the floor is taken as work plane height.

Illuminance level, which is required for visual task performing depends on two parameters:
- Apparent target size (function of the smallest size of target and viewing distance).
- Degree of visual task difficulty (depends on task contrast and reflectance).

The degree of visual task difficulty increases when the task contrast and reflectance decreases. For very difficulty visual tasks and a small apparent target size the required illuminance level should be higher. In accordance with the Weber-Fechner Law and lighting practice (illuminance values increase by a factor of ~1.5, that is

the smallest increment needed to give a significant difference in the subjective lighting effect) the following scale of illuminances is recommended: 20–30–50–75–100–150–200–500–750–1000–1500–2000–3000–5000 lx.

In practice the required illuminance level for a given task (work plane) refers to the average illuminance on the task surface (work plane). Average illuminance on a given surface is calculated as an arithmetic average of at least three measured illuminances at evenly distributed points on that surface. This means that measured illuminances at each point of surface need not be equal or higher than required by standard illuminance values. Only average values of illuminance are compared with required values.

Most lighting standards for different areas, tasks or activities gives, as seen in Table 1, ranges of three illuminances. Higher values in the range can be used as follows:
- When unusually low reflectances and/or contrast are present.
- When visual performance is critical.
- When the visual capacities of the workers are below normal (e.g. most workers are 40 years old or more).
- When accuracy or higher productivity is important.

According to ISO (ISO 8995, 1989) the lower values in the range can be used as follows:
- When reflectance or contrasts are unusually high.
- When accuracy or higher productivity is not important.
- When the task is executed occasionally only.

Many working interiors are dim at illuminances lower than ~200 lx. That is why 200 lx is recommended as the minimum illuminance for workspace where work is executed for long periods.

3.2. Illuminance Uniformity

Variation of illuminance should be considered on two areas: on and around the visual task area and in the whole interior. Excessive changes of illuminance in the visual field can be distracting and cause changes in adaptation during the movement of the fixation point from a bright to a dark surface. Transient adaptation problems can reduce visual performance, so uniformity is important. Uniformity of illuminance is calculated as a ratio of minimum measured illuminance to the average illuminance of a given work plan (task surface). It should be pointed out that task area is not usually identical with the entire area of the workstation and during the measurement this should be taken into account.

Table 1. Illuminances ranges for typical areas, tasks or activities (from ISO 8995, 1989: 3)

Illuminance range, lx	Type of area, task or activity
20–30–50	Outdoor circulation and work areas
50–100–150	Circulation areas, simple orientation or short temporary visits
100–150–200	Rooms not used continuously for working purposes
200–300–500	Tasks with simple visual requirements
300–500–750	Tasks with medium visual requirements
500–750–1000	Tasks with demanding visual requirements
750–1000–1500	Tasks with difficult visual requirements
1000–1500–2000	Tasks with special visual requirements
above 2000	Performance of very exacting visual tasks

Recommended uniformity values differ in different countries. For example, British recommendations require the uniformity of illuminance over any task area and immediate surround be > 0.8 (that does not necessarily have to apply to the entire room). Polish recommendations are less restrictive and uniformity of illuminance on the work plane should be > 0.65 for prolonged work and 0.4 for casual work and communication areas.

In most work interiors there are different visual tasks with various degrees of difficulty, and some tasks require a higher level of illuminance. Often, especially when most tasks require a lower level of illuminance, it is waste to illuminate the whole interior to a higher level of illuminance. In those cases general lighting should ensure a lower level of illuminance and some workstation should be additionally lighted by localized or local lighting. However, in those cases it is important for the average illuminance difference between two adjacent interiors (corridor and room) and between different work planes in the same room to be < 1:5.

3.3. Luminance Distribution

The luminance distribution in the field of view controls the adaptation level of the eyes, which affects task visibility. The luminance distribution in the field of view also affects visual comfort. For a given lighting level, differences in luminance will be caused by differences in surface reflectance. It can happen that the illuminance is appropriate for the visual task but luminance balance in the whole interior is unacceptable. Luminance distribution can be described either by luminance ratios of adjacent surfaces and surfaces viewed in sequence or by reflectances of major interior surfaces. Too high luminance ratios (> 1:10) will cause fatigue and/or glare but too low luminance ratios result in a dull and non-stimulating working environment. To improve the visual performance of a task the luminance of the task surround should be lower than task luminance, but > 1:3.

3.4. Glare

The lighting system should provide a luminous environment that is free of the glare phenomenon. This phenomenon exists when parts of the visual field are excessively bright in relation to the brightness of the general surroundings to which eyes are adapted. From the psychophysiological point of view there are two kinds of glare: disability glare and discomfort glare. Glare is experienced when the excessive brightness of an object (light source) is in or close to the direction of the line of sight (direct glare) or when such light sources are reflected in glossy surfaces (reflected glare).

3.4.1. Disability glare

Disability glare impairs the vision of details or objects (visibility and visual performance are reduced) without necessarily causing discomfort. It occurs when a large source of low luminance or a small source of high luminance is seen close to the line of sight to the visual task. Disability glare is connected with a veiling luminance produced in the eye on the retina by scattered light in ocular media inhomogeneities. This luminance is superimposed over the well-focused image on the retina and, as a result, reduces its contrast that results in decreased visual performance and visibility. Veiling luminance, or in other words disability glare, depends on illuminance from the glare source at the eye and the angle between the primary object and the glare sources. It is largely independent of the luminance of the source.

3.4.2. Discomfort glare

Discomfort glare causes a feeling of discomfort without necessarily impairing the vision of details or objects. It is normally experienced as a feeling of discomfort, annoyance or pain, which increases with time, and it may contribute to fatigue. It has been found that discomfort glare depends on the luminances of the glare sources, the adaptation luminance level (background luminance), the number of glare sources, the source area and the angle between the primary object and the glare sources. But the degree of discomfort glare also depends on the kind of activity being performed. A more demanding visual task will result in a stronger feeling of discomfort. The limitation of discomfort glare is one of the major quality aspects of lighting because when discomfort glare is properly controlled disability glare does not exist. Discomfort glare must be controlled at the design level (stage) and in fact it is control of direct discomfort glare from lamps and luminaries by controlling the luminance of those in the direction of the observer's eyes. The degree of luminance control differs according to the type of task and activity. The CIE has classified task and activities into five groups (i.e. Quality Classes) according to the degree of the required luminance control. There are quality classes: A, B, C, D, E, where class A means the most restrictive glare limitation for very exacting visual tasks and class E means very low glare limitation for low visual tasks demands. There are various glare systems and different countries have adopted different design procedures to evaluate the degree of discomfort glare in a given situation.

Table 2. Lamp colour appearance group (from ISO 8995, 1989: 3)

Colour appearance groups	Colour appearance	Correlated colour temperature, K
1	Warm below	3 300
2	Intermediate	3 300 to 5 300
3	Cold	above 5 300

group 1: is appropriate for residential type areas

group 2: is most widely used in working interiors

group 3: should be used only for high lighting levels, for special tasks (like colour matching) or in warm climates

3.5. Color of Sources

The quality of the color of light emitted by lamps is characterized by two properties:
- color appearance
- color rendering capabilities.

Both properties of a light source are determined by the spectral composition of the light emitted. Different spectral compositions of light can result in a similar color appearance but they can produce different color rendering.

3.5.1. Color appearance

The color appearance of a lamp refers to the apparent color (chromaticity) of the light it emits. Its correlated color temperature in Kelvins (K) describes it. Lamps can be divided into three main groups according to their color temperature (Table 2).

The color appearance of an object depends on the spectral distribution of the light that illuminates it, the chromatic adaptation of the observer and the spectral reflection characteristics of the surface.

3.5.2. Color rendering

Color rendering describes the appearance of object colors under a given light source compared with their appearance under a reference source. A general color-rendering index R_a has been introduced for objective identification of color rendering properties. Its maximum of 100 corresponds with excellent color rendering (i.e. a test source gives exactly the same effects as the reference illuminant).

Color rendering of light sources depends on the spectral distribution of light. To simplify specifications of the color rendering indices, color rendering groups have been introduced (Table 3).

3.6. Flicker

The light emission of all AC lamps has a cyclic variation. It is small for filament lamps but more visible for fluorescent and discharge lamps. Flicker usually increases with the lamp's age (especially for fluorescent lamps) and can be avoided by replacing old lamps with new ones. Flicker from high-pressure lamps is more noticeable in lamps with transparent than in those with fluorescent coating on the bulb.

Rotating parts of rotating machinery which are lighted by

Table 3. Lamp color rendering groups (from ISO 8995, 1989: 3)

Colour rendering group	Colour rendering index range	Colour appearance	Example of use	
			Preferred	Acceptable
1A	$R_a \geq 90$	Warm Intermediate Cold	Colour matching, clinical examinations	
1B	$80 \leq Ra < 90$	Warm Intermediate	Offices, hospitals	
		Intermediate Cold	Printing, paint and textile industries, demanding industrial work	
2	$60 \leq Ra < 80$	Warm Intermediate Cold	Industrial work	Offices
3	$40 \leq Ra < 60$		Rough industries	Industrial work
4	$20 \leq Ra < 40$			Rough industries

fluroescent or discharge lamps can appear as stationary, or as reduced speed, or as in reverse rotation. This phenomenon is called the stroboscopic effect and it might be a potential risk factor at this type of workstation. It can be avoided by lighting the rotating part of machinery by incandescent lamps or by dividing the lamps between three phases.

Using lamps with high frequency supplies can more effectively reduce flicker and stroboscopic effects.

4. LIGHTING SYSTEMS

4.1. Classification According to Luminaries Layout

The visual field of a worker is different depending on whether worker is concentrating on a task or looking away for relaxation. For this reason a distinction between task lighting and the lighting of the environment should be made. According to the luminaries' layout in the interior, the following methods of electrical lighting can be distinguished:

- General lighting – lighting, which provides a uniform level of illumination on the work plane in the whole of an area without any provision for special local requirements;
- Localized lighting – lighting, which illuminates an interior and at the same time provide a higher illuminance over a particular part or parts of the interior (using luminaries above the visual task)
- Local lighting – additional lighting providing illuminance over a small area, for special task requirements, which does not contribute in general surrounding lighting and is controlled separately from general lighting

The method of lighting selection depends on the required illuminance level, room characteristics and work stands layout. General lighting provides the same required illuminance level at each workstation in the room. It is appropriate for small offices (small room, one workstation) or for rooms, where on all the workstation the required illuminance is at the same level and task lighting is inappropriate. Localized general lighting provides a required illuminance level at work stands (using task luminaries above the workstation) and a lower level of general illumination. It is desirable in open-plan arrangements. Local lighting provides a required level of illuminance only at that workstation, when additional task luminaire is mounted. Local lighting should be designed to use only for use with general lighting, because it illuminates only a small area of the workstation. The general lighting level should be at least 20% of the required illuminance level (i.e. 20% of total illuminance from general and local lighting). This method of lighting is appropriate in big interiors, where work stands differ considerably in illuminance requirements, and work stands are unstable.

4.2. Classification According to Light Distribution

- direct lighting – lighting by luminaries distributing 90–100% of the emitted light directly reaching the reference plane (i.e. surface to be illuminated) (Rea 1993). This term usually refers to light emitted downward.
- indirect lighting – lighting by luminaries distributing only 0–10% of the emitted light directly reaching the reference plane (i.e. surface to be illuminated). It means that 90–100% of emitted light is distributed upward.
- semidirect lighting – lighting by luminaries distributing 60–

90% of their emitted light downward and the other 40–10% upward.
- semi-indirect lighting – lighting by luminaries distributing 60–90% of their emitted light upward and the otehr 40–10% downward.
- general diffused lighting - lighting by luminaries distributing 40–60% of the emitted light downward and the other 60–40% upward.

A broad range of different possibilities of creating general lighting like direct lighting, indirect lighting and combinations of direct and indirect lighting gives lighting designers a lot of variants of lighting designs for specific interiors. Choosing one of those lighting systems has its consequences in the appearance of the luminous environment and in energy consumption. Direct lighting emphasizes horizontal planes like the work plane and the floor and also vertical surfaces (like walls) when the luminaries have wider light distribution or when they are placed near the wall. There is a wide range of direct-type luminaries with a variety of light distribution characteristics from very broad (using translucent diffusing shielding) to narrow concentrated (using specular reflectors and louvers). Special attention should be paid to glare limitation, especially in VDT rooms. This lighting system has the lowest energy consumption (appropriate illuminance level on the work plane at the lowest energy consumption).

Indirect lighting illuminates the ceiling. So the ceiling is the brightest surface in the visual field. The illuminationi of the ceiling should be evenly distributed to avoid too high values of luminance. Indirect lighting can provide diffuse lighting similar to the light of an overcast day without creating veiling reflections and shadows (incidence of light from many directions). This lighting system has the highest energy consumption (appropriate illuminance level on the work plane at higher energy consumption).

A combination of direct and indirect lighting can produce excellent results. The indirect portion should not create excessive luminances of the ceiling and the direct portion should not produce glare.

5. MAINTENANCE

Two main factors, that is a decrease of the lamp's light emission in time (lamp aging) and accumulation of dirt on lamps, luminaries and other surfaces in the interior, result in a progressively decreasing illuminance level in time. Good maintenance of the lighting system means periodical cleaning of luminaries, lamps replaced before the end of their lifetime and periodical cleaning or painting of the walls and ceiling. The lighting designer should plan appropriate intervals of clearning, with the frequency depending on the type of luminaire, the rate at which dirt accumulates in a given interior and the cost of cleaning. There are two kinds of lamp replacement, called individual and group replacement. Individual replacement (replacement of all lamps in a given interior) is preferred in large interiors and in busy areas (i.e. corridors).

REFERENCES

ISO 8995: 1989, *Principles of Visual Ergonomics – The Lighting of Indoor Work Systems*.

NORTH, R.V., 1993, *Work and the Eye* (New York: Oxford University Press).

prEN 12464: 1996, *Lighting Applications – Lighting of Work Places*.

REA M.S., 1993, *Lighting Handbook Reference & Application* (8th ed) (New York: Illuminating Engineering Society of North America).

Illumination: Basic Definition

P. R. Boyce

Lighting Research Center, Rensselaer Polytechnic Institute, Troy, NY 12189, USA

1. PHOTOMETRIC QUANTITIES

Illumination is the use of light to make an object visible. Light is electromagnetic radiation in the wavelength range 380 nm to 760 nm. Electromagnetic radiation within these wavelength limits produces a response from the human visual system. Electromagnetic radiation outside these wavelength limits does not.

The total electromagnetic radiation emitted by a source is quantified as the radiant flux. Radiant flux is the rate of flow of energy and is measured in watts. The total amount of light emitted by a light source in all directions is quantified as the luminous flux. Luminous flux is radiant flux multiplied by the relative spectral sensitivity of the human visual system and is measured in lumens. The relative spectral sensitivity conventionally used is the Standard Photopic Observer of the Commission Internationale de l'Eclairage (CIE).

While luminous flux is a useful term for describing the light output of a light source, it does not describe the distribution of light from a luminaire. To describe such a distribution a measure of the amount of light emitted in a specific direction is necessary. Luminous intensity is used for this purpose. Luminous intensity is the luminous flux emitted / unit solid angle, in a specified direction, and is measured in candelas.

Both luminous flux and luminous intensity have density measures associated with them. The luminous flux falling on unit area of a surface is called the illuminance and is measured in lumens / meter^2 or lux. The luminous intensity emitted per unit projected area in a given direction is the luminance and is measured in candelas / meter^2 . Table 1 summarizes these photometric quantities.

Table 1. The photometric quantities

Measure	Definition	Units
Luminous flux	That quantity of radiant flux which expresses its capacity to produce visual sensation	Lumens (1m)
Luminous intensity	The luminous flux emitted in a very narrow cone containing the given direction divided by the solid angle of the cone, i.e. luminous flux/unit solid angle	Candela (cd)
Illuminance	The luminous flux/unit area at a point on a surface	Lumens/meter^2
Luminance	The luminous flux emitted in a given direction divided by the product of the projected area of the source element perpendicular to the direction and the solid angle containing that direction, i.e. luminous flux/unit solid angle/unit area	Candela/meter^2
Reflectance	The ratio of the luminous flux reflected from a surface to the luminous flux incident on it	

For a matte surface: Luminance = (Illuminance x Reflectance) Π

2. COLORIMETRIC QUANTITIES

The photometric quantities do not differentiate between light made up of different wavelength combinations, i.e. light with different spectral distributions. The spectral distribution of the light reaching the eye is largely but not entirely responsible for what color is perceived. Two three-dimensional color spaces, L_{ab} and L_{uv}, have been developed by the CIE as maps of color. L_{ab} is generally used for object colors and L_{uv} for self-luminous colors. The position of an object in these color spaces is calculated from the spectral distribution of the light reflected from or emitted by the object. The coordinates of an object in color space indicate what color it will be perceived to be. Objects that have the same coordinates in color space will appear to be identical in color, even if they have different spectral distributions. Objects that have different coordinates in color space will be perceived as different in color.

Although the L_{ab} and L_{uv} color spaces are the most sophisticated maps of color, other simpler color systems are used by the lighting industry. The CIE 1964 uniform color space is used in the calculation of the CIE General Color Rendering Index, an index taken to classify the color rendering capabilities of light sources. The positions in the color space of eight test colors under a reference light source and under the light source of interest are calculated. The differences between the positions of each test color under the two sources are obtained, adjusted for chromatic adaptation, averaged and modified so that a maximum value of 100 is obtained when there are no differences.

The CIE General Color Rendering Index is a rather course measure of the ability of different light sources to render colors. Different light sources of interest have different reference light sources. Further, because of the averaging, two light sources that render colors differently can have the same CIE General Color Rendering Index. This means that the CIE General Color Rendering Index should only be used for separating light sources into broad classes of color rendering capabilities.

A two-dimensional color surface, the CIE 1931 chromaticity diagram, is used to characterize the color appearance of light sources and to define the acceptable color characteristics of light signals (figure 1). The boundary of the surface represents colors created from single wavelengths. The point at the center, called the equal energy point is where a spectrally neutral surface will be located. Any other color will be located somewhere between the equal energy point and the edge of the diagram. The closer the coordinates of the color are to the edge of the diagram, the greater the strength of the color. Figure 1 also contains two other pieces of information; the areas in which the chromaticity coordinates of signal lights need to fall if the signal is to be perceived as the specified color, and the full-radiator locus. This latter represents colors generated by a black-body radiator. The color appearance of a nominally-white light source is conventionally described by its correlated color temperature. This is the temperature of the black-body radiator which is closest to the coordinates of the light source on the CIE 1931 chromaticity diagram.

3. THE GENERATION OF LIGHT

Light is generated naturally, in the form of daylight, and artificially, by converting various forms of energy into electricity and then into light. Natural light is primarily characterized by its variability.
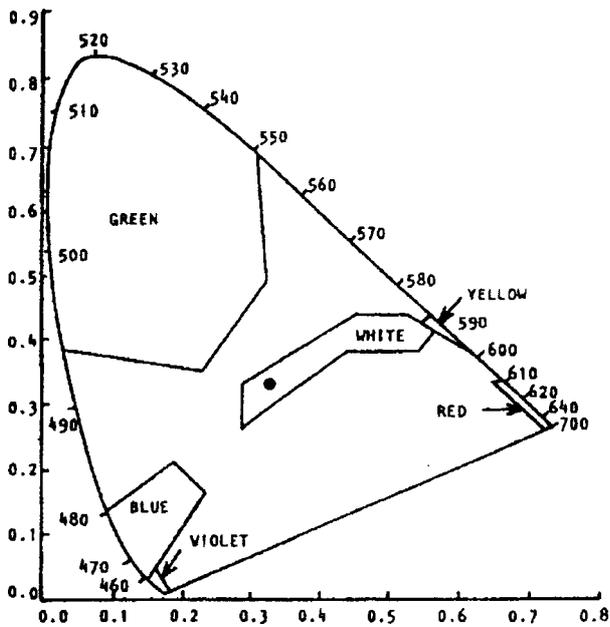
Figure 1. The CIE 1931 Chromaticity Diagram. The filled circle is the equal energy point. The boundary curve is the spectrum locus with the individual wavelengths marked. The enclosed areas indicate the chromaticity coordinates of light signals that will be identified as the specified colors.

Depending on the meteorological conditions, time of year, time of day, and latitude, daylight will vary in amount, spatial distribution and spectrum. Daylight has two components, sunlight and skylight. Sunlight is light received directly from the sun. Skylight is light from the sun scattered in the atmosphere. The nature of the atmosphere and the distance that the light passes through it determine the balance between sunlight and skylight. The greater is the path length and the more scattering centers there are in the atmosphere, the higher is the proportion of skylight.

Electric light sources used for illumination can be divided into two classes; incandescent lamps and discharge lamps. Incandescent lamps produce light by emission from a hot wire. Discharge lamps produce light by emission from an electric discharge in a gas. Incandescent lamps do not require any other equipment to operate. Discharge lamps require additional equipment, called control gear, to operate because the electrical

conditions needed to initiate a discharge are different from those needed to sustain it.

Table 2 summarizes the performance characteristics of two incandescent lamp types and four discharge lamp types and gives the most common applications for each lamp type. The values in the table 2 should be treated as indicative only. Details about the characteristics of any specific lamp should be obtained from the manufacturer.

4. CONTROL OF LIGHT OUTPUT

To produce illumination it is necessary to both generate light and control its distribution. For daylight illumination, the distribution of light is determined by the size, shape, orientation, placement and shielding of windows and skylights. For electric illumination, light distribution is determined by the luminaire the light source is operated in and the arrangement of the luminaires in the space. A luminaire provides mechanical support for the light source, connects the light source to the electricity supply, houses any control gear necessary, and controls the light distribution by some combination of diffusing, reflecting or refracting elements.

As for the control of light output, daylight through a window can be limited by blinds. Wherever the sun, or a very bright sky, is likely to be directly visible through a window, blinds are desirable to avoid glare. The most acceptable form of blinds are those that shield the viewer from a direct view of the sun or bright sky while preserving some view out.

Switching or dimming is used to control the light output of electric light sources. Time switches are useful where the space is occupied on regular schedule. Occupancy sensors can be used to avoid waste of electricity by switching off lighting when there is nobody in the space. However, if such sensors are insufficiently sensitive they may be disliked because they switch lighting off when people are present. In addition, frequent switching may shorten lamp life. Manual switching is the most common form of switching. Manual switches are more likely to be used if the control panel is labeled so that the operator knows which luminaires are being switched.

Dimming systems exist for incandescent light sources and for tubular and compact fluorescent light sources. The factors to consider when assessing a dimming system are the range of light output over which dimming can be achieved; the extent to which the color properties of the lamp may change; and whether lamp

Table 2. Summary of characteristics of some common electric light sources

Light Source	Luminous Efficacy (1m/W)	Correlated Color Temperature (K)	CIE General Color Rendering Index	Lamp Life (hrs)	Applications
Tungsten	8–19	2700	100	750–2000	Domestic Retail
Tungsten-Halogen	8–20	2900	100	2000–4000	Display
Tubular Fluorescent	60–110	3000–5000	50–95	9000–20,000	Commercial
Compact Fluorescent	50–70	2700–4100	80–85	9000–20,000	Commercial Retail
Mercury Vapor	30–60	3200–7000	15–50	16,000–24,000	Agricultural Industrial
Metal Halide	50–110	3000–6500	65–95	3000–20,000	Industrial Commercial Exterior
High Pressure Sodium	60–140	2100–2500	20–70	10,000–24,000	Industrial Exterior
Low Pressure Sodium	100–180	1800	n.a	16,000–18,000	Exterior

life is altered. A common technique for dramatically increasing the life of an incandescent lamp is to operate it slightly below its rated voltage.

5. THE HUMAN VISUAL SYSTEM

The human visual system involves both eye and brain working together. An image of the world is formed on the retina of the eye. At the retina, the photons of light forming the image are absorbed and converted to electrical signals. Some image processing occurs at the retina, and at various stages up the optic nerves and in the visual cortex. The processed images produce perceptions based partly on past experience.

There are two types of photoreceptors in the retina, called rods and cones. Rods and cones differ in their sensitivity to different wavelengths and in their absolute sensitivities to light. Rods show a peak sensitivity to light at 507 nm while cones show a peak sensitivity at 555 nm. Rods are approximately 100 times more sensitive than cones. In simple terms, rods provide night vision while cones provide day vision. Further, rods and cones are distributed differently across the retina. Although there are cones spread across all of the retina, they are more dense in a small central area of the retina called the fovea. Rods are also spread throughout the retina apart from the fovea, where there are none. Rods reach their maximum concentration about 20∞ from the fovea. This difference between photoreceptor types in the fovea and the peripheral retina is magnified by the connections between photoreceptors and optic nerve fibers. The ratio of photoreceptors to optic nerve fibers is close to one in the fovea but increases rapidly as the deviation from the fovea increases. This pattern is reflected in the way in which the parts of the retina are used. Basically, the peripheral visual field is used to detect changes in the visual environment which are then examined by turning the head and eyes so that the location where the change occurred falls on the fovea. The fovea is the part of the retina where the finest visual acuity and contrast sensitivity occur. It is what we use when we look directly at something.

The human visual system can function to some degree from starlight to bright sunlight, a range of luminance of about 12 log units. But it cannot operate over the whole of this range simultaneously. To adjust the sensitivity of the visual system to different prevailing luminances, a process of adaptation occurs. Adaptation to a sudden change of about 2 to 3 log units of luminance is fast, a few seconds at most, but adaptation over a larger range can take much longer, of the order of minutes. While adaptation is in progress, visual capabilities are reduced. Even when completely adapted, the capabilities of the visual system depend on the luminance to which it is adapted. Three distinct ranges of luminance can be distinguished; the photopic, mesopic and scotopic. Table 3 summarizes these ranges and the associated visual capabilities.

Interior lighting and decor usually produce luminances which ensure the visual system is in the photopic state. Exterior lighting on roads and urban areas usually provides luminances high enough to ensure the visual system is operating around the photopic–mesopic boundary. It is only in areas completely without lighting that the scotopic state is reached.

Table 3. Summary of the operating states of the visual system

Operating State	Luminance Range Photoreceptor	Active Wavelength	Peak Sensitivity	Characteristics
Photopic	>3cd/m²	Cones	555 nm	Fine resolution Good color vision
Scotopic	<0.001cd/m²	Rods	507 nm	No color vision Poor resolution Fovea "blind"
Mesopic	>0.001 and <3cd/m²	Cones and Rods	Between 555 and 507 nm	Reduced color Reduced resolution relative to photopic

6. TASK PERFORMANCE AND VISUAL PERFORMANCE

Illumination is often provided to facilitate the performance of tasks. Visual tasks usually have three components; visual, cognitive and motor. The visual component uses the sense of sight to collect information from the visual environment. The cognitive component is carried out in the brain and ascribes meaning to the visual information collected. The motor component is the physical response to cognitive output. Of course, these three components interact to produce a complex pattern between stimulus and response. Further, every task is unique in its balance between visual, cognitive and motor components and hence in the effect illumination has on task performance. This makes it impossible to derive a single relationship between illumination and task performance that can be generalized to all tasks.

However, it is possible to establish a general relationship between illumination and visual performance. Figure 2 shows the relative visual performance of a task requiring the detection of a stimulus in a known position, for four different size stimuli, for a range of luminance contrasts and retinal illuminances. The form of the relationships shown in Figure 2 for each target size have been called the plateau and escarpment of visual performance. What this means is that visual performance changes only a little over a wide range of visual sizes, luminance contrasts and retinal illuminations (the plateau). However, if visual size or luminance contrast or retinal illumination is reduced sufficiently, visual performance deteriorates rapidly (the escarpment). The existence of a plateau of visual performance implies that for a wide range of task sizes and contrasts, visual performance is relatively insensitive to illumination but when conditions reach the edge of the plateau, performance declines rapidly. This pattern of performance has been found for many real tasks with a large visual component.

The relationships shown in Figure 2 are derived from tasks which are viewed directly, i.e. which are imaged on the fovea. However, there are a whole class of tasks that involve visual search because the location of the object to be found is not known. A visual search is made using a series of eye fixations, until the object to be found is detected and identified. Usually, the object, or something like it, is detected in the peripheral visual field and then identified after the fovea has been brought to bear on the object. The speed with which a visual search can be carried out is dependent on both the visibility and conspicuity of the object
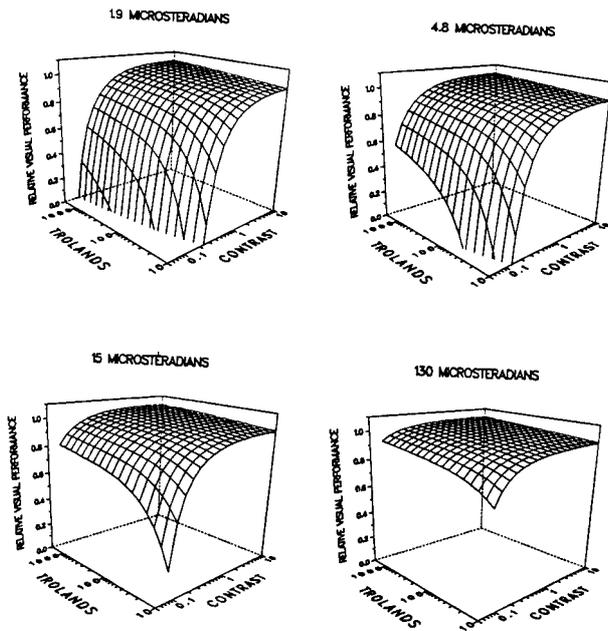
Figure 2. Four relative visual performance surfaces for stimuli of four different sizes, measured in microsteradians. Each surface shows relative visual performance plotted against retinal illumination in trolands and luminance contrast. Retinal illumination in trolands is calculated as the product of the luminance of the surface being viewed and the associated pupil size. (After Rea, M.S., and Ouellette, M.J., Relative visual performance: A basis for application, Lighting Research and Technology, 23, 135, 1991)

being sought. The visibility of the object will be determined by its size, luminance contrast and color difference from the background. The conspicuity of the object is the extent to which it is different from all the other objects in the area being searched. It is possible to have an object which is highly visible but difficult to find because it is not conspicuous.

The level of visual performance that is achievable with a given task depends on the stimulus the task presents to the visual system and the operating state of the visual system. There are five characteristics of a task and its lighting which determine these conditions. These are its visual size, luminance contrast, color difference, retinal image quality and retinal illumination.

6.1. Visual Size
A large stimulus is the easier to detect than a small one. The size of a stimulus is always expressed in some form of angular measure. The visual size of a stimulus for detection is often expressed as the solid angle the stimulus subtends at the eye. For resolution of detail, the visual size is usually given as the angle the critical dimension of the stimulus subtends at the eye. For complex stimuli, the spatial frequency distribution is used to characterize the stimulus. Lighting can do little to change the visual size of two-dimensional objects but shadows can be used to enhance the visual size of some three dimensional objects.

6.2. Luminance Contrast
A stimulus with a high luminance contrast is easier to detect than a low luminance contrast stimulus. Luminance contrast is expressed in a number of ways. For stimuli which are seen against a uniform background, e.g. print, luminance contrast is defined as:

$$C = (L_t - L_b) / L_b$$

where
C = luminance contrast
L_b = Luminance of the background
L_t = Luminance of the detail

For stimuli which have a periodic pattern, e.g., a grating, the luminance contrast or modulation is given by:

$$C = (L_{max} - L_{min}) / (L_{max} + L_{min})$$

where

C = Luminance contrast
L_{max} = maximum luminance
L_{min} = minimum luminance

It should be noted that these definitions of luminance contrast have different ranges of possible values. The luminance contrast of a stimulus can be modified by lighting either because the distribution of light leads to veiling reflections occurring on specular surfaces, or because light is scattered inside the eye over the retinal image, or, where color stimuli are involved, because different light sources with different color rendering properties are used

6.3. Chromatic Difference
A stimulus can be discriminated from its background because it is different in color. Such color differences can be used for detection even when there is zero luminance contrast. Chromatic difference can be expressed in terms of positions on an equal-luminance plane of a color space. Light sources with different spectral emissions can alter chromatic differences.

6.4. Retinal Image Quality
The quality of the retinal image is influenced by the optical quality of the object being imaged, the amount of light scatter that occurs in the eye and the ability of the eye's optical system to focus the image on the retina. The choice of light source can influence pupil size and hence depth of field but having the correct optical refraction in the form of glasses or contact lenses is usually a better approach.

6.5. Retinal Illumination
The state of adaptation of the visual system is determined by the retinal illumination. The retinal illumination in turn is determined by the luminances in the visual field, the pupil size and the absorption of light in the eye.

By altering any or all of these characteristics, the achievable level of visual performance can be moved to the plateau. How effective making any of these changes will be depends on the initial condition. The closer the task is to threshold, the bigger

the effect of changing any of the characteristics will be. Above threshold, the most effective characteristic to change is the one which is limiting performance. For example, increasing the luminance contrast of a stimulus which is very small in size will help but increasing the visual size would be more effective. Similarly, enhancing color difference will have little effect if the task has a high luminance contrast unless the color can be used to distinguish the stimulus from other non-stimulus objects. Often it is not possible to change the characteristics of the task. In this situation, only the lighting can be changed but it is always desirable to match the lighting to the physical characteristics of the task. For example, if the task is two-dimensional and of matte reflectance located on a matte background, increasing the retinal illumination is about the only option. However, if the object is three-dimensional and is a specular reflector, the appropriate light distribution can increase the visual size by casting shadows, and increase the luminance contrast by producing highlights. If the object is distinguished from its background primarily by color, the light spectrum used is an important consideration

7. VISUAL COMFORT

Designing good illumination almost always involves a consideration of both visual performance and visual comfort. Illumination that is inadequate for the performance of a task will lead to visual discomfort, but visual discomfort can also occur when the lighting of the task is adequate. This is because visual discomfort can be produced from anywhere in the lit space but visual performance is usually only affected by the lighting in the immediate area of the task.

Red, sore, watering eyes; headaches and migraines, gastro-intestinal problems and muscular aches and pains can all be due to inappropriate illumination, although all can have other causes. There are several different aspects of illumination that can cause discomfort. They are:

- Flicker — Whether a regular, visible fluctuation in luminous flux will be considered uncomfortable depends on the context. In the entertainment industry flicker is often associated with an attempt to generate excitement. However, in functional spaces, flicker will almost certainly be strongly deprecated. Whether a fluctuation of luminous flux will be visible as flicker will depend on its frequency and modulation. The occurrence of flicker can be made less likely by using high-frequency control gear for discharge lamps and/or by the mixing of light from lamps powered from different phases of the electricity supply.
- Glare — Glare can be both absolute and relative. It is possible to have absolutely too much light. Too much light is rare indoors, although it can happen where there are many high-reflectance surfaces, but is common in full sunlight. Glare can also occur when the range of luminance in a visual environment is too large. Conventionally, this form of glare is divided into two types, disability glare and discomfort glare. Disability glare reduces visual performance because light scattered in the eye reduces the luminance contrast of the retinal image. Disability glare is most commonly experienced from oncoming headlights when driving at night. Most forms of disability glare also cause discomfort but it is possible to have disability glare without discomfort when the area of the source causing the disability is large.

- Discomfort glare causes discomfort without any effect on visual performance. Discomfort glare increases with increasing luminance and solid angle of the glare source increase and decreases with increasing luminance of the background and deviation from the glare source. Most luminaires are designed to limit the discomfort glare likely to occur for common directions of view.
- Shadows — Shadows can cause discomfort when they make task performance more difficult. How and if they make task performance more difficult depends on their size and location. If a large area is in shadow, the effect is the same as if the illumination has been reduced. If the reduction is sufficient to move visual performance off the plateau of visual performance, discomfort is likely to be experienced. If shadows are cast over and within parts of an object and those shadows hide important details, then again visual performance is likely to be reduced and visual discomfort experienced. However, shadows can also be valuable in revealing the form of three-dimensional objects. Shadows can be reduced by increasing the amount of interreflected light in the space or by providing local lighting which can be adjusted in position.
- Veiling reflections — When a source of high luminance, such as a luminaire, is reflected from a specular surface, such as a computer monitor, the luminance contrast of the display will be reduced. This reduction is said to be due to veiling reflections. The magnitude of veiling reflections are determined by the specularity of the material being viewed and the geometry between the observer, the object and the high-luminance source. Veiling reflections only occur when the object has specularly reflecting surfaces and when the geometry conforms to the laws of specular reflection. Veiling reflections can also be considered of benefit in some situations, e.g. when displaying silver plate, but in this situation the reflections are conventionally called highlights although physically they are the same phenomenon as veiling reflections.

8. LIGHT AND AGING

The human visual system deteriorates with age. Specifically, by forty the ability to focus at close distances is reduced; by sixty the amount of light reaching the retina is markedly reduced and the amount of light scattered in the eye is markedly increased; and by eighty the color of the light is distorted by increased absorption of the short wavelength light and a decline into partial sight is increasing likely. The ages given at which these changes occur are at best approximate but the trend is consistent. These changes in the optical characteristics of the eye with age diminish the visual capabilities of older people. Usually visual acuity, contrast sensitivity and color discrimination capabilities are reduced, the time taken to adapt to large and sudden changes in luminance is increased, and glare is perceived more frequently.

The detrimental effects on visual capabilities can be partially offset by the careful use of light. More light will help provided it is delivered without glare, shadows and veiling reflections. Adjacent areas should be lit to similar illuminances, or if this is not possible, there should be a transition zone lit to an intermediate illuminance. It should also be noted that the judicious use of decor to attach high luminance contrast to salient

detail in the visual environment can be make a space much easier for people with poor visual capabilities to understand.

9. TISSUE DAMAGE

Like all forms of electromagnetic radiation, exposure to light in sufficient quantities can cause tissue damage to both the eye and the skin. Tissue damage can be caused by either photochemical and thermal effects. Different types of damage to the eye and skin have different wavelength sensitivities The probability of different types of tissue damage occurring usually increases as the product of the spectral irradiance and the duration of exposure increases.

Tissue damage to the eye is minimized in many situations by the usual behavioral response to bright light exposure, that being to blink and look away. This involuntary response is only effective for light sources which appear bright. Sources which produce large amounts of ultraviolet or infrared radiation without much visible radiation are dangerous because they do not appear bright. Tissue damage to the skin takes one of two forms, photochemical damage caused by ultraviolet radiation and burns caused by visible and infrared radiation.

The sun is probably the light source with the greatest potential for tissue damage because it produces such large amounts of ultraviolet, visible and infrared radiation. Exposure to the sun commonly leads to skin reddening and sunburn and, if frequently indulged in, is associated with skin aging and skin cancer. Voluntary staring at the sun can also lead to retinal burns. Exposure to the sun clearly needs care but so does exposure to some powerful electric light sources. Such light sources are typically used in such applications as floodlighting large sports stadia while others are a source of optical radiation for industrial processes.

The hazard represented by a light source in a given situation can be evaluated by applying the recommendations given in publication RP 27 of the Illuminating Engineering Society of North America "Recommended Practice for Photobiological Safety for Lamps and Lamp Systems". Application of these standards to various commonly used electric light sources have shown that electric light sources used for conventional interior lighting do not usually represent a hazard.

REFERENCES

ILLUMINATING ENGINEERING SOCIETY OF NORTH AMERICA, 2000, *Lighting Handbook, 9th Edition* (IESNA, New York)

INTERNATIONAL COMMISSION ON NON-IONIZING RADIATION PROTECTION, *Measurements of Optical Radiation Hazards*, ICNRIP, 6/98, ICNRIP, Oberschleibheimm Germany, 1998.

REA, M.S. and OUELLETTE, M.J., 1991, Relative visual performance: a basis for application, *Lighting Research and Technology*, 23, 135-144.

WYSZECKI, G., and STILES, W.S., 1982, *Color Science: concepts and Methods, Quantitative Data and Formulae* (John Wiley and Sons, New York)

Lighting Equipment

A. Wolska

Safety Engineering Department, Central Institute for Labor Protection, Warsaw, Poland

1. LAMPS

There are two main lamp categories: incandescent and discharge lamps. Within each category there are some lamp groups with a range of lamps that differ in construction, wattage, luminous efficacy, color properties, etc.

1.1. Incandescent Lamps

The incandescent lamp group consists of two main subgroups: normal incandescent (or tungsten) lamps and tungsten halogen (or halogen) lamps.

1.1.1 The tungsten incandescent lamp

This is the oldest electric light source that produces light by electrical heating of a tungsten filament to high temperature (incandescence point). Visible radiation is emitted. The filament is sealed in a glass bulb (clear, opalizing, frosting, coloring or mirror coating), which is usually filled with an inert gas like argon or krypton. The most common types are known as general lighting service (GLS), decorative (e.g. candle lamps) and reflector (e.g. PAR, crown-silvered lamps). A normal incandescent lamp has a low color temperature of about 2700 K and emitted light is perceived by the eye as warm. It is almost yellow. The advantages of a filament lamp are:

- low initial costs,
- immediate full light output,
- continuous spectral emission,
- excellent color rendering (R_a = 100),
- simple operation,
- ease of dimming.
  The disadvantages are:
- low luminous efficacy (8 $\prod$ 21 lmW^{-1})
- short life (about 1000 hours)
- life sensitive to voltage fluctuation and vibration

Filament lamps are mainly used for domestic and display lighting.

1.1.2 Tungsten halogen lamps

like normal tungsten lamps, produce light by an incandescent filament but halogen (i.e. iodine, chlorine, bromine) is added to the normal gas filling. The halogen regenerative cycle prevents evaporated tungsten from blackening the bulb. Halogen lamps have an increased efficacy (about 10% higher than normal tungsten lamps) and extended life compared with normal tungsten lamps (about 2000 hours). These lamps almost perfectly maintain light output through life. Tungsten halogen lamps for normal lighting purposes have a color temperature of between 2800 and 3400 K and the emitted light is whiter (compared with normal tungsten lamps) and has correspondingly cooler color appearance. This type of lamp has the same advantages and disadvantages as normal tungsten lamps. There are two main types of halogen lamps: mains-voltage types (i.e. double-ended, single-ended, double envelope, and reflector lamps) and low-

voltage types (reflector and capsules lamps). Halogen lamps are used either for domestic and display lighting, or for projectors and vehicle headlamps.

1.2. Discharge Lamps

Discharge lamps can be divided into two main groups: low pressure (fluorescent lamps and low pressure sodium lamps) and HID — high pressure (high-pressure sodium, high-pressure mercury, and high pressure metal halide lamps).

1.2.1 Fluorescent lamps

These are low-pressure mercury discharge lamps in which light is produced predominantly by fluorescent powders (phosphor) activated by the ultraviolet energy of the discharge. The spectral light distribution depends on the mix of phosphors. A wide range of color appearance (depending on correlated color temperature) and color rendering group is available. The main advantages are:

- high efficiency (up to 104 lmW^{-1})
- long life (up to 12 000 hours)
- wide range of correlated color temperatures (2700 $\prod$ 6500 K, i.e. from warm-white to cool-daylight)
- wide range of color rendering index (R_a = 50 $\prod$ 98)
- relatively low cost.
  The disadvantages of fluorescent lamps are:
- decrease of luminous flux during life time (after 8000 hours it will be about 70% of the initial value)
- blackening of the tube wall (especially at its ends)
- efficacy sensitive to ambient temperatures (especially in exterior applications, when lamp wall temperature is below optimum value)
  There are a wide range of fluorescent lamps:
- tubular fluorescent lamps (standard, retrofit, high-frequency, miniature, special, rapid start)
- compact fluorescent lamps (with and without outer envelope)

Fluorescent lamps are used either for interior lighting (offices, schools, hospitals, factories, etc.) or for exterior lighting.

1.2.2 Low-pressure sodium lamps

In these visible radiation is produced by the sodium discharge directly, but emitted light is only in the yellow part of the spectrum, close to the maximum sensitivity of the human eye. Efficacy is the highest of all lamp types (up to 200 lmW^{-1}) but the monochromatic light results in no color rendering. The average life is up to 6 000 hours.

Low pressure sodium lamps are used where color rendering is of minor importance and mainly contrast recognition counts, e.g. roads and security lighting, harbors, marshalling yards.

1.2.3 High-pressure sodium lamps

These radiate energy across a good part of the visible spectrum. This increases the color temperature (1950—2500 K) and improves color rendering (up to R_a = 85). Luminous efficacy is high (up to 140 lmW^{-1}) and life is up to 12 000 hours).

This type of lamp is used for road lighting especially in city centers and industrial interior lighting.

1.2.4 High-pressure mercury lamps

These emit part of the radiation from the discharge in the visible region and part in the ultraviolet region. Fluorescent powder

Table 1. Lamp parameters – performance comparison.

Lamp	Power	Luminous flux	Luminous efficacy	Color temperature	Color rendering index	Life
	[W]	[lm]	[lmW^{-1}]	[K]	[R$_a$]	[hours]
Tungsten incandescent	15 ÷ 1000	90 ÷ 18000	8 ÷ 21	2700	100	1 000
Tungsten halogen	5 ÷ 2000	60 ÷ 48400	12 ÷ 33	2800 ÷ 3400	100	2 000
Tubular fluorescent	4 ÷ 140	120 ÷ 8350	25 ÷ 104	2700 ÷ 6500	50 ÷ 98	12000
Compact flourescent	5 ÷ 25	200 ÷ 1500	40 ÷ 65	2700 ÷ 6000	85 ÷ 96	10000
High pressure mercury	50 ÷ 1000	1800 ÷ 58000	36 ÷ 60	3400 ÷ 6000	33 ÷ 52	15000
Metal halide	70 ÷ 2000	2400 ÷ 20000	68 ÷ 95	3000 ÷ 5600	65 ÷ 92	6000
High-pressure sodium	50 ÷ 1000	1300 ÷ 130000	37 ÷ 140	1950 ÷ 2500	23 ÷ 85	12000
Low-pressure sodium	18 ÷ 180	1800 ÷ 32000	100 ÷ 200	---	none	6000

coating of the inner surface of the outer bulb converts the ultraviolet radiation into visible radiation. Luminous efficacy is medium (up to 60 lmW^{-1}), life is up to 15 000 hours and color temperature is in the range of 3400 to 6000 K. The color rendering index is up to 52.

Mercury lamps are used for road lighting, industrial lighting and for some commercial applications.

1.2.5 Metal halide lamps

These are high pressure mercury lamps with other metallic elements, introduced in the form of halides, which operate in the arc tube. The mix of elements gives a wide range of efficacy (up to 95 lmW^{-1}), color appearance (from 3000 to 5600 K) and color rendering (R$_a$ = 65 – 92).

Metal halide lamps are used mainly in commercial interiors and industry, as well as for color TV lighting in studios.

The characteristic parameters of above mentioned lamps are listed in table 1.

2. LUMINAIRES

A luminaire (light fixture or lighting fitting) is an apparatus which controls the distribution of light given by a lamp or lamps and which includes all the components necessary for fixing and protecting the lamps and for connecting them to the supply circuit (Yarham 1994: 2).

Luminaires can take many different forms depending on the type and number of lamps, application form and place, and efficiency or aesthetic priorities.

2.1. Optical Control

The following are the main methods of optical control that can be used in luminaires: reflection, refraction, diffusion, polarization, and interference. Light emitted by a lamp is directed to the work plane by different elements of the optical system like reflector, louver, refractors, diffuser, and shades. The design of the optical system affects the luminaire efficiency (light output ratio), light distribution, degree of brightness control (glare index), appearance, and utilization factor.

2.1.1 Reflection

This process is when the incident light leaves a surface from the incident side without change in frequency. Reflection may be specular (polished surfaces), spread (rough surface), diffuse (matt surface), mixed, selective and nonselective depending on the kind of surface. This technique of light control is very efficient and light can be distributed very precisely thanks to either different shapes of reflectors (circular, parabolic, elliptical), shapes and dimensions of the cells in a grid of louvers or different finish of surface (polished, matt, rough). Reflectors are used in all kinds of projectors, spotlights and floodlights and very often in office luminaires together with another optical element like a louver. Luminaires with parabolic polished louvers can precisely control light output and brightness in all directions. They have a high efficiency, good glare control, and can be used in VDT offices.

2.1.2 Refraction

This process is when the direction of light changes during its transmission from one medium to another in which its speed is different. This technique usually uses prisms (reflecting or refracting) and lenses (cylindrical, prismatic, and stepped). Those kinds of light directors are made of glass or plastics. Luminaires with refractors made of large numbers of small prisms (named prismatic controllers) are suitable for general office lighting because of their good glare control and efficiency. However, they are not suitable for VDT rooms, because they do not reduce glare sufficiently to prevent reflections in VDT screen.

2.1.3 Diffusion

This process occurs when incident light is redirected by uniform scattering in all directions (according Lambert's cosine law), primarily by the process of diffuse transmission or diffuse reflection. Luminance of the surface is the same in all directions. A diffuser scatters the light emitted by a lamp before it leaves the luminaire. For example, a luminaire with an opal diffuser attachment does not significantly change the light distribution of an installed lamp but reduces its overall brightness. It is not recommended for open-plan offices because the average luminance of diffusers is still rather high and constant in all directions. In small rooms it can be acceptable if the luminaire is not visible at viewing angles during normal work positions.

2.1.4 Polarization

This process takes place when unpolarized light is oriented in a defined direction. Polarized light is mostly obtained by transmitting light through multiple refractive layers. Polarized light can reduce veiling reflections and reflected glare, so that it can be used in VDT rooms.
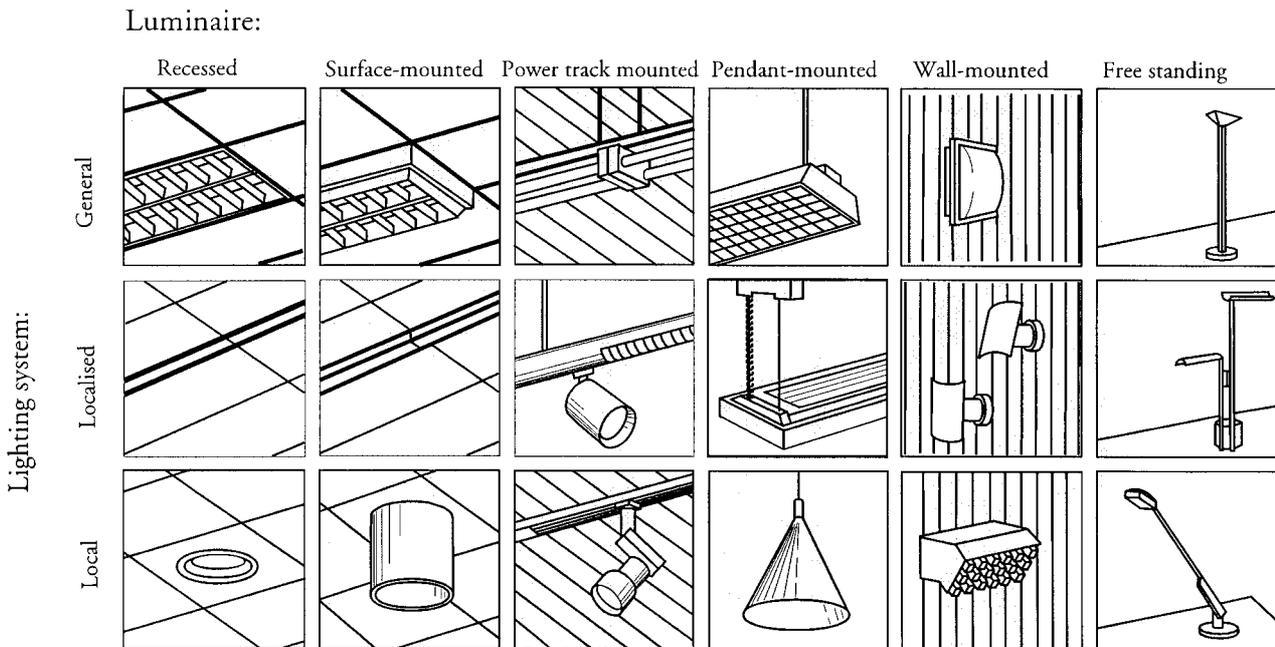
Luminaire:

Figure 1. Luminaire classification according to mounting position and lighting system (Dybczyñski, 1998: 3).

2.1.5 Interference phenomenon

This is used in reflector coatings to increase the reflectance and in luminaire glass or plastic attachments coatings to increase their transmittance.

2.2. Luminaire Characteristics

The main luminaires characteristics can be listed as follow:

mounting position - luminaires can be mounted in several ways: fixed on the ceiling (surface mounted), recessed into the ceiling, suspended from the ceiling (pendant mounted), free standing and wall mounted (see figure 1).

- light distribution — presented graphically by a polar curve shape which is a schematic illustration of the luminous intensity distribution of the luminaire in characteristic vertical planes. Those curves present the way in which the luminaire controls the light from the lamp. Figure 2 shows different kinds of light distribution, from narrow to very broad, and figure 3 shows classification of luminaires according to upward and downward direction of luminous flux distribution.
- light output ratio — the ratio of luminous flux emitted by a luminaire to that emitted by lamps used in that luminaire. The luminaire is more efficient as the light output ratio is bigger and closer to 1.0.
- utilization factor — the ratio of luminous flux incident on a work plane to the overall luminous flux emitted by luminaire lamps (all luminaires in an interior). For chosen luminaires and their layout in a given interior the utilization factor should be at least 0.5.
- operating condition — the degree of protection against the ingress of dust or moisture is classified according to the Ingress Protection (IP) System. In this system a luminaire is described by a two-digit number, e.g. IP64. The first digit classifies the degree of protection against the ingress of solid bodies, from fingers and tools to dust. The second digit

classifies the degree of protection against the ingress of moisture.
- electrical protection — luminaire classes: 0, I, II, III, according to the type of protection provided against electric shock (where class III indicates the safest usage — the best electrical protection against electric shock).

3. LUMINAIRE SELECTION

The efficiency of the lighting system is a combination of luminaire efficiency and the organization of work environment, such as task difficulty and location, reflectances of main surfaces in interior and dirt depreciation. Although efficiency is very important, it must be considered together with luminance distribution of the luminaire according to glare control. Higher efficiencies are often obtained by wider light output at angles that can cause the glare. Sometimes a chosen luminaire is either efficient or with good glare control, but does not integrate esthetically within the environment. Luminaires should be incorporated into the environment as an integral part of interior design — more as a decorative element than a disturbing part of the equipment. The luminaire should also integrate mechanically, electrically, and acoustically with the interior design (Rea 1993: 1).
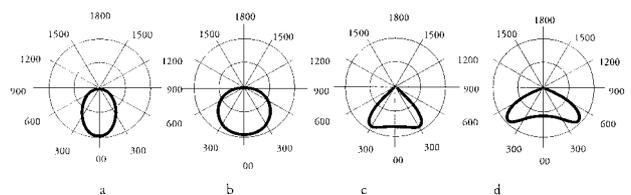
Figure 2. Typical polar curves: (a) narrow distribution, (b) diffuse distribution, (c) dark-light distribution (usually used in VDT rooms), (d) wide distribution.
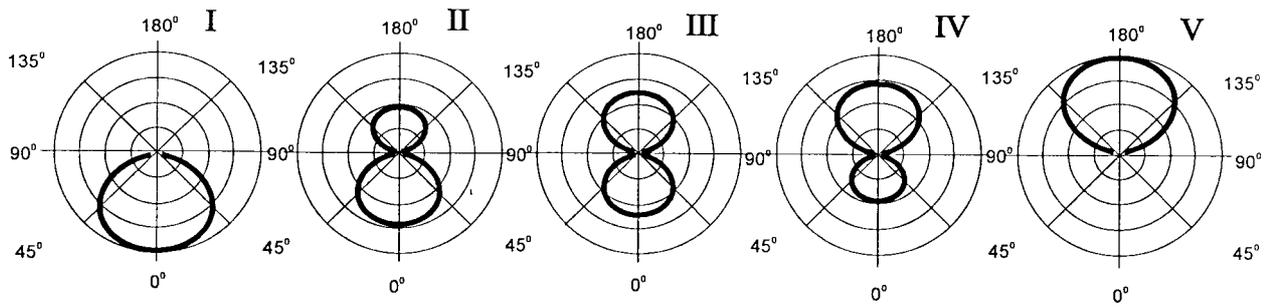
Figure 3. Classification of luminaires according to upward and downward direction of luminous flux distribution: I — direct, II — semi-direct, III — diffuse, IV — semi-indirect, V — indirect.

REFERENCES

DYBCZYŇSKI. W., 1998, *Technika Świetlna '98: Poradniki — Informator, part 1* (Warszawa: Polski Komitet Oświetleniowy SEP).

REA, M.S., 1993, *Lighting Handbook: Reference and Application* (8th ed.) (New York: Illuminating Engineering Society of North America).

YARHAM, R., 1994, *Code for Interior Lighting 1994* (London: Chartered Institution of Building Services Engineers (CIBSE)).

Mental Workload under Thermal Stress

B. Chase

Department of Engineering, University of San Diego, CA, USA

1. INTRODUCTION

Human performance of cognitive tasks is affected by a variety of psychological, physiological and environmental variables. Exposure to either excessive cold or heat can produce performance decrements on simple cognitive tasks and on more complex perceptual motor tasks. The majority of research in this area has focused on cognitive performance in heat and this will also be the focus here.

Mental workload can be thought of as the amount of mental effort required for task performance. The assessment of mental workload is a process of applying models of human cognition to the analysis of tasks and has as its goal maintaining mental workload at a level that allows acceptable task performance.

For over 50 years researchers have been examining the effects of thermal stress on human cognitive performance. The research thus far paints a somewhat fuzzy picture of the effects of thermal stress on mental workload.

2. THERMAL ENVIRONMENT

2.1. Measurement

Past studies of human performance under thermal stress have employed a variety of measurement techniques. Many studies either did not specify or did not fully specify the environmental conditions. Recent research employs wet bulb globe temperature (WBGT).

2.2. Dry–Bulb Temperature

Most people are familiar with standard household thermometers. Dry bulb temperature is simply the reading of a typical mercury thermometer. The result is a measure of the temperature of the surrounding air. However, there are often other factors that can affect the heat stress of a human.

2.3. Wet–Bulb Temperature

Wet-bulb temperature is obtained by placing a wet wick over the mercury bulb and when air passes over the wick, evaporation and consequently cooling occurs. The rate of cooling from evaporation is dependent upon the humidity of the surrounding air. Consequently, the wet- and dry-bulb temperatures can be used to calculate the relative humidity.

2.4. Globe Temperature and WBGT

A well-known index is WBGT, which attempts to combine the effects of air temperature, water vapor pressure, air speed and radiant temperature into a single, robust value. WBGT is calculated using:

WBGT (outdoor conditions) = 0.7 wbt + 0.2 gt + 0.1 db

WBGT (indoor conditions) = 0.7 wbt + 0.3 gt,

where wbt is wet bulb temperature, dbt is dry bulb temperature and gt is globe temperature. Globe temperature is obtained by taking a thin copper sphere, painted flat black, and placing a thermometer at the center and allowing the thermometer to reach equilibrium.

3. ATTENTION AND COGNITIVE PERFORMANCE

3.1 Attention

The general topic of attention has been investigated since the turn of the century. Scientific inquiry into the subject of attention, while popular around the turn of the century, lost recognition during the first half of the twentieth century (possibly due to the influence of Behaviorism). During modern applied research, attention was of interest, as the devices humans had to operate became more complex, and consequently more demanding. Of interest was the ability to attend to many objects of stimuli at once, such as in guiding a land, air or sea vessel. These activities involve elements of attention like tracking, vigilance, decision-making, communicating, etc.

Fundamental aspects of attention can be described such as selective attention, or the ability to concentrate on certain elements of one's environment while ignoring others. Divided attention describes the ability to divide attention among various tasks, which may be competing for our attentional resources. The idea of *mental workload* is meant as a measure or assessment of the attentional demands a task may place on a human operator.

3.2. Models of Attention

There is a variety of competing models in the area of attention. The current available models can be divided into two types as they pertain to human factors. The first type is often termed a bottleneck model, which refers to a point in the process where a limit of attention may occur. This type of model may be contrasted with a resource model, where attention is viewed as a capacity that has limits. Further, these two models may each be divided into categories. Bottleneck models can be divided into early selection and late selection. Resource models can be divided into single resource and multiple resources theories.

3.2.1 Bottleneck models

In 1958 Broadbent proposed a model known as filter theory. In filter theory, an early selection model, stimuli enter in the central processor individually and are filtered according to whether they are wanted or extraneous items. Later, other models were proposed in which the filter did not serve to block unattended stimuli, but rather attenuated their signal. Others argued for a late selection model, which would place the screening process later in the sequence of events. The main impetus for this approach was the belief that even unwanted stimuli were at some point being identified. While the issue is quite difficult, in an attempt to deal with conflicting evidence, some researchers felt that the bottleneck was not fixed. Rather there could be a shift from early to late selection in the attentional process.

3.2.2 Resource models

One of the main proponents of the resource model was Kahneman (1973). In his model, attention is viewed as a unitary limited capacity resource. Owing to difficulties in experimental findings, some were led to propose a multiple resource model, which holds that there is no single attentional resource. Rather, there are
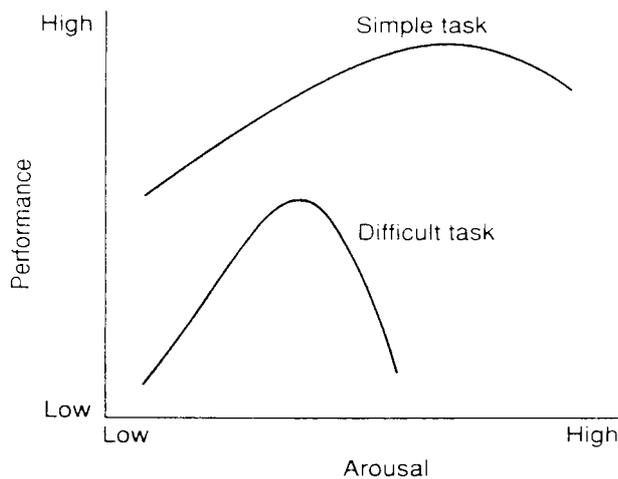
Figure 1. Yerkes–Dodson Law.

multiple resources from which we may draw, which may or may not overlap (Proctor and van Zandt 1994).

Basic aspects of attention can be distinguished, such as that of *divided attention*. This refers to the ability to divide attention between two (dual) or more simultaneous tasks. Dual task performance has been studied in great depth in the human factors/ ergonomics literature. Dual tasks are more difficult and demand perceptual motor skills, more closely approximating industrial and military tasks (Ramsey 1995).

3.3. Attention and Arousal

The restriction of attention that occurs during high periods of arousal has been termed perceptual narrowing (Kahneman 1973). An individual's attentional capacity is influenced by his or her arousal. The level of arousal may exert an influence on the amount of attentional resources available. This relationship between arousal and performance underlies a fundamental law known as the Yerkes–Dodson Law (Yerkes and Dodson, 1908). This inverted U-shaped relationship is illustrated in Figure 1.

The Yerkes–Dodson Law states that as arousal increases, performance initially increases. At the point where arousal, presumably from some stressor, exceeds the organisms ability to cope, performance begins to deteriorate. Separate curves can be drawn for different classes of tasks. Simple cognitive tasks appear to be less sensitive to the level of arousal, whereas complex perceptual motor tasks seem to show a greater sensitivity to the level of arousal.

3.4. Dual-Tasks

In situations where an operator's attention must be divided between two competing tasks, the tasks may not have the same priority or emphasis level. In cases such as these, performance on one task can be plotted along with performance on the second task.

The performance trade off or dual-task cost can be captured graphically with a performance operating characteristic (POC) curve. POC is similar to the receiver operating characteristic (ROC) curve, which can plot hit rates against false-alarm rates in signal detection experiments. POC curves allow one to examine such aspects of performance as allocation of attention, divided attention, dual-task cost and an operator's functional performance

region. With these metrics, relative divided attention cost measures can be obtained and compared, which can control for individual differences in task performance.

4. MENTAL WORKLOAD METHODS

4.1. Measurement

There are multiple methods for the assessment of mental workload. The results of these assessments can prove useful in system design and redesign. Further, as the system is broken down into the constituent parts and the workload requirements of the system components analyzed, predictive models of performance can be constructed.

4.2. Assessment Techniques

4.2.1. Primary task measures

Measures of the primary system task can reflect the experimenter's interest and overall system goals determine workload requirements and the choice of these measures on the task of interest.

4.2.2. Secondary task measures

It is assumed that performance on the primary task consumes a certain amount of cognitive resources and a secondary task will use any resources left over from performance of the primary task. Consequently, if the primary task leaves few resources, performance will suffer on some secondary task and performance on this secondary task becomes a measure of mental workload.

When using this method of task analysis, it is important to keep in mind the fact that the secondary task should not seem artificial and distracting.

5. PHYSIOLOGICAL MEASURES

Physiological measures can often remove subjective elements from the analysis. Heart rate analysis can prove quite useful in estimating mental workload. At lower levels of workload, the heart rate frequency can fluctuate, whereas under higher workload heart rate demonstrates greater constancy.

6. SUBJECTIVE MEASURES

On the more subjective side, an easy method for estimating workload is simply to ask the operator how difficult the task. While a highly subjective measure, questionnaires can be constructed in a manner that can allow for greater accuracy in rating. The rating scales should be anchored by thorough descriptions of the extreme and middle scale values and should attempt to analyze the task on multiple dimensions.

6.1. Heat Stress Effects

Studies on the effect of heat have generally yielded inconsistent results. This is partly due to the fact that different kinds of cognitive work are differentially sensitive to thermal stress. The majority of the work in heat and human performance has been concerned with sustained attention or vigilance tasks. However, when analyzing performance on various tasks, one must consider the type of task being performed.

A comprehensive study that took into account the type of task involved and converted, where possible, all temperature measures over to WBGT is Ramsey (1995). By dividing task types

into two categories, Ramsey acknowledged the differentially sensitive nature of task type. The two categories were: mental, cognitive, very simple perceptual motor, sensory, time estimation, reaction time, etc.; other perceptual motor tasks, including tracking, vigilance, vehicle or machine operation, complex or dual tasks, etc.

Category 1 can be thought of as simple cognitive tasks; and category 2 can be referred to as complex or dual-tasks. Ramsey's conclusion in regard to the category 2 tasks (above) is that there is an onset of a statistically significant decrement in performance in the range 30–33∞C WBGT. Simple cognitive tasks are less sensitive to the effects of thermal stress.

Studies in this area have employed a variety of temperature scales (e.g. effective temperature, WBGT, dry bulb temperature); have required the performance of multiple tasks that may draw from different cognitive resources; were often ambiguous about task emphasis, and rarely obtained a baseline of subject performance.

While many past surveys of the literature have focused on the first two difficulties mentioned above (i.e. temperature index and task type), the issues of task emphasis and individual differences in mental capacity have not been properly dealt with in the literature.

The idea of equating baselines is not absent from the literature on attention in general. Past studies have employed this important methodology (e.g. Somberg and Salthouse 1982). A close analogy to equating the baselines of performance is found in the concept of training. If subjects are trained to a specified level of proficiency, one could argue that the task is equally difficult for all participants. This, however, is not persuasive. The method of extended practice has been criticized as an invalid method of equating. Participants will invariably differ in the time taken to achieve a threshold of performance and this difference in the amount of practice may exert significantly on the ability to detect overall differences in performance. Any differences may, in fact, be due to differences in the level of automaticity. Further, it has been argued that heat stress limits based as they currently are, upon physiological responses to stress, ignore the true goal of productive and efficient system performance. Continuing exposure after behavioral performance and efficiency begin to fail, but before physiological limits are reached, is not appropriate for system performance and efficiency (Hancock and Vasmatzidis 1998).

6.2. Reducing Heat Stress

The second Industrial Revolution is bringing about a transition in the type of work being performed. Work activities, in industrial and military settings, are becoming more cognitively demanding. Therefore, the purely physiological approach to environmental stress is no longer sufficient. A cognitive performance approach to evaluating the interface of the worker and the environment is required.

Reduction of stress due to heat in industrial and military settings can be accomplished through a variety of mechanisms. The most effective approach is to reduce heat at the source or to lower, where possible, the environmental temperature and humidity. The use of air conditioners, fans, and/or dehumidifiers can accomplish heat reduction at the source. Along the path from the source to the individual, shielding may be used. At the individual level, interventions could include: loose fitting clothes,

reduction in work rate, enforced rest breaks with provided drinking water and other fluids, and performing outdoor work at cooler times of the day.

Workers should be trained in proper work habits for work in hot environments. Use of equipment such as air-cooled, water-cooled and ice-bag vests can ameliorate the effects of heat.

In terms of heat exposure limits put forth by OSHA (Occupational Safety and Health Administration), one must look at the air velocity, the temperature in WBGT units, and the workload of the constituent tasks.

7. CONCLUSION

As industrial work becomes more cognitive in nature and less physical, safety issues need to evolve to account for the hazards that environmental stress may impose on the workforce. There is a need for universal standards in regard to environmental stressors for different categories of work. Heat, in particular, is a complicated issue and when multiple stressors are imposed, things become quite complex. Data concerning the effect of environmental heat on human cognitive performance in military and industrial settings is one clear future need. A further area of fruitful investigation would be the derivation of predictive models of human behavior under stress. Such models could be quite valuable in the simulation of performance and human reliability in both extreme and more common work environments.

Once the decrement in performance due to heat is observed, one must determine the point of decrement onset by testing at a variety of temperatures. Further aspects of performance can also be investigated, such as ability to allocate attention in varying emphasis levels. By requiring participants to switch attention between two tasks, it can be determined whether heat affects the ability to ignore irrelevant information, attend solely to one task, switch the emphasis level, and equally to share attention among two equally discriminable tasks.

By examining behaviorally based measures rather than physiologically based measures, heat standards and predictive models can be developed that focus upon behavioral performance. In this way, exposure would not continue past the point of performance break down, which may occur before physiological system limits are reached. This area of investigation has come a long way, but future efforts will no doubt lead to more useful findings concerning heat and human performance.

REFERENCES

KAHNEMAN, D., 1973, *Attention and Effort* (Englewood Cliffs: Prentice Hall).

HANCOCK, P.A. and VASMATZIDIS, I., 1989, Human occupational and performance limits under stress, the thermal environment as a prototypical example. *Ergonomics*, 41, 1169–91.

PROCTOR, R.W. and VAN ZANDT, T., 1994, *Human Factors in Simple and Complex Systems* (Boston: Allyn & Bacon).

RAMSEY, J.D., 1995, Task performance in heat: a review. *Ergonomics*, 38, 154–65.

SOMBERG, B.L. and SALTHOUSE, T.A., 1982, Divided attention abilities in young and old adults. *Journal of Experimental Psychology: Human Perception and Performance*, 8, 651–63.

YERKES, R.M. and DODSON, J.D., 1908, The relation of strength of stimulus to rapidity of habit-formation. *Journal of Comparative Neurology of Psychology*, 18, 459–82.

Noise at Work

P.J. School

1. INTRODUCTION

Undesirable sound in the 20 Hz to 20 000 Hz (20 kHz) range is called noise. Noise may be constant or may vary in duration, frequency, or magnitude. Kryter (1970) and Price *et al.* (1989) provide convincing evidence of the debilitating effects of noise on humans. Noise impairs human performance, and high levels cause temporary hearing impairment or permanent hearing loss. Therefore, noise is a Human Factors engineering consideration.

The following is a succinct description of essential concepts that may be used to quiet freestanding items such as pumps, motors or engine-powered equipment. A very extensive coverage of noise and noise control is provided by Harris (1991). Lord *et al.* (1980) is a readable text that contains several examples of practical designs. Hirshorn (1989) is a concise reference.

2. NOISE REDUCTION

2.1. Feasibility

Feasibility determination involves methods that permit the prediction of the magnitude of noise reduction and the cost of noise reduction effort. Basically, data are gathered on a predecessor or mock-up item. Those data are used to generate a notional design and a cost estimate to implement that design. That estimate is compared with medical, worker compensation or other costs that would accrue if the item was not quieted.

If noise reduction is infeasible, measures such as posting warning notices requiring use of hearing protectors or prohibiting access to a noisy area may be used. Because most hearing protection devices impair speech intelligibility and some devices cause discomfort, personnel tend to avoid their use. Also, people occasionally forget to wear hearing protectors. Consequently, the effectiveness of such measures relates to training, monitoring and management control.

2.2. Reductions Achieved

Over the past 15 years, quieting efforts on several free standing, diesel engine powered items produced 5–20 dB noise reductions. Those noise reductions led to attractive cost and human performance benefits.

This experience showed that 6 dB reductions are often readily achievable. Twelve-decibel reductions require more effort and insight. In a few cases, reductions of > 18 dB were achieved. Those large noise reductions involved relocating noisy components, adding vibration isolation, adding parts such as intake and cooling air silencers or balancing rotating parts.

2.3. Sound Level Measurement

Subjective judgement of sound levels is unreliable; therefore, noise assessment or abatement requires sound level measurement and first decimal place accuracy. Instruments that comply with the American National Standards Institute (ANSI) S1.4 Type 0 or Type I sound meter requirements provide sufficient accuracy (ANSI S1.4-1983, para. 3.2). A calibrated Type 0 meter is accurate to ± 0.4 dB; the Type I is accurate to ± 0.7 dB.

2.4. Fundamentals

2.4.1. Sound levels

Sound levels in decibels may be obtained from sound power or sound pressure measurements. Sound power levels (L_w) are expressed as acoustic power in watts (W). Sound pressure levels (L_p) are expressed as the magnitude of air pressure changes in pascals. The term "sound level" indicates that the data are log base 10 transformed.

2.4.2. Sound power level

Sound power level in decibels (dB L_w) is defined as:

$$L_w = 10 \log_{10} (W/W_{ref}), \text{ where } W_{ref} \text{ is } 10^{-12} \text{ watt}$$

or

$$L_w = 10 \log_{10} W + 120.$$

Thus, sound power decibels equals 10 log based 10 of the ratio of a sound power level to the 10^{-12} watt reference level.

2.4.3. Sound pressure level

Sound pressure level in decibels (dB L_p) is defined as:

$$L_p \text{ dB} = 10 \log_{10} (P/P_{ref})^2, \text{ where } P_{ref} \text{ is } 20^{-6} \text{ pascals}$$

or

$$L_p \text{ dB} = 20 \log_{10} (P/P_{ref}).$$

Thus, sound pressure level in decibels equals 20 log based 10 of the ratio of a sound pressure to the 20^{-6} pascals reference level.

2.4.4. Decibels are logarithmic

The sound level algorithms contain a $\log_{10}$ term; therefore, sound levels must be logarithmically added or subtracted. The following decibel addition and subtraction nomographs provide approximations. The algorithms that follow the nomographs are more accurate than the nomographs.

2.4.5. Decibel addition

To add two sound decibel levels ($L_1 + L_2$), subtract L_1 from L_2 to yield X; then arithmetically add the corresponding Y to the larger of the two levels.

Add $X =$	0	1	2	3	4	5	6	7	8	9
$Y =$	3	2.5	2.1	1.8	1.5	1.2	1	0.8	0.6	0.5

Add $X =$	10	11	12	13	14	15 to 19	20
$Y =$	0.4	0.3	0.3	0.2	0.2	1	0.04

Example: If L_1, L_2 are both 100 dB the difference is 0; thus, $X = 0$ and $Y = 3$ dB; so, 100 dB + 100 dB = 103 dB.

The preceding nomograph is based on the following algorithms:

$$\log^{-1} (L_t/10) = \log^{-1} (L_1/10) + \log^{-1} (L_2/10)$$

or

$$L_t = 10 \log_{10} [\log^{-1} (L_1/10) + \log^{-1} (L_2/10)]$$

where

L_t = total sound level
L_1 = a sound level
L_2 = another sound level.

2.4.6. Decibel subtraction

If one has a known overall sound level (L_t) which is the sum of two levels, a known level (L_1) and an unknown level (L_2), then L_2 can be determined by arithmetically subtracting L_1 from L_t to yield "X" for the following nomograph. Then, the "Y" that corresponds to the "X" may be subtracted arithmetically from L_t to yield the unknown level L_2.

Subtract									
X =	0.1	1	2	3	4	5	6	7	8	9
Y =	16.4	6.9	4.3	3	2.2	1.7	1.3	1	0.8	0.6

Subtract						
X =	10	11	12	13	14	15 to 19	20
Y =	0.5	0.4	0.3	0.2	0.2	0.1	0.04

Example: If the overall sound level (L_t) is 100 dB and one of the two emitters (L_1) is 95.0 dB, then X = 5 and Y = 1.7. Therefore, the L_2 level is 100 dB − 1.7 = 98.3 dB.

The decibel subtraction algorithms are

$$L_2 = 10 \log_{10} [\log^{-1} (L_t/10) - \log^{-1} (L_1/10)] \text{ and}$$
$$\text{delta} = \log^{-1} (L_t/10) - \{10 \log_{10} [\log^{-1} (L_t/10) - \log^{-1} (L_1/10)]\},$$

where

L_t = total sound level
L_1 = measured sound level of one emitter
L_2 = an unknown sound level produced by another emitter
delta = the amount to be subtracted from the overall (L_t) level.

2.5. Inverse Distance Relationship

Sound waves from a point source that is surrounded by nothing but uniform density air (a free field) radiate in a spherical pattern. Under those circumstances, sound pressure levels are inversely proportional to the distances from the emitter.

A way to extrapolate sound pressure level over distance in a free field is called the "6 dB rule." According to the rule sound pressure levels increase or decrease by 6 dB if the distance to the source is halved or doubled. Using the rule, if the sound pressure level emitted by a point source is 100 dB at 10 m then the level at 5 m would be 100 + 6 dB = 106 dB; and at 20 m the level would be 100− 6 dB = 94 dB.

Sound waves radiating from a sizeable emitter do not become spherical until they travel some distance from the emitter. The area where sound waves have not become spherical is called "the near field." Extrapolations to or from the near field are likely to be inaccurate. Therefore in the near field, sound level measurements are preferred to extrapolations. Textbooks suggest that the near field extends four to seven times the largest item dimension. Thus, if an item's largest dimension is 2 m then the near field could extend 8–14 m from the item.

2.6. Octave Data

An audio spectrum is needed to decide the feasibility of noise reduction and the design. Full octave or fractional octave measurements can provide a sound magnitude versus sound frequency spectrum.

Noise measurement octaves are approximate frequency multiples. Commonly used full octave band center frequencies are 31.5, 62, 128, 250, 500 Hz, etc. One-third octave noise measurements provide a more complete audio spectrum. The common one-third octave center frequencies are 31.5, 40, 50, 63, 80, 100, 125, 160, 200, 250 Hz, etc.

2.7. Sound Level Weighting

A widely published family of curves (Fletcher and Munson 1933) shows that human hearing is non-linear with respect to both frequency and sound pressure. "A-Weighted Decibel Level" (dB(A)) measurements take the non-linearity of human hearing into account. Therefore, much noise reduction work is done with dB(A) data.

2.8. Sound Wave Fates

Though other phenomena may occasionally be of interest, three sound wave fates are almost always of interest. A sound pressure wave may

1. radiate in all directions from an emitter;
2. be reflected from a surface; or
3. be absorbed by a surface.

If a sound wave strikes a surface, some of it may be transmitted through the surface.

For simplicity, Figure 1 depicts a single set of wave paths as straight lines rather than a spherical pattern. If a sound wave

1. does not strike a surface but instead radiates through uniform density air, the sound pressure level will decrease 6.02 dB with each doubling of the distance; or
2. if the wave strikes a mostly reflective surface, most of the sound will be reflected toward the emitter, some of the sound will be absorbed in the surface and the rest will be transmitted through the reflective surface; or
3. if the wave strikes a mostly absorptive surface, most of the sound will be absorbed (converted to heat) in the surface,
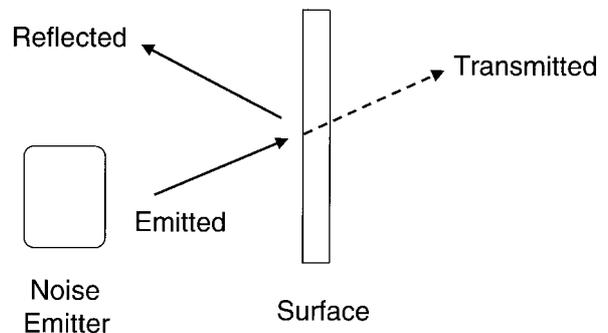
SIMPLIFIED SOUND PATHS

Figure 1.

some of the sound will be reflected by the surface, and the rest will be transmitted through the absorptive surface.

Most surfaces reflect, absorb and transmit sound. Designing a noise-reducing surface typically involves using those characteristics in ways that result in sufficient noise reduction while avoiding unacceptable cost, weight or size gains.

2.8.1. Transmission loss

Sound level reductions that occur from one side of a surface to the other are called "transmission loss." Reflecting sound or absorbing sound can produce transmission loss. The noise spectrum of interest and design limits (such as weight and size) determine whether one uses reflection, absorption, or the combination of both to achieve transmission loss.

3. NOISE REDUCING AN ITEM

3.1. Noise Levels Add Logarithmically

The overall sound level produced by a group of emitters is the sum of the levels produced by each emitter in the group. For example, if an emitter had a sound level of 100 dB and a second emitter was added, then the sound level would increase to 103 dB. On the other hand, if two emitters are > 10 dB apart, the combined level will increase < 0.5 dB. Thus, noise-reduction effort usually focuses on emitters that are < 10 dB of the overall level.

3.2. General Noise-reduction Strategies

Two common strategies for reducing noise are (1) surrounding the noisy item with a noise attenuating enclosure or (2) quieting emitters that comprise the item.

If one cannot change the design of the noisy item itself, noise reduction usually involves an enclosure. With engine-powered items, enclosure design usually requires the use of absorption and reflection. Many items require cooling airflow and that may be provided by sound-attenuating ducts.

3.3. Noise Emitter Determination

Noise reduction efforts begin with overall noise and significant emitter noise measurements. To focus attention productively, one ranks significant emitters (emitters that are within 10 dB of the overall noise level) from highest level to lowest level. Significant emitters may be located by measuring levels at short distances, by making suspected components temporarily inoperative or by touching suspected areas to detect vibrations or rattles. Suspected high frequency (> 1 kHz) emitters may be confirmed by surrounding them with home insulation type glassfibre. To confirm suspected low frequency (< 1 kHz) emitters, surround them with lead-filled foam polymers or other dense material such as lead roof flashing.

3.4. The Ideal Mass Law

Thin, lightweight absorptive materials produce large transmission losses at frequencies > 1 or 2 kHz. At frequencies < 1 kHz, absorptive materials are usually impractical because the thickness needed would be unacceptable. So, reflection is used to achieve low frequency transmission losses.

Reflection-caused transmission loss for a single wall surface can be estimated with an idealized approximation of "the mass law." Hirshorn (1989: D-2) provides the following algorithms for ideal transmission loss sound pressure calculations.

3.4.1. Ideal mass law for English units

Transmission loss (T_L) = 20 log W + 20 log F – 33.5,

where

W = mass density of the reflecting surface (lb per ft^2)
F = frequency (Hz).

3.4.2. Ideal mass law for metric units

Transmission loss (T_L) = 20 log W + 20 log F – 47.5,

where

W = mass density of the reflecting surface (kg per m^2)
F = frequency (Hz).

Figure 2 shows ideal transmission loss for two surface mass densities. Please note that each doubling of either the mass density or the frequency of sound impinging on the surface increases the transmission loss 6 dB.

A comparison of ideal mass law calculations to data gathered with 10 lb/ft^2 material shows that calculated values deviate from measured transmission losses by 5–8 dB at frequencies < 1 kHz (Hirshorn 1989: D-3). Low frequency deviations from ideal mass law numbers are caused by factors described in some detail by Lord *et al.* (1980; 242–7).

To adjust for Ideal Mass Law deviations from empirical data, we estimate low frequency transmission losses by subtracting 6 dB from each of the calculated ideal transmission loss numbers < 1 kHz. Then, we use measurements made on prototyped enclosures to determine if the low frequency transmission losses lead to sufficient overall noise reduction. If not, we adjust the design by changing the surface density or by adding damping to the surface.

Damping increases low frequency transmission losses by converting sound to heat or by shifting the surface resonance to a lower frequency. Thus, damping can avoid some of the weight gain associated with increasing the surface mass density of an untreated surface. Damping may be achieved by adhesively
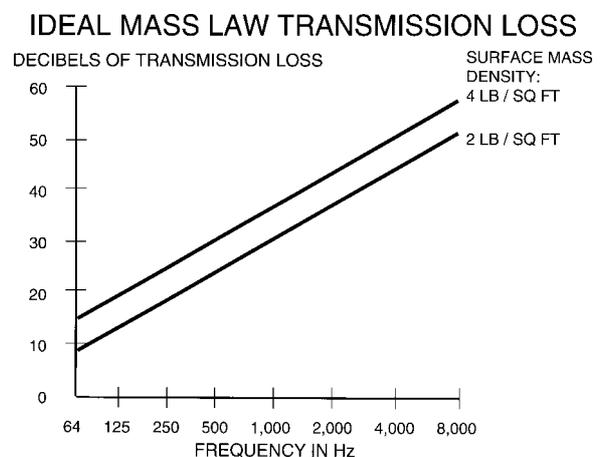

Figure 2.

attaching damping materials to the surface or by making the surface with composite materials that provide good damping.

3.5. Calculating Absorption

Many acoustics texts list sound absorption coefficients for commonly used materials. Additionally, manufacturers of acoustic materials provide lists of absorption coefficients.

According to Hirshorn (1989; C-6), the sound absorption coefficient equals sound energy absorbed (converted to heat) divided by the incident sound energy. Therefore, absorption-caused transmission losses are calculated by multiplying the sound power level by the appropriate (frequency dependent) absorption coefficient of the absorptive material. To do that, sound pressure levels (L_p) are converted to sound power levels (L_w) that are multiplied by the appropriate (frequency dependent) absorption coefficients.

According to Harris (1991; 1.12, 1.13), the metric units L_w to L_p to algorithm is:

$$L_w = L_p + 20 \log_{10} r + 10.9 - C,$$

where

> r = distance from the emitter (m)
> $L_w = 10^{-9}$ watts reference level
> C = an air density correction term.

The English units algorithm is:

$$L_w = L_p + 20 \log_{10} r + 0.6 - C,$$

where

> r = distance from the emitter in feet.

C varies from +0.55 at 1100 millibars, −10°C, to −0.73 at 900 millibars, 55°C. Harris provides a C = 0.6 at 1000 millibars, 20°C.

Those algorithms apply to situations in which sound can radiate in a free (unobstructed) field. An enclosure that surrounds a machine is anything but a free field. Therefore, sound reverberates inside an enclosure and "builds up."

Estimating reverberative buildup with calculations is arduous and error prone. Instead of attempting to calculate reverberative buildup, we develop a design by degrading the calculated transmission losses for frequencies < 1 kHz by 6 db each. Then, we construct a mock-up enclosure and change the thickness or type of liner material as measurements on the mock-up indicate.

4. NOISE ENCLOSURE DESIGN RECOMMENDATIONS

4.1. In General

1. Use a calibrated sound meter to obtain overall dB(A) sound levels and overall full or fractional octave data. Then identify the high level emitters and gather the same data at those loci.

2. Rank significant emitter levels from greatest to least dB levels.

3. Focus effort on emitters that are within 10 dB of the overall sound level, and do whatever is practical to quiet each.

4. Select a surface mass density that causes the transmission losses (each minus 6 dB) < 0.75 kHz to meet design goals.

5. Initially line the interior enclosure surface with thick absorptive material.

6. Assure that the total area of all unsealed openings through the enclosure does not exceed 1% of the total enclosure surface area.

7. Iterate the design to optimize tradeoffs of weight, noise reduction, and component accessibility.

4.2. Recurrent Design Elements

Over several enclosure design efforts (1) sufficient low frequency transmission loss was usually obtained with 1–2 lb/ft² surface material; and (2) an absorptive liner consisting of 1–3 inches of glass fiber usually was needed.

On unquieted engine-powered items, noise emitted from the engine fan, crankcase, engine air intake, hydraulic components or the power train was usually within 10 dB of the overall noise level. Exhaust noise was usually not within 10 dB.

4.3. Tips

- For early design mock-ups, a low cost enclosure can be made of common 0.5-inch thick gypsum wallboard mounted and sealed with duct tape. To reduce reverberative buildup, the inside should be lined with glass fiber.
- Commonly available spreadsheet software can be used to aid calculation.

REFERENCES

AMERICAN NATIONAL STANDARDS INSTITUTE, 1983, *Specification for Sound Level Meters*. ANSI S1.4-1983 (New York: Standards Secretariat, Acoustical Society of America).

FLETCHER, H. and MUNSON, W.A., 1933, Loudness, its definition, measurement, and calculation. *Journal of the Acoustical Society of America*, 5, 82–108.

HARRIS, C.M. (ed.), 1991, *Handbook of Acoustical Measurements and Noise Control* (3rd ed.) (New York: McGraw-Hill).

HIRSHORN, M., 1989, *Noise Control Reference Handbook* (New York: Industrial Acoustics Co.).

KRYTER, K.D., 1970, *The Effects of Noise on Man* (New York: Academic Press).

LORD, H.W., GATLEY, W.S. and EVENSEN, H.A., 1980, *Noise Control for Engineers* (New York: McGraw-Hill).

PRICE, G.R., KALB, J.T. and GARINTHER, G.R., 1989, Toward a measure of auditory handicap in the Army. *Annals of Otology, Rhinology and Laryngology*, 98 (suppl.) 140.

Noise: Definitions

Z. Engel, D. Augustynska, J. Koton and A. Kaczmarska

Katedra Mechaniki i Wibroakustyki, Wydzial Inzynierii Mechanicznej i Robotyki, Akademia Gorniczo-Hutnicza im. St. Staszica, Al. Mickiewicza 30, paw. D-1, 30-059 Krakow, Poland

1. BASIC DEFINITIONS

All disturbing, annoying, strenuous and hazardous sounds influencing the hearing organ and other senses of a human body are considered as noise.

From the point of view of physics all sounds are mechanic vibrations of an elastic medium (gas, liquid or solid). Those vibrations can be considered as an oscillating movement of medium particles around the state of equilibrium, causing the change of the medium pressure versus the static (atmospheric) pressure. Out of this change of pressure (it means, the disturbance of equilibrium), which in a form of successive local condensations and thinnings of medium particles is transferred into the space surrounding the vibration source, an acoustic wave is being formed.

The difference between the pressure instantaneous value at the acoustic wave passing pressure value is called the acoustic pressure, p (expressed in Pa). Considering the wide range of an acoustic pressure from $2 \_ 10^{-5}$ to $2 \_ 10^2$ Pa the logarithmic scale is used and as a consequence the notion of an acoustic pressure level, L, expressed in dB is applied — according to the equation:

$$L = 10 \lg \frac{p^2}{p_0^2} \qquad (1)$$

where p is root mean square pressure value (Pa), p_0 is threshold of rms pressure value, called the reference sound pressure value, equal to 20 mP.

This is a nominal value equivalent of an acoustic pressure of the single tone (sinusoidal vibration) of 1000 Hz frequency at which the hearing impression forms, it means, the so-called threshold of audibility for that frequency occurs.

Ranges of an acoustic pressure occurring in an environment are shown in Figure 1.

All values characterizing the noisiness of work environment — which are going to be discussed in the further parts here — maximal sound A level, peak sound C level, A-weighted sound level, the level of personnel exposure to a noise related either to an 8-h work period or to a week work period are quantities based on the acoustic pressure level.

The acoustic pressure levels corrected according to the frequency-weighting characteristics A and C sound level meters are called the sound A level and the sound C level respectively.

The maximal sound A level means the maximal effective level of sound A occurring during the observation while the peak sound C level the maximal instantaneous level of sound C occurring during the observation.

The equivalent A-weighted sound level, $L_{Aeq,T}$ (the value used either for the description of the noise varying in time or the variable exposure time) is defined as a mean value of the sound A level varying in time being an equivalent to the reaction of a hearing organ exposed to the influence of a constant level noise in a weighted period of time. It is expressed in dB and given by the formula:

$$L_{Aeq,Te} = 10 \lg \left[\frac{1}{T_e} \int_0^{T_e} \left(\frac{p_A(t)}{p_0} \right)^2 dt \right] \qquad (2)$$
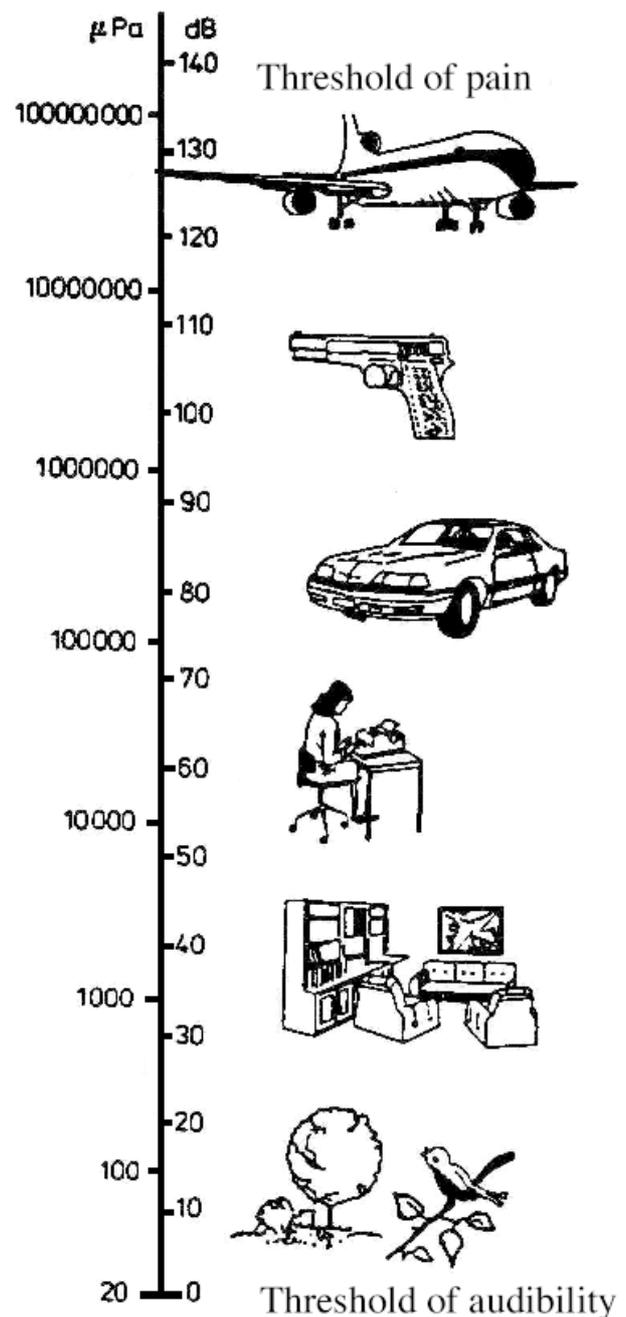

Figure 1.

where T_e is time of exposure (s), p_A is instantaneous value of an acoustic pressure, corrected according to the frequency-weighting characteristics A (Pa).

The A-weighted sound levels, in dB, related to a normal work period (either 8 h or a week) are called the sound exposure level related to the 8-h work period — $L_{EX,8h}$ or related to the week work period — $L_{EX,w}$ and are given by formulae:

$$L_{EX,8h} = L_{Aeq,Te} + 10\lg\frac{T_e}{T_0} \qquad (3)$$

$$L_{EX,w} = 10\lg\left[\frac{1}{5}\sum_{i=1}^{n}10^{0,1(L_{EX,8h})}\right] \qquad (4)$$

where $L_{Aeq,Te}$ — equivalent A-weighted sound level related to the sound exposure time T_e (dB), T_0 is reference time = 8 h = 28 800 s, i is consecutive workday in a specified week, n is number of workdays in a specified week (can be different than 5).

Instead of the exposure time level related to the workday or week another expression called the daily or weekly noise exposure is in use, $L_{A,Te}$ (Pa^{2}s). The interrelation between the noise exposure and the noise exposure level related to 8 h work period is as follows (2–4):

$$L_{A,Te} = 1.15\cdot10^{-5}\cdot10^{0,1L_{EX,8h}} \qquad (5)$$

The knowledge of several additional acoustic phenomena is needed for the effective application of technical means of noise control. They are:

- Velocity of an acoustic wave propagation (speed of sound), c, it means the velocity of the propagation of disturbance of the medium equilibrium defined as the ratio of the distance traveled by the disturbance in an elementary time interval to the value of that interval. For example this velocity equals 340 m/s in the air of temperature 20°C and at the normal atmospheric pressure.
- Acoustic vibration period, T is the shortest time interval after which the same state of an observed phenomenon (vibrations or disturbances) is repeating itself.
- Acoustic vibration phase, f, is the value of a vibrating particle deviation — in the given point and at the given moment — from the mean value of the particle location.
- Acoustic vibration frequency (sound frequency), f, is the number of vibration periods in the time unit.
- Acoustic wave length, l, is the distance between two successive points, measured in the direction of the disturbance propagation, in which the vibrations are in the same phase (or the distance traveled by the wave front during one full period).

The acoustic wavelength is given by the expression: $l = c/f$, in m, where c is the sound velocity in m/s and f is the frequency (Hz). For the audiofrequency range: f = 16–16 000 Hz and the wavelength: l = 21–0.021 m.

In simplification, one can say that the noise is the sum of a large number of sinusoidal vibrations. Factoring the complex vibrations into a sum of single vibrations is called the spectrum determination or the spectrum (frequency) analysis of noise.

A transmission of the disturbing energy is connected with an acoustic wave propagation in a medium. The acoustic wave energy is characterized by the following notions and values:

- Sound power of a source, N_a, in W is it is the measure of the energy radiated by the source in the unit of time:

$$N_a = \frac{E_a}{t} \qquad (6)$$

where E_a is acoustic energy of a source (W◊s), t is time (s).

- Sound intensity, I_a in W/m^2, the value of an sound power transmitted by a unit area perpendicular to the direction of an acoustic wave propagation:

$$I_a = \frac{N_a}{S} \qquad (7)$$

where N_a is sound power (W), S is surface area (m^2).

The following relationship exists between the sound intensity, I_a (W/m^2) and the acoustic pressure (p):

$$I_a = \frac{p^2}{\rho\cdot c} \qquad (8)$$

where r is medium density (kg/m^3), c is sound velocity (m/s).

Similarly to the acoustic pressure, because of the wide range of an acoustic power and sound intensity variations, the logarithmic scale is in use and notions of:

- sound power level, L_N (dB):

$$L_N = 10\lg\frac{N_a}{10^{-12}} \qquad (9)$$

- sound intensity level, L_L (dB):

$$L_i = 10\lg\frac{I_a}{10^{-12}} \qquad (10)$$

The sound power level is the basic value characterizing the emission of noise from its source and therefore it is applied for the estimation of a machinery noisiness. In practice, the sound power level corrected according to the frequency-weighting characteristics A, called a sound power level A is the most often used.

Taking into consideration the type of noise influence on people it can be either called the annoying noise — not causing any permanent effects in humans, or the harmful noise — which either causes permanent effects or creates a certain risk of their occurrence.

In respect of its time dependency the noise is either steady — when the sound A level in a certain point varies not more than by 5 dB — or transient (variable in time, interrupted) — when the sound A level in a certain point varies by more than 5 dB. One kind of the transient noise is the impulse noise consisting of one or more sounds, each of a duration < 1 s (PN-N-01307:1994).

DIVISION OF SOUNDS CONSIDERING THEIR FREQUENCY

Inaudible sounds	Sounds audible for humans			Inaudible sounds
	Audible low tones, sensitivity of human organs	Middle tones, sharpened sensitivity of human hearing organs	Overtones, weaker sensitivity of humans	Ultrasounds
1 10	10 100 20	0 10	00 40	00 10000	20000 Hz

Audibility range for dogs

Audibility range for bats

Machine noise

Wind noise
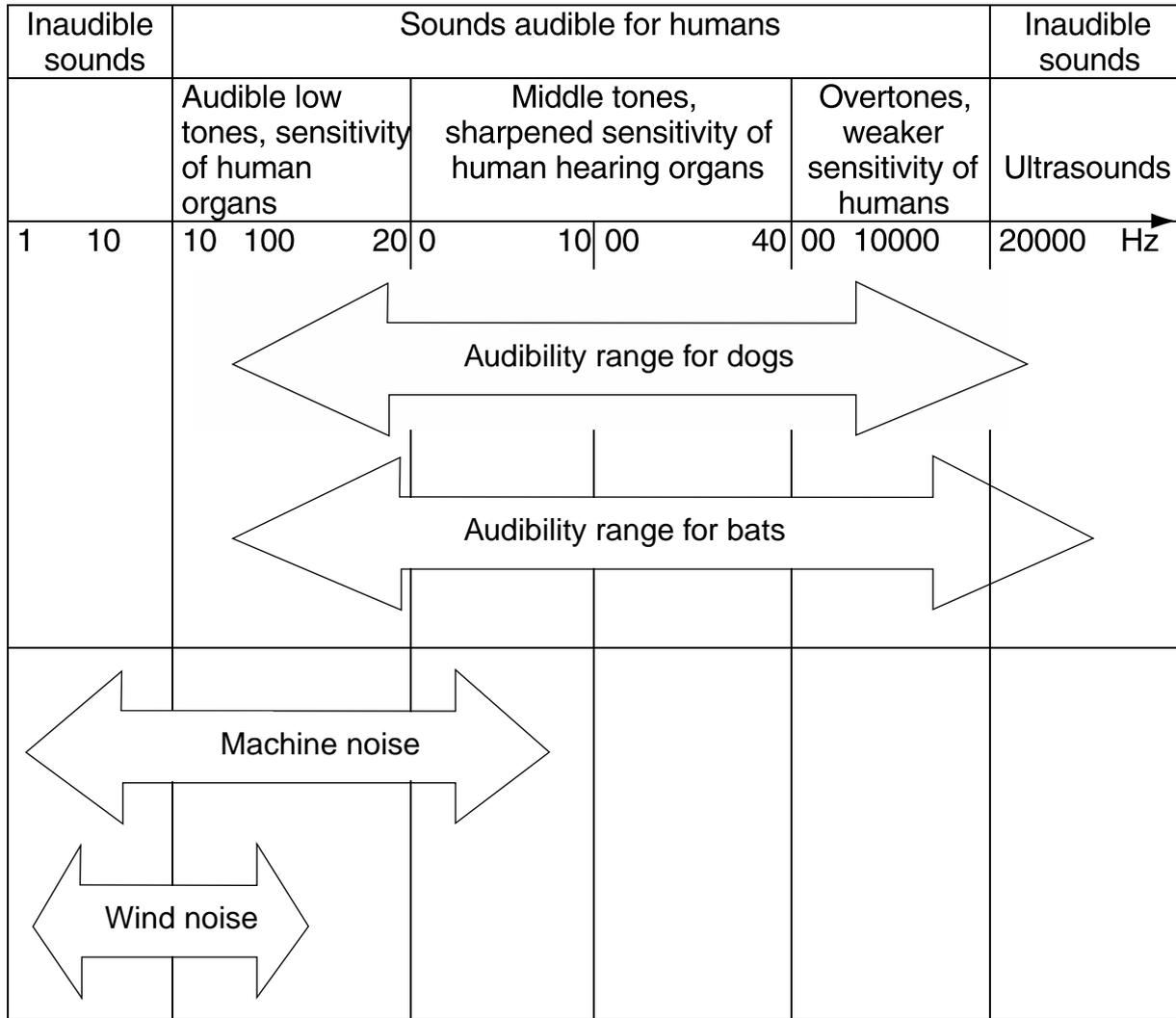
Figure 2.

Taking into account its frequency range the noise can be divided into three groups:

- Infrasound noise, in which spectrum there are components of infrasound frequencies from 2 to 16 Hz and of audible ones up to 50 Hz (PN-N-01338:1986) [26].
- Audible noise, in which spectrum there are components of audible frequencies from 16 to 16 000 Hz.
- Ultrasound noise, in which spectrum there are components of audible and ultrasound frequencies from 10 to 100 kHz.

There are also some other methods of differentiations of noise, e.g. taking into consideration the cause of its formation and the classification of its sources. There is, for example, an aerodynamical noise — formed as a result of the air or other gas movement, and a mechanical noise caused by friction and collisions of solids, mainly elements of machines. There is also a way of dividing noise by considering an environment in which it occurs. A noise in industry is called an industrial noise, a noise in housing areas, public utilities and rest areas is called a communal noise while the noise in means of transportation a traffic noise.

The schematic presentation of the sound division when taking into account their frequencies is given in Figure 2.

1.2. Loudness and Noisiness

The relation between physical values characterizing sound and a subjective feeling is not simple, however strictly defined. Subjective characteristic values of a sound are loudness — depended on its intensity, and a timbre — depended on the spectrum frequency and timing (duration, time variation of the intensity — sound building up and decay).

Interrelations of the sound loudness and intensity in the whole audibility range is presented in Figure 3. Narrow lines represent curves of the equal loudness level, giving the isophone contours. The intensity range is very broad, for example for the tone of 1000 Hz frequency (standard value in acoustics) it spreads out from the threshold value of $2 \_ 10^{-12}$ W/m², hardly noticeable by an ear, (equivalent to the threshold of audibility of an acoustic pressure of $2 \_ 10^{-5}$ Pa) to the 10 W/m² being the threshold of pain.

Taking into account the wide span of the intensity of an

Figure 3.

Figure 4.

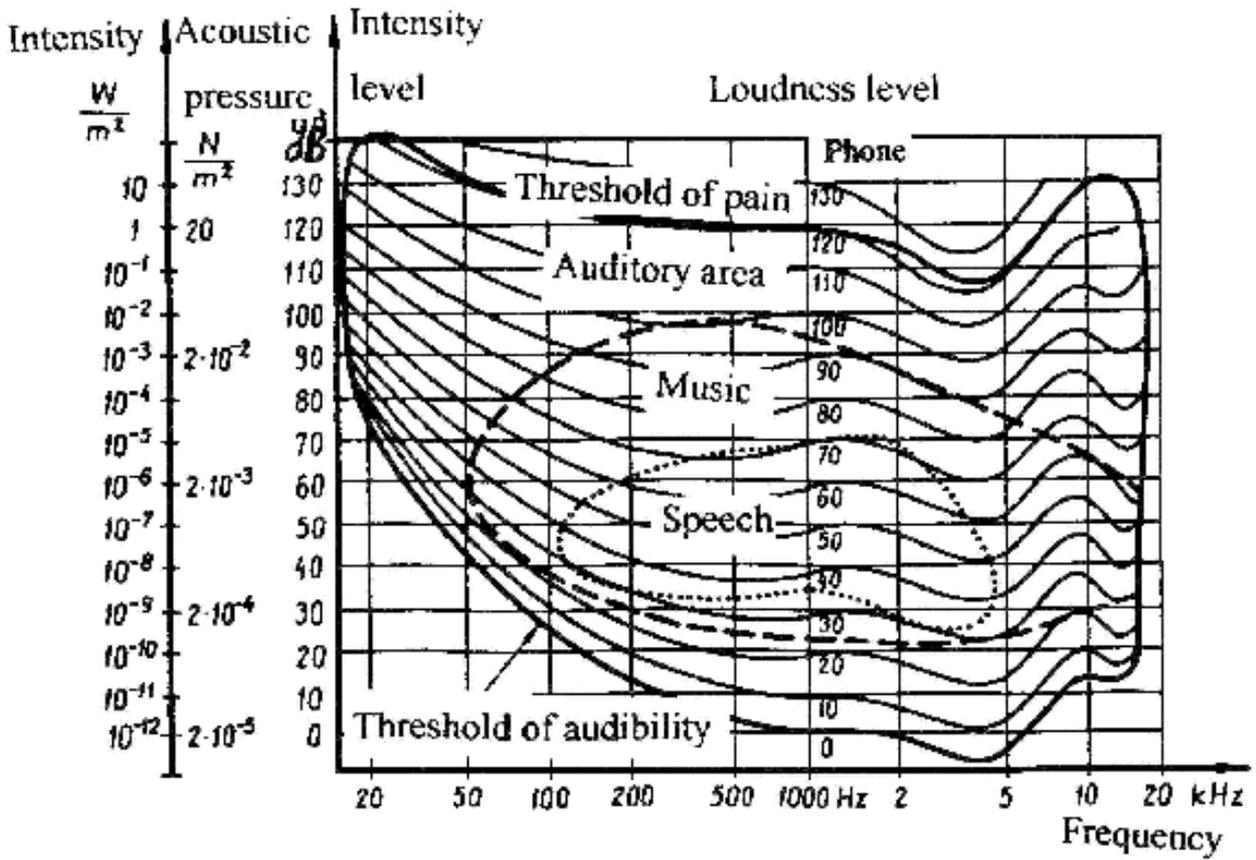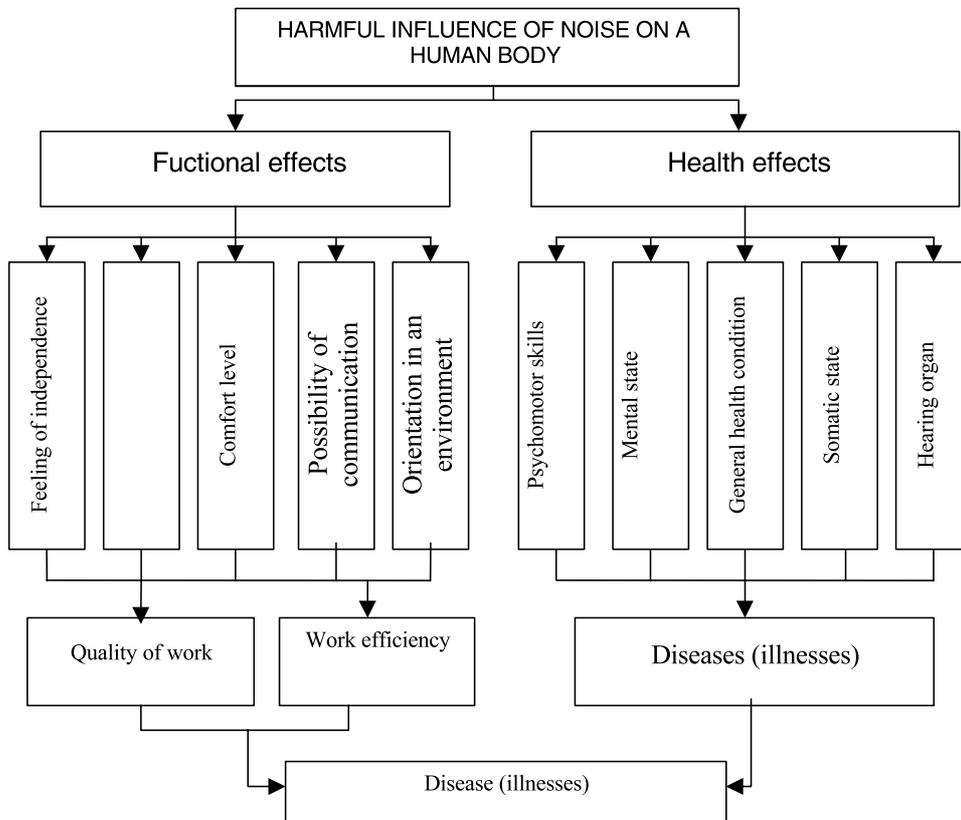auditory sensation area as well as the fact that the impression of loudness is (in agreement with the Weber–Fechner's law — valid for all senses) proportional to the logarithm of the sound intensity, the level of the sound intensity (equation 40) related to the threshold of audibility of the 1000 Hz frequency tone has been introduced: $I_0 = 10^{-12}\,W/m^2$.

In Figure 3 the straight horizontal lines represent equal sound levels from 0 to 130 dB in 10-dB intervals (marked on the axis of ordinates) and acoustic pressures marked on the left hand side of the same axis. The frequency given on the axis of abscissae is also in the logarithmic scale.

2. INFLUENCE OF NOISE AND ITS EFFECTS ON A HUMAN BODY

The negative influence of noise on a human body under the condition of occupational risk can be divided into two kinds (Figure 4):

- health effects (hearing organ, general health condition, etc.),
- functional effects (non-hearing-related influence on a human body).

2.1. Influence of Noise on a Hearing Organ

The harmful influence of noise on a hearing organ is connected with the following features and risk occurrences:

- A-weighted sound level or sound A level (for a steady noise) > 80 dB. Weaker stimuli are not harmful for the hearing organ even when acting for a long time without any interruption (Table 1).
- Long time of the noise exposure. The effects accumulate in time, they depend on the acoustic energy doze transferred into a body in a certain period (Table 1).
- Continuous exposure to noise is more harmful than interrupted. Even short brakes allow for the regeneration process of the hearing organ. The impulse noise is especially harmful. It is characterized by such rapid and high increase of the acoustic pressure that the defending mechanism of the hearing organ preventing the energy penetration into the ear has no chance to react.

Table 1. The risk of noise induced impairment of hearing as a function of the A-weighted sound level and the time of exposure (ISO 1999: 1975)

A-weighted sound level, dB	Risk of impairment of hearing,% Time of exposure in years							
	5	10	15	20	25	30	35	40
< 80	0	0	0	0	0	0	0	0
85	1	3	5	6	7	8	9	10
90	4	10	14	16	16	18	20	21
95	7	17	24	28	29	31	32	29
100	12	29	37	42	43	44	44	41
105	18	42	53	58	60	62	61	54
110	26	55	71	78	78	77	72	62
115	36	71	83	87	84	81	75	64

- Noise spectrum with the majority of components of medium and high frequencies. Noise of such spectrum is more dangerous for the hearing than the one with the maximal energy in the low frequency range. That is connected with the fact that the sensitivity of a human ear is the highest in the frequency range 3–5 kHz.
- Special, individual susceptibility to impairment of hearing. It depends on the inherited features as well as acquired ones (e.g. as a result of past diseases).

Table 1 illustrates the individual susceptibility to noise. Of persons who worked for 40 years in the environment where the equivalent sound level equaled 90 dB, 21% would suffer from hearing loss. Lowering the level to 85 dB decreases the percentage of sufferers to 10%. Persons of a special susceptibility to the harmful influence of noise will be mainly in that group.

The effects of noise influence on the hearing organ can be divided as follows:

- Impairment of anatomic structures of the hearing organ (perforations, diminution of eardrum membrane) being mainly the result of a single and short exposure for noise of the peak acoustic pressure > 130–140 dB.
- Impairment of a hearing efficiency manifesting itself as the increase of the threshold of audibility. It is normally the result of a prolonged exposure to noise of the A-weighted sound level > 80 dB. This increase of the threshold can be reversible (temporary threshold shift) or permanent (permanent hearing loss).
- The audiometric examinations reveal the permanent hearing loss. The average permanent hearing loss of 30 dB at frequencies of 1000, 2000 and 4000 Hz on the side of a better ear, after making allowances for age, is considered the critical loss. It is the base for the diagnosis and the medical certificate of the occupational deafness as the occupational disease.
- The occupational hearing loss (the occupational deafness), the permanent disability without the possibility for any improvement has been for years ranked high on the list of occupational diseases. There are ~3000 new cases yearly what constitutes one-third of all cases registered in Poland.

2.2. Non-Hearing-related Influence of Noise

The non-hearing-related influence of noise on a human body is not fully recognizable. The anatomical connection of a nervous hearing system with a cortex enables hearing stimuli to influence other brain centers (especially the central nervous system and the endocrine glands system) and as a consequence a state and functioning of many internal organs.

It has been experimentally proofed that significant disorders of a body physiological functions can occur when the acoustic pressure level > 75 dB. Weaker acoustic stimuli (55–75 dB) can cause distraction of concentration, hindrance of work and the reduction of its efficiency. It can be stated that the non-hearing-related effects are generalized responses of a human body on a sound influence as the factor contributing to the progress of several illnesses (high blood pressure, gastric ulcers, neurosis, etc.).

Among non-hearing-related effects their influence on understanding and concealing of speech and sound security signals should be mentioned. Difficulties in a verbal communication in a noisy environment (of 80–90 dB level) and concealing of warning

signals not only increase the strenuousness of working conditions and reduces the work efficiency but also can be the reason of injury at work. The criterion of speech understanding is one of the main criteria of the noise estimation in an environment.

3. MEASUREMENTS AND THE ESTIMATION OF NOISE IN AN ENVIRONMENT

Taking into consideration the purpose of measurements the methods are divided into:

- methods of measuring machine noise; and
- methods of measuring noise in places occupied by people (at work places).

3.1. Methods of Measuring Machine Noise

Those methods are applied for the determination of values characterizing noise of a machine considered as a separate noise source at certain experimental and exploitation conditions. Such values are: an acoustic power level A or an acoustic power level in frequency bands (calculated either on the bases of the acoustic pressure level or the sound intensity), the A-weighted sound level and the sound C peak level at work places (for machines equipped with the work place fixed the machine), a noise source direction indicator (if necessary), an acoustic pressure level and a sound A level — measured at specified measuring points (e.g. in case where the acoustic power level is impossible to be measured). (In an agreement with standard EN ISO series 11200 they are called the averaged in time pressure level of an acoustic emission corrected by characteristics A, and the peak level of an acoustic emission corrected by characteristics C.)

The measured values are applied for:

- Comparison of measured levels with the established boundary values.
- Designing, implementation and the estimation of noise control and reduction means.
- Comparison of machine noises either of the same kind or of different kinds.
- Establishing the sound A level or the acoustic pressure level at a specified distance from the source.

All those values should be determined according to the European and international standards EN ISO 3740 and EN ISO 11200.

3.2. Methods of Measuring and Estimating Noise in Places Occupied by People

Those methods are used for establishing the degree of risk to be noise exposed for people at their work places and in specified distances from the noise source or sources. The results are mainly used for the comparison of existing acoustic conditions with the conditions required by standards and hygienic regulations. They are also helpful in the estimation and choosing the proper enterprises for noise control.

Methods of measuring noise in the work environment are given in standards. The instrumentation system, noise dosimeters or integrating sound level meters of a type 2 or better accuracy and at least 53 dB impulse range should be applied (they should meet the requirements of IEC 804:1985 and IEC 1252:1993).

There are two basic methods of measurements: direct and indirect.

The direct method consists of the continuous measurement during the whole time of worker exposure and of reading the meters, e.g. the noise dosimeter or the integrating sound level meter. Those results precisely describe the workers' risk.

The indirect method is based on the noise measurement done in the shorter time than the one being assessed and the use of mathematical formulae for the estimation of values needed.

The mode and frequency of taking measurements, the way of their registration, storing and making available for workers are given in the regulation of the Ministry of Health and Social Security.

The estimation of the occupational risk is based mainly on the comparison of measured or estimated noise values with the permissible ones (the level of noise exposure, the maximal sound A level and the peak sound C level) being in force simultaneously. Even if only one of those values is higher than allowed the noise is regarded as exceeding the permissible value.

3.3. Permissible Noise Values in Work Environment

Allowed noise values in regard of a hearing protection — according to standards of many countries — are as follows:

- Noise exposure level related to 8-h work day ($L_{EX,8h}$) should be < 85 dB and the equivalent daily exposure should not overcome $3.64 \_ 10^3\,Pa^2\,s$ or (only when the noise is affecting a human body unevenly in separate days of a week).
- Noise exposure level related to the work week ($L_{EX,w}$) should be < 85 dB and the equivalent week exposition should be < $18.2 \_ 10^3\,Pa^2\,s$.
- Maximal sound A level (L_{Amax}) should not exceed 115 dB.
- Peak sound C level (L_{Cpeak}) should not exceed 135 dB.
- The last two values are important for the estimation of short and impulse noises.
- Given above standard regulations are compulsory if any special rules are not stating still lower values (e.g. at a juvenile work place — $L_{EX,8h}$ = 80 dB and at a pregnant woman work place — $L_{EX,8h}$ = 65 dB).
- Taking into account the possibility of fulfilling the basic tasks by workers (it means considering the non-hearing-related noise effects — the strenuousness criterion) the noise values should not be higher than listed in Table 2.

Table 2. A-weighted sound level $L_{Aeq,Tc}$ for employees present at their working places (in accordance with PN-N-01307:1994)

No	Work place	A–weighted sound level, $L_{Aeq,Tc,}$ dB
1.	Direct control cabins without a phone connection, laboratories with noise sources, rooms with computers, typing machines, fax, and other rooms of similar destination.	75
2.	Dispatcher's cabins, observation and remote control cabines with a phone connection used in a process control, rooms for precision works and other rooms of similar destination.	65
3.	Administration rooms, design offices, rooms for theoretical works, data processing and other rooms of similar destination.	55

Notice: Values in Table 2 should not be related to 8 h work day. They concern nominal time, T_e, during which an employee is present at the certain work place.

3.4. Noise Risk and Sound Sources in Work Environment

Large number of employees of basic branches of the economy and industry, in many countries, work under the risk condition because of annoying and stressful factors of the work environment, among them noise of an exposure level > 85 dB.

The most endangered are workers of industry, building, forestry and transport, metallurgy, machine industry, chemical industry, mining and electric power plants.

Main noise sources in machinery and equipment are shown in Figure 5.

Assuming that the main noise sources at work places are machines, equipment and technological processes, one can specify the basic groups of noise sources [5]:

- Machines being the energy sources, e.g. engines (maximum sound A level up to 125 dB), compressors (up to 113 dB).
- Tools and pneumatic engines, e.g. hand operated pneumatic tools: hammers, chisels, grinders (up to 134 dB).
- Machines for comminution, crushing, sieving, cutting, cleaning, e.g. ball mills (up to 120 dB), vibration sieves (up to 119 dB), crushers (up to 119 dB), shaking grids (up to 115 dB), circular saws for metals (up to 115 dB).
- Machines for plastic forming, e.g. pneumatic hammers (up to 122 dB), presses (up to 115 dB).

NOISE SOURCES IN MACHINES AND EQUIPMENT
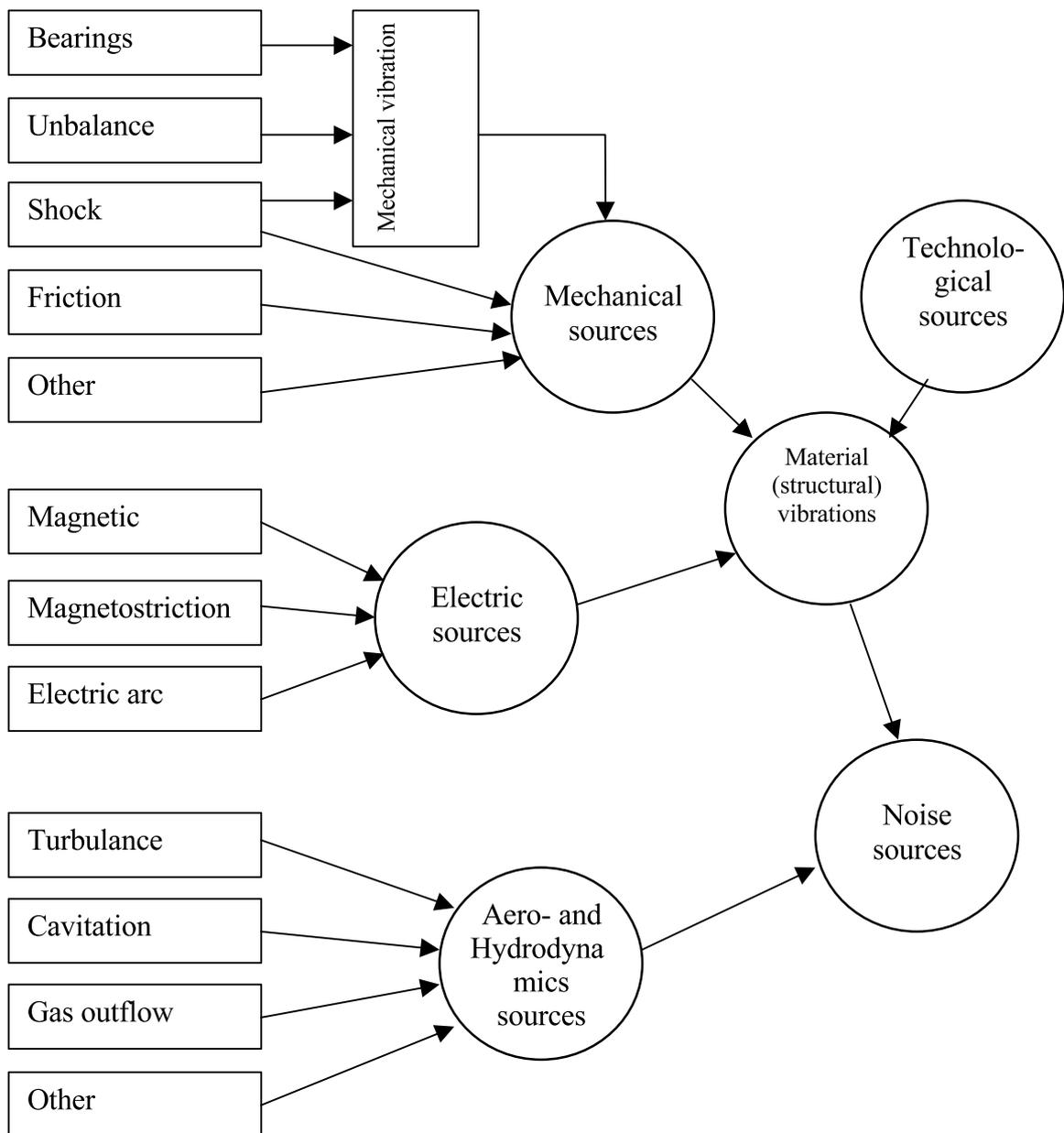
Figure 5.

- Metal working machines, e.g. grinders, automatic lathes, boring machines (up to 104 dB).
- Woodworking machines, e.g. chisels (up to 108 dB), planers (up to 101 dB), moulders (up to 101 dB), circular saws (up to 99 dB).
- Textile machines, e.g. winders (up to 114 dB), looms (up to 112 dB), spinning frames (up to 110 dB), drawing frames (up to 104 dB), twisters (up to 104 dB), carding machines (up to 102 dB), flow devices, e.g. valves (up to 120 dB), vents (up to 114 dB), works transport devices, e.g. overhead cranes, conveyers, feeders (up to 112 dB).

4. METHODS AND MEANS OF THE NOISE PROTECTION

According to the European and national regulations an employer is obliged to provide the protection against hazard connected with the noise exposure and ensure the application of:

- Technological processes not generating any excessive noise.
- Machines and other technical equipment causing as small as possible noise, not in excess of the quantity allowed.
- Solutions lowering the noise level at work places (with the priority for means of the noise reduction at its source).

At those work places, where regardless of applying all possible means of the noise reduction, the noise level is exceeding the limit, an employer is supposed to:

- Find out the reason of such noise and to design the program of technological and organizational activities aimed at lowering that noise.
- Provide employees with hearing protections, matching the noise characteristics as well as individual features of workers and to enforce the use of such protections.
- Limit the noise exposure time and to introduce work breaks.
- Mark the noise hazard zones and, when reasonable and feasible, to limit the access to such zones by fencing them.

Persons employed at those work places must be provided with the information concerning:
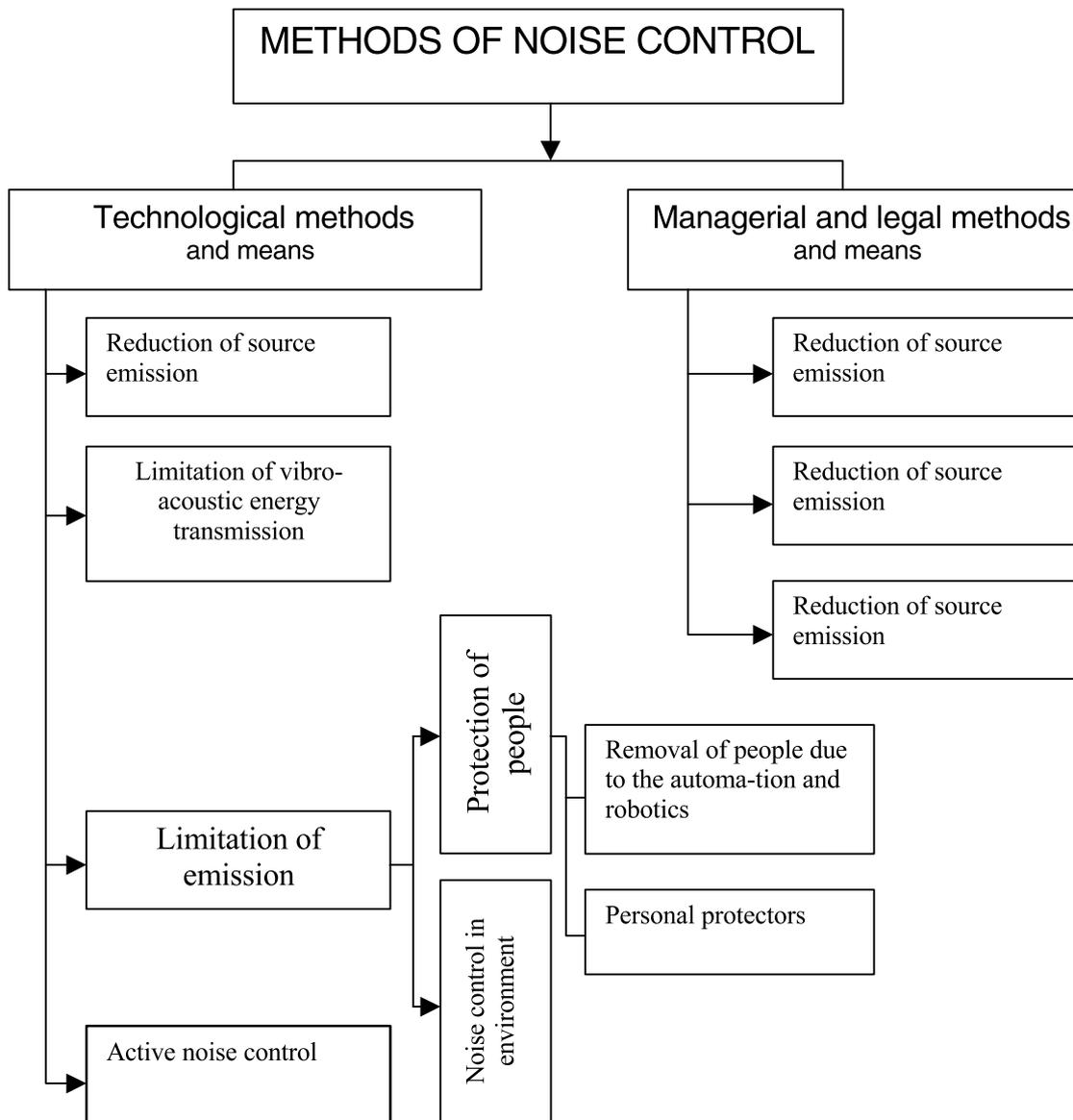
Figure 6.

- Noise measurements results and the health hazard they are facing.
- Activities undertaken in connection with noise exceeding the limit.
- Proper selection of individual hearing protection devices and the way of their use.

When the noise at work places is exceeding the limit employees are undergoing periodic medical examinations. In case of significant and dynamic progress of hearing defects the frequency of audiometric examinations should be increased and the time between successive tests shortened to 12 or even to 6 months. In case of an impulse noise hazard or noise of the A-weighted sound level > 110 dB the audiometric examination must be done at least once a year.

Methods of noise control are schematically presented in Figure 6.

4.1. Technical Means of Noise Reduction

4.1.1. Change of a noisy technological process into a less noisy

The loudest technological processes can be substituted by less loudly ones, e.g. forging by hammer can be substituted by rolling and forming while machining by hand operated tools can be substituted by electric and chemical treatment and mechanized tools.

4.1.2. Mechanization and automation of technological processes

Mechanization and automation of technological processes together with the sound insulated control cabins for the personnel is one of the most modern, futuristic and in the same time the most efficient way of noise and other hazards (e.g. dustiness, high temperature, injuries) elimination.

Majority of cabins used nowadays in industry ensures the noise reduction by 20–50 dB at frequencies > 500 Hz.

4.1.3. Constructing and implementing of silent-running machines, equipment and tools

Changes of technological processes and the introduction of mechanization and automation require longer periods of

realization and cannot be implemented for small-series or an atypical production.

Very efficient silencing of sound sources in machine (the reduction of a sound emission) can be done by:

- Reduction of forces causing vibrations (and restricting their spectra), e.g. by careful balancing of machine elements, changes in a stiffness and system structures, changes in a frictional resistance.
- Change of aerodynamic and hydrodynamic conditions (e.g. by a change in geometry of intake and exhaust of energetic media and a velocity of their flow).
- Reduction of a radiation efficiency coefficient (e.g. by changes of dimensions of vibro-acoustic energy radiating elements, change in materials, and an insulation of boards in a system).

4.1.4. Acoustically correct layout of a plant and utilization of rooms

The following guidelines should be observed at the design stage of industrial buildings:

- Buildings and rooms requiring silence (e.g. laboratories, design offices, places for the scientific research) should be isolated from buildings where a noisy technological process is being performed.
- Machines and equipment should be grouped, if feasible in separate rooms, according to their loudness.
- Noise can be significantly raised by not proper utilizing of a room, especially by too dense location of machines. The shortest recommended distance between machines is 2–3 m.

4.1.5. Acoustic dampers

Reduction of noisiness in the air and other gas ducts (ventilation systems, intake and exhaust systems of fluid-flow machines, e.g. compressors, blast machines, turbines, engines) can be achieved by the application of acoustic dampers. Modern constructions of acoustic dampers do not cause any losses of a machine power. They create a high resistance against noisy transient flows, while not choking steady flows that transport the air or gases. Well-known dampers of that kind are reflexive dampers, it means the
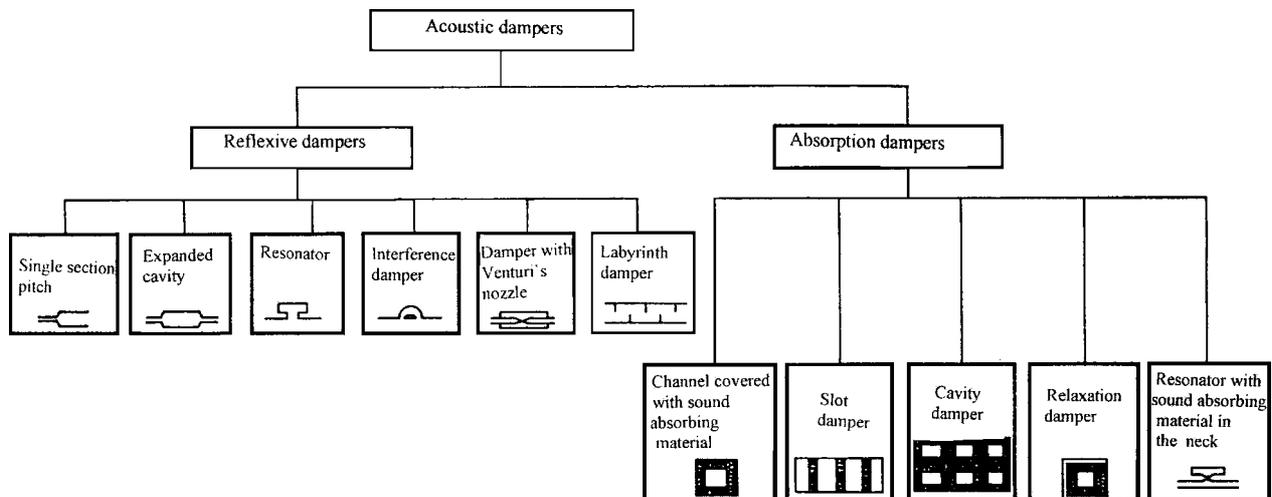
Figure 7.

acoustic wave filters and the absorptive dampers containing the sound absorbing material.

The reflexive dampers operate on the bases of the reflection and interferention of acoustic waves and have very good damping qualities in the range of low and medium frequencies. They are applied where the flow velocities are significant and the temperatures high, mainly in engines, blowers, compressors and sometimes also in ventilators.

The absorption dampers counteract the acoustic energy transmission along ducts by absorbing majority of that energy by the sound absorbing material. They damp mainly medium and high frequencies and therefore are used in ventilation ducts. In practice both kinds of dampers are often used because industrial noise sources emit the energy in a wide range of frequencies.

A different kind of dampers comprise active dampers. In Figure 7 the classification of dampers together with their schematic presentation is shown.

4.1.6. Sound insulated enclosures

An attenuation of the noise source can be achieved by enveloping it, as a whole or a part of it, in an enclosure. Sound insulated enclosures should efficiently damp sound waves emitted by the sound source while not disturbing the normal work and maintenance of enveloped machines.

Typical, the most often used enclosures, have sound insulated walls made out of steel sheets covered from the inside by a
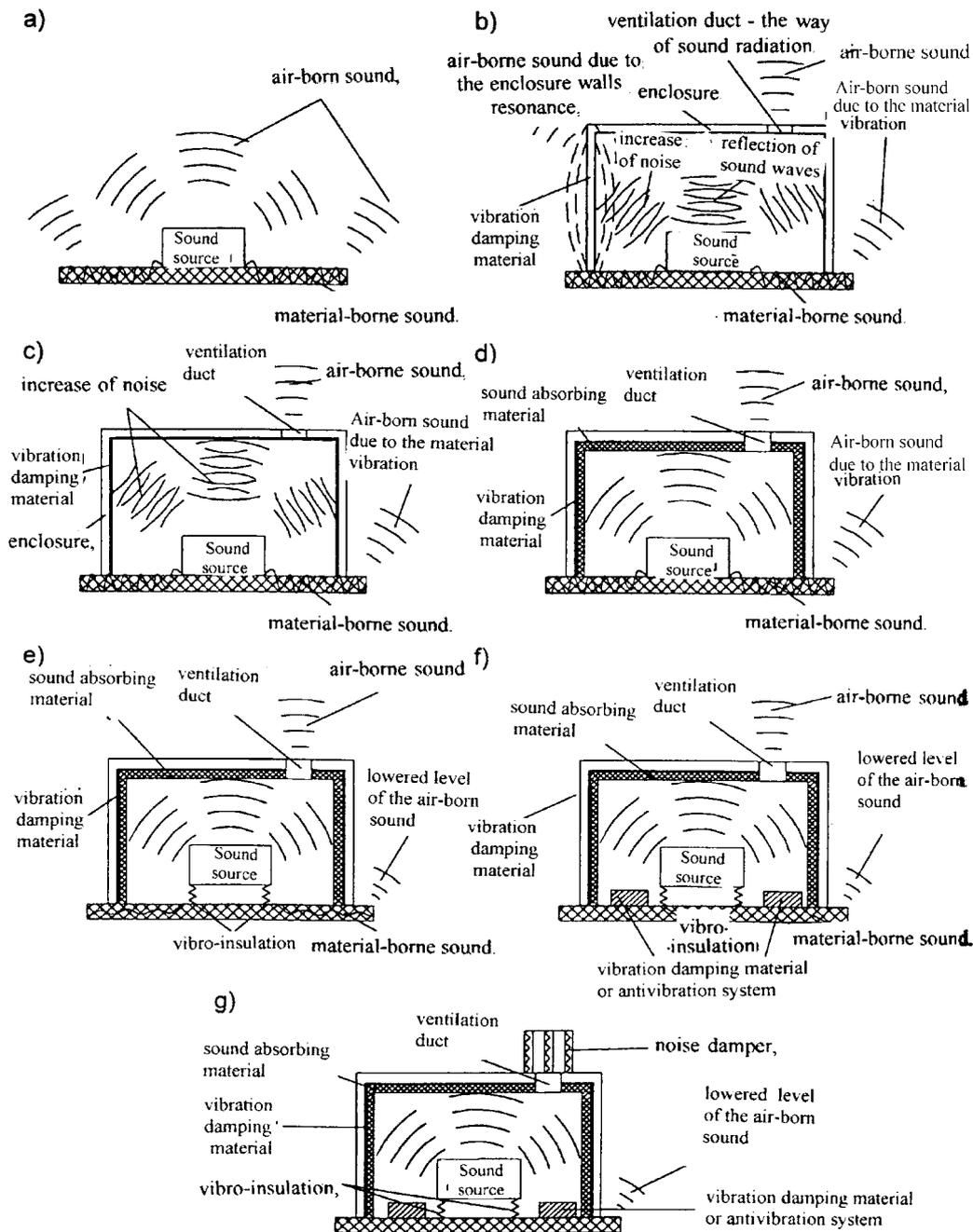
Figure 8.

damping mass or the sound insulating materials. There are also enclosures with multilayered walls.

Correctly made enclosures can attenuate the sound A level by 15–20 dB. Partial enclosures are much less efficient.

Ventilation ducts and other openings, necessary for the technological process, lower the performance of the enclosure. Application of acoustic silencers, e.g. in the form of channels covered with the sound absorbing material, is needed for openings.

Examples illustrating the principle of operation of the enclosure are given in Figure 8.

4.1.7. Sound absorbing screens

Sound absorbing screens are used as shields of work places. Their aim is to damp the noise emitted by different machines in the direction of a work place as well as to damp the emission from that place into an outside space. To achieve the maximum efficiency the screen should be placed as close as possible to the sound source or the work place.

Basic elements of a screen are: an insulation layer inside (usually a steel plate of a proper thickness) and sound absorbing layers outside (plates of mineral or glass wool shielded by perforated sheets).

When the screen is being used in the interior it should be a part of the whole acoustic system to cooperate with other elements damping the reflected waves energy (sound absorbing materials). Effectiveness of properly installed sound absorbing screens is estimated to be 5 — 15 dB in the distance of 1.5 m behind the screen on the axis perpendicular to its surface.

The principle of operation of the acoustic screen is shown in Figure 9. The acoustic energy source in an open space is shown in Figure 9a. The same source with the screen — in Figure 9b

(also in an open space). The acoustic energy source with the screen located inside the room with the ceiling reflecting acoustic waves — Figure 9c, and absorbing them — Figure 9d.

4.1.8. Sound absorbing materials

Sound absorbing materials applied on walls and a ceiling of the room increase its absorbing capacity. Achieved in such a way the lowering of the reflected waves sound level leads to the diminishing of the total noise level in the room.

Porous materials such as textiles, mats of mineral and glass wool, plates and porous wall plasters, plates and porous plastic mats, plastics sprayed under pressure, are the most often applied.

The maximum of the adsorption coefficient of the properly chosen sound absorbing material should occur in the same frequency range where there is the maximum component of the noise spectrum and the allowed noise values are exceeded.

As the experience shows satisfactory damping results (lowering the noise level by 3–7 dB) could be achieved in such rooms where the original noise absorption was small.

Sound absorbing systems such as ceilings, partitions, protective walls, produced in Poland or imported from abroad are available nowadays off-the-shelf.

4.1.9. Hearing protectors

In all cases when the sound level near the work place exceeds the allowed value and all means of the noise reduction are not available immediately the hearing protectors should be used.

Such protectors are also used when an excessive noise occurs seldom or a worker maintaining a noisy machine only occasionally has to enter the place when this machine is located.

The basic value describing the acoustic performance of hearing protectors is their damping of the sound (called previously
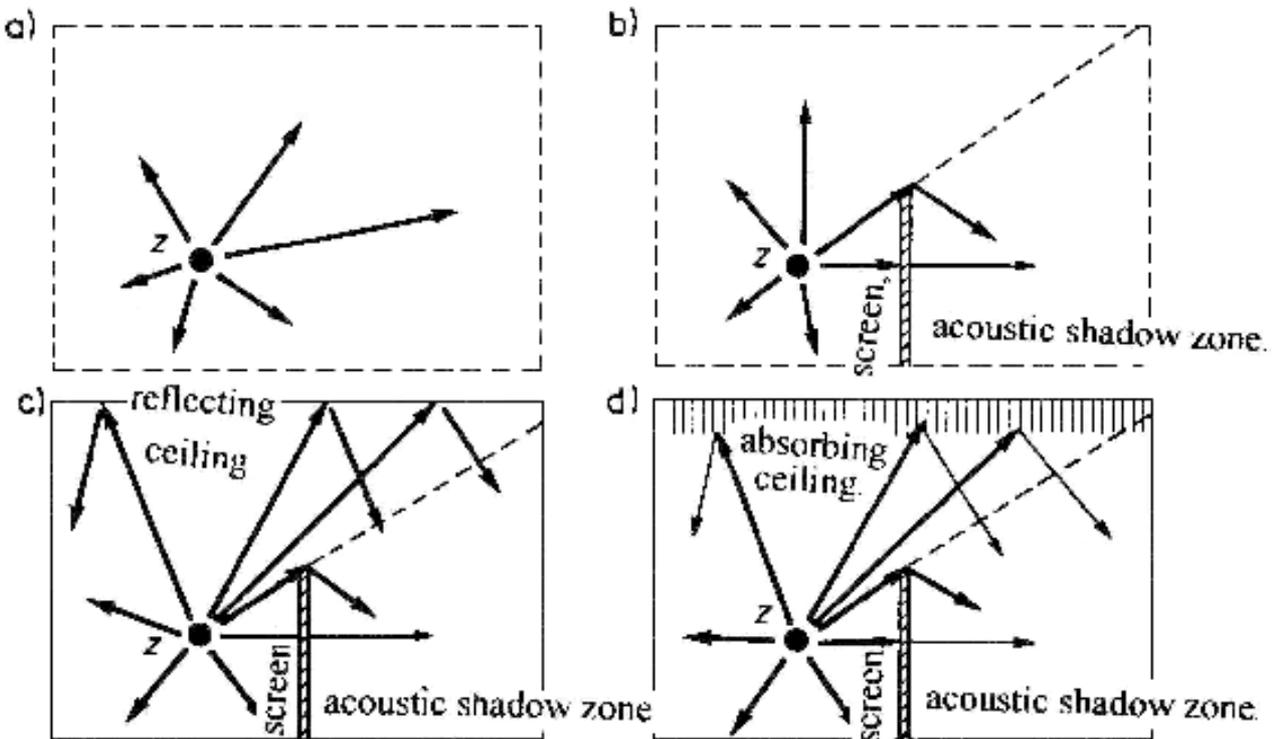

Figure 9.

the acoustic effectiveness) [8]. The hearing protectors are considered satisfactory when the A-weighted sound level under the protector is lower than the admissible value (85 dB).

Considering their construction the hearing protectors can be divided into: antinoise pads (expendable or reusable), antinoise ear-protectors (either with an on-head hold down spring or an on-helmet one) and anti noise helmets.

Selection of protectors for each real acoustic condition should be done with the consideration if they are going to give the satisfactory protection to the hearing organ. Choosing protectors for each work place is done on the bases of acoustic pressure levels in octave frequency bands or sound A and C levels measured at those places and parameters of hearing protectors granted the certificate of the safety mark B.

4.1.10. Active methods of noise reduction

The low frequency noise is especially difficult to limit. Known and used for years traditional (passive) noise reduction methods for the frequencies below 500 Hz are not effective and very costly. In recent years more and more often the active methods are being used. Their characteristic feature is the compensation of noise by sounds from the additional external energy sources. Those sources are controlled by signals of the specially shaped amplitudes and phases. Matching the control factors it is possible either to provide or to absorb the vibro-acoustic energy from specified places in the system.

The general principle of the active compensation of the acoustic field parameters is as follows:

- Primary sound source, called the compensated source, generates the acoustic wave, called the primary wave or the compensated wave.
- Secondary sound source, called the compensating source, generates the secondary acoustic wave, called the compensating wave.
- Destructive interferention of both waves takes place at the specified observation point.

To get the satisfactory noise reduction, strictly described relations between amplitudes and phase shifts of the compensated and the compensating signals must occur at the observation point. In an ideal case the full reduction of the compensated signal will be reached. It will happen only when the compensating signal is the ideal inversion of the compensated one.

Active noise reduction systems, the most often applied in practice, are the active noise dampers of fluid-flow machines and internal combustion engines (the damping efficiency is 15–30 dB for the frequency up to 600 Hz). Active systems are also used often in hearing protectors, they allow to increase the damping effectiveness by 10–15 dB in the frequency range 50–300 Hz.

5. INFRASOUND NOISE

5.1. Factor characteristics

Noise in which spectrum there are components of the infrasound frequencies (2–16 Hz) and the audible frequencies (up to 50 Hz) is called the infrasound noise. At present another expression — low frequency noise (comprising the range from 10–250 Hz) — is more and more often used.

Infrasounds, contrary to the popular believe of their inaudibility, are received by the body by the special hearing way (mainly by the hearing organ). Their audibility depends on the acoustic pressure level. The individual hearing perception especially of low frequency sounds is very different. The lower frequency the higher their audibility threshold, e.g. for frequencies 6–8 Hz the threshold is at ~100 dB, for frequencies 12–16 Hz at ~90 dB.

Apart from the special hearing way the infrasounds are also received by receptors of the vibration feeling. Thresholds of that perception are ~20–30 dB higher than audibility ones.

When the acoustic pressure level > 140 dB the infrasounds can cause permanent harmful changes in the body. The resonance of internal organs of the human body can occur, what is subjectively felt as an unpleasant internal vibration (from 100 dB already). Apart from the heaviness in ears it is one of the most typical symptoms identified by persons under the influence of infrasounds. However the dominant effect of infrasounds, occurring even at the small excesses of the audibility threshold, is their strenuousness. Over-tiredness, discomfort, sleepiness, loosing of balance and psychomotor skills, disorder of physiological functions are symptoms senses in a different degree by individuals. Changes in the central nervous system characteristic for the weakening of a vigil state (especially dangerous for drivers and machine operators) are the objective proof of those subjective feelings.

Main infrasound sources in a work environment are fluid-flow low-rotation machines (compressors, ventilators, engines), electrical equipment (mills, boilers, chimneys), metallurgical furnaces (especially electric-arch furnaces) and casting devices (moulding machines, shakeout grits).

5.2. Permissible Values and Measuring Methods

The infrasound noise at work places is characterized by the acoustic pressure level in octave bands with mid-band frequencies of 8, 16 and 31.5 Hz. The permissible values established for the health protection of employees are listed in Table 3.

In case of protected work places for juvenile and pregnant women the permissible values are lower (Table 4) [18,19].

Taking into account the possibility of performing basic tasks by employees (the criterion of the infrasound

Table 3. Permissable values of the infrasound noise established for the health protection of employees

Mid-band frequency of octave bands [HZ]	Acoustic pressure level for 8 hr noise exposure, [db]	Maximal acoustic pressure level, [db]
8; 16	110	137
31.5	105	132

Table 4. Permissable values of infrasound noise at the juvenile and pregnant women work places

Mid-band frequency of octave bands [HZ]	Acoustic pressure level for 8 hr noise exposure, [dB]
8; 16	85
31.5	80

Table 5. Permissable values of the infrasound noise - with the possibility of performing basic tasks by employees' taken into account

No	Work place	Mid-band frequency of octave bands, [HZ]	Acoustic pressure level for 8 hr noise exposure, [HZ]
1	Dispatcher's cabins, remote control cabins, rooms for precision works and others of similar destination	4; 8; 16; 31.5	90 85
2	Administration rooms, design offices, rooms for theoretical work, data processing and others of similardestination	4; 8; 16; 31.5	85 80

strenuousness) acoustic pressure levels should not exceed values given in Table 5.

Methods of the infrasound noise measuring are described in normatives. Measurements can be done in a direct way (utilizing a mobile frequency analyzer) or in an indirect way (when noise is measured at industrial conditions and the frequency analyses done in a laboratory).

Presently, in accordance with the international standard ISO 7196: 1995 [34], measurements of a sound G level, it means the acoustic pressure level corrected to the frequency G characteristics, are also recommended. The G-characteristics corresponds roughly to the threshold of the infrasound hearing perception and — as the experimental results show — correlates well with the subjective estimation of the sound strenuousness.

According to the recommendation included in the newly established standards: ISO 9612: 1997, ISO 7196: 1995, the proposal of the American Conference of Government Industrial Hygienists (ACGIH) as well as with regard to a progress in the knowledge of the infrasound influence on the human body and with the development of measuring techniques the discussions on the verification of the national NDN infrasound noise standards are planned.

5.3. Methods of Hazard Limitations

In the prevention of the infrasound noise harmful influence the same rules and regulations as in case of a normal noise apply. However the protection against infrasounds is more complicated because of the significant length of the infrasound waves (20–170 m) for which traditional walls, partitions, screens and sound absorbers are not very effective. In some cases infrasound waves are increased because of the resonance of rooms, structure elements and the whole buildings.

The best protection against the harmful infrasound noise gives its suppression at the generating source, it means directly in machines and devices.

Other solutions are:
- Use of sound dampers at inlets and outlets of the air (or gas) in fluid-flow machines.
- Proper foundation engineering (with the vibro-insulation) of machines and devices.
- Stiffening of walls and building structures in case they are getting into a resonance.
- Applying sound insulated cabins of a heavy structure (made of bricks) for machine and installation operators.

- Applying active methods of the noise reduction (connected with the active absorption and sound compensation).

6. ULTRASOUND NOISE
6.1. Factor Characteristics

An ultrasound noise is the noise which spectrum contains components of audible and ultrasound frequencies — from 10 to 100 kHz.

Air ultrasounds being a part of the ultrasound noise can penetrate the human body through the hearing organ as well as the whole body surface. Investigations of the ultrasound noise influence on the hearing organ condition are rendered difficult because in an industrial environment the ultrasounds are usually accompanied by the audible noise. Then it is difficult to assess if the deterioration of hearing is the result of the audible or the ultrasound components only or simultaneous action of both kinds.

However the prevalent opinion is nowadays that as a result of non-linear phenomena occurring inside the ear the subharmonic components are formed under the influence of the ultrasounds. Their acoustic pressure level is of the same order of magnitude as the basic ultrasound partial. In consequence of that phenomenon the hearing impairment concerns the subharmonic frequencies of the ultrasounds.

Apart from its harming the hearing, the negative influence on the vestibular organ in the internal ear manifested by headaches and dizziness, balance disorders, nausea, drowsiness, over-tiring has been found.

Examinations of the non-hearing connected influences have shown that the occupational exposure at the ultrasound noise of levels > 80 dB — in the range of high audible frequencies and > 100 dB — in the range of low ultrasound frequencies, causes changes of the vaso-vegetative nature.

Main sources of the ultrasound noise in a work environment are technical ultrasound devices which generate ultrasounds indispensable for the performing of the technological process. The ultrasound noise can be also generated by high-speed engines: compressors, pneumatic devices, metal working machines and some textile machines.

6.2. Permissible Values and Measurements Methods

The ultrasound noise at work places is described by the acoustic pressure levels in one-third-octave bands of mid-frequencies from 10 to 100 kHz.

Permissible values, established with a view to protect the employees' health ca not surpass the values given in Table 6.

Table 6. Values of the acoustic pressure level of the ultrasound noise - allowed having employees' health in view

Mid-band frequencies of one-third-octave bands KHZ	Acoustic pressure level for 8 hr noise exposure dB	Maximal acoustic pressure level dB
10; 12.5; 16	80	100
20	90	110
25	105	125
31.5; 40; 50; 63; 80; 100	110	130

Table 7. Permissable values of the acoustic level at work places for the juvenile and pregnant women

Mid-band frequencies of one-third-octave bands KHZ	Acoustic pressure level for 8 hr noise exposure dB	Maximal acoustic pressure level dB
10; 12.5; 16	75	77
20	85	87
25	100	102
31.5; 40; 50; 63; 80; 100	105	107

At the protected work places (juvenile, pregnant women) lower values are allowed. They are presented in Table 7.

6.3. Methods of Risk Limitation

The same rules and requirements apply for the prevention of the harmful influence of the ultrasound noise as for a normal noise. The medical examination of workers should be performed every 2 years.

Taking into consideration the short-waves propagation (of low frequency ultrasounds) in the air (wave lengths from 3 mm to 2 cm) it is relatively easy to reduce its harmful influence on humans, e.g. by airtight sealing and enclosing of sources, remote control of the technological process in which ultrasounds are being applied, avoiding direct contact with an ultrasound transducer, using individual protection means, etc.

REFERENCES

AMERICAN NATIONAL STANDARD ANSI S 12.12.1002 (1992).

BERANEK, L.L. and VER, I.L., 1991, *Noise and Vibration Control Engineering: Principles and Applications* (New York: Wiley).

COMMISSION PROPOSAL FOR A COUNCIL DIRECTIVE ON THE MINIMUM HEALTH AND SAFETY REQUIREMENTS REGARDING THE EXPOSURE OF WORKERS TO THE RISK ARISING FROM PHYSICAL AGENTS, 93/C77/01, OJ No. C77, 1993, p. 11, 94/C130/03, OJ No. C130/3, 1994.

CROCKER, M.J. (ed.), 1997, *Encyclopaedia of Acoustics* (New York: Wiley).

EN ISO 11200-11204 (1995) *Acoustics — Noise Emitted by Machinery and Equipment …*.

EN ISO 3743 — 1,2: 1995, 3744: 1995, 3746 (1995) *Acoustics — Determination of Sound Power Levels of Noise Sources Using Sound Pressure …*.

EN ISO 116900 — 1, 2: 1996 (1996) *Acoustics — Recommended Practice for the Design of Low-Noise Work Places Containing Machinery …*.

ENGEL, Z., 1993, *Noise and Vibration Control in Environment* (Warsaw: PWN). [in Polish]

FASOLD, W., KRAAK, W. and SCHIRMER, W., 1984, *Taschenbuch Akustik.* Bd. I, II (Berlin: VEB). [in German]

HARRIS, C.M. (ed.), 1979, *Handbook of Noise Control* (New York: McGraw-Hill).

IEC 804: 1998 (1998) *Integrating — Averaging Sound Level Meters.*

IEC 1252: 1993 (1993) *Electroacoustics — Specifications for Personal Sound Exposure Meters.*

ISO 1999 (1975, 1990 2nd ed.) *Acoustics — Determination of Occupational Noise Exposure and Estimation of Noise-induced Hearing Impairment.*

ISO 7196 (1995) *Acoustics — Frequency-weighing Characteristic for Infrasound Measurements.*

ISO 9612 (1997) *Acoustics — Guidelines for the Measurement and Assessment of Exposure to Noise in a Working Environment.*

Performance Effects of High G Environments

W.B. Albery

Biodynamics and Acceleration Branch, Air Force Research Laboratory, Wright-Patterson AFB, OH 45433, USA

1. INTRODUCTION

Since man first flew in a powered aircraft in 1903, he has had to perform flying tasks in a > 1 G environment — macrogravity. Macrogravity poses problems to the human physiologically, ergonomically and cognitively. Physiologically, the human loses eye-level blood pressure at 22 mm Hg/G as G increases from 1.0 to 9.0, for example in the F-16 aircraft. To maintain consciousness and perform in the high G environment, the human must be protected against the effects of high G. Ergonomically, the human gets heavier as G increases. A 200 lb man weighs 1800 lb at 9 G! Arms, legs and head are heavier, and tasks such as flipping switches on the instrument panel require trajectory strategy and practice with a 10 lb arm that now weighs 90 lb. Cognitively, the loss of eye-level blood pressure and reduced eye–brain-level arterial oxygen saturation (< 80%) relate to flying performance decrement. What can be done to counter these effects of high G? Modern air forces have developed countermeasures to these performance effects of high G environments.

2. FUNDAMENTALS OF THE HIGH G ENVIRONMENT

The body axes for sustained acceleration are shown in Figure 1. The inertial resultants of body acceleration are listed in Table 1. Centrifugal acceleration is expressed in terms of angular velocity and the radius of turn:

$$G = (w^2 r)/g,$$

where w = angular velocity in radians/s and G = unitless measure of acceleration, r = radius of turn (m) and g = gravitational constant (9.8 m s^{-2}). A typical human centrifuge with a 20-ft radius (6.1 m) can generate 9 G by turning at 38 revolutions/min, which is ~4 radians s^{-1}.

The easiest axis for the human to tolerate sustained acceleration is in the chest-to-back (+Gx) direction. This is why astronauts are launched on their backs. Volunteers have been exposed up to +16.5 Gx unprotected for brief periods. Subjects completely immersed in water have endured > 30 G (von Gierke *et al.* 1991). The least tolerable direction is the –Gz or eyeballs-down direction. Some pilots have been exposed to brief periods of -Gz, but the increased blood pressure at the eye–brain level is very uncomfortable. The lateral direction (±Gy) is less tolerable than ±Gx exposures. Most high G aircraft routinely expose the pilot(s) to +Gz. Man, protected and straining, can achieve +9 Gz for short periods and perform complex piloting tasks.

3. PHYSIOLOGICAL EFFECTS OF HIGH G

The high G environment is also characterized by its effect on human physiology. The accepted model of human G tolerance is the hydrostatic column model of man (Burton and Whinnery 1996). In this model, the eye–heart distance results in a change in blood pressure measured at heart level and at eye level (Figure 2).

It is shown that, as one sits in the normal 1 G environment, there is a differential in the eye-to-heart blood pressure. This is illustrated in the seated pilot at the left. The effect of additional +Gz is illustrated where the pilot's neck is elongated. Since each G costs 22 mmHg blood pressure, at 6 G the pilot has –12 mmHg blood pressure at eye level. This –12 mmHg blood pressure must be raised > 0 to keep the eyes and brain perfused. The loss of eye-level blood pressure is typically countered by three protective techniques:

- Cardiac compensation: the baroceptors in the circulatory system sense a reduced blood pressure and signal an increase in heart rate. The increased heart rate increases blood pressure.

Table 1. Nomenclature for inertial resultant of body acceleration.

Linear motion	Physiologic descriptive	Physiologic standard	Vernacular descriptive
Forward	chest-to-back G	+Gx	eyeballs-in
Backward	back-to-chest G	–Gx	eyeballs-out
Upward	positive G	+Gz	eyeballs-down
Downward	negative G	–Gz	eyeballs-up
To right	left lateral G	–Gy	eyeballs-left
To left	right lateral G	+Gy	eyeballs-right

Figure 1. Body axes of sustained acceleration.

Figure 2. Hydrostatic column equivalents.

- Straining maneuver: a self-initiated strain against a closed glottis that raises intrathoracic pressure, which in turn increases eye-level blood pressure.
- Anti-G suit/positive pressure breathing: artificial means of raising one's blood pressure; the anti-G suit inflates and increases pressure to the legs and abdomen. This increased pressure aids in venous return to the heart, which improves G tolerance. The US Air Force has developed a positive pressure breathing system that inflates the pilot's lungs as a function of increasing G (12 mmHg/G to a maximum of 60 mmHg). This system, called COMBAT EDGE, has helped protect humans to +12 Gz for brief periods.

4. ERGONOMIC EFFECTS OF THE HIGH G ENVIRONMENT

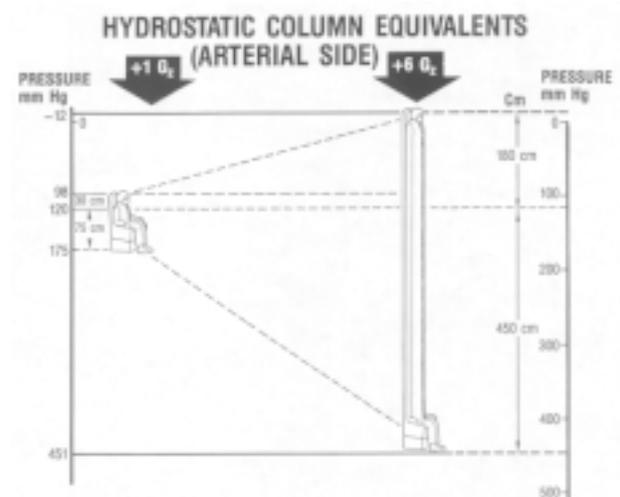The pilot becomes heavier in the seat as G increases. Arms, legs, torso and head all become heavier, and with the same motor control and musculoskeletal systems used at normal 1 G, man initially has problems controlling movements of the head, arms and legs at high G.

4.1. Ability to Move Under Increased G

Increased gravitational fields increase the weight of body parts. Body parts can become elongated or compressed under the G vector; this can affect the shape and function of the soft internal organs including the heart, lungs, kidneys, liver, etc. Higher muscle forces are required to keep the head, torso and limbs in desired positions. At G forces of ~+2 Gz there is increased pressure on the buttocks, drooping of the face and noticeably increased weight of all body parts; at this level of G force it is difficult to raise oneself, and at +3–4 Gz it is nearly impossible. Experiments were conducted in the Wright-Patterson AFB (WPAFB) Dynamic Environment Simulator (DES) centrifuge where subjects arose from a seated position and performed a whole-body jump at Gz levels up to 1.8 Gz as the centrifuge rotated. Although most subjects had no problem standing up at 1.8 Gz, jumping and leaving one's feet was very difficult (Albery and Woolford 1997). Above +3–4 Gz, controlled motions require greater effort, adaptation and learning to offset loss of fine motor control. While seated in a high-performance aircraft, one typically cannot raise the arm at forces > 8 Gz or legs at > 3 Gz. Head pitching is difficult at > 4 Gz, and some individuals who get their heads pitched forward at high Gz (> 6 G) are unable to right themselves in the seat until the acceleration is unloaded. The hand can be raised slightly at 25 Gz. One can barely slide the feet on a floor > 5 Gz. Speech is severely affected, yet possible, up to +9 Gz if the operator is utilizing protective techniques properly (see above).

4.2. Arm Trajectory at High G

When reaching for a switch on the instrument panel at high G, the tendency of the pilot is to complete the movement below the switch because the arm is now heavier due to increased +Gz. The pilot learns after several attempts to initiate the trajectory of the arm–hand at a location above the switch. A re-calibration of gross motor movement occurs during long exposure to even low-level acceleration (Albery and Woolford 1997).

5. COGNITIVE EFFECTS OF HIGH G

Very rarely does high-sustained acceleration affect only one factor such as speech, vision or manual dexterity. Task performance is manually measured through subjective reports, physiological and behavioral techniques. Cognitive performance is significantly affected when eye-level arterial oxygen saturation is < 80%. Percent arterial oxygen saturation ($\%S_aO_2$), as measured with a pulse oximeter, can drop to < 80% when the ventilation/perfusion ratio is not evenly matched. This can occur when subjects are exposed to high G. Complex task performance has been shown to deteriorate when $\%S_aO_2 < 80$ (Albery and Chelette 1998). Percent cerebral oxygen saturation ($\%rO_2$) is also used as a physiological measure in centrifuge research (Chelette et al. 1998).

5.1. Characterizing Human Performance at High G — Exposures up to 3 Minutes

In studies at the Air Force Research Laboratory, WPAFB, human performance during simulated aerial combat sorties has been characterized via "performance surfaces" (Figure 3). These 3-D projections allow the observer to look at the cognitive and physiological performance of the subject as a function of brief exposures to sustained acceleration (+ Gz).

In figure 3, the performance task score is shown on the vertical axis, time spent at Gz > 8 is shown on the one horizontal axis, and the physiological variable, minimum cerebral oxygen saturation (min rO_2), is shown on the other horizontal axis. One can see from this performance surface that the best performance in terms of score was when the subjects spent only 10 s at 8 G and when rO_2 was at its highest, 65%. Chelette et al. (1998) compared male and female performance during a 3-min simulated air-to-air combat situation that exposed subjects to 9 Gz. They found no significant differences between male and female flying performance.

5.2. Characterizing Human Performance at High G – Exposures from 3 to 11 Minutes

Some pilots and many centrifuge subjects have endured G exposures that included peaks to 9 G for several minutes. While evaluating new protective garments, humans have endured 10 Gz continuously for 3 min. Subjects performing a dual-task simulated aerial combat maneuver in the DES centrifuge endured up to 27 peak exposures to 9 G over 11 min and still maintain
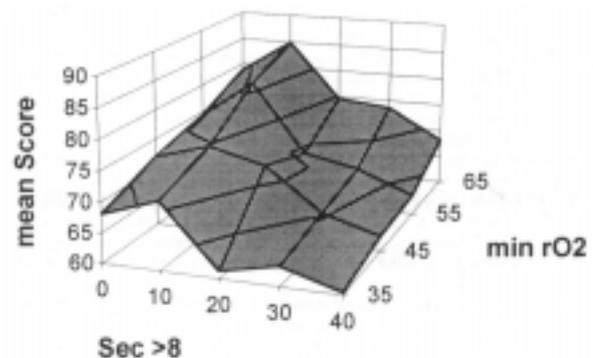

Figure 3. Performance surface.

excellent tracking and secondary task performance. This exposure is much longer than what a typical dogfight would probably last, but illustrates the enhanced protection of humans wearing new COMBAT EDGE-type G protection garments.

5.3. Effect of High G on Orientation Perception

Sustained acceleration (G >1) can also have a profound effect on human perception. Some of these effects are discussed in the Previc chapter. It is believed that the otolith organs of the vestibular system are gravity transducers (graviceptors) and, as such, play an important role in human perception of motion, especially in the absence of visual cues.

5.3.1. Somatogravic illusion – high G pitch-up sensation

Sustained acceleration experienced by pilots during catapult launching of Navy aircraft off the decks of aircraft carriers has been implicated as a contributing factor to aircraft mishaps. The catapult launch generates a + Gx acceleration on the pilot which can elicit a somatogravic illusion (a false sensation of body tilt that results from perceiving as vertical the direction of a non-vertical gravito-inertial force). Pilots of carrier-launched aircraft need to be especially wary of the somatogravic illusion. These pilots experience pulse accelerations lasting 2–4 s and generating peak inertial forces of +3–5 Gz. Although the major acceleration is over quickly, the resulting illusion of nose-high pitch can persist for > 30 s afterward, resulting in a particularly hazardous situation for the pilot unaware of the phenomenon. Navy scientists reproduced on the centrifuge the pitch-up sensation experienced by pilots during catapult launches, and warnings were issued to all pilots taking off from aircraft carriers.

5.3.2. G-excess illusion

The G-excess illusion is a false or exaggerated sensation of body tilt that can occur when the G environment is sustained at > 1 G. This illusion has been implicated in a number of low-level turning and looking mishaps in the USAF. It is believed that the otoliths "over report" a roll or pitch when the pilot and the aircraft are in a > 1 G maneuver. This over reporting occurs because the otoliths are "tugged" more than usual when G > 1 and this is perceived by the pilot as an increased roll or pitch attitude. If a pilot experiences the G-excess illusion, the tendency is to compensate for the illusion, which can cause overbanking, loss of altitude and even controlled flight into the terrain. Chelette et al. (1998) duplicated the G-excess illusion on the WPAFB DES centrifuge (Albery et al. 1997). Rolling and/or pitching the head while in an excess G environment (> 1) can cause an illusory sensation of vehicle tilt. This illusion occurs in the pitch axis if the head is forward, but translates to the roll axis as the head is turned toward one shoulder.

5.4. Time and Mass Estimation in Macrogravity

Time, mass and reaction time perception experiments have been conducted under high levels of G stress (Albery et al. 1997). G

levels up to and including +8 Gz affect time perception. The longer the time estimation, such as > 15 s, the greater the error. Subjects tended to underestimate the length of longer tasks, as if they wanted to finish the task early. Likewise, for mass discrimination, it was found that performance is impaired during increased +Gz. The ability to discriminate between the heaviness of several objects became more difficult under G stress. In several studies involving choice reaction time (more than one possible selection) in acceleration stress, it was found that reaction times actually decrease under G stress. This effect of improved reaction time performance under acceleration can be possibly attributed to the alerting cue phenomenon (subjects are more "on their toes" when exposed to short periods of sustained acceleration).

6. CONCLUSIONS

Most humans can tolerate and perform in G levels up to 4 G for tens of seconds without straining and unprotected. Above 4–6 Gz, the loss of eye-level blood pressure must be countered for the pilot to remain conscious. Performance is degraded. By straining and raising one's eye-level blood pressure, unprotected pilots can be trained to endure +9 Gz and to perform complex tasks for brief periods of exposure (10–15 s).

Using anti-G protective pants, vest and positive pressure breathing, pilots can be trained to endure +9 Gz and perform complex flying tasks for minutes at a time. Human task performance tends to decrease under high G as a function of decreasing eye-level blood pressure and reduced arterial oxygen saturation, which directly affect cognition.

No gender differences have been found in male–female flying performance in the high G environment, when the subject/pilot is well protected.

High G can have profound effects on vision, orientation perception, and time and mass perception.

REFERENCES

ALBERY, W. and WOOLFORD, B., 1997, Design for macrogravity and microgravity experiments. In Salvendy, G. (ed) *Handbook of Human Factors and Ergonomics* (2nd ed.) (New York: Wiley), pp. 935–63.

ALBERY, W. and CHELETTE, T., 1998, Effect of G-suit type on cognitive function. *Aviation Space and Environmental Medicine*, 69, 474–9.

BURTON, R. and WHINNERY, J., 1996, Biodynamics: sustained acceleration. In DeHart, R. (ed) *Fundamentals of Aerospace Medicine* (2nd ed.) (Baltimore: Williams & Wilkins), pp. 201–60.

CHELETTE, T., ALBERY, W., ESKEN, R. and TRIPP, L., 1998, Female exposure to high G: performance of simulated flight after 24 hours of sleep deprivation. *Aviation Space and Environmental Medicine*, 69, 862–8.

VON GIERKE, H., MCCLOSKEY, K. and ALBERY, W., 1991, Military performance in sustained acceleration and vibration environments. In Gal, R. and Mangelsdorff, D. (eds) *Handbook of Military Psychology* (New York: Wiley), pp. 335–64.

Physiological Costs of Noise Exposure: Temporary Threshold Shifts

H. Irle, J.M. Hesse, and H. Strasser

Ergonomics Division/University of Siegen

Paul-Bonatz-Str. 9-11/57068 Siegen/Germany

1. INTRODUCTION

Annual statistical surveys of mutual industrial accident insurance associations show that noise induced hearing loss – as has always been the case – is the number one of the occupational diseases. Approximately 1/3 of all known cases originates from detrimental acoustic stress in the working environment of industrial countries. That is, however, only the tip of the iceberg of aural and extra-aural effects of noise, i.e., of acoustic stress that is unwanted, disturbing, aggravating, and frustrating; that impedes conversation; and that, in some circumstances, can cause damage to the hearing. Expenditures in the form of monetary compensation for the irreversible hearing damage associated with noise are still relatively low in comparison with the damage to the economy of more than 20 billion US $ per year, as, e.g., calculated by the German Association for Noise Control. For these reasons, "noise" is being attacked on various levels. The great variety of brochures, books, and journals as well as technical regulations, guidelines, standards, and safety regulations reflects both the broad and deep knowledge of "noise" and the technical possibilities for noise control. But it should not be forgotten that even tolerable noise with a rating level of 85 dB(A) / 8 h in the production sector is associated with essential "physiological costs" which the ear has to pay.

2. TEMPORARY AND PERMANENT THRESHOLD SHIFTS

High noise levels, at least, lead to Temporary Threshold Shifts (TTS) which are dependent on the preceding exposure and last for varying time periods. If such TTS are not subsided completely at the beginning of a subsequent noise exposure, the hearing threshold shifts can become permanent. Moreover, high noise exposures beyond a level of 120 dB inhere a high risk for acute damage in the inner ear.

Permanent Threshold Shifts (PTS) as the long-term aural effects normally do not occur after just days, weeks, or months of risky noise exposures but are the then irreversible consequences of noise exposures over years. PTS can be regarded as the accumulated sum of the temporary threshold shifts which still remain 16 h (i.e., 1000 min) after the end of an 8 h-working day. Therefore, they may be represented in the equation PTS = S TTS$_{1000}$. As TTS can be regarded as a predecessor of PTS, this measure has some prognostic value and is used in ergonomics to evaluate noise exposures with respect to "physiological costs" the ear has to pay for noise with varying levels, exposure times, frequencies, and time structure (e.g., continuous noise, impulse noise).

3. AUDIOMETRIC MEASUREMENTS TTS$_2$, T(0 DB), AND IRTTS

Figure 1 shows what can be measured audiometrically when subjects have been exposed to noise. The lower curve represents the hearing threshold of otologically normal young subjects. It means, e.g., that a minimum signal amplitude of about 4 dB is necessary to hear a 1 kHz tone. Test tones with a lower frequency have to be increased in level according to the hearing threshold to, e.g., about 10 dB at 250 Hz or 30 dB at 64 Hz, and lowered to smaller values in the area of about 4 kHz which represents the most susceptible frequency range of the human ear.

As the normal hearing threshold is the result of averaging over hearing thresholds of a representative large group of subjects, the individual hearing threshold of a subject must not be congruent with this curve but can take its course below or (as shown in Figure 1) some decibels above the normal threshold. When the physiological responses of noise exposures have to be quantified it is always necessary to determine the individual resting hearing threshold as a reference base. After an acoustic exposure, threshold shifts can be measured in a standardized procedure as represented in the upper curve. Normally within 2 minutes after the end of an exposure the maximum threshold shift has to be quantified in dB audiometrically. It is advisable to concentrate on the maximum relative hearing threshold shift above the individual resting hearing threshold (in this case at 4 kHz). This parameter, the so-called TTS$_2$, is a classic characteristic value of audiometric examinations.

Furthermore, it is advisable with this frequency of the maximum threshold shift to monitor the restitution of the hearing threshold shift (back to the resting threshold) at exactly predetermined points in time (Figure 2). This point in time t(0 dB) is also an important characteristic value of the acoustic strain analysis. When a linear time scale is used, the shape of the restitution of a temporary hearing threshold shift resembles an exponential function. If, however, it is plotted against a logarithmic time scale (see the block in the upper right corner of Figure 2), the regression function TTS(t) is a straight line.

In order to allow an overall assessment and a comprehensive
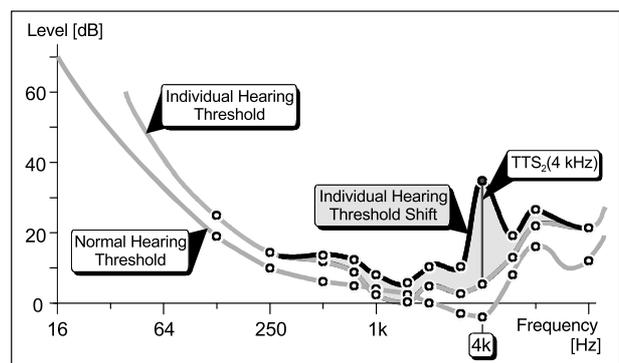
Figure 1. Hearing threshold of otologically young subjects (lower line), example of a resting hearing threshold (middle line) and threshold shift of an individual after a noise exposure (upper curve) with a maximum in the area of 4 kHz, measured within 2 min after the end of the acoustic exposure.
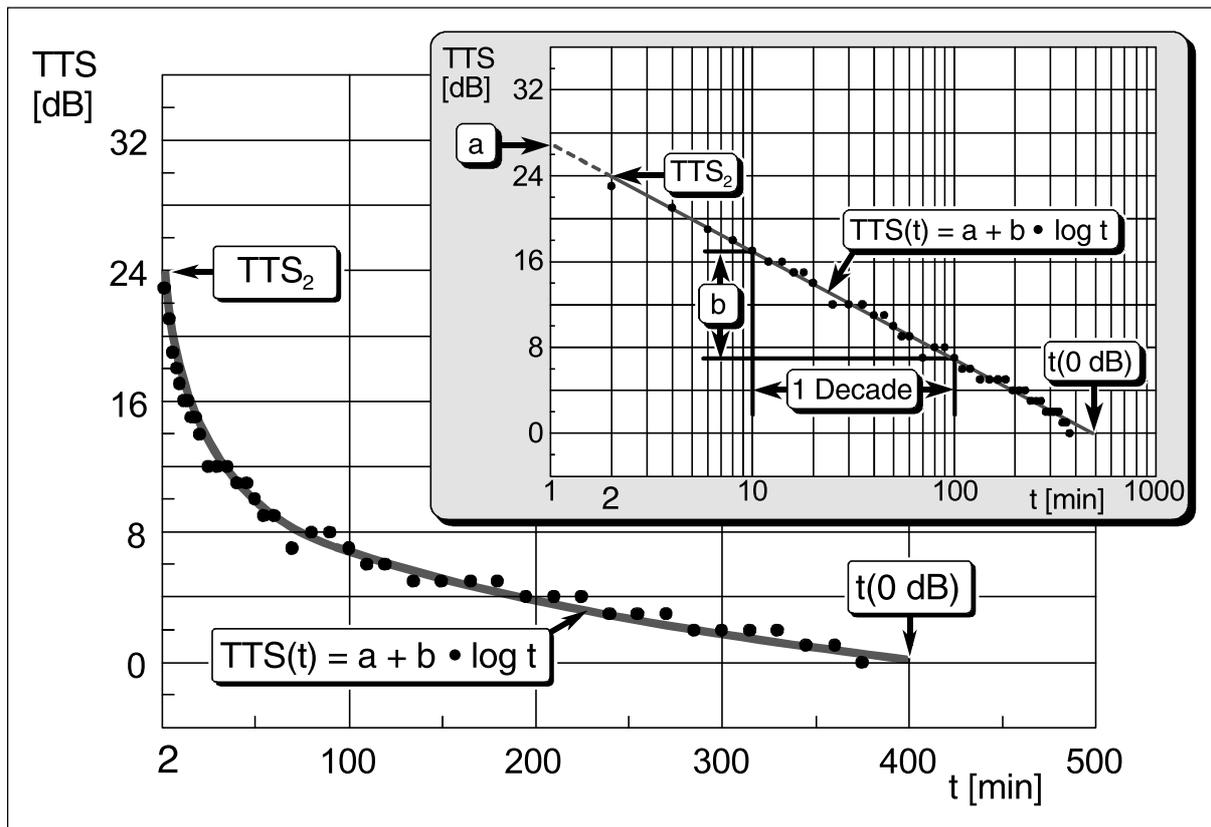
Figure 2. Example for registered TTS-values for the recovery period with a linear and a logarithmic time scale as well as representation of the regression function TTS(t).

statistical analysis, the area under the regression line has to be determined, as can be seen in Figure 3. This Integrated Restitution Temporary Threshold Shift (IRTTS) is computed as the integral of the regression function TTS(t) from 2 min after the exposure to the point t (0 dB). The IRTTS is a numeric value for the total threshold shift (in dB x min) which has to be "paid" by the hearing in physiological costs for the exposure. This global characteristic value is similar to the sum of work-related increases of heart rate which can be measured after finishing dynamic muscle work until total recuperation from cardiovascular stimulation.
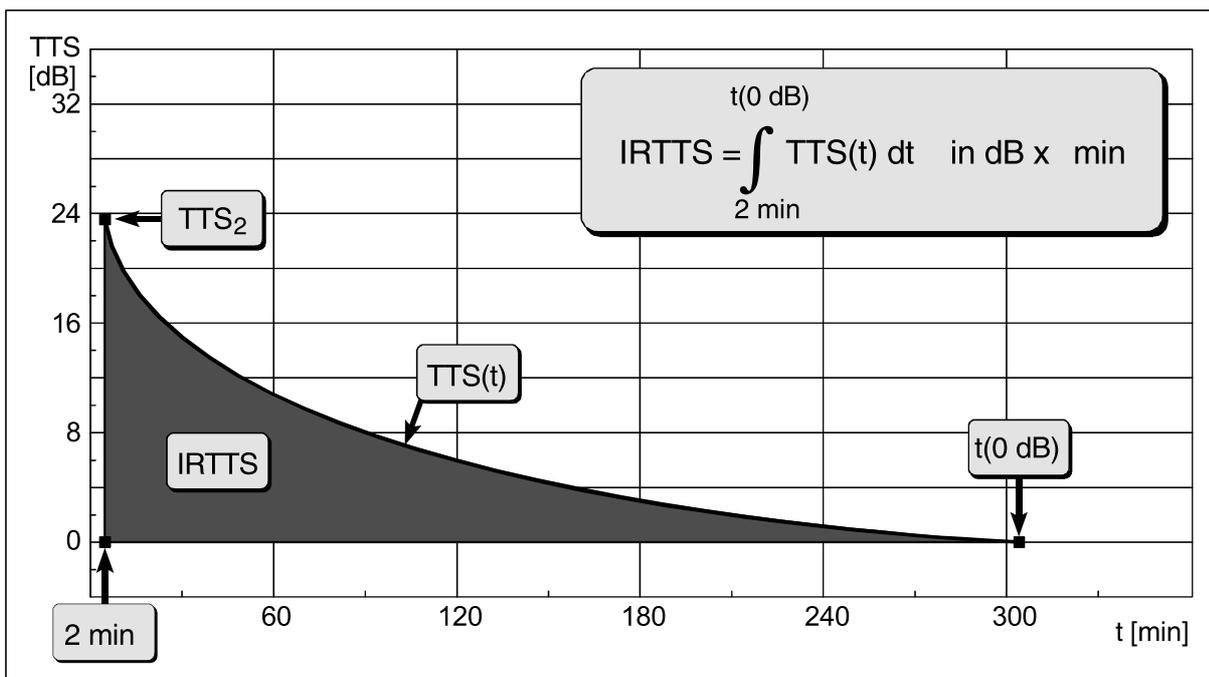
Figure 3. Exemplary representation of the IRTTS (Integrated Restitution Temporary Threshold Shift).

In the following, results of comprehensive investigations will show that essentially varying physiological responses to equally rated noise exposures do exist.

4. PHYSIOLOGICAL COSTS ASSOCIATED WITH CONTINUOUS AND IMPULSE NOISE WITH THE SAME AMOUNT OF ENERGY

During revisions of Occupational Health and Safety Guidelines in Europe and the US (N.N. 1986, N.N. 1990, N.N. 1998), at least linearly evaluated noise levels of more than 140 dB (or levels L_{AI} >130 dB(A)) have been declared as harmful to hearing. Yet, impulse noise with levels below that is not necessarily considered critical, as long as the daily noise dosage, resulting from the number of single impulses and their respective exposure times, does not exceed an energy equivalent rating level of 85 dB(A). However, from an ergonomics point of view, the equating of impulse and continuous noise (on the basis of the 3-dB-exchange rate) must be questioned (see, e.g., Strasser and Irle 2000). The energy equivalent measuring and rating procedure assumes that noise with the same energy inheres the same long-term damage. According to Figure 4 this should be valid for 85 dB(A) / 8 h and, e.g., 94 dB(A) / 1 h, as well as 113 dB(A) / 45 s and for the splitting-up of this short-time continuous exposure into energy equivalent series of impulses with the same level of 113 dB(A).

For safety purposes and with respect to ethical and moral aspects, and to absolutely insure that no harm is done to test subjects, acutely "dangerous" exposures of impulse noise with levels over 120 dB(A) never should be provided in laboratory studies, but 113 dB(A) is still considered acceptable. According to Miller (1974), similar threshold shifts can be expected from 94 dB(A) / 1 h and 85 dB(A) / 8 h. Therefore, for economical reasons, studies with 94 dB(A) over 1 h can be carried out instead of providing exposures over 8 h.

In cross-over-trial laboratory studies Ss were exposed to 94 dB(A) for 1 h, 113 dB(A) for 45 s, and impulse noise with a level of 113 dB(A) (180 x 250 ms-impulses, 450 x 100 ms-impulses, 1800 x 25 ms-impulses, and, finally, 9000 impulses for 5 ms, each. Five-ms impulses with noise levels between 110 and 120 dB are common in practice, e.g., while hammering metal plates. All impulses were separated by 3-s noise intervals so that test duration lasted up to 7 h 30 min for the 5-ms-impulse exposure. Simultaneous to the acoustic exposure of the Ss via headphones, the noise exposure was also checked with an identical pair of headphones with an artificial head measuring system (Figure 5). For the audiometric measurements which have been described in section 3, the Ss were placed in a soundproof cabin in order to reduce disturbing outside noise and guarantee the replication of environmental conditions. For details see Hesse *et al.* (1994).

As an example, Figure 6 shows the measured values of all subjects, the arithmetically averaged values, as well as the regression lines for 4 of altogether 6 noise exposures. Also, the statistically calculated data are summarized next to the real measured values in the upper right corner of each part of the figure. For continuous noise of 94 dB / 1 h (Figure 6a), e.g., the regression function TTS(t) is 28.5-12.45 · log t, the TTS_2 = 24.8
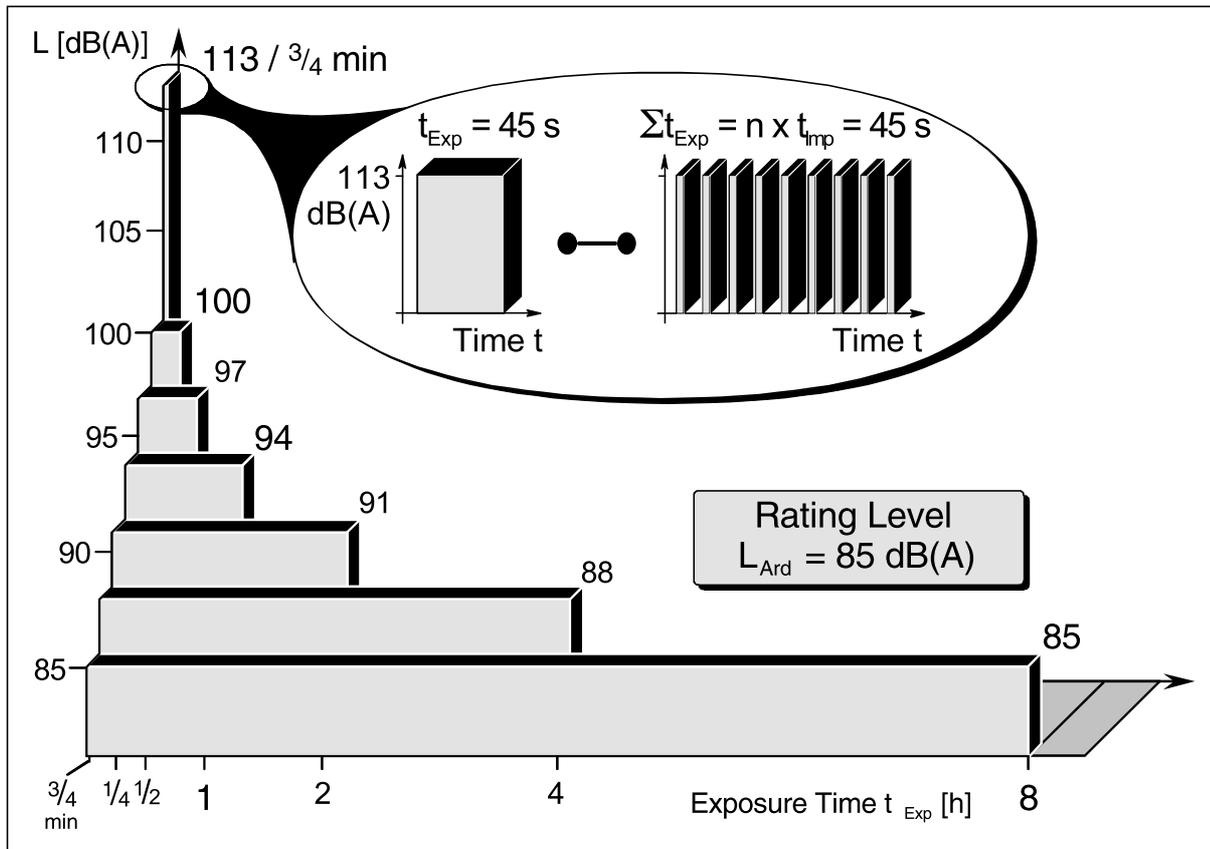
Figure 4. Varying levels and exposure times representing an energy equivalent noise exposure with a rating level L_{Ard} of 85 dB(A).
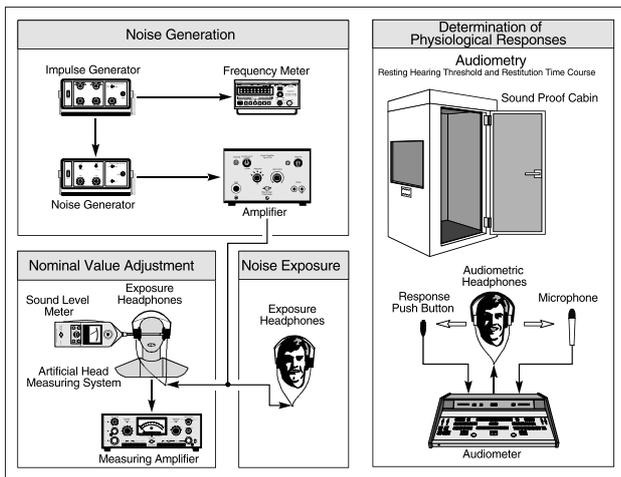
Figure 5. Schematic test set-up for noise exposure and audiometric measurements in a soundproof cabin.

dB which was calculated via this regression function, and the length of time t(0 dB) it took to reach the resting hearing threshold is 195 minutes. The physiological responses which vary from subject to subject can be seen in the real measured values of the TTS_2 – with an interindividual range from 13 to 32 dB – and in the time required to reach the resting hearing threshold t(0 dB) – with a range from 70 to 210 min – thus explaining the statistical correlation value (r^2) of only 0.75. If the arithmetically averaged values and their course along the regression line are examined, the approximation is clearly excellent, which can also be seen in the statistical correlation value (r^2) of 0.99.

Figure 6b shows the results of the short-time continuous noise exposure of 113 dB for 45 seconds. It can be seen that an increase of the level of continuous noise together with a corresponding shortening of exposure time leads to a weakening of physiological responses. The TTS_2 values are essentially lower and the restitution times are substantially shorter.

Figure 6c shows the restitution after the exposure to 180 impulses of 250 ms, each, and shows that the physiological responses already have the tendency to be higher than those resulting from continuous noise at the same level. Both TTS_2 – with values around 15 dB – and the restitution time – with values up to 100 min – are increased. This tendency becomes even more pronounced with 100-ms impulses, whereby the threshold shifts reach values of more than 16 dB and last on average up to more than 300 min (not graphically shown here). When the length of the impulses is reduced to 25 ms, the physiological cost again grows substantially.

Finally, 9000 5-ms impulses result in an essential increase of physiological costs. As can be seen in Figure 6d, the TTS_2 exceeds 20 dB and the restitution time is up to 600 min, i.e., 10 h.

Figure 7 shows a simplified plotting of all restitution time courses TTS(t) which reveals a substantial increase of physiological costs associated with the splitting-up of continuous noise (113 dB(A) / 1 x 45 s) into shorter and shorter impulses. If the damage-risk based on the energy equivalence principle – with its roughly comparable consequences of strain – is accepted, then the position of the regression lines for all test conditions would have to be at least similar. This, however, not at all can be confirmed by the results of the investigations.
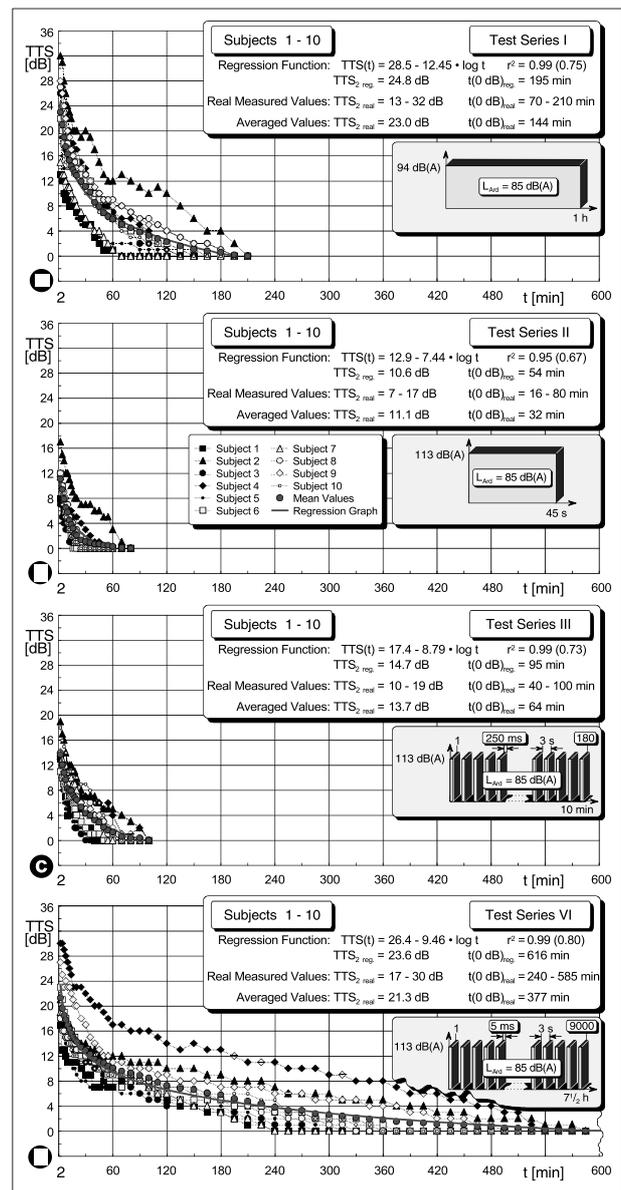
Figure 6. Individual and mean restitution time course for all 10 Ss after the exposure to 94 dB / 1 h (a), 113 dB / 1 x 45 s (b), 113 dB / 180 x 250 ms (c), and 113 dB / 9000 x 5 ms (d). All exposures contain the same sound energy.
(**Source: Hesse et al. 1994**)

The overall physiological costs in the form of IRTTS also differ greatly with varying exposures. The ratio of IRTTS resulting from the two extremes "113 dB(A) for 45 s" (147 dBmin in TS II) and the energy equivalent exposure to 5-ms impulses (2473 dBmin in TS VI) amounts to 1:16. According to the results of significance tests, the consistent increases of physiological costs are in a strong causal-deterministic relation with the shortening of the length of the impulses.

In order to quantify these results with regard to the damage-risk based on the energy equivalence principle, the IRTTS values of all five test series with 113 dB were standardized to the total physiological costs resulting from the exposure to 94 dB for 1 hour (TS I) (Figure 8).
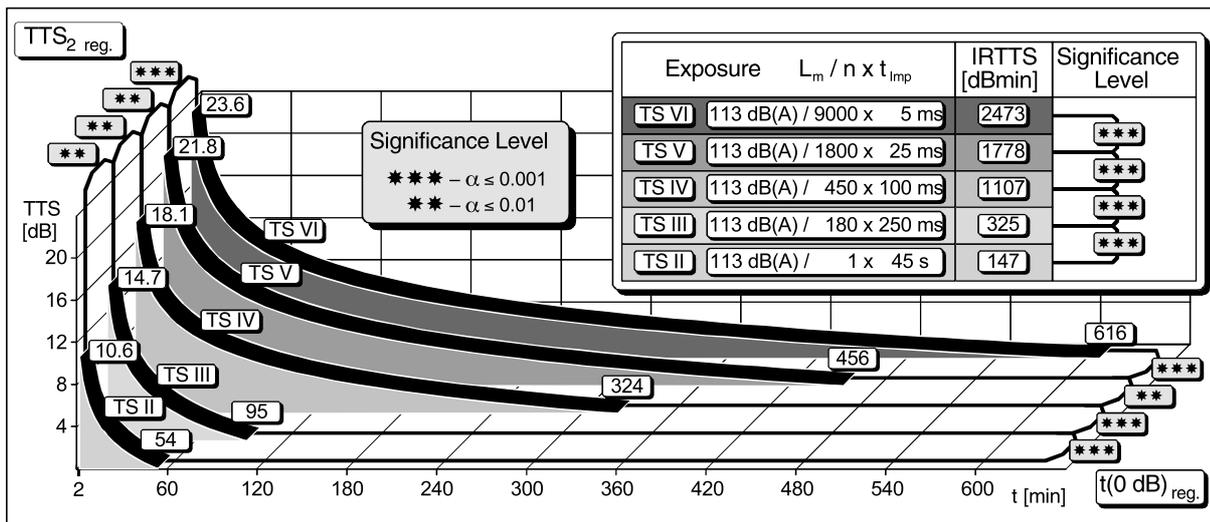
Figure 7. Restitution time course TTS(t) with characteristics TTS$_{2\ reg.}$, t(0 dB)$_{reg.}$, and physiological cost IRTTS as well as symbolic labeling of significance levels for the differences between the responses to the exposures with equal exposure level.

L$_m$ = 113 dB(A) and exposure time t$_{Exp}$ (n x t$_{Imp}$) (according to the one-tailed WILCOXON test) (Source: Strasser *et al.* 1995)

Values of this ratio as a "risk factor" > 1 thus stand for an increased danger whereas values < 1 mean a lower risk for hearing. Compared to impulse noise based on IRTTS of 94 dB/1 h, the risk with 5-ms impulses is 2.5 times higher. Compared to continuous noise of 113 dB / 45 s, the risk with 5-ms impulses is even 16 times more dangerous. The results of prior studies, which were carried out with other subjects and with continuous noise of 100 dB / 15 min, 110 dB / 1.5 min, and 113 dB / 45 s, also energy equivalent to 94 dB / 1 h or 85 dB / 8 h, confirm the high reliability of the results of the studies; e.g., the IRTTS values for 94 dB(A) / 1 h are almost identical.
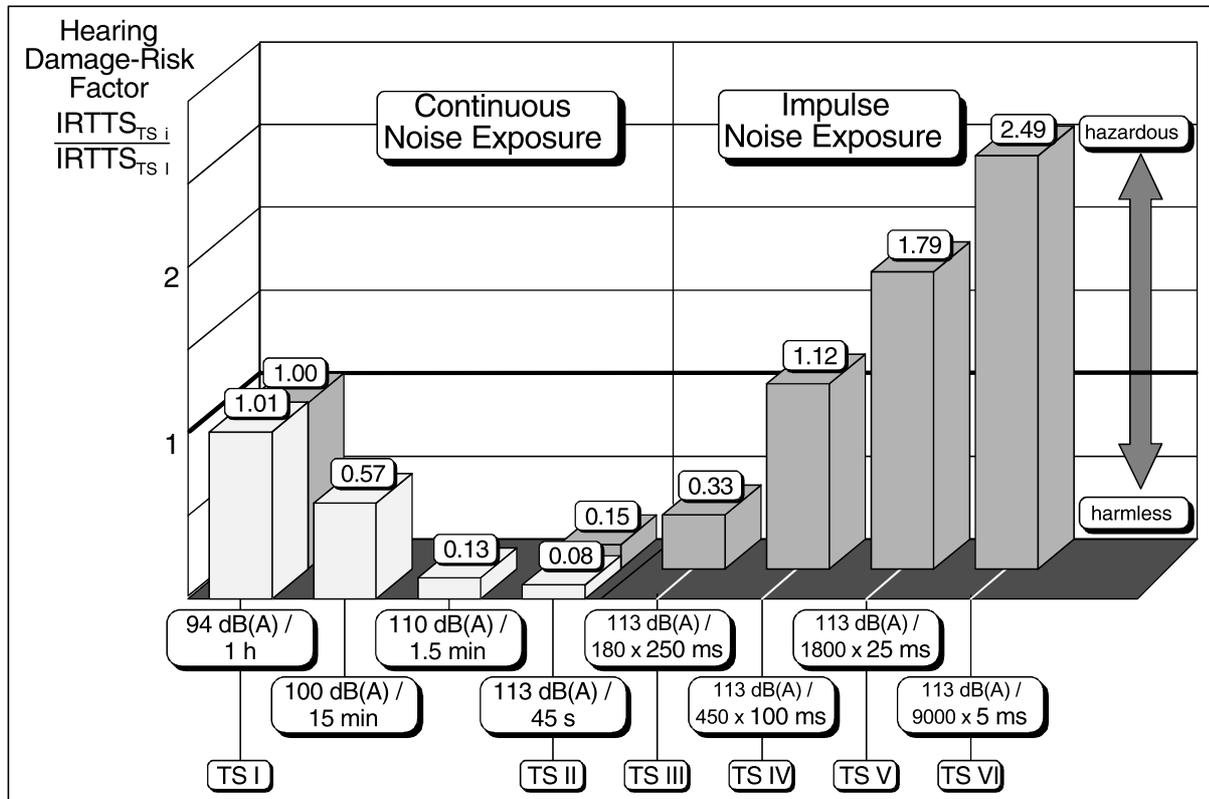
Figure 8. Hearing damage-risk criterion of energy equivalent continuous and impulse noise exposures determined via the comparison of IRTTS values with IRTTS (94 dB(A) / 1 h) = 1.00 (Source: Hesse *et al.* 1994).

5. PHYSIOLOGICAL COSTS OF ENERGETICALLY NEGLIGIBLE CONTINUOUS AND IMPULSE NOISE

Noise exposures – even if they do not exceed the rating level of 85 dB(A) – are often annoying, disturbing, and performance-reducing, and, furthermore, they can cause considerable hearing threshold shifts. As a result of a continuous noise of 94 dB(A) for 1 hour – energy equivalent to 85 dB(A) / 8 h – hearing threshold shifts (TTS$_2$) of approximately 20 to 25 dB must be expected immediately after the exposure as shown above. Usually, the restitution time for these threshold shifts is at least 2 hours. The sum of the temporary threshold shifts which can be audiometrically monitored represents a measure of the "physiological costs" of noise for which the organism – in addition to the psychological effects – must "pay." Such hearing threshold shifts, if and when recovery is completed before another acoustic exposure occurs, still may represent a "normal" fatigue of hearing. Yet, if restitution time is greatly prolonged, a daily exposure to noise can lead to a Noise Induced Permanent Threshold Shift (NIPTS). The same may occur if recovery from a temporary threshold shift as a response to continuous noise cannot take place in noisy surroundings. That happens in situations where, e. g., a continuous noise exposure with a limited duration is followed by a lower noise of longer duration. Furthermore, often impulse or continuous noise does not occur exclusively; rather, it is a combination of both.

If low level continuous noise or impulse noise is added to an already existing high level continuous noise, numerically no essential increase in the rating level results. Yet, it cannot be expected that the aural strain of these exposures is always negligible. Therefore, according to Figure 9, in cross-over test series, subjects were exposed to noise of 94 dB(A) for 1 hour (TS I), energy equivalent to an 8 h-rating level L_{Ard} of 85 dB(A). In a second test series (TS II), the same exposure was combined with 900 energetically negligible 5-ms impulses with a level of 113 dB(A) which increased the rating level by only 0.4 dB. Whereas the noise exposure of TS I and TS II was followed by an idealized resting phase in a soundproof cabin, in a third test series (TS III) the continuous noise of 94 dB(A) / 1 hour was followed by 3 hours of white noise at 70 dB(A). Such an additional load increases the L_{Ard} by merely 0.1 dB to 85.1 dB(A). In all three test series the noise-induced temporary threshold shift (TTS$_2$) and its restitution (see upper part of Figure 9) were measured.

As shown in Figure 10, the continuous noise exposure of 94 dB(A) for 1 hour was associated with a TTS$_2$ of somewhat more than 20 dB which disappeared completely after about two hours. The additional impulse noise caused only a small increase in the TTS$_2$. Yet, more importantly, the restitution time increased from about 2 h to 3 h. The TTS$_2$-values of TS III did not differ from those resulting from TS I. That was expected as the conditions up to that point in time were identical. But due to the additional subsequent energetically irrelevant exposure, the mean restitution time increased considerably from 126 min up to 234 min (about 4 h). The mean total physiological costs represented by the Integrated Restitution Temporary Threshold Shift (IRTTS) increased in TS II by approximately 40 % (902 versus 660 dBmin) and in TS III even by 140 % (1613 versus 660 dBmin).

6. GENERAL CONCLUSIONS

The exemplary results presented above have shown that energy equivalent impulse noise exposures that differ in their time structure lead to quite different physiological costs. Whereas the difference in aural strain from impulse noise compared to continuous noise does show fewer substantial effects in the temporary threshold shift two minutes after the end of the exposure (TTS$_2$), the effects are distinctly pronounced in the different restitution times, i.e., the strain compensation. Taking
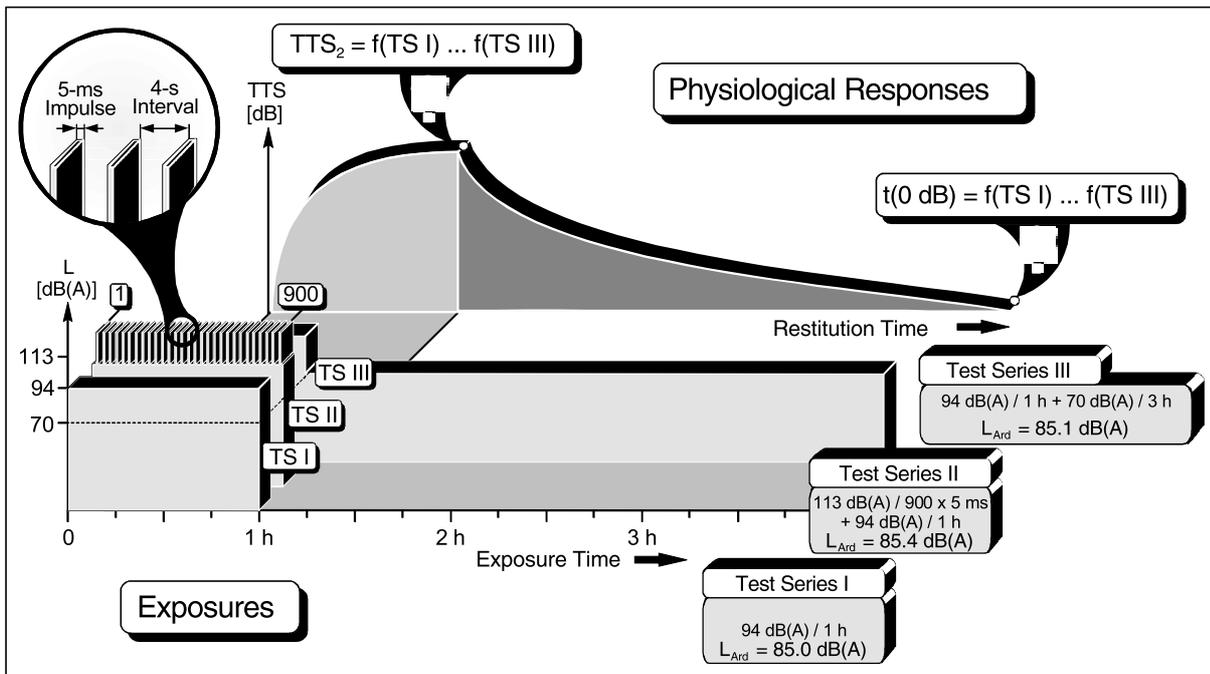
Figure 9. Hypothetical restitution time course of the temporary threshold shift (TTS) associated with three noise exposures with nearly equal rating level L_{Ard} per day.
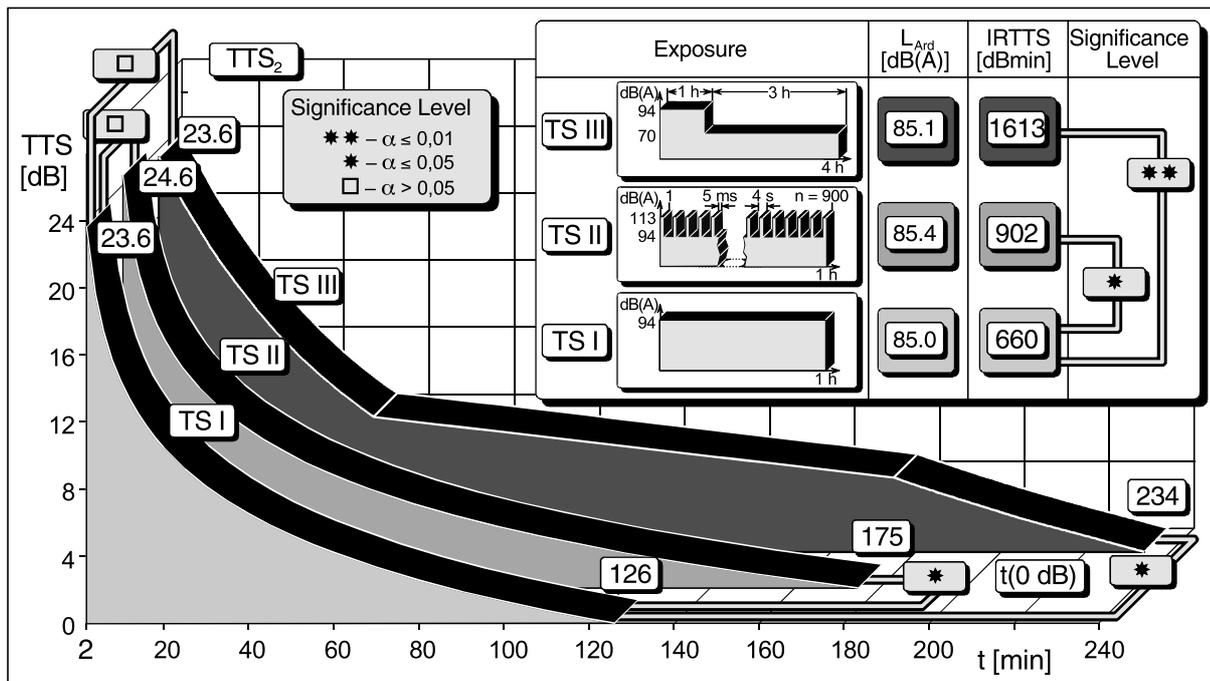
Figure 10. Restitution time courses TTS(t) and IRTTS (in dBmin) as physiological cost associated with preceding noise exposures (94 dB(A) / 1 h in TS I, 94 dB(A) / 1 h + 9000 x 113 dB(A) / 5 ms in TS II, and 94 dB(A) / 1 h + succeeding 70 dB(A) / 3 h in TS III). All test series inhered the same noise energy of legally tolerated 85 dB(A) / 8 h (source: Irle *et al.* 1998).

this into consideration, the IRTTS as an integral parameter proved to be an indicator for potentially long-term hearing-physiological damage more valid and reliable than commonly used TTS_2. Because reversible auditory fatigue is always of certain prognostic value for irreversible hearing impairments, the results disprove the validity of the damage-risk based on the energy equivalence hypothesis. Especially in the areas close to the lower and upper end of the exposure continuum which is still legally allowed, it must be explicitly stated again that hearing cannot be energy equivalent.

The results reported also show that levels of noise which have no influence on the rating level which traditionally is calculated according to the energy equivalence principle are often of great importance, as they can lead to considerably prolonged restitution times. Therefore, the purely energy equivalent determination of the rating level of both impulse noise and low sound levels can lead to an underestimation of latent problems so that over time a reversible TTS can evolve into a permanent threshold shift. This is also of importance for the acoustic design of break rooms for noise-exposed workers. Conditions which allow an undisturbed restitution of hearing should always be present.

REFERENCES

HESSE, J.M., IRLE, H. und STRASSER, H., 1994, Laborexperimentelle Untersuchungen zur Gehörschädlichkeit von Impulsschall. *Z.Arb.wiss.* 48 (20 NF) 4, 237-244

IRLE, H., HESSE, J.M., and STRASSER, H., 1998, Physiological cost of energy-equivalent noise exposures with a rating level of 85 dB(A): Hearing threshold shifts associated with energetically negligible

continuous and impulse noise. *Int. Journal of Industrial Ergonomics* 21, 451-463

MILLER, J.D., 1974, Effects of noise on people. *J. Acoust. Soc. Am.* 56 (3) 729-764

N.N., 1986, *Directive on the protection of workers from the risk related to noise at work* (86/188/EEC). Richtlinie des Rates vom 12. Mai 1986 über den Schutz der Arbeitnehmer gegen Gefährdung durch Lärm am Arbeitsplatz (86/188/EWG). *Amtsblatt der Europäischen Gemeinschaft* Nr. L, 28-34

N.N., 1990, *Accident Prevention Regulation-Noise*, UVV-Lärm, Unfallverhütungsvorschrift der gewerblichen Berufsgenossenschaften (VBG 121), C. Heymanns Verlag, Köln

N.N., 1998, *Criteria for a recommended standard – Occupational noise exposure*, revised criteria 1998. DHHS (NIOSH) Publication No. 98-126. U.S. Department of Health and Human Services, Cincinnati, OH

STRASSER, H., HESSE, J.M., and IRLE, H., 1995, Hearing threshold shift after energy equivalent exposure to impulse and continuous noise. In: Bittner, A.C. and Champney, P.C. (Eds.): *Advances in Industrial Ergonomics and Safety VII*, Taylor & Francis, London/New York/ Philadelphia, 241-248

STRASSER, H. and IRLE, H., 2000, Noise: Measuring, evaluation and rating from an ergonomics point of view. In: KARWOWSKI, W. (Ed.). *International Encyclopedia of Ergonomics and Human Factors*. Taylor & Francis, London, New York

Tolerance to Sustained +G_z Acceleration

R. R. Burton

Human Effectiveness Directorate, Air Force Research Laboratory, Brooks AFB, TX 78235–5118, USA

1. INTRODUCTION

The environments produced by accelerating a human body are grouped into three main categories: (1) transient, (2) sustained and (3) chronic. Acceleration takes place when the velocity of a body changes either in magnitude or direction. Changes in acceleration that last 1 or 2 s and do not achieve steady-state are termed transient, abrupt or impact. Sustained acceleration is usually a steady-state environment with short periods of changing accelerations at the beginning (onset) and ending (offset) of the acceleration environment. The duration of this exposure is usually £ 1 h. Chronic acceleration is a steady-state exposure that lasts for several hours, days or even years. Not surprisingly with these types of definitions that rely on the duration of exposure, there are not clear distinctions where transient acceleration ends and sustained acceleration begins. The same can be said about sustained and chronic accelerations. Because of the very short duration of transient acceleration, the environment is developed using sleds or drop towers. This article deals with sustained acceleration that uses the centrifuge to provide the acceleration as does chronic acceleration studies.

Reasons for studying these types of accelerations are to develop information on human effects as they primarily relate to the following activities: (1) transient for impact injuries, (2) sustained for pilot tolerance during fighter aircraft maneuvers and (3) chronic for long-term space flight.

For more information on transient acceleration, see Brinkley and Raddin (1996) and for sustained and chronic acceleration, see Burton and Whinnery (1996) and Burton and Smith (1996).

2. PHYSICAL PRINCIPLES

The human, in the sustained acceleration environment, physiologically responds to the inertial force generated by the acceleration. The forces of acceleration (a) are described by Isaac Newton's three laws of motion. The first law states that a body remains at rest or in motion in a straight line until acted upon by a force. The second law of motion defines the force (F) required to overcome the first law, i.e. F = mass (m) x acceleration (a). The third law states that for every action there is an equal reaction, thus establishing the physical principles of acceleration in circular motion as generated by a centrifuge, i.e. centripetal (acceleration) force (a) establishes an equal centrifugal (inertial) force (G). Thus, a centrifuge or an inside loop of an aircraft produces the same acceleration with an equal inertial force acting in the opposite direction.

Equation 1 predicts the centripetal acceleration resulting from the motion involved:

$$a = r v^2, \tag{1}$$

where a is centripetal acceleration (m s^{-2}), v^2 is radians s^{-1} and r is the radius of the circle (m).

Radian degree tables are available but for simple calculations, 360∞ = 6.28 radians. The product of equation 1 is "a" (m s^{-2}) that for convenience is divided by Earth's gravitational constant (9.81 m s^{-2}) providing a value that is in multiples of Earth's gravity (1 g). Thus, "g" is frequently used in place of "a." Hence, $g = G$, the inertial force (G) resulting from the acceleration (g) defined in equation 1 as "a" is also measured and is equal.

"G" resulting from "g" is the parameter of importance in defining human tolerance to acceleration levels (intensities). The physical and physiological effects on the body come from G. Of particular importance for the human tolerating increasing levels of sustained G, is the distribution of blood within the body.

Fluid pressures within an acceleration field obey the hydrostatic pressure law (equation 2):

$$P_H = h \, d \, g, \tag{2}$$

where P_H is hydrostatic pressure, h is column height, d is specific density of the liquid and g is the level of acceleration.

Since $g = G$, equation 1 predicts the effect of G on arterial pressure (P_a) and blood flow within the body. Hence, it has been adapted for use as a human G-level tolerance prediction model (equation 3; Burton and Whinnery 1996):

$$G_{LT} = (P_a \sum d)/h, \tag{3}$$

where G_{LT} is G-level tolerance of a person seated upright, h is eye-heart vertical distance (350 mm), P_a is mean arterial blood pressure at eye level is 100 mmHg and d is specific density of mercury (13.6) used in blood pressure measurements.

3. AXIAL SYSTEM NOMENCLATURE

Since equation 3 predicts the G level that a human can tolerate, it becomes clear that "h" (eye–heart *vertical* distance) plays a major role in G-level tolerance determinations. Thus, the orientation of the body in the acceleration environment must be identified when defining G-level tolerances. Thus, a standard axial system of nomenclature has been adopted to identify this relationship for all acceleration-related studies (table 1).

3.1. Relaxed G Tolerance

There are two definitions of human tolerance to sustained acceleration. First there is the level of acceleration that can be tolerated with a cardiovascular basis, using visual criteria, known as G-*level tolerance*. The other tolerance is the duration of acceleration that can be tolerated with a metabolic basis, using symptoms of fatigue G-*duration tolerance*. Obviously both of these tolerances have common dimensions of G level and duration, i.e. the difference is in the dimension criteria that determine the tolerance.

3.2. G-level Tolerance

Natural, control or basic human G-level tolerance (+G_z) is measured on relaxed subjects on a centrifuge with G onset rates of ≥ 1 G s^{-1} (Burton *et al.* 1974). This tolerance measurement is called "rapid onset relaxed" (ROR). The subject is seated upright with "g" directed towards the head and, of course, corresponding G directed towards the feet. The G level is increased until a

predetermined tolerance criterion is reached. Several criteria exist: (1) greyout, (2) blackout or (3) loss of consciousness. The first two criteria are visual, recognized by the subject who terminates the centrifuge exposure by activating a hand brake. Greyout (also known as peripheral light loss or tunnel vision) is a dimming of the vision while blackout (central light loss) is total loss of vision. Greyout visual symptoms occur because blood flow in peripheral regions in the retina of the eye ceases as systolic P_a at eye-level falls from a normal level of 125 mmHg to ~40 mmHg. Blackout occurs with eye-level P_a = 20 mmHg. With blackout, subjects are still conscious, thus they can hear and are aware. This visual symptom is possible because intra-ocular pressure is 20 mmHg that causes the retinal blood vessels to collapse with ~20 mmHg P_a in the head. Recovery from these visual symptoms is immediate upon reestablishing blood flow to the eye.

A subject loses conscious, known as *G*-induced loss of consciousness (G-LOC), when blood flow to the brain ceases. Recovery from this condition is not immediate as with visual symptoms, but requires 15–30 s. Thus, G-LOC is an extremely dangerous condition as the aircraft remains totally without control while the pilot is recovering consciousness. *G*-level tolerance can be predicted solving equation 3 and is ~4 *G*, i.e. each *G* requires ~25 mmHg P_a at heart level (Burton and Whinnery 1996).

If the subject is not physically relaxed, *G* tolerance will increase, e.g. simply gripping one hand will increase it by 0.6 *G*. Contraction of arm, leg and abdominal muscles can increase *G* tolerance by several *G* thus forming the basis for a maneuver that is used by pilots to increase their *G* tolerance during high *G* exposures (see below).

The use of slow-onset of *G* rates (i.e. 0.1 *G* s^{-1}), known as gradual onset relaxed (GOR) exposures, defines a *G* tolerance that is ~1 *G* higher than ROR tolerances. GOR tolerance is increased above ROR because of the activation of a physiologic reflex (baroreceptor initiated) that increases heart rate, causes arterial vessel constriction thus raising P_a. This reflex is more pronounced in fighter pilots than subjects not regularly exposed to *G* suggesting that a type of environmental adaptation has occurred.

3.3. G-duration Tolerance

G-duration tolerances for relaxed subjects are measured using a constant *G* level that continues until the subject becomes fatigued and ends the exposure by applying a hand brake that stops the centrifuge. *G*-duration tolerances for *G* levels (£ 5 *G*) are predicted using:

$$G_D = 2666\ e^{-1.265G}, \qquad (4)$$

where G_D is *G*-duration tolerance (min) and *G* is *G*-level exposures ≤ 5 *G*.

4. METHODS TO INCREASE G TOLERANCE

Sustained *G* levels above those naturally tolerated by a human, is an environment that is studied because it is generated by fighter planes during maneuvers that are used in air combat, i.e. aerial combat maneuvers (ACM). The ACM is a very dynamic activity as pilots rapidly change their aircraft positions using inside loops as they attack the enemy or for defensive purposes. Therefore, fighter aircraft are designed to develop very high *G* levels (i.e. 9

G) because highly maneuverable aircraft, that can make tighter turns or make them more rapidly, have a great advantage over their adversary during ACM.

Thus, pilots are exposed to *G* levels far in excess of their natural *G*-level tolerances of ~4 *G*. To allow pilots to tolerate these high *G* levels for considerable periods without having G-LOC, methods have been developed that provide protection against *G*. These methods are called "anti-*G*" or "*G* protection". However, it is important to understand that they do not truly "protect" a pilot against the effects of *G*. Anti-*G* methods do not block the effects of *G*, but only increase pilots' *G* tolerances thus exposing them to higher *G* levels and increasing the potential hazards of *G* exposures.

4.1. Anti-G Suit

Various types of anti-*G* suits were developed in World War II by the USA, Canada and Australia. The US-design anti-*G* suit with a few modifications is used today by fighter pilots from all nations that have high-performance aircraft. This garment is made of fire-retardant material that covers pneumatic bladders that apply pressure to the legs and supports the abdomen (Burton and Whinnery 1996). This suit is inflated with pressurized air, controlled by an anti-*G* valve at 1.5 psi/*G* to a maximum of 10 psi. This suit increases *G* tolerance by 1.0–1.5 *G* by increasing P_a and elevating the position of the heart during *G* – the latter reducing "h" of equation 3.

4.2. Anti-G Straining Maneuver

Also during World War II, a physical maneuver, called the M-1, was perfected in conjunction with the anti-*G* suit. M-1 uses the combination of muscle contractions and a forced exhalation against a closed or partially closed glottis (Wood *et al.* 1946). This maneuver can increase pressure within the lungs by as much as 100 mmHg resulting in a similar increase in P_a. This method is now called an Anti-*G* Straining Maneuver (AGSM) that with the increase of P_a at eye-level of 100 mmHg increases *G*-level tolerance by 4 *G* (equation 3). By combining the AGSM with the anti-*G* suit, pilots can tolerate 9 *G*, to match the acceleration-level capability of their aircraft (Burton and Whinnery 1996). On the other hand, performing the AGSM is extremely fatiguing to the pilots thus limiting their *G*-duration tolerance capability.

Obviously this maneuver is very important to pilots as it forms the basis for 50% of their *G* tolerance. Since it must be learned, training programs have been developed that teach pilots, on a centrifuge, how to perform an effective AGSM.

Since muscle contractions form the basis for the AGSM, physical conditioning programs have been developed for pilots to increase their muscle strength, thus performing this maneuver with less effort and slowing the onset of fatigue.

4.3. Positive Pressure Breathing During G (PBG)

Positive pressure breathing (PPB) systems were first developed to protect pilots flying in low-pressure environments from hypoxia. This system forces air into the lungs above the ambient environmental pressure thus increasing oxygen levels in the arterial blood. To assist in breathing against this increased pressure in the lungs, a counter-pressure is applied outside the chest at the same pressure level, using a type of inflatable vest. This

PPB system has been developed recently as a *G*-protection method known as positive pressure breathing during *G* (PBG). Pressures as high as 60 mmHg are used that require little physical effort by the pilot. Since PBG increases pressure in the lung like the AGSM and with much-less effort, pilots fatigue less rapidly thus increasing their *G*-duration tolerance. An operational version of PBG is now commonly used internationally by pilots flying high-performance aircraft (Burton and Whinnery 1996).

4.4. Reclining or Pronation of Pilots

Subjects using seats that have reclined backs are placed in a supine position and exposed to +*G*$_x$. Subjects can also be pronated resulting in –*G*$_x$ exposures (table 1). These positions reduce "h" of equation 3 thus significantly increasing *G* tolerances. A *G*$_x$ level tolerance as high as 15 *G* has been realized. Although this method for increasing *G*-level tolerance is very effective, cockpit design complexities have prevented its use in fighter aircraft.

5. HIGH SUSTAINED G (HSG) TOLERANCE

Anti-*G* methods described above have increased *G* level and duration tolerances far above normal relaxed levels. This environment is known as high-sustained *G* (HSG; Burton *et al*. 1974). Advanced anti-*G* methods allow > 9 *G* to be tolerated by pilots thus providing them more than adequate protection for flying 9 *G* fighter aircraft. These high-*G*-level tolerances are measured using the same visual and G-LOC criteria that determine relaxed *G*-level tolerances (see above).

G-duration HSG tolerances are limited because of pilot fatigue that results from performing the AGSM. These tolerances are measured using continuous *G* exposures that can be at a constant *G* level or can be variable. Variable *G* profiles are designed to simulate ACM that are used by pilots in combat and training thus the name Simulated ACM (SACM; Burton and Shaffstall 1980). The SACM is most commonly used to measure HSG duration tolerances. *G*-duration tolerances for a continuous and constant *G* profile are predicted using:

$$G_D = 14.08 \, e^{-0.328G}, \qquad\qquad (5)$$

where *G*$_D$ is as in equation 4 and *G* is *G*-level exposures > 5–9 *G*.

This equation can be used also to approximate SACM tolerances by increasing the predicted tolerance by 20–30% (Burton 1998).

6. PERFORMANCE WITH G EXPOSURE

Pilot must have excellent awareness during ACM, known as Situation Awareness (SA). Good SA is a very high priority item for fighter pilots. Without a doubt their most critical SA is during ACM. To preserve their SA, pilots must maintain good vision during *G* exposures. The occurrence of tunnel vision (greyout) or blackout with *G* exposures will greatly reduce their performance thus anti-*G* methods are very important for pilots' SA. Of course, G-LOC with loss of awareness eliminates SA and all performance capability for at least 30 s.

Subject performance during exposures to increased *G* has been measured using subjects on the centrifuge. Unfortunately values obtained from these studies are difficult to evaluate relative to pilot performance during ACM (Burton *et al*. 1974). These studies generally show an exponential decay in motor function (i.e. target tracking) beginning at 3 *G* with increasing reductions in performance to 60–70% at 7 *G*. These levels of performance reduction during HSG appear to be unrealistically high for pilots that train to achieve near perfection in their performance during ACM.

The major problem in measuring performance relevant to the fighter pilot is the complexity of the visual and cognitive requirements of the ACM while interacting with *G*. An operational ACM can not be accurately duplicated on the centrifuge. Perhaps in the future improved visual displays and the use of Virtual Reality may provide the means to measure pilot performance more accurately in the laboratory.

REFERENCES

BRINKLEY, J.W. and RADDIN, J.H., JR, 1996, Biodynamics: transient acceleration. In DeHart, R. L. (ed) *Fundamentals in Aerospace Medicine* (2nd ed) (Baltimore: Williams & Wilkins), pp. 163–201.

BURTON, R.R. and WHINNERY, J.E., 1996, Biodynamics: sustained acceleration. In DeHart, R. L. (ed) *Fundamentals in Aerospace Medicine* (2nd ed) (Baltimore: Williams & Wilkins), pp. 201–60.

BURTON, R.R., LEVERETT, S.D. JR and MICHAELSON, E.D., 1974, Man at high sustained +*G*$_z$ acceleration: a review. *Aerospace Medicine*, 45, 1115–36.

BURTON, R.R. and SMITH, A.H., Adaptation to acceleration environments. In Fregly, M. J. and Blatteis, C. M. (eds) *Handbook of Physiology, Section 4: Environmental Physiology*, vol. II (New York: Oxford University Press), pp. 943–70.

BURTON, R.R. and SHAFFSTALL, R.M., 1980, Human tolerance to aerial combat maneuvers. *Aviation Space and Environmental Medicine*, 51, 641–8.

BURTON, R.R., 1998, Mathematical models for predicting human tolerance to sustained acceleration. Presented at the Research and Technology Agency Conference on "Computer Models for Aircrew Safety Assessment: Uses, Limitations and Requirements," 26–28 October, Wright-Patterson Air Force Base, Ohio.

WOOD, E.H., LAMBERT, E.H., BALDES, E.J. and CODE, C.F., 1946, Effects of acceleration in relation to aviation. *Federation Proceedings*, 5, 327–44.

Toxicology

S. Konz

Department of Industrial and Manufacturing Systems Engineering, Kansas State University, Manhattan, KS 66506, USA

1. TOXINS

Toxicology deals with the long-term effects of foreign chemicals upon the body – it considers health. A toxicology problem might be the effect of 10 years' exposure to a solvent; industrial hygienists are the technical specialists for toxicology problems.

Chemicals *affect* the *body* with *doses* producing a *response* (Konz 1995: chapter 24). For their "effect," the threshold limit values (TLV) discussed below usually use non-reversible functional changes in an organ (usually the kidney or liver). These effects are more like "corrosion" than "short circuits"; that is, they take place gradually over a long period.

Initially, the "body" is an animal body, not a human body. Animals are given various doses and results are extrapolated to humans (using a safety factor). Of course, there can be questions whether the appropriate animal was used and whether the safety factor is sufficient.

TLV need to consider dose–response. Part of the problem is that the body has the ability to transform toxins to less toxic forms and eliminate them from the body. This transformation ability varies greatly with people as does their individual susceptibility. TLV assume a 70 kg adult.

2. TOXIN ROUTES

From a toxicology viewpoint, a toxin must enter the blood. Three potential entrance points are the skin, mouth and lungs.

The skin is not permeable to *most* compounds. Unfortunately, there are a few that do penetrate the skin (they have "skin" next to them in TLV tables). A potential problem is entry through breaks in the skin (cuts, abrasions). The mouth (eating, drinking) is another potential entry route. But the lungs are the primary source of toxin entry to the blood. Most retained particles are < 5 mm in diameter; larger particles tend to be filtered out in the nose and upper airways.

3. TOXIN TARGETS

Targets are divided into exterior (skin) and interior (organs).

The skin is the largest organ system of the body; toxicology problems of the skin are called dermatoses or dermatitis. Dermatitis is divided into atopic and contact dermatitis. In allergic contact dermatitis, the person takes 14–21 days for an allergy to develop following initial contact; then, however, the person is sensitized and the response occurs very quickly. One can even become sensitized by various medications (e.g. benzocaine, neomycin, merthiolate and some topical antihistamine creams).

3.1. What to Do?

Protective clothing helps – especially aprons and gloves. Naturally the material should not be permeable to the chemical it is defending against. Also use splashguards on machines. Personal cleanliness is important; be sure employees have the proper workplace facilities to keep clean. Avoid solvents and strong soaps;

add skin moisturizers after washing the hands. Work clothing washed at home may transfer the toxin to the worker's family.

The first interior target is the lungs/respiratory system. Problems range from an increase in airflow resistance (asthmatic response) to chronic interstitial lung disease. The lungs have a defense system (macrophages: "big eaters") that disposes of the particles that settle in the lungs. Inert particles, however, present problems to the macrophages – killing them and forming scar tissue (and thus lack of active lung surface area) or causing cancer.

If the toxin passes into the blood, it attacks other organs. Examples are lead (nervous tissue), cadmium (kidneys) and carbon monoxide (hemoglobin). An unborn baby is especially sensitive to some of these toxins as it does not have a good ability to eliminate poisons; the most critical time is the first 3 months of pregnancy.

4. TOXIN ELIMINATION

The primary elimination organs are the liver and the kidney.

The liver transforms fat-soluble compounds to water-soluble compounds and re-inserts the compounds into the blood (although some bile goes to the small intestine). In the kidney, the blood gives up the water-soluble compound, which then is eliminated in the urine.

Since not all the blood goes directly to the liver, some blood (with the compound) goes to various "target organs." That is, there is a "race" between the kidney de-toxifying the compound and the compound affecting a target organ.

Sometimes there are two compounds competing for the same liver enzyme, thus slowing biotransformation. Examples are alcohol and barbiturates, or cigarette smoke, and cotton dust or asbestos. It has been reported that asbestos workers who smoke cigarettes have a 92 times greater chance of lung cancer than the general population. Heavy metals (lead, methyl mercury) inhibit enzymes.

5. THRESHOLD LIMIT VALUES

Each year, the American Conference of Governmental Industrial Hygienists (ACGIH, 6500 Glenway Avenue, Building D-7, Cincinnati, OH 45221, USA; tel.: +1 (513) 661 7881) issues updated TLV. (Although the ACGIH calls its limits TLV-TWA, the US government calls its limits Permissible Exposure Limits (PEL), the National Institute of Occupational Safety and Health calls its limits Recommended Exposure Limits (REL), and the American Industrial Hygiene Association (AIHA) calls its limits Work Environment Exposure Limits (WEEL)).

The ACGIH gives three different types of TLV: a time-weighted average (TLV-TWA), a short-term exposure limit (TLV-STEL) and a ceiling (TLV-C).

TLV-TWA is the concentration, for a normal 8 h workday and 40 h week, to which nearly all workers may be repeatedly exposed, day after day, without adverse effects. It primarily recognizes chronic (long-term) effects. For administrative simplicity, the TLV assume that concentration x time is a constant. That is, you would get the same effect from eight aspirins taken once in 8 h as from one aspirin taken each hour for 8 h. What an assumption!

TLV-STEL is concerned with acute (short-term) effects of (1) irritation, (2) chronic or irreversible tissue damage and (3) narcosis of sufficient degree to increase the likelihood of accidental

injury, to impair self-rescue or to reduce work efficiency. STEL is a 15 min time-weighted exposure that should not be exceeded any time during the day, even if TLV-TWA is met. There should not be more than four STEL exposures of 15 min/day; there should be at least 60 min between each STEL exposure. For TLV with no STEL listed, ACGIH recommends "STEL should exceed 3 times the TWA for no more than 30 min/day; under no circumstances should STEL exceed 5 times TWA, provided that TWA is not exceeded."

TLV-C is the concentration that should not be exceeded during any part of the working day.

Excursions (peaks) are the recorded values of the toxin. They are permitted above the TWA and STEL (but not above the C). For example, acetic acid has a TWA = 10 ppm. An exposure of 12 for 4 h can be balanced by an exposure of eight for the remaining 4 h. The TWA would even permit an exposure of 80 for 1 h, if balanced by exposure of 0 for the remaining 7 h. This, however, is where the STEL comes in.

STEL for acetic acid is 15. Thus the exposure of 80 for 1 h is too high.

TLV needs to be adjusted if (1) the concentration varies during the day, (2) the day is not 8 h and (3) exposure is to more than one substance.

First, assume a worker was exposed to acetone for 4 h at 500 ppm, 2 h at 750 ppm and 2 h at 1500 ppm. The equivalent exposure is:

$$\text{TWA} = (C_a t_a + C_b t_b \ldots C_n t_n)/8,$$

where TWA is the time-weighted average (equivalent 8 h exposure), C is concentration of $a, b, c \ldots$ (ppm or mg/m^3) and t is time of exposure to concentration $a, b, c \ldots$ (h). For the example:

$$\text{TWA} = [(500)(4) + (750)(2) + (1500)(2)]/8$$
$$= 6500/8 = 812.$$

Since 812 is greater than the TLV-TWA of 750, the exposure is not acceptable.

Next, assume a worker was exposed to 600 ppm of acetone during the entire working day but worked 12 h/day. Then:

$$\text{TWA} = [(600)(12)]/8$$
$$= 7200/8$$
$$= 900.$$

Since 900 > 750, the exposure is not acceptable. Assuming the worker worked only 4 h/day yields 600 (4)/8 = 300. Since the 300 < 750, the exposure is acceptable.

Finally, consider exposure to a mixture of substances. The following formula assumes the effects are additive, not independent or multiplicative:

$$\text{TWA}_{\text{mixture}} = (C_1/\text{TLV}_1) + (C_2/\text{TLV}_2) + (C_3/\text{TLV}_3),$$

where $\text{TLV}_{\text{mixture}}$ is the equivalent TWA mixture exposure (maximum of 1 permitted), $C_{1,2,3}$ is the concentration (8 h) for a specific substance and $\text{TLV}_{1,2,3}$ is the TLV for a specific substance.

Assume exposure for 8 h to 500 ppm of substance 1 (TLV = 750), of 45 ppm of substance 2 (TLV = 200) and 40 ppm of substance 3 (TLV = 100). Then:

$$\text{TWA}_{\text{mixture}} = 500/750 + 45/200 + 40/100$$
$$= 0.667 + 0.225 + 0.400$$
$$= 1.292.$$

Since 1.292 > 1, the exposure is not acceptable.

6. CONTROLS

Controls are divided into engineering, administrative and personal protective equipment.

6.1. Engineering Controls

6.1.1. Substitute a less harmful material
Use solvents with higher TLV (e.g. toluene with TLV = 100, not benzene with TLV = 10). Use water-base cleaning compounds instead of organic base; use latex paints instead of oil (solvent) based paints; use glues without solvents.

6.1.2. *Change the material or process*
Reduce carbon monoxide by using electric-powered fork trucks, not gasoline-powered. Use electrostatic spray painting. Remove grinding particles or solder fumes with a vacuum system instead of blowing them into the room using compressed air. Reduce use of cleaning solvents by cleaning before the substance has time to harden or oxidize; an analogy is dishes in a dishwasher.

6.1.3. Enclose (isolate) the process
As a first approximation, cost of air handling is proportional to the volume moved. The operating cost is twofold: (1) the direct cost of electrical power for the fans and (2) the hidden cost of replacing the conditioned air (heated or cooled and humidified and purified to desired values). Reduce the volume of air handled by capturing substances and vapors before they are dispersed.

Two guidelines are: (1) Physically enclose the process or equipment and (2) remove air from the enclosure (hood) fast enough so that air movement at all openings is into the enclosure (i.e. negative pressure).

One example of enclosure is using plastic strips (such as used for strip doors) around solvent tanks; even better is a lid on the tank. A sealer coat on a concrete floor reduces dust; less dust means less air changes/h and thus lowers heating/air-conditioning costs.

6.1.4. Use wet methods
Wet floors before sweeping; steam cotton; use moistened flint in potteries; clean castings with a high-pressure water jet instead of abrasive blasting.

6.1.5. Provide local ventilation
The flow sequence should be input air, worker, contaminant, exhaust air. That is, the worker is upwind of the contaminant.

Local ducts are much more efficient (25–50% increase in duct velocity) if they have a flange (flat plate perpendicular to the duct axis) equal to duct diameter at the duct entrance.

Dumping the exhaust "out the window" is not satisfactory. Clean it with filters, cyclones, vapor traps, precipitators, etc. Discharge exhaust so it escapes from the "cavity" that forms around buildings; otherwise it may become the input air.

6.1.6. Provide general (dilution) ventilation
Use general ventilation when the contaminant is released from non-point sources. Forced ventilation (fans, blowers) is preferable to natural ventilation (open doors, windows) since air direction, volume and velocity can be controlled.

6.1.7. Use good housekeeping

Remove dust from the floor and ledges to prevent dust movement by air currents, traffic and vibration. Fix leaking containers. Clean up chemical spills.

6.1.8. Control waste disposal

Each disposal problem needs to be considered separately. Establish specific procedures for safe disposal of dangerous substances, toxic residues, contaminated wastes and containers with missing labels. Drain systems become chemical storage systems; be sure you know what is going to mix in your drains.

6.2. Administrative Controls

6.2.1. Reduce exposure time

Reduced exposure time not only reduces exposure but also it increases recovery time. For example, if exposure is reduced from 8 to 4 h/day, the body's daily recovery time goes from 16 to 20 h. Even more important than the absolute recovery time is the relative recovery time. At 8-h exposure, recovery is 16 h, or 2 h recovery/h of exposure. At 4-h exposure, recovery is 20 h, or 5 h recovery/h of exposure.

6.2.2. Periodically monitor employees

The TLV concept is to monitor the *air* around a person. "Biological monitoring" is monitoring the inside of a person's body (e.g. blood, urine). The ACGIH now publishes Biological Exposure Indices for ~40 substances.

6.2.3. Train supervisors, engineers and workers

Knowledge is power. Material Safety Data Sheets (MSDS) are required by the Hazardous Communication Standard (29 CFR 191.1200). Manufacturers and distributors of hazardous chemicals must provide MSDS; employers are responsible (1) to inform their employees about the chemicals and (2) to train employees in safe use of the chemicals.

6.2.4. Screen potential employees

The goal is to avoid use of workers who are hypersensitive to a substance. Examples would be people with allergic contact dermatitis and pregnant women who would be exposed to teratogens. Unfortunately, due to discrimination claims, screening employees has many potential problems.

6.3. Personal Protective Equipment

The third (and last) line of defense is personal protective equipment. Examples include respirators and protective clothing (aprons, gloves). Respirators protect versus airborne toxins while protective clothing protects the skin. Dentists should wear a mask over their nose to reduce the breathing of particles from the drilling of fillings. Note that toxins, which normally could not penetrate the skin, can penetrate the skin if it is broken (cut, rash).

REFERENCE

KONZ, S., 1995, *Work Design: Industrial Ergonomics* (4th ed.) (Scottsdale: Holcomb-Hathaway).

Working Clothing – Thermal Properties and Comfort Criteria

R. Ahasan† and J. Salovaara‡

†Division of Work Science, University of Oulu, Finland‡Department of Industrial Design, Helsinki University of Arts and Design, Helsinki, Finland

1. INTRODUCTION

Owing to unsuitable clothing, worker's physiological reactions are often reflected as deterioration of their work performances. A high level of protection by clothing can also severely impede heat exchange through the sweat evaporation. Thus, reliable measures to evaporative heat transfer through the design of suitable clothing should be developed. In this context, the thermal properties of clothing are reviewed, and corrective measures are emphasized that can match with worker's physiology, and on the effects of thermal balance, work performance and the comfort criteria.

Work clothing has a major contributing role to the workers' physiological response. It complicates the emission of body heat and cause physiological reactions through increasing of sweating, heart rate, heat cramps and frequently impedes worker's performance (Beshir 1994). Physiological changes and seasonal acclimation can lead thermal effect if clothing does not match properly with the working environment. Working conditions also may be stressful due to incorrect or insufficient clothing, and for high metabolic rate of the workers. The effect of air temperature (T_a °C) and humidity is mainly reflected in skin temperature (T_{sk} °C) and in the clothing fabrics (Woodcock 1962). Higher humidity also causes excessive thermal stress more than in the dry conditions (Candas et al. 1979, Boisvert et al. 1993). Sweat rate increases occurs more with the humidity in clothing with higher insulation, and it can lead to an increase in T_{sk}. Thus, increasing concern has been growing from the concerned professionals about undesirable exposure to working clothing. Kenney et al. (1987) demonstrated workers' physiological limits in varying thermal environments. They illustrated the effects of moisture absorption in increased thermal load integrating the coefficients of properties of heat exchange. Lotens (1993) derived a clothing model that handles most of the physical phenomenon. Crockford (1994) illustrated especial garments, which can reflect 40% incoming radiation, and can reduce the convection loss by 25% and if the clothing have 20% of maximum evaporative capacity (E_{max}). Sen and Dutta (1985) discussed about heat radiation, and they proposed aluminum-foiled clothing for workers who are involved with furnace and steel millwork. Crockford et al. (1984) mentioned about the design and suitability of protective clothing for workers. In a hot, humid climate, radiative heat load may > 400 W/m² and metabolic heat load may range > 300 W/m², and thus, evaporative cooling is to be maintained (Kähkönen 1994). Physical work in hot, humid climate is heavy and characterized by static and dynamic loads of variable duration and intensity (Ahasan 1996). Workers with low physical capacities are usually less able to tolerate heat, and

they need good clothing. Environmental control measures are also important to reduce the risk of heat illness. In this regard, ISO 7933 and NIOSH (1986) has been devised some principles that work rate (or workload) should be compensated for by a reduction of heat load. Crockford (1994) added that lower E_{max} might limit the physical workload. By shortening the time of work exposure or changing the climate, thermal load can be reduced.

The limiting values (i.e. warning level of acclimatized persons) in accordance with ISO (International Standard Organization) and NIOSH (National Institute of Occupational Safety and Health, USA) standards provide the maximum level of safety for majority of employees (Ilmarinen 1978). However, reduction of climatic and metabolic stress, and improvement of protective clothing remain a major challenge on the selection of fabrics. In this regard, several thermal indices and heat exchange model have been developed by Holmer (1995) to select clothing ensembles (Gagge et al. 1986). For this, individuals' metabolic condition, stature, sex, age and even social, religion and culture are very important considerations. Often clothing insulation is not adjusted to meet the requirements of actual activity level. Thus, in the design and purchase of working clothing, local criteria should be considered.

Working clothing must be wearable that can balance evaporative cooling, and be really available in the local market. The design has to be simple and easy to manufacture locally. It must be comfortable and likely to be functional for effective sweat evaporation through its fabrics. Many developing countries are poor in economy, where physical work combines with heat stress and cause acclimation through sweating. In this sense, thermal properties of tropical clothing (Ahasan et al. 1996) are very important that determine heat exchange rate, and as such, act as a resistance to heat loss through the evaporation of sweats. Therefore, correct choice of working clothing depends on the heat stress index, permissible limit values, thermal properties and local climatic condition.

2. THERMAL PROPERTIES OF WORKING CLOTHING

Working clothing affect heat transfer from and to the human body mainly through the insulation and permeability of the fabrics. The subjects can feel pleasant due to optimum insulation and ventilation (i.e. concentration gradients) in the clothing across the textile layers (Olesen and Nielsen 1983, Nielsen et al. 1989). For instance, in tropical countries, heat radiation and humidity have a direct effect for which clothing has to be selected carefully. Solar heat radiation is continuos a serious heat problem for the workers in tropics, even moderate physical workload could be heavy and stressful in that situation (Sen 1965). Holmer (1985) also has reported the detrimental effects of moisture absorption in working clothing and insulation characteristics. Havenith and Lotens (1994) and Shapiro et al. (1995) discussed the effects of sweat absorption in working clothing.

Water absorption, transporting properties and drying rate of sweats are also evaluated (Nordon and David 1967, Nielsen and Endrusick 1990, Nielsen 1992) from the selected clothing ensembles. Evaporated sweat may condense and accumulate and thereby influence the distribution of humidity in working clothing (Madsen 1976). With evaporative heat transfer, it has been possible to reduce up to 50% using cotton clothing (Kenney et

al. 1987). High relative humidity (RH) can also affect clothing layers which is prevalent in the tropics. Ha and Tokura, (1995) noted that sweats are significantly higher in polyester than cotton and a wet fabric in direct contact with skin produce an unpleasant sensation as *chilly effect* or *clammy feelings*. Radiative energy in tropical climate is an additional heat load while wearing color-absorbing and reflecting clothing (Lotens and Pieters 1995). That is why thermal properties of clothing ensembles and fabric materials are to be given priority.

2.1. Clothing Insulation

Insulation characteristics are important for working clothing to select proper clothing ensembles. It plays a vital role for moisture distribution and sweat absorption and, has a leading influence on comfort of clothing (Fanger 1972, Givoni and Golman 1972, Holmer and Elnäs 1981, ISO-9920, Nielsen and Toftum 1992). Fabric surface and its thickness, i.e. amount of air trapped inside fabrics, mainly determine thermal insulation of clothing (or Clo). Clo (e.g. thermal insulation or resistance of fabrics are necessary to maintain the thermal balance between T_{sk}, textile materials and T_a) for different clothing ensembles can be found from DIN standard 33403, Olesen and Nielsen (1983) and the ISO's draft standards. It can control the heat exchange rate between skin and environment. Clothing insulation is varied to the individual's physiological response, and to the combination of workload, RH, T_a, air velocity (V_a) and clothing ensembles. Olesen *et al.* (1982) has evaluated Clo by movable thermal manikin on the effects of body posture and activity level.

Permeability of clothing ensembles indicates the ability of water vapor to pass through the fabrics (i.e. ventilation). High permeability means more sweat will be allowed to evaporate and high ventilation can enhance cooling of skin through sweat evaporation. Light and moderate insulated, permeable with good ventilated clothing (long sleeve shirt, trouser, socks and shoes having 0.6 Clo) are recommended where no risk of high contamination (Holmer and Elnäs 1981). Fabric thickness and clothing layers also reflect clothing insulation (Larose 1947). Typical clothing ensembles (Holmer and Elnäs 1981, ISO-9920, Oohori *et al.* 1985, McCullough *et al.* 1989) are listed in Table 1, which are comparatively available.

2.2. Clothing Ensembles and Fabric Materials

Clothing ensembles and fabric materials have a significant impact on thermal properties and comfortability. Nielsen (1993) emphasized on dissipation of metabolic heat that can be reduced
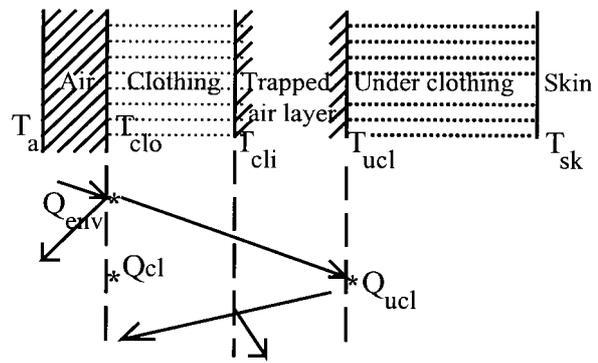
Figure 1. Radiation effect on clothing ensemble consisting outer garment and under clothing.

up to 25% depending on proper fabrics. High humidity and solar radiation cause excessive sweating and, thus, Hall (1971) developed a simplified method for clothing ensembles. Lotens and Pieters (1995) evaluated radiation effect on two-layer clothing (Figure 1).

Heat is absorbed first in the outer layer (Q_{cl}) and partly transmitted to underclothing (Q_{ucl}), until the radiation energy is completely distributed over the clothing layers and then to the environment (Q_{env}). The resultant temperature in the fabrics (T_{ucl} for underclothing, T_{cli} for inside of outer clothing and T_{clo} outside of outer clothing) depend on the distribution of Q_{env} over the clothing layers, and on the heat resistance of fabric layers and the boundary conditions (T_{sk}, T_a). Absorption properties of working clothing show a greater or equal transit in T_{sk} and may have beneficial effect on thermal sensation. During absorption, vapor concentration moves upward (Figure 2) along the line, and in a downward direction for desorption (Havenith and Lotens 1994). Sweating is related with fabrics and clothing ensemble are also needed to be formulated from the required sweat rate (SW_{req}) and Clo which depend on physical properties of fabrics (Table 2): thickness (density, course, number/inch), specific weight (sp. wt.), thermal resistance, water absorbency; thermal conductivity; moisture transfer, air permeability (AP), water vapor resistance (WVR) and heat flow transmission rate (HFTR).

Absorption capacity for actual wearer has seldom quantified, but Nordon and David (1967) developed a numerical method including the rate of equilibrium between fabrics and surrounding air. To simulate clothing ensembles, Havenith and Lotens (1994) developed a computer model based on geometry of clothing and activity of worker. They focused on the prediction of thermal

Table 1. Clothing insulation and permeability index of clothing ensembles.

Clothing ensembles		Clo–values		permeability index		TER
($R_{T,evp}$)		I_{cl}	I_T	i_m	i_{cl}	(Pa m²/w)
1. Cotton shirt, trouser, underwear, shoes	0.65	1.25	0.39	0.32		30
2. T-shirt, underwear, socks, shoes	0.84	1.41	0.39	--		34
3. Shirt, pant, underwear, belt, socks		1.00	1.51	0.30	0.27	44
4. Aluminized coat, CST uniform	1.36	1.89	0.33	--		53
5. Aluminized coat (midnight) CS-uniform	1.74	2.19	0.30	0.27		68
6. Chemical protective clothing (Goretex)	1.40	1.90	--	--		180

(I_T is total insulation; i_m and i_{cl} are permeability index and/or boundary layer, and clothing layer respectively under the standard conditions. $R_{t,evp}$ is the total evaporative resistance)

H₂O -Vapor concentration
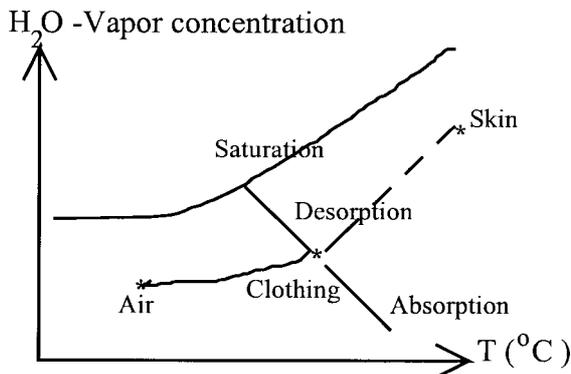
Figure 2. Effects of H₂O vapor absorption in clothing, at skin and in surrounding air.

effect on heat balance through the process of vapor absorption in different fabric's layer. David (1965) reported the effects of saturated sweat absorption in clothing ensembles, using a thermal manikin. More recently, De Dear *et al.* (1989) reported the thermal effects, moving from one environmental humidity to another with absorbing and non-absorbing clothing.

2.3. Heavy Protective Clothing for Working in Dangerous Conditions

Heavy protective working clothing is important for pesticide sprayers, in chemical plants (or wastes disposal), food industries and in nuclear power plants (Raven *et al.* 1979, Mihal 1981). Flameproof clothing is used for fire brigades, and nuclear, biological and chemical (NBC) clothing is used for particular situation to reduce the danger of work hazards. Heavy protective over-garment and water-repellent fabrics are also used in heavy protection; however, this clothing is very expensive. Protective clothing with microporous and ventilated fabrics significantly can improve heat exchange and reduce physiological strain. An adjustable valve must be efficient in NBC clothing to washout most of the metabolic heat. NBC clothing poses heavy constraints to work even with moderate work, particularly with high humidity. Therefore, proper schedule of work and better ventilation should be provided (Ohnaka *et al.* 1993). Protective clothing should design that heat tolerance in terms of dynamic changes to inside body temperature must not be above the tolerance level, as well as to improve work tolerance and can protect against chemical liquids, dusts, aerosols and poisonous substances. For better comfort and performance, temperature of space between skin and clothing layer (T_L) should be monitored carefully. Cortili *et al.* (1996) found a reliable match of T_L between 36.2 and 36.4°C (when V_a = 0.1 and 0.5 respectively) that may valid mainly in the tropical climates (T_a = 35°C, RH = 85%).

2.4. Typical Working Clothing

Typical clothing and its model depend on multifactors such as sex, age, health, stature and body composition. Even social status, culture, religion and region can affect thermal responses. For instance, Thai female workers are shorter than Tanzanian women are, and their clothing is not similar. Females in SAARC (Southeast Asian Association for Regional Co-operation: India, Nepal, Bangladesh, Pakistan, Bhutan; Maldives, Sri Lanka, etc.) countries use *sari* (one piece of long cloths that is to be trapped and cover all the whole body, blouse and petticoats). *Sallower* and *kamiz* (long legged trouser and shirts) are popular for young women. *Lungis* (hollow type long legged clothing) and sandals are very common in men's wearing in these regions. In the big cities, working clothing are usually pants, shirts, aprons and shoes. Women in the ASEANS (Association of Southeast Asian Nations) use pants, skirts and tops. African and Arab women use hollow type dresses. Owing to religion, shorts and light clothing are forbidden in some Muslim countries (e.g. Malaysia, Saudi Arabia), and even there are State rules that clothing must be worn for complete cover of all sexual parts. However, working clothing should be selected so that it can reduce physiological and climatic stress. Generally, cotton, nylon and polyester are used as light working clothing. Traditional or regular working clothing is generally light dresses that includes trousers, T-shirt, sandals or canvas shoes (0.5–0.7 Clo).

ACGIH (1991) devised summer clothing (0.6 Clo); Ramsey (1978) suggested normal clothing (0.5–1.0 Clo); Goldman (1981) added regular clothing (0.5–0.7 Clo); and Bernard *et al.* (1986) preferred cotton coveralls (0.7–1.0 Clo). Jogging shoes, socks, shorts, single jacket or T-shirt are athletic clothing. Aluminized garments are used for workers in the steel and re-rolling mills to protect from radiant heat. Further, proper working clothing should be considered from the available fabrics, which are cheaper and thermally matched with metabolic reactions, because ISO and NIOSH recommended clothing are not easily suitable and cheap in some circumstances.

3. COMFORT CRITERIA

Comfort criteria are very important for the design of working clothing. To assess proper working clothing, thermal properties of fabrics, and the effects of metabolic reactions should be considered. Feeling of clothing discomfort also depends on the type of tasks, jobs, work environment and body heat balance on clothing ensemble (Wyon *et al.* 1975, Hafez and Beshir 1987). ISO standards have been evaluated thermal sensation and discomfort levels using various clothing ensembles. Galbraith *et al.* (1962) evaluated comfort sensation while wearing cotton clothing, e.g. light gray, pure white or mixed cotton, linen or polyester (65% cotton plus 35% polyester or linen) are also

Table 2. Physical characteristics of textile materials [ISO 5084; ISO 5085; CAN2-4.2M77 (1977)]

Fabrics	Thickness (mm)	SP.wt (g/m²)	TR (Km²/w)	AP (1/m²/s)	WVR (mm air)	HFTR (%)
Cotton	1.52	153	0.028	1741	*	62
Polyester	0.73	280	0.016	104	1.4	69
Polyester/ cotton mixed	2.40	210	0.058	983	2.3	55

preferred. Ramsey (1978) described physiological responses for general guidelines, but this approach does not contribute to the heat balance equation. Fanger (1972) developed mean thermal sensation on clothing effects. Neilsen *et al.* (1989) predicted comfort level using subjective and physiological rating scales. Havenith and Lotens (1984) compared semipermeable and impermeable garments to heat dissipation and discomfort of thermal sensation. Hensel (1981) showed the relationship between thermal discomfort and skin wettedness by various ensembles. For thermal comfort, seasonal variation in T_a, V_a, heat radiation and humidity are to be selected properly (Vokac *et al.* 1976) because physiological change and seasonal acclimation can lead to thermal effect if activity and clothing ensembles do not match properly.

Working clothing effect on climatic and physiological parameters is to be integrated for acceptable thermal sensation (Sprague *et al.* 1994). In general, subjects in tropical climates feel comfortable using light clothing due to solar heat and higher humidity (e.g. RH $\cong$ 70–85%). Moisture absorption is generally regarded as a prime factor in thermal comfort that depends on fabric rigidity, wicking, clinging, liquid buffering, drying ability and other thermal properties of working clothing. Condensation of evaporated sweat is highly relevant with comfort criteria (Gonzalez *et al.* 1986); and the color of clothing and the wavelength of radiation effect on clothing are very important (Nielsen 1990).

4. DISCUSSION

Recommendations for proper working clothing should rely on the understanding of physiological effects of heat exposure to clothing by environmental control that can enhance safety and work performance. Kenney *et al.* (1986) emphasized acclimatization control advocated by specific measures (using natural and cross ventilation, heat barriers, reduction of physical work, rest breaks or provision of driving water). Local contexts are to be applied for physiological adaptations with working clothing and climatic variables, because threshold limits (TLV) or recommended standards as national legislation in tropics are yet to be established. In most cases, workers or entrepreneurs know those parameters, but they do have some sort of thermal sensation from which clothing can be selected. Acceptable working clothing must be choice on physiological and climatic conditions which match properly. The workers are to be advised to wear proper clothing and trained through metabolic heat reduction scheme (Ramsey 1995) and overall environmental control) by certain control (NIOSH 1986, ISO 7730, ASHRAE 1992, Olesen and Madesen 1995). To protect from the ill-effects, TLV or certain standards are important for clothing control that keep the body heat balance and that can bring a massive advantage to worker's health, safety, efficiency and productivity. Following standards are important for recommendations: general guidelines for metabolic heat exposure limits, and selection of normal clothing, unacclimatization and environmental control (Kodak 1983, NIOSH 1986); appropriate Clo and design of clothing (Ramsey 1978, 1995, Bernard *et al.* 1991); modification of wet-bulb glove temperature (WBGT) index and correction factors for various clothing (ISO 7933, ACGIH 1991); estimation of thermal insulation of clothing (ISO 9920); evaluation of moderate thermal work environment, and prediction of thermal sensation and

discomfort (ISO 7730); guidelines for evaluation of thermal strain (ISO (9886); subjective judgement of thermal sensation (ISO 10551) and ergonomics of thermal environment (ISO 12894).

This established standard means a situation or measures above which work should not be resumed wearing working clothing. According to ASHRAE (1977), comfort range is ~23–27°C. Fanger (1970) added that clothing having 0.5–0.7 Clo is better for sedentary work during lower V_a (> 0.3 m/s). Sen (1982) advised that T_{room} @ 23.5∞C is comfortable at lower humidity levels (e.g. 50% RH). He found a comfort zone in the Indian climate to be at 26°C in corrective effective temperature for those engaged in light or moderate work. Air-conditioning systems are used in many commercial offices, market complexes or banks where a thermal comfort zone is taken into consideration.

Common ergonomics approaches are the arrangement of proper clothing materials with less thermal impact (Epstein *et al.* 1986). Fabric thickness affects insulation and can influence mean skin temperature because of vapor diffusion and moisture transportation from skin to the clothing ensemble (Bakkevig and Nielsen 1994). For example, a fishnet textile allows more air to sweep the sweating and at the same time can act as effective ventilation (Nielsen and Endrusick 1990). Normally cotton, gabardine and mixed fabrics (cotton, gabardine, linen) are a benefit in the tropics (Kogi and Sen 1987). They added that fabrics should be flame-resistant, flexible and washable. Foot and underwear are to be suitable. Sandals are extensively used for light wearing, especially in the tropical developing countries, but to ensure more safety, canvas shoes are recommended. Soft rubber or plastic shoes can be used for moderate work. Gym shoes and boots are recommended for heavy work where protection (e.g. from burning or cutting) of feet is a vital. Sweat accumulation is generally higher in underwear that can be affected by fiber types. Excessive sweat may feel uncomfortable, and thus underwear clothing (i.e. fabric construction) should select carefully (Bakkevig and Nielsen 1994).

It has been evident that most of the protective clothing severely restrict evaporative heat loss and increase thermal strain, but microporous fabric significantly can protect against chemical liquids, dusts or aerosols, and impermeable fabrics can protect against gases and vapors. NBC clothing restricts heat loss by evaporation when radiation is direct or omnidirect. So, adjustable valves for air outlet must be provided to transmit sweating and heat through the clothing fabrics. Smith *et al.* (1994) emphasized on TLV and exposure time that physical fatigue can be normalized. Madsen (1976) suggested that workers in cold storage, slaughtering house or in fish processing factory need multilayered clothing (I_T = 1.20 Clo). He added that outer layer should be cotton, middle layer a mixed polyester and cotton, and inner layer could be pure cotton. Makambaya (1994) added special working clothing (outer wind-proof and waterproof garments) for high wind and rain. Hygroscope and damped humidity-gradient fabrics are also suitable to transfer saturated sweat from the skin. Anyway, open and thin garment-knits are preferred especially for tropical countries.

5. CONCLUSION

Most of the entrepreneurs do not afford costly clothing for their workers; however, working clothing should be selected on the basis of thermal properties (e.g. suitable and adjustable Clo),

available in the market and long lasting with reasonable price. Generally, cotton fabrics are the cheapest and comfortable in the tropics. NBC and/or heavy protective clothing are used in the cold climate, which are bit expensive. But modifications of heat exchange should be adjusted carefully whatever working condition is. Moreover, it has been evident that impaired heat exchange with heavy protective clothing contributed to most of the physiological strain. The workers' general opinion is that protective clothing is sometimes more of a handicap than a protection. Therefore, it is important to know a reliable indicator of thermal load in a wide range of different combination of physiological and climatic variables on these clothing.

Physiological effect and significance of clothing on tropical human are not yet completely known. In this regard, subjective assessment is to be evaluated where local influences, such as asymmetric radiation, draught, vertical rays or direct radiation are existed. Moreover, control of heat exposure must satisfy national health and safety standards (or legislation); however, many countries are yet to establish legislation. Clothing must be worn to reduce the thermal impact as well as to protect physical and environmental work hazards. Thus, thermal properties of clothing should evaluate on the effects of local population (i.e. on physiological adaptation with local climate) that can match with thermal comfort sensation. The practitioners normally estimate metabolic rates following heat stress standards, but thermal comfort reflect on the integrated state of thermoregularity system, where T_{sk} and internal body temperature (T_{body}) must be integrated together. Metabolic rate and clothing values must be keyed in by ergonomist to design a proper choice of working clothing. So, careful control of exposure duration, thermal indices (T_a, RH, V_a, radiant heat flux, solar radiation, etc.) and metabolic factors (sex, age, obesity, health status, physical fitness, skin temperature and wetness, etc.) are important to be considered (Lee 1980). Wearing NBC clothing results in increased respiratory strain and elevated T_{body}, and inspiratory or expiatory effect. But distributing it in multilayered ensembles can reduce sweat accumulation. Sweat may not evaporate completely in these clothing, but if inner layers are absorbing in nature then it will perhaps be benefited. Outer layers must be easy to open to allow sweat be exposed to the air. Latent heat of condensed sweat must be available for conduction to clothing surface and to the environment. Under certain conditions, evaporative sweat may condense before escaping to the environment, giving off its heat to the surrounding clothing layers. The concept of basic insulation (I_{cl}) is valid for a limited range of ensembles and thermal conditions. In this regard, effect of condensation evaporation (efficiency and conductive heat exchange) is needed further experimental work along with consideration of work posture, body motion and wind effects.

Recent modifications by ISO standards are rational but not promising for low-cost ensembles. Sweat rate prediction by the WBGT index may not reflect the actual heat stress for clothing design if there is high T_r, T_a or V_a. It also seems that climate condition in thermal work environment for which ISO standards are applicable is too narrow in the tropical climate. The estimating errors of I_{cl} may not be always correct in evaluating thermal environment and selection of proper clothing. Tropical countries are experienced with heavy physical work, and wearing protective clothing can increase sweating and vapor concentration that

results in uncomfortable feelings. In most of the developing countries workers are less educated and are lacking knowledge about work hazards. Environmental control should relate with the use of proper clothing to enhance safety and work performance. A model should derive for standard clothing especially for SAARAC, ASEANS, African, Arab and other tropical countries that handle most of the physical phenomena influencing heat transfer criteria. Hence, further R & D should focus on the ability to interact sweat transportation, heat dissipation, and on how such effects can modify Clo for cheaper and available clothing. Standard clothing recommended by ISO, NIOSH or ASHRAE is also very expensive and, therefore, ergonomics of low-cost clothing should be a great achievement.

REFERENCES

ACGIH, 1991, *Threshold Limits Values for Physical Agents in the Work Environment and Heat Stress* (Cincinnati: American Conference on Governmental and Industrial Hygienists).

AHASAN, M.R., QUDDUS, R., VÄYRYNEN, S., VIROKANNAS, H. and MOHIUDDIN, G., 1999, Work-related problems in metal handling tasks in Bangladesh: obstacles to the development of safety and health measures. *Ergonomics*, 42, 385–96.

AHASAN, M.R., VÄYRYNEN, S. and VIROKANNAS, H. 1996, Low-cost protective clothing for tropical climate. In L. Straker and S. Pullock (eds), http://www.curtin.edu.au/conference/cyberg

AHASAN, M.R., VÄYRYNEN, S. and VIROKANNAS, H., 1997, Mathematical model in the estimation of heat stress and sweat loss. In C. Constanda, S. Seikkala and J. Saaranen (eds), *Analytical Methods* (Harlow: Addition Wesley Longman), vol. 1, 21–6.

ASHRAE, 1992, *Thermal Environmental Conditions for Human Occupancy* (American Society of Heating, Refrigeration and Air Conditioning Engineers).

ASHRAE, 1977, *Physiological Principals, Comfort and Health* (American Society of Heating, Refrigeration and Air Conditioning Engineers), 1–36.

BAKKEVIG, M.K. and NIELSEN, R., 1994, Impact of wet underwear on thermoregulatory responses and thermal comfort. *Ergonomics*, 37, 1375–89.

BERNARD, T.E., KENNEY, W.L., BALINT, L. and O'BRIEN, J.F., 1986, *Heat Stress Management Program for Nuclear Power Plants* (Palo Alto: Electric Power Research Institute).

BERNARD, T.E., KENNEY, W.L., HANES, L.F and O'BRIEN, J.F., 1991, *Heat Stress Management Program for Power Plants* (Palo Alto: Electric Power Research Institute).

BESHIR, M.Y., 1994, Thermal work environment in Africa. *African Newsletter on Occupational Health and Safety*, 4, 3.

BISHOP, P.A., SMITH, G., RAY, P., BEAIRD, J. and SMITH, J., 1993, Empirical prediction of physiological response to prolonged works in encapsulating clothing. *Ergonomics*, 34, 1503–12.

BOISVERT, P., NAKAMURA, K., SHIMAI, S., CANDAS, V. and TANAKA, M., 1993, Sweat rate and body temperature during exercise in dry and humid environment. In R. Nielsen and K. Jorgensen (eds), *Advances in Industrial Ergonomics and Safety V* (London: Taylor & Francis), 399–409.

CAN2-4.2-M77 (Method-49), 1977, *Method for the Test of Resistance of Materials to Water Vapor Diffusion (Control-Dish Method)*.

CANDAS, V., LIBERT, J.P. and VOGT, J.J., 1979, Influence of air velocity and heat acclimation on human skin wettness and sweating efficiency. *Journal of Applied Physiology*, 47, 1194–1200.

CORTILI, G., MOGNONI, P. and SAIBENE, F., 1996, Work tolerance

and physiological response to the thermal environment wearing protective clothing. *Ergonomics*, 39, 620–33.

CROCKFORD, G.W., 1994, Work wear and protective clothing for hot workplaces in tropical countries. *African Newsletter on Occupational Health and Safety*, 4, 16–17.

CROCKFORD, G.W., SEN, R.N. and SPENCER, J., 1984, The design of work wear and protective clothing for use in the tropics. *Indian Journal of Physiology and Allied Sciences*, 38, 125–33.

DAVID, H.G., 1965, The effects of changing humidity on the insulation of hygrospic clothing. *Textile Research Journal*, 35, 820–6.

DE DEAR, R.J., KNUDSEN, H.N. and FANGER, P.O., 1989, Impact of air humidity on thermal comfort during step changes. *ASHRAE Transactions*, 95 (pt 2), 336–50.

DIN-33403, 1986, *Klima am Arbeitsplatz und in der Arbeitsumge bung* (Berlin).

EISSING, G., 1984, *Protective Clothing: Study in Induced Thermal Strain and in Protection to Working Conditions in a Sub-tropical Country* (Geneva: ILO).

EPSTEIN, Y. SHAPIRO, Y. and BRILL, S., 1986, Comparison between different auxiliary cooling devices in a severe hot and humid climate. *Ergonomics*, 29, 41–8.

FANGER, P.O., 1970, *Thermal Comfort: Analysis and Applications in Environmental Engineering* (New York: McGraw-Hill).

FANGER, P.O., 1972, *Thermal Comfort* (New York: McGraw Hill).

GAGGE, A.P., FOBELETS, A.P. and BERGLUND, L.G., 1986, A standard predictive index of human response to the thermal environment. *ASHRAE Transaction*, 92, 709–31.

GALBRAITH, R.L., WERDEN, J.E., FAHNESTOCK, M.K. and PRICE, B., 1962, Comfort of subjects clothed in waters repellent clothing and Orlon suits. *Textile Research Journal*, 32, 236–42.

GIVONI, B. and GOLDMAN, R.F., 1972, predicting rectal temperature response to work, environment and clothing. *Journal of Applied Physiology*, 32, 812–22.

GOLDMAN, R.F., 1981, Prediction of heat strain: revised 1979–1980. In F.N. Dukes-Dobos and A.H. Hensel (eds), *Proceedings of a NIOSH Workshop on Recommended Heat Stress Standards*. No. 81-108 (Cincinnati).

GONZALEZ, R.R., BRECKENRIDGE, J.R., LEVELL C.A., KOLKA, M. and PANDOLF, K.B., 1986, *Efficacy of Heat Exchange Using Wetable Cover Over Chemical Protective Garments* (ASTM), 515–34.

HA, M. and TOKURA, H., 1995, Effects of two kinds of clothing made from hydrophobic and hydrophilic fabrics and local sweating rates at 37°C. *Ergonomics*, 38, 1445–55.

HAFEZ, H.A. and BESHIR, M.Y., 1987, Effects of thermostat temperature on human performance. In S.S. Asfour (ed.), *Trends in Human Factors IV* (Amsterdam: Elsevier Science), 325–32.

HALL, S.A., 1971, Heat stress in outdoor manual workers in East Africa. *Ergonomics*, 14, 91–4.HARDY, J.D., 1970, Thermal comfort: skin temperature and physiological thermoregulation. In J.D. Hardy, A.P. Gagge and J.A.J. Stolwijk (eds), *Physiological and Behavioural Temperature Regulation* (Springfield: Thomas), 856–73.

HAVENTIH, G. and LOTENS, W.A., 1984, *What Actually is the Advantage of Semi-Permeable Over Impermeable Rainwear?* Report IZF-6 (Soesterberg: TNO Institute for Perception).HENSEL, H., 1981, *Thermoreception and Temperature Regulation* (London: Academic Press).

HOLMER, I., 1985, Heat exchange and thermal insulation compared in woollen and nylon garments during wear trials. *Textile Research Journal*, 55, 511–18.

HOLMER, I., 1995, Protective clothing and heat stress. *Ergonomics*, 38,

166–82.HOLMER, I. and ELNÄS, S., 1981, Physiological evaluation of resistance to evaporative heat transfers. *Ergonomics*, 24, 63–74.ILMARINEN, R., 1978, Influences of clothing ensembles on several physiological parameters during work in hot and humid environments. Dissertation, Deutsche Sporthochschule, Cologne.

ISO CD-12894, 1993, *Ergonomics of Thermal Environment — Medical Supervision of Individuals Exposed to Hot and Cold Environments* (Geneva: International Standard Organization).

ISO DIS-10551, 1993, *Assessment of Influence of Thermal Environment Using Subjective Judgement Scales* (Geneva: International Standard Organization).

ISO TR-11079, 1993, *Evaluation of Cold Environments — Determination of Required Clothing Insulation* (Geneva: International Standard Organization).

ISO-5084, 1977, *Textiles — Determination of Thickness of Woven and Knitted Fabrics* (Geneva: International Standardization Organization).

ISO-5085, 1987, *Textiles — Heat Retention Properties Fabrics by Means of Heating Body* (Geneva: International Standard Organization).

ISO-7730, 1993, *Moderate Thermal Environment — Determination of PMV and PPD Indices and Specifications of the Conditions of Thermal Comfort* (Geneva: International Standard Organization).

ISO-7933, 1989, *Hot Environments — Analytical Determination and Interpretation of Thermal Stress Using Calculation of Required Sweat Rate* (Geneva: International Standard Organization).

ISO-9886, 1992, *Evaluation of Thermal Strain by Physiological Measurements* (Geneva: International Standard Organization).

ISO-9920, 1993, *Ergonomics — Estimation of Thermal Characteristics of Clothing Ensembles* (Geneva: International Standard Organization).

KÄHKONEN, E., 1994, Thermal work environment and ISO Standards. *African Newsletter on Occupational Health and Safety*, 4, 4–6. KENNEY, W.L., LEWIS, D.A., ANDERSON, R.K. and KAMON, E., 1986, A simple tests for prediction of relative heat tolerance. *American Industrial Hygiene Association Journal*, 47, 203–6.

KENNEY, W.L., LEWIS, D.A., HYDE, D.E., DYKSTERHOUSE, T.S., ARMSTRONG, C.G., FOWLER, S.R. and WILLIAMS, D.A., 1987, Physiologically derived critical evaporative coefficients for protective clothing ensembles. *Journal of Applied Physiology*, 63, 1095–9.

KERSLAKE, D.M.K., 1972, *The Stress of Hot Environment* (Cambridge: Cambridge University Press).KODAK, 1983, *Ergonomic Design For People At Work*, vol. 1 (New York: Eastman-Kodak).

KOGI, K. and SEN, R.N., 1987, Third World ergonomics. In D.J. Oborne (ed.), *International Reviews of Ergonomics* (London: Taylor & Francis), vol. 1, 77–118.LAROSE, P., 1947, The effects of wind on thermal resistance of clothing with special reference to protection by coverall fabric of various permeability. *Canadian Research Journal*, 25, 169–90.

LEE, D.H.K., 1980, Seventy-five years search for heat index. *Environmental Research*, 22, 331–56.LOTENS, W.A., 1989, The actual insulation of multi-layer clothing. *Scandinavian Journal of Work, Environment and Health*, 15, 66–75.LOTENS, W.A., 1993, *Heat Transfer from Humans Wearing Clothing* (Soesterberg: TNO-Institute for Perception).LOTENS, W.A. and HAVENITH, G., 1994, Effects of moisture absorption in clothing on the human heat balance. *Ergonomics*, 38, 1092–1113.

LOTENS, W.A. and PIETERS, A.M.J., 1995, Transfer of radiation heat through clothing ensembles. *Ergonomics*, 38, 1132–5.MADSEN, TL 1976, Thermal manikin for measuring the thermal insulation capacity for human clothing. Thermal Insulation Laboratory (paper no. 48) (Technical University of Denmark).

MAKAMBAYA, S., 1994, Work in cold environment. *African Newsletter on Occupational Health and Safety*, 4, 4–6.

MᴄCULLOUGH, E.A., JONES, B.W. ZBIKOWSKI, P.J. ARPIN, E.J., KONZ, S.A. and ROHLES, F.H. 1989, *ASHRAE Transactions*, 89, 327–52.

MIHAL, C.P., 1981, Effects of heat stress on physiological factor for industrial workers performing routine work and wearing impermeable vapor-barrier clothing. *American Industrial Hygiene Association Journal*, 42, 97–103.

MILLICAN, R., BAKER, R.C. and CONK, G.T., 1981, Controlling heat stress — administrative verses physical control. *American Industrial Hygiene Association Journal*, 42, 411–16.

MORRIS, L.A., 1995, Practical issues in the assessment of heat stress. *Ergonomics*, 38, 183–92.NIELSEN, B., 1990, Solar heat loads: heat balance during exercise in clothed subjects. *European Journal of Applied Physiology*, 60, 452–6.NIELSEN, B., 1992, Heat stress cause fatigue: exercise performances during acute and repeated exposure to hot and dry environment. *Medicine and Sport Science*, 34, 207–17.NIELSEN, B., 1993, Evaporation of sweat from dressed man. In R. Nielsen and K. Jorgensen (eds), *Advances in Industrial Ergonomics and Safety* V (London: Taylor & Francis), 429–34.

NIELSEN, R. and ENDRUSICK, T.L., 1990, Thermo-regulatory response to intermittent exercise are influenced by knit structure of underwear. *European Journal of Applied Physiology*, 60, 15–25.

NIELSEN, R. and TOFTUM, J., 1992, *Udvikling af beklaedning til kolde arbejdspladser* (Copenhagen: Arbejdsmiljofondet).

NIELSEN, R., GAVHED, D.C.E. and NILSSON, H., 1989, Thermal function of a clothing ensemble during work: dependency on inner layer fit. *Ergonomics*, 32, 1581–94.

NIOSH, 1986, *Criterion For A Recommended Standard, Occupational Exposure to Hot Environments*. DDHS 86–113 (Washington, DC: National Institute for Safety and Health).

NORDON, P. and DAVID, H.G., 1967, Coupled diffusion of moisture and heat in hygroscope textile materials. *International Journal of Heat and Mass Transfer*, 10, 853–66.OHNAKA, T. TOCHIHARA, Y. and MURUMATSU, T., 1993, Physiological strains in hot-humid conditions while wearing disposable protecting clothing commonly used by the asbestos removal industry. *Ergonomics*, 36, 1241–50.

OLESEN, B.W. and MADSEN, T.L., 1995, Measurement of physical parameters of the thermal environments. *Ergonomics*, 38, 138–53.

OLESEN, B.W. and NIELSEN, R., 1983, Thermal insulation of clothing measured on a movable thermal manikin and on human subjects. *ECSC Program Research*.

OLESEN, B.W. SLIWINSKA, E., MADSEN, T.L. and FANGER, P.O., 1982, Effects of body posture and activity on thermal insulation of clothing: measurement by movable thermal manikin. *ASHRAE Transactions*, 88, 791–805.

OOHORI, T. BERGLUND, L.G. and GAGGE, A.P., 1985, Comparison of eminent two-parameter indices of vapor permeation of clothing — as factor governing thermal equilibrium and human comfort. *ASHRAE Transactions*, 90, 85–101.

RAMSEY, J.D., 1978, Abbreviated guidelines for heat stress exposure. *American Industrial Hygiene Association Journal*, 39, 491–5.

RAMSEY, J.D., 1995, Task performance in heat: a review. *Ergonomics*, 38, 154–63.

RAVEN, P.B., DODSON, A.T. and DAVIS, T.O., 1979, The physiological consequences of wearing individual respirators: a review. *American Industrial Hygiene Association Journal*, 40, 515–34.

SEN, R.N. and DUTTA, N.C., 1985, An ergonomic design of low-cost protective work clothing for the furnace workers. In *Proceedings of the International Symposium of Ergonomics In Developing Countries*, Jakarta, 18–21 November (Geneva: ILO).

SEN, R.N., 1982, Certain ergonomic principles in the design of factories in a hot climate. In *Proceedings of the International Symposium On Occupational Safety, Health and Working Conditions In Relation to Transfer of Technology to the Developing Countries* (Geneva: ILO), 123–47.

SHAPIRO, Y., MORAN, D. and EPSTEIN, Y., 1995, Validation and adjustment of mathematical prediction model for sweat rate response to outdoors environmental condition. *Ergonomics*, 38, 981–6.SEN, R.N., 1965, Physical and environmental factors affecting health of workers in the industry. *Indian Labour Journal*, 6, 735–46.

SMITH, G. BISHOP, P. BEAIRD, J. RAY, P. and SMITH, J., 1994, Physiological factors limiting work tolerance in chemical protecting clothing. *International Journal of Industrial Ergonomics*, 13, 147–55.

SPRAGUE, D.H. and MUNSON, D.M., 1994, A complete ensemble method for estimating thermal insulation values of clothing. *ASHRAE Transactions*, 80, 120–9.VOKAC, Z., KÔPKE, V. and KEUL, P., 1976, Physiological responses: thermal humidity and comfort sensation in wear with cotton and polypropylene vests. *Textile Research Journal*, 46, 30–8.

VOKAC, Z., KÔPKE, V. and KEUL, P., 1976, Physiological responses: thermal humidity and comfort sensation in wear with cotton and polypropylene vests. *Textile Research Journal*, 46, 30–8.

WOODCOCK, A.H., 1962, Moisture transfer in textile systems. *Textile Research Journal*, 32, 628–33.WYON, D.P. FANGER, P.O. OLESEN, B.W. and PEDERSEN, C.J., 1975, The mental performance of subjects clothed for comfort at two different air temperature. *Ergonomics*, 18, 359–74.

Part 8

System Characteristics

Accident Analysis and "Human Error"

E. Hollnagel

Graduate School of Human–Machine Interaction Linköpings Universitet, S-581 83 Linköping, Sweden

1. INTRODUCTION

An accident can be defined as an unexpected event with an undesirable outcome. Both criteria are important, since neither an expected unpleasant event, nor an unexpected pleasant one qualify as an accident. As long as accidents have happened, people have tried to understand them, both to satisfy themselves that the causes were not in conflict with the accepted wisdom of the world, and to help finding ways to prevent that the same — or a similar — accident should occur again.

The concern for accidents and their causes has attracted considerable interest in the last half of the twentieth century, mainly due to a number of spectacular accidents in complex industrial systems. The more famous of these are the explosion at the Flixborough plant (1974); the accident at the nuclear power plant at Three Mile Island (1979); the explosion of the Challenger space shuttle (1986); the meltdown at the nuclear power plant in Chernobyl (1986); the aircraft collision at Tenerife (1977); and the multitude of problems on the space station MIR (1998).

In accidents such as these, failures of human action and judgement have often been seen as part of the causes, and in a growing number of accidents the main cause has been attributed specifically to "human error". In the 1960s the number was around 30%, but grew during the following decades so that the number at present often is put as high as 70–90%. This development is, of course, not the expression of a simple fact or relation. Rather, there are a number of reasons for this trend, for instance: (1) that technological systems have become more complex, hence more difficult to control; (2) that improved models and methods for "human error" analysis have made this cause more likely; (3) that technological systems have become more reliable, hence raising the relative number of other causes; or (4) that it is sometimes cheaper and more convenient to put the blame on a human than to redesign an entire system.

2. MODELS AND METHODS FOR ACCIDENT ANALYSIS

An accident model is a generalized description of how an accident may have happened. Such models are invariably based on the principle of causality, which states that there must be a cause for any observed event. The early accident models tended to see accidents as caused either by failures of the technology or incorrect human actions ("human errors"). This view was gradually extended to recognize both the contribution of latent system states, and the complexity of conditions that could end in an incorrectly performed human action — even leading to the extreme notion of "error forcing" conditions. In contemporary accident models, a distinction is made between actions at the "sharp" end, which often are the initiating events, and actions at the "blunt" end, which create the conditions that either make an action failure near inevitable or turn minor mishaps into major disasters.

Despite these developments, specifically the increasing sophistication in accounting for the organizational determinants of accidents, there is an almost intransigent preference to refer to "human error" as a singular concept. The history of accident analysis clearly demonstrates that the notion of a cause itself is an oversimplification, since a cause represents a judgement in hindsight rather than an unequivocal fact. This acknowledgement, notwithstanding, accident models seem to be firmly entrenched both in the idea that a "true" or root cause can be found, and in the idea that "human errors" necessarily must be part of the explanations. This view should be contrasted with the pragmatic, so-called ecological, view which points out, first, that action failures are both an unavoidable and necessary element of efficient human performance and, second, that the same types of action failures occur across tasks and domains. Without making shortcuts or using heuristics people would be unable to work effectively, and without failing every now and then they would not be able to learn. The challenge for cognitive systems engineering and artifact design is, of course, to make sure that the systems are so robust that minor variations in performance do not lead to fatal consequences, yet so sensitive that users have the freedom necessary to create and optimize control strategies.

3. THE NATURE OF "HUMAN ERROR"

Since the 1960s many attempts have been made to provide a technical definition of the concept of error. Yet despite the fact that the term "error" has a relatively simplistic meaning in everyday life, the term is extremely difficult to pin down precisely when considered from a technical point of view. One reason is that the same term is used to denote either (1) an outcome or consequence, (2) the act or event itself, or (3) the possible cause. Such ambiguity is clearly not conducive for developing accounts of causes and effects. A further reason is that analyses of the nature and origins of error from different professional points of view, often have quite different and, at times, incompatible premises. Thus, an engineer might prefer to view the human operator as a system component subject to the same kind of successes and failures as equipment. Psychologists, on the other hand, often begin with the assumption that human behavior is essentially purposive and can only be fully understood with reference to subjective goals. Finally, sociologist have traditionally ascribed the primary error modes to features of the prevailing sociotechnical system and in a sociological analysis items such as management style and organizational structure are often hypothesized as the mediating variables that influence error rates.

Irrespective of the above differences there seem to be at least three intuitive parts to any definition of error:

- First, there must be a clearly specified *performance standard* or *criterion* against which a response can be measured. Engineering reliability analysts have traditionally used objective measures, such as system parameters, as the standard for acceptable performance. In contrast, investigators working from the standpoint of cognitive psychology have tended to rely on subjective criteria such as the momentary intentions, purposes, and goal structures of the acting individual. According to this view there are two basic ways that an action can go wrong. In one case the intention to act is adequate but a subsequent act does not

go as intended. Here the "error" is conventionally defined as a slip. In another case, actions proceed according to plan but the plan is inadequate. Here the "error" is usually classed as a mistake (Norman 1981).

- Second, there must be an *event* which results in a measurable *performance shortfall* such that the expected level of performance is not met by the acting agent. Many researchers of human performance ascribe to a pessimistic view according to which "human errors" provide strong evidence of design defects of the human information-processing system. Considerable research has therefore aimed at identifying the design defects of the human mind. Once identified the assumption is that guidelines can be developed to determine in which situations the human operator can and cannot be trusted with the care of complex plant and equipment. A more optimistic viewpoint is found in a line of research that emphasizes that most "human errors" have their origins in processes that perform a useful and adaptive function. This approach takes a much more beneficial view of the variability of human cognition and performance, and relate ''error mechanisms'' to processes that underpin intelligent functioning and especially the human ability to deal with complex data that are characterized by a high degree of ambiguous and uncertainty.

- Third, there must be a degree of *volition* such that the actor has the opportunity to act in a way that will not be considered erroneous. Thus, if something is not avoidable by some action of the person, it is neither reasonable nor acceptable to speak of an error. This, presumably, also includes the so-called error forcing conditions, even though the full implications of that approach have not yet been realized. Factors that occur outside the control of the individual, for example, "acts of God" or acts of nature, are therefore better defined as accidents.

4. ACCIDENT ANALYSIS

Accident analysis denotes the set of methods and principles that are used to find, in a systematic manner, the cause — or causes — of an accident. The engineering practice of accident analysis

is based on the assumption that there is a true or real cause of an accident — sometimes referred to as the root cause. This should be contrasted with the more recent view of, for instance, cognitive systems engineering, according to which the cause is a social construct (Woods *et al.* 1994). Formally speaking, a cause is the identification, after the fact, of a limited set of aspects of the situation that are seen as the necessary and sufficient conditions for the effect(s) to have occurred. In order for a cause to be an acceptable explanation, it must be possible to associate it unequivocally with a system structure or function (people, components, procedures, etc.), and to do something to reduce or eliminate the cause within accepted limits of cost and time.

Accident analysis methods are usually based on a hierarchy of causes, although most hierarchies represent convenience rather than a taxonomy. The classification causes has undergone a significant development during the last 30 years or so (Figure 1). In the early days of accident analysis, i.e. the 1950s and 1960s, causes were roughly seen as belonging to three groups called technical failures, "human error", and anything else – called "other". The development in the categories, paralleled by a development in accident analysis methods, has in particular taken place within the groups of "human error" and "other". As far as the latter is concerned, the major new categories have described various causes of organizational failures, including safety culture, quality assurance, and pathogenic organizations. As far as the former is concerned, the development has been to distinguish between causes in different types of work (management, maintenance, design, and operation), and further to increase significantly the categories for "human errors", going from omissions and commissions to detailed information processing failures or ''cognitive errors''.

4.1 Traditional Human Factors and Ergonomic Approaches

Models of operator error developed within a human factors and ergonomic tradition generally provide a useful framework the development of practical tools to classify errors in terms of their external manifestations. The objective of such analyses is usually to predict the probability of *simple* error forms, which may occur
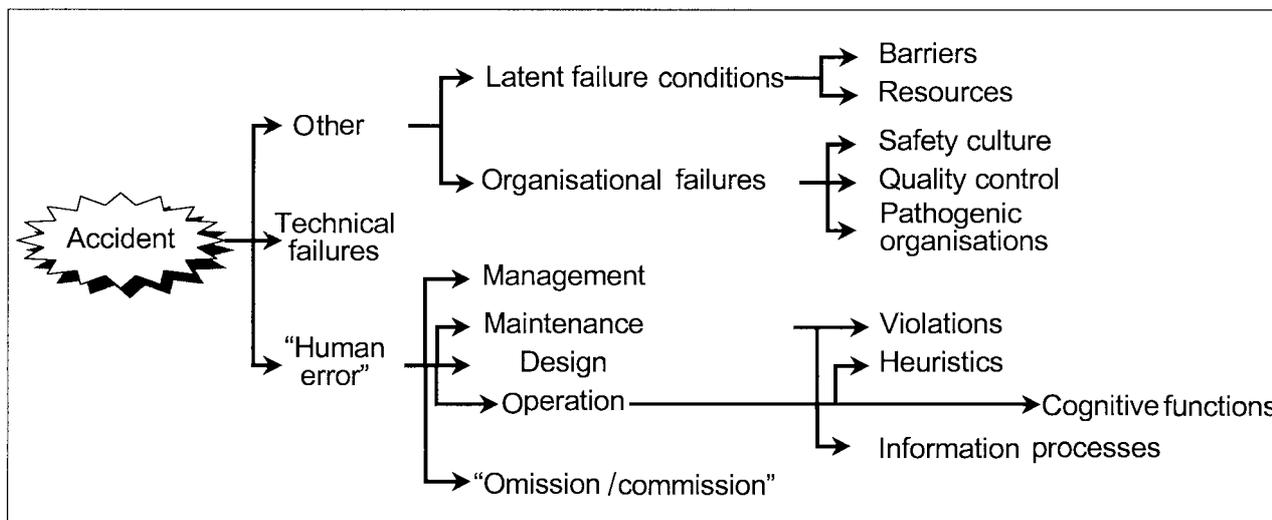

Figure 1: Development in the classification of causes.

in human—machine interactions, such as the distinction between "errors of omission" and "errors of commission". Although more complex models of human error have been specified at the level of observable behavior, the basic limitation of the approach is that it provides little information regarding the possible psychological causes. Thus, the analytic capability of traditional ergonomic models is typically quite low and resultant models contain few general principles that can be used to construct a description of failures in terms of underlying psychological functions.

4.2 Information-processing Models

In contrast to traditional human factors approaches, information-processing models have a highly developed basis and analytic capability, but cannot always easily be converted to useful and practical tools for performance analysis or prediction. The analytic capability of information-processing models derives mainly from the large number of general statements relating to error tendencies or mechanisms that can be incorporated into such models. Since the models often are of a rather loose nature, many of these explanations are unfortunately ad hoc and lack a clear theoretical foundation. (For example, it is commonly assumed that short-term memory is capable of processing a strictly limited number of chunks of information and that demands in excess of this amount typically cause the short-term memory system to fail.) Furthermore, when information-processing models permit the formulation of general predictions that certain types of error will occur, the extent to which such predictions transfer to a real-world setting is unclear. Analyses of accident and near-miss reports, for example, frequently describe situations where information-processing failures occurred although the situations were well within the presumed performance capabilities of the human operator. Conversely, there are many well-documented instances where operators have succeeded to control a situation where information-processing models would have predicted failure.

Information-processing models also face the problem that the capabilities of the human are far from being stable or constant. In analogy with physical artifacts, each of the components of the human information processing system are assumed to have identifiable capabilities, hence also boundaries (high and low). Yet despite heroic attempts to define and measure the basic capabilities, success has been strictly limited. This fact could be taken as an indication that the underlying approach is wrong, i.e. that the human is not really an information processing system — although people in many ways can be described as such with considerable benefits.

4.3 Cognitive-systems Engineering

Models of erroneous actions generated from the standpoint of cognitive systems engineering tend to do quite well in terms of analytic capability, predictive capability, technical basis, relation to existing taxonomies, practicality, and cost/effectiveness. Currently, the main shortcoming is that they fall short of the ideal with regard to their ability to predict likely error forms. They are, however, not bettered in this respect by either ergonomic models or information-processing models. Moreover, cognitive-systems models are particularly strong in terms of their technical content because they are based on viable and well-articulated

models of human action. Cognitive-systems engineering focuses more on how human performance is influence by the internal and external context than on how it can be explained by hypothetical models of cognitive functions and structures. The models therefore provide a better basis for the development of practical tools for error analysis and subsequent reduction.

4.4 Sociotechnical Approaches

Since the late 1980s, a further approach to human error modeling has been on the rise, which places the focus on the organizational and operational context in which the work takes place (Reason 1997). The main background for this approach is the recognition that a common factor in a large number of accidents is the organizational condition. As the view on human error gradually became more refined it was acknowledged that incidents evolved through a conjunction of several failures and factors, and that many of these were conditioned by the context. This evolved into the previously mentioned notion of people at the "sharp end" of a system as contrasted with people and events at the "blunt end". The basic idea is that human performance is a compromise between the demands from the monitored process and the resources and constraints that are part of the working environment. The demands and resources meet at the sharp end, and the result shows itself in the actions that people make. The demands come from the process, but by implication also from the organizational environment in which the process exists. The resources and the constraints are more explicitly given by the organizational context, e.g. in terms of procedures, rules, limits, tools, etc.

The sociotechnical approach is in many ways the inverse of the classical information processing approach, which concentrates on the internal human information processing mechanisms. The sociotechnical approaches therefore run the same danger of focusing on one aspect of the situation, the context, thereby neglecting the other. The sociotechnical approaches are nevertheless valuable as reminders of the need to consider both sides — cognition and context. At present the sociotechnical approaches have not been developed to the stage where they can provide a potential explanation for human erroneous actions. As a matter of fact, they tend to reduce the relative contribution of individuals and of the operator's cognitive functions. It is not inconceivable that in the future more mature sociotechnical approaches will be developed, which can enrich the explanations of human erroneous actions. At present, however, they have not reached this stage.

5. CONCLUSIONS

One lesson of almost 40 years of research into "human error" is that it is not a specific category. Rather, everything is a "human error" if only one goes back far enough. The logical consequence of that is that the analysis may also go a step beyond the "human error", and thereby avoid simple-minded explanations. Although the relativity of the analysis is an unresolved issue, one inescapable conclusion is the realization that human actions are not a special type of causes. Human erroneous actions are rather a symptom of the conditions of work, and of the demands from the organization and the environment (the end users). Technology is often used in clumsy ways that are not well adapted to the needs of the people at the "sharp end". This results in unnatural and

unusual demands on the practitioner that tend to congregate at the higher tempo or higher criticality periods of activity. The manifestation of this has been an apparent epidemic of failures labeled as "human error". Yet it is only by seeing erroneous actions as an indication of missing adaptation, that the findings from accident analyses can be used constructively to understand and prevent accidents.

REFERENCES

NORMAN, D.A., 1981, Categorization of action slips. *Psychological Review*, 88, 1–15.

REASON, J.T., 1997, *Managing the risks of organizational accidents* (Aldershot, UK: Ashgate).

WOODS, D.D., JOHANNESEN, L.J., COOK, R.I., and SARTER, N.B., 1994, *Behind human error: Cognitive systems, computers and hindsight* (Columbus, Ohio: CSERIAC).

Adaptive Automation

M. W. Scerbo

Department of Psychology, Old Dominion University, Norfolk VA, 23529, USA

1. INTRODUCTION

Automation refers to the allocation of functions usually performed by a human to a machine. Automated machines or systems enable humans to increase their span of operation or control, perform functions that are beyond their normal abilities, maintain performance for longer periods of time, and perform fewer mundane activities. Automation can also help reduce operator workload and human error.

On the other hand, automation changes the nature of work and can create new types of problems. Excessive trust in automation can lead to complacent behavior and a diminished ability to notice system failures. Skills can deteriorate when activities traditionally performed by operators become automated. Further, automation can lead to increases in workload when system and operator goals are in conflict.

Adaptive automation can address some of these shortcomings through a more dynamic approach to function allocation. In adaptive automation, the level of automation or the number of systems operating under automation can be modified in real time. In addition, changes in the state of automation can be initiated by *either* the human or the system. Consequently, adaptive automation enables the level or modes of automation to be tied more closely to operator needs at any given moment.

2. BACKGROUND

Adaptive automation has its roots in artificial intelligence. In the 1970s, efforts were directed toward developing adaptive aids to help allocate tasks between humans and computers. By the 1980s, adaptive aiding concepts were being applied to advanced fighter aircraft. A collaborative effort among the Defense Advanced Research Projects Agency (DARPA), Lockheed Aeronautical Systems Company, McDonnell Aircraft Company, and the Wright Research and Development Center, called the Pilot's Associate program, was aimed at providing pilots with an intelligent system that could supply them with appropriate information in the proper format when it was needed. The Pilot's Associate was a network of cooperative knowledge-based subsystems that could monitor and assess the status of its own systems as well as events in the external environment (Hammer and Small 1995). The information could then be evaluated and presented to the pilot. The Associate could also suggest actions for the pilot to take. Thus, the system was designed to function as an assistant for the pilot.

During the same period, researchers and developers in the computer industry were developing adaptive interfaces. For instance, Wilensky, Arens, and Chin (1984) describe their experience with intelligent help for the UNIX operating system. Their UNIX Consultant (UC) was designed to provide general information about UNIX, procedural information about executing UNIX commands, as well as debugging information. The UC had several subsystems designed to analyze the user's queries, deduce the user's goals, monitor the interaction history of the user, and plan and present the system's response. Thus, the UC system can both respond to user queries and offer suggestions when user difficulties are detected.

3. EXAMPLES OF ADAPTABLE AND ADAPTIVE SYSTEMS

Research on adaptive technology has led to some confusion regarding the nature of systems that are *adaptable* and those that are *adaptive*. Table 1 presents a taxonomy of adaptive technology. One dimension addresses the underlying source of flexibility in the system; that is, whether the information displayed or the functions themselves are flexible. The other dimension concerns how the changes are invoked. In adaptable systems, changes among presentation modes or function allocation are initiated by the user. On the other hand, in adaptive systems both the user and the system can initiate changes in the state of the system.

Oppermann and Simm (1994) point to EXCEL as an example of a system where the presentation of information can be modified to meet the user's needs. EXCEL is a spreadsheet program that permits the user to change the structure and layout of dialog boxes. Specifically, the user can substitute a dialog box for the default data entry mode.

An example of an adaptive information system is described by Mason (1986). This system provides a flexible version of the on-line UNIX Programmer's Manual based on the idea of adaptive prompting. Initial requests for the Manual are answered with a default set of instructions that assume the user is a novice. Subsequent queries are monitored and evaluated and the system changes the nature of the information presented to match the level of expertise exhibited by the user. Thus, the system modifies its presentation of information automatically to provide the appropriate level of detail for the determined skill of the user.

The Naval Dynamic Allocation Research Testbed (N-DART) described by Morgan, Cook, and Corbridge (1999) is an example of a system in which functions can be adapted to the user's needs. The user plays the role of an anti-air warfare coordinator whose job is to counter incoming threats quickly and efficiently. The system compiles the tactical picture and evaluates the threats automatically. The operator is responsible for addressing all threats and assigning appropriate weapons to counter the different types of threats. The task of assigning weapons can be allocated to the system. Morgan and her colleagues found that performance was improved when weapon assignment was allocated to the system during periods of high workload.

Researchers at NASA have developed an adaptive system that

Table 1. Taxonomy of Adaptive Technology.

Source of Flexibility	Invocation	
	Adaptable	Adaptive
Information Presentation	MS EXCEL Oppermann and Simm (1994)	Adaptive on-line Programmer's Manual Mason (1986)
Functionality	N-DART Morgan *et al.* (1999)	Biocybernetic, Closed Loop System Pope *et al.* (1995)

uses an index of EEG to moderate task load (Pope, Bogart, and Bartolome 1995). In their system, EEG signals recorded from the scalp are sent to a LabView Virtual Instrument (VI) that computes an index of engagement. Based upon the value of this index, the VI determines whether a change in task mode is required and sends the appropriate signal to a PC running the operator's task. The index is continually updated and comparisons between successive values are assumed to reflect changes in engagement and to initiate switches among task modes. Thus, this system is one in which the functionality is modified in real time to meet the needs of the user.

4. ADAPTIVE STRATEGIES

Recently, Morrison and Gluckman (1994) described a taxonomy aimed at understanding how adaptive automation *ought* to be implemented. One dimension is based upon different methods of modifying system functionality. For instance, entire tasks can be allocated to either the system or the operator, or a specific task can be partitioned so that the system and operator each share responsibility for unique portions of the task. Further, a task could be transformed to an alternative format to make it easier for the operator to perform.

The second dimension described by Morrison and Gluckman (1994) concerns the triggering mechanism for shifting among modes or levels of automation. They describe three ways that this can be accomplished. With goal-based strategies, changes among modes or levels of automation are triggered by a set of criteria or external events. For example, a goal-based adaptive system might automate certain activities only during specific phases of a task typically associated with high workload or when the occurrence of critical events or emergency situations is detected. A second method would use physiological measures that reflect operator workload to trigger changes among modes like that of the Pope *et al.* (1995) biocybernetic system described above. A third strategy would use an index of the operator's performance as a basis for change among modes of operation. This could be accomplished by utilizing models of operator performance. Under this approach, the system would change modes of operation based upon information about the current state of the system, external events, and expected operator actions predicted from patterns of user activity. Alternatively, real-time measures of operator performance could also be used to invoke the changes in automation.

5. AUTHORITY AND INVOCATION

One important issue associated with adaptive automation concerns authority and invocation, i.e. who should have control over changes among modes of automation? One could argue that the operator should always have authority over the system because he or she is ultimately responsible for the behavior of the system. In addition, it is possible that operators may be more efficient at managing resources when they can control changes in the state of automation.

On the other hand, there may be times when the operator needs to change automation modes at the precise moment he or she is too busy to make that change. Further, the operator may not be the best judge of when and if automation is needed. Finally, there are situations where it would be very beneficial for the system to have authority over automation invocation. If lives were at stake or the system was in jeopardy, allowing the system to intervene and circumvent the threat or minimize the potential damage would be paramount. Thus far, this debate is still unsettled.

6. A TEAM PERSPECTIVE

Scerbo (1996) has discussed some of the unique issues facing users of adaptive technology. For instance, both the operator and the system will have to learn the capabilities and limitations of one another. Efficient performance with the system may require extensive training. Although the user and the system will have unique responsibilities, they may have to collaborate on others. During periods of high workload, the system should be capable of stepping in and assuming some of the user's responsibilities. In addition, it is critical that the operator and system are capable of exchanging information freely and effortlessly.

Activities like collaboration, backing up one another, and communication suggest that knowledge of team performance may be useful in designing adaptive technology. There are several factors that are important for efficient team performance including: activities associated with the acquisition and distribution of information to team members, the need to distribute work equitably, setting the proper pace of work and coordinating and sequencing activities among members, setting team objectives and monitoring team performance. Many of the same kinds of functions apply to adaptive technology. For instance, the operator and system need to know the capabilities and limitations of one another, functions must be allocated between the system and operator so as to stabilize workload, and system performance must be continually monitored and evaluated. In fact, Hammer and Small (1995) described the Pilot's Associate as "an electronic crew member".

7. CONCLUSION

Adaptive technology represents the next step in the evolution of automation. These systems will be qualitatively different from those available today. The users' experiences with these systems are apt to be less like working with a tool or machine and more like interacting with a co-worker. Thus, training, communication, and social, organizational, and motivational factors will be far more critical than they are in present-day systems. In addition, the researchers and designers of adaptive automation will also be faced with new challenges. Current system analysis, design, and evaluation methodologies are apt to be woefully inadequate. Standard human factors methods and techniques will not be applicable to many of the human performance issues with this technology. Instead, an understanding of how humans share tasks and information will be more beneficial. Fortunately, the evolution of adaptive automation is still in its infancy. Consequently, designers, cognitive engineers, and psychologists still have time to begin to address the broad range of issues that surround adaptive automation before the technology is widely implemented.

REFERENCES

HAMMER, J.M. and SMALL, R.L., 1995, An intelligent interface in an associate system. In W.B. Rouse (ed.) *Human/technology interaction in complex systems*, Vol. 7, Greenwich, CT: JAI Press, pp. 1–44.

MASON, M.V., 1986, Adaptive command prompting in an on-line documentation system. *International Journal of Man-Machine Studies, 25,* 33–51.

MORRISON, J.G. and GLUCKMAN, J.P., 1994, Definitions and prospective guidelines for the application of adaptive automation. In M. Mouloua and R. Parasuraman (eds) *Human Performance in Automated Systems: Current Research and Trends* (Hillsdale, NJ: Lawrence Erlbaum Assoc.), pp. 256–63.

MORGAN, C., COOK, C.A. and CORBRIDGE, C., 1999, Dynamic function allocation for naval command and control. In M.W. Scerbo and M. Mouloua (eds) *Automation Technology and Human Performance: Current Research and Trends* (Mahwah, NJ: Lawrence Erlbaum Assoc.), pp. 134–38.

OPPERMANN, R. and SIMM, H., 1994, Adaptability: User-initiated individualization. In R. Oppermann (ed.) *Adaptive User Support: Ergonomic Design of Manually and Automatically Adaptable Software* (Hillsdale, NJ: Lawrence Erlbaum Assoc.), pp. 14–66.

POPE, A. T., BOGART, E. H. and BARTOLOME, D., 1995, Biocybernetic system evaluates indices of operator engagement. *Biological Psychology,* 40, 187–96.

SCERBO, M.W., 1996, Theoretical perspectives on adaptive automation. In R. Parasuraman and M. Mouloua (eds) *Automation and Human Performance: Theory and Applications* (Mahwah, NJ: Lawrence Erlbaum Assoc.), pp. 37–63.

WILENSKY, R., ARENS, Y. and CHIN, D.N., 1984, Talking to Unix in English: An overview of UC. *Communications of the ACM,* 27, 574–93.

Affordances

M. J. Dainoff and L. S. Mark

Center for Ergonomic Research, Department of Psychology, Miami University, Oxford, OH 45056, USA

1. DEFINITION

Affordance refers to a fundamental relationship in ergonomics — between the user and those elements in the world with which the user interacts during the performance of goal-directed actions. Insofar as the basic problem in ergonomics can be regarded as elucidating the *fit* between the human being and those things (tools, workplaces, environments) with which he or she interacts (Grandjean 1982, Dainoff and Dainoff 1986), affordances become a conceptual tool for ergonomists to assess goodness of fit.

The concept of affordance was developed by the perceptual psychologist J. J. Gibson (1979) to refer to the mutuality between physical attributes of people and their environments with respect to the behavioral acts required to function in that environment. For example, consider the simple act of reaching for an object across a table. The object can be said to afford reaching without standing if the actor's arm and trunk are large enough relative to the distance of the object. Gibson was primarily interested in a theory of perception and action; for him the challenge for the actor was to perceive whether the existing layout of surfaces in the environment would support a particular goal-directed action. This work evolved into a clearly defined perspective, complete with its own journal, called ecological psychology. The focus of ecological psychology has been rooted in evolutionary biology (relating the perceptual and motor capabilities of animals — including humans — to physical attributes of the environments in which they evolved). Recently ergonomists have attempted to utilize this approach, and particularly the concept of affordance, to examine the environments and tools constructed by humans.

Affordances were introduced to a much broader audience by Norman (1988) in his *Psychology of Everyday Things* (albeit with a different theoretical perspective). Affordances have been explicitly included as a component of Helander's (1997) overview of the field in his Presidential Address to the International Ergonomics Association. There is now a distinct perspective which can labeled ecological ergonomics (Flach 1995).

In particular, for complex work environments, we can conceptualize hierarchies of affordances, in which the goals at one level of the hierarchy serve as means for accomplishing goals at a higher level of the hierarchy (Vicente and Rasmussen 1990, Vicente 1999). For example, the higher-order goal may be wanting to purchase a book. At the first level of the hierarchy, signing on to the Internet affords locating an online bookseller's web site. At the web site, finding the book in a catalogue (second level of the hierarchy) affords making the purchase.

Affordances can be defined as those properties of the environment of an animal that have consequences for the animal's (goal-directed) behavior (Turvey 1992). This definition can serve as a starting place for unpacking the concept of affordance as a tool for assessing fit. (See Karwowski's 1991 discussion of compatibility for a complementary theoretical perspective on this issue.) Ultimately, we want to ensure that the designed environments are able to support goal-directed actions that are healthy, safe and productive.

2. CONSTRAINTS ON ACTION CAPABILITIES

What are the determinants of a person's action capabilities in a given environment? We can define two classes of constraints. Personal constraints refer to both body dimensions (anthropometry) and biodynamic constraints (body mass, strength, flexibility, etc.) related to the individual acting within an environment. Environmental constraints include both size and shape of objects and surfaces as well their physical properties relevant to the action. Thus, the highest step a person can climb bipedally is determined by his/her leg length and the ability of leg muscles to generate sufficient force to lift the center of mass over the surface of support (Warren 1984). Similarly the surface of support must support the action. It can't be too high relative to the person's leg length, it must be sufficiently flat, strong enough to support the person's body weight, and not too slippery.

It should be emphasized this ratio between leg length and step height is a constant proportion across people of different sizes (Warren 1984). Mark and Vogele (1988) showed a similar relationship with leg length for the perception of maximum sitting height. Other such intrinsic relationships have been found for reaching, stepping through apertures, stepping across gaps, and grasping; see Dainoff *et al.* (1999) for a more extended discussion.

3. PERCEIVING AFFORDANCES

It appears that information about affordances relating to simple actions can be scaled in terms of relevant dimensions of the user's body. Moreover, there is evidence (Mark 1987) that people actually use this information in determining their capabilities to perform goal-directed actions. However, as Mark *et al.* (1990) indicate, active exploration by the user of his/her environment may be essential to revealing the critical information that specifies the affordance. In their study, they perturbed users' judgement of maximum seat height by having them wear blocks on their shoes. This artificial change in leg lengths was appropriately compensated for by those users who were permitted actively to explore their environment while wearing the blocks. However, users who did not explore the environment did not compensate for their new "leg" lengths.

What is the role of cognition in all of this? First, the same object or environment may be an affordance for action and one situation and not in another. What may change is the nature of the goal-directed activity (i.e. task). Thus, a skilled carpenter will use one grade of sandpaper to prepare a surface for painting, and a finer grade for finishing. This user must understand which affordances (sandpaper density) are relevant to achieving the goal, and be able to focus attention on the relevant information specifying those affordances (e.g. numeric printed code, visual and tactile appearance). This is made more or less difficult depending on how well the designer understands the importance of such affordance-specifying information to the user. Finally, there are typically many different action patterns leading to a given goal; some of those are more likely to enhance health, comfort, and productivity than others. Part of the user's task to select from among multiple affordances, those which support more desirable action patterns.

4. AN EXAMPLE: THE ECOLOGY OF SITTING DOWN TO A MEAL

Consider the simple problem of sitting down to eat a meal. We assume that the user will require both a chair and a table. (This assumption includes an implicit recognition of a broader social/cultural context including the act of eating being accompanied by eye contact and conversation among a group of diners.) These two elements of the environment (chair and table) should afford the possibility of the user being able to attain a reasonable posture while manipulating eating tools (knife, fork, spoon) and food objects. The question is: can a typical dining room chair be considered an affordance for the particular action sequence called sitting down to eat? The answer is yes only if the user in question is an adult. It would not be an affordance if the user in question is a 3-year-old child. Thus, within the definition of affordance we are required to simultaneous assess the physical characteristics of the chair and table along with the capabilities of the user — in this case, anthropometric attributes such as leg height and seated elbow height.

The situation is worth examining in more detail. The action sequence of eating at a table (within the context of modern Western culture) entails a certain set of postural requirements. The food must be manipulated with either tools (knife, fork, spoon) or, for certain foods, directly with the fingers. There is an envelope of postural configurations within which these actions can occur with some degree of comfort and effectiveness.

The goal is for food items to be divided into portions whose dimensions are compatible with the size of the mouth and jaw, and to be transported from the plate to the mouth (with minimal spillage). To accomplish this goal, the trunk must be not far from vertical, the forearms and hands not far from horizontal and slightly above the plate. For a adult whose relevant anthropometric dimensions fall within a certain range, these postural demands can be achieved; hence, the table–chair *system* is an affordance for this particular action sequence for this particular group of users, in the sense that its properties have behavioral consequences for the these users.

It should be noted that, even within the envelope of acceptable postural configurations, there will be variability in comfort and effectiveness. People with shorter trunks and longer upper arms will find the table height a bit too high; people with longer legs will have to angle them downward and backwards to keep from bumping into the underside of the table. In both cases, comfort and even efficiency may be reduced.

However, at some point, the user's postural capabilities will fall outside of the envelope of acceptability, and the system will no longer afford this action sequence. This is, of course, the case with small children. The seat height may be too tall for the child to climb onto without assistance. However, even if the child is helped into the chair, the table will be so far above the location of the elbows, that the postural requirements for food manipulation and transport cannot be met.

In considering common solutions to this problem, an additional user capability must be considered; namely, the fine motor control of fingers required to achieve effective manipulation of food tools and food objects. *Effective* in this case can be operationally defined in terms of absence of spilled food, and amount of food actually transported. (This could be further elaborated in much greater detail within the social context of 'table manners' or etiquette.)

One solution is to provide a high chair with detachable tray. This eating system is essentially a miniature version of the adult size chair and table, in which the distance between seat surface and eating surfaces corresponds to a child's anthropometric dimensions. Thus, children within a certain range of seated elbow heights will have the capability of postural adjustments within the required envelope. At the same time, this system has the flexibility of fitting into adult social action patterns. It can either be made an adjunct to the adult table — allowing the child to be part of the social interaction — or it can be independent. Thus, by raising the child's eye height to the approximate level of the adults, this solution affords the possibility of the child either eating in the direct presence of adults or at least having the possibility of eye contact with surrounding adults even if not at the table. On the other hand, the high chair (usually) requires that the child be lifted into and out of the chair by an adult.

A second solution involves child-sized furniture. This affords appropriate postural configurations necessary for eating, but at the cost of social interaction with adults. There are, of course, circumstances where this particular definition of the action pattern sitting and eating is desirable (e.g. children's parties, large family meals with several children).

5. ANALYSIS OF CONSTRAINTS

The above example represents the application of affordance analysis to a specific case of practical ergonomic design. In particular, theory and practice come together in considering how the relationship between a goal-directed user and his/her environment is integrated with respect to constraints (personal and environmental) on the user's actions. The sitting and eating example can be used to elaborate these constraints.

With respect to personal constraints, we can first consider anthropometric dimensions scaled with respect to corresponding environmental dimensions; i.e., the child is too small to use an adult chair. Biodynamic constraints refer to the agility and strength necessary to carry out the manipulative actions of handling tools and food while maintaining the appropriate supporting posture (i.e. trunk and head upright and arms and hands extended).

For the individual action capabilities (for eating) to be expressed in *performance*, support is required from relevant properties of the environment. That is, design of furniture dimensions (environmental constraint) must be scaled to human body dimensions (personal constraints). In the pre-industrial era, tools and clothing were crafted individually for each person, so that these scaling relationships were natural. With the advent of mass production, sizing becomes an issue. A limited number of products must be designed to fit a variable population of users. In the clothing industry, fit is approximated by a limited number of standard sizes which can, in certain circumstances, be modified by alterations. Whitestone and Robinette (1997) provide an excellent discussion of fit in the design of military helmets). In transportation, and, more recently, office furniture, fit is approximated through adjustable dimensions.

However, household furniture, such as dining room tables and chairs, are not adjustable. Why is this, since, fit in terms of mapping of individual and environmental constraints, is rarely perfect? In the case of sitting at a table to eat, the consequences

of a moderate degree of lack of fit (reduced comfort and performance) are typically not serious. The meal does not last long enough for the lack of affordance to matter. (Western travelers who have had to eat while kneeling through a traditional Japanese banquet may disagree!)

6. AFFORDANCES FOR COMPUTER WORKSTATIONS

Suppose the dining room table is converted into a computer workstation. This work system, which previously afforded a relatively comfortable dining experience, will not, for many users, support the same level of comfort and performance for prolonged periods of computer use. The postural envelope will be somewhat different, but the duration and intensity of action sequences will be considerably longer. At this point, the nature of the task becomes important.

Tasks are an *expected* set of goal-directed actions that are to be executed by the individual within some physical environment. Inappropriate tools (environmental constraints) for a given task result in poor fit. If the mismatch among individual, environmental and task constraints are such that the postural envelope is exceeded (awkward posture and excess force) for prolonged periods of time without adequate rest pauses, the result may be a class of musculoskeletal disorders which have been given the misnomer of "ergonomic illnesses" (see National Research Council 1998 for a review of these issues).

The now relatively widespread availability of adjustable "ergonomic" chairs, keyboard/input device support surfaces and worksurfaces for computer workstations reflect the general realization of the level of demands associated with prolonged computer work. The result is that user is now faced with the problem of coordinating multiple degrees of freedom of adjustability (Dainoff 1998). In a modern task chair, the seat surface may be tilted upward and downward, raised up and down, and slid backwards and forwards. The backrest may be inclined and each armrest may be raised up and down and angled in and out. Keyboard support surfaces are available which also move up and down, side to side, and in and out. Monitor support surfaces do the same.

Consequently, such workstations should represent an effective affordance for prolonged computer work in that the multiple degrees of freedom of adjustment will allow task-appropriate postural configurations for a wide range of individual users. Unfortunately, the user is often either not aware of the range of adjustment mechanisms available, or if aware, does not understand why such adjustments are important (Mark and Dainoff 1988). Thus, the added adjustability is not either used or even misused. In short, the user does not perceive the affordances of the system.

The affordance research of Mark *et al.* (1990) discussed earlier can be easily related to the problem of instructing users of adjustable chairs and worksurfaces. Users must be shown how to actively explore the ranges of adjustability and encouraged to do so as to reveal the potential ranges of postural orientations relative to their own body dimensions. Thus, in a real sense, the thoughtful designer of office furniture is not only creating affordances, but enhancing the user's opportunities for perception of the information specifying those affordances.

7. GENERALIZATION

We have seen that, to the extent that the combination of personal and environmental constraints afford goal-direction action patterns which are safe, healthy, and productive, fit can be said to have been achieved.

The preceding discussion has involved constraints that have been primarily physical; however, the same kinds of considerations apply to more cognitive domains.. Getting lost while navigating through a web site might more clearly be expressed in terms of lack of perceivable affordances. A case in point is the emerging discipline of ecological interface design (Vicente and Rasmussen 1990, Vicente 1999). These authors have demonstrated that is possible to design displays of complex industrial processes which reveal the underlying affordances for operator control. This work provides ample documentation of the usefulness of affordance theory in designing practical solutions to complex problems utilizing hierarchies of mean–ends relationships.

REFERENCES

DAINOFF, M.J. and DAINOFF, M.H., 1986, *People and Productivity: A Manager's Guide to Ergonomics in the Electronic Office* (Cincinnati: MDA).

DAINOFF, M.J., MARK, L.S. and GARDNER, D.L., 1999, Scaling problems in the design of work spaces for human use. In P.A. Hancock (ed.), *Human Performance and Ergonomics: Handbook of Perception and Cognition* (2nd. ed.) (New York: Academic Press).

DAINOFF, M.J., 1998, Ergonomics of seating and chairs. In W. Karwowski and W. Marras (eds), *Occupational Ergonomics Handbooks* (Boca Raton: CRC Press).

FLACH, J., 1995, The ecology of human–machine systems: a personal history. In J. Flach, P. Hancock, J. Caird and K. Vicente (eds), *Global Perspectives on the Ecology of Human–Machine Systems* (Hillsdale: Lawrence Erlbaum).

GIBSON, J.J., 1979, *The Ecological Approach to Visual Perception* (Boston: Houghton Mifflin).

GRANDJEAN, E., 1982, *Fitting the Task to the Man: An Ergonomic Approach* (London: Taylor & Francis).

HELANDER, M., 1997, A systematic definition of ergonomics. In *Proceedings of the 13th Triennial Congress of the International Ergonomics Association*, Tampere, Finland.

KARWOWSKI, W., 1991, Complexity, fuzziness, and ergonomic incompatibility issues in the control of dynamic work environments, *Ergonomics*, 34, 671–86.

MARK, L., 1987, Eye height-scaled information about affordances: a study of sitting and chair climbing, *Journal of Experimental Psychology: Human Perception and Performance*, 13, 361–70.

MARK, L.S. and DAINOFF, M.J., 1988, An ecological framework for ergonomic research. Innovation, *Journal of the Industrial Designers Society of America*, 7, 8–11.

MARK, L.S. and VOGELE, D., 1988, A biodynamic basis for perceived categories of action: a study of sitting and stair climbing, *Journal of Motor Behavior*, 19, 367–84.

MARK, L.S., BALLIET, J.A., CRAVER, K.D., DOUGLAS, S.D. and FOX, T., 1990, What an actor must do in order to perceive the affordance of sitting, *Ecological Psychology*, 2, 325–66.

NATIONAL RESEARCH COUNCIL (NRC), 1998, *Work-related Musculoskeletal Disorders: A Review of the Evidence* (Washington, DC: National Academy Press).

NORMAN, D., 1988, *The Psychology of Everyday Things* (New York: Basic).

TURVEY, M.T., 1992, Affordances and prospective control: an outline of the ontology, *Ecological Psychology*, 4, 173–87.

VICENTE, K., 1999, *Cognitive Work Analysis: Towards Safe, Productive, and Healthy Computer-based Work* (Hillsdale: Lawrence Erlbaum).

VICENTE, K.J. and RASMUSSEN, J., 1990, The ecology of human–machine systems II. Mediating 'direct perception' in complex work domains, *Ecological Psychology*, 2, 207–49.

WARREN, W.H., 1984, The perception of affordances: visual guidance of stair climbing, *Journal of Experimental Psychology: Human Perception and Performance*, 10, 683–703.

WHITEHEAD, J.J. and ROBINETTE, K.M., 1997, Fitting to maximize performance of HMD systems. In J. Melzer (ed.), *Head-mounted Displays: Designing for the User* (New York: McGraw-Hill).

Automation in Vehicles

N.A. Stanton and M.S. Young

Department of Psychology, University of Southampton, SO17 1BJ, UK

1. INTRODUCTION TO VEHICLE AUTOMATION

The theme of vehicle automation is likely to receive increased attention from ergonomists and psychologists in the coming years as the gap between concepts, prototypes, and production vehicles narrows. Although much is written on ergonomic issues connected with vehicle operation, such as occupant packaging, driver comfort, safety, design of controls and displays (see Peacock and Karwowski 1993) little attention is given to vehicle automation. Even texts which address future vehicles tend to concentrate on Road Traffic Informatics and not automation (Parkes and Franzen 1993).

The need for vehicle automation has been well rehearsed by the engineering community. Stanton and Marsden (1996) have introduced three main arguments in favor of vehicle automation. The first argument assumes that driving is an extremely stressful activity and consequently, the suggestion goes, automating certain driving activities could help make significant improvements to the driver's well-being. The second argument is based on the fact that human error constitutes a major cause of road accidents; thus it could be reasonably suggested that the removal of the human element from the control loop might ultimately lead to a reduction in accident statistics. The final argument is based on economic considerations and presumes that automation will enhance the desirability of the product and lead to substantial increases in unit sales.

Examples of new vehicle control systems that replace drivers include Adaptive Cruise Control (ACC) and Active Steering (AS). ACC controls both speed and headway of the vehicle, slowing the vehicle down when presented with an obstacle and restoring target speed when the obstacle is removed. In this way ACC differs from traditional Cruise Control (CC) systems. In traditional cruise control, the system relieves the driver of foot control of the accelerator only (i.e. relieving the driver of some physical workload), whereas ACC relieves the driver of some of the decision-making elements of the task, such as deciding to brake or change lanes (i.e. relieving the driver of some mental workload), as well as physical demands of accelerator control. Likewise AS replaces the driver by guiding the vehicle along the road and maintaining the vehicle within the lane. This is achieved with the aid of onboard cameras that are able to detect the lane markings and therefore the position of the vehicle within the lane. This system should reduce the incidence of lane sharing (driving across both lanes) and lane creep (weaving in and out of lanes).

Potentially, then, automation as embodied by ACC and AS are welcome additional vehicle systems that will add comfort and convenience to the driver. However, certain ergonomics issues do arise when considering any form of automation and these need to be addressed to improve overall system performance (Stanton and Marsden 1996). It is envisaged that although these systems will behave in exactly the manner prescribed by the designers and programmers, this may lead to some scenarios in which the driver's perception of the situation is at odds with the system operation (Stanton and Marsden 1996). There is also little known about the effects of combining advanced vehicle systems — for example, the operation of ACC and AS together — and the relative merits compared to either system operated alone.

2. COMPARING STUDIES OF VEHICLE AUTOMATION

To our knowledge, there are only three empirical studies in the public domain that report upon the effects of vehicle automation on driving performance. We will briefly review these studies in this section.

The first study to be considered was conducted by Nilsson (1995) who sought to compare driver behavior and workload in critical scenarios in manual and ACC conditions. The driving simulator was a moving base Saab 9000 with automatic gear box. Nilsson claims that the simulator was able to evoke 'impressions, reactions and actions which are very close to those experienced by the driver in real driving' (Nilsson 1995: 1255). Nilsson compared drivers' behavior between the manual and ACC conditions in critical traffic situations. All of these scenarios required intervention by the participant. Workload was measured using the NASA Task Load Index, which is a subjective rating scale.

The second study sought to examine the ability of drivers to reclaim control under an ACC failure scenario (where the ACC system fails to detect a vehicle in its path), and compare the level of mental workload with that under manual control of the vehicle (Stanton, Young, and McCaulder 1997). A fixed-base driving simulator based on a Ford Orion with automatic gear box was used (the Southampton Driver Simulator). On the basis of the first study, it might be reasonable to expect drivers to have some difficulty in detecting ACC failure. Measures were collected of all primary driving task performance data and secondary task (using the rotated figures task) data was collected to provide a measure of driver workload and compare task demand in manual and ACC scenarios. In addition, the automated condition was designed to present a failure situation that is anticipated in ACC operation. Participants were instructed to follow a lead vehicle at a speed and distance that felt comfortable so that driving demands could be held constant in the manual and automated conditions.

The third study was devised to investigate driver behavior and workload demanded by the driving task in manual and automated scenarios (Young and Stanton 1997). The automated scenarios were Active Steering (AS), Adaptive Cruise Control (ACC) and combined automation (AS and ACC together). Studies I and II only considered the implementation of a single system, whereas study III considered the effects of combining systems. Study III used the Southampton Driver Simulator (as used in study II). Drivers were instructed to follow a vehicle at a comfortable distance for the each of the experimental trials and were also asked to attend to the secondary task whenever they could. After completing each trial, participants were asked to complete a workload questionnaire.

Taken together, the three studies seem to show that automation will have an impact upon driving performance. First, it seems the parameter under automated control is held at a more consistent value than when it is under manual control; for example, ACC holds the target speed of a vehicle more constant than manual accelerating and braking, and AS keeps the vehicle

more consistently in the center of the lane than manual steering. Second, for some functions (e.g. AS and combined ACC and AS) automation is accompanied by reductions in driver workload. Third, some drivers fail to intervene effectively in automation failure scenarios. Finally, subjective reports suggest that the driving public judge automated systems in a fairly positive manner.

3. PSYCHOLOGICAL ISSUES

An in-depth analysis of the psychological factors associated with the operation of automated systems is required to enable recommendations to be developed. This study plans to investigate the seven main psychological issues in detail. First, is the issue of *locus of control*, the extent to which removal of control from the driver affects performance of the vehicle? Second, is the issue of *trust* that the driver has in the automated systems? Third, is the *situational awareness* of the driver about the operational status of the technological system and the driving context? Fourth, and connected to the third issue, is the issue of *mental representation* that the driver builds up of the automated systems. Fifth, is the issue of *mental and physical workload* associated with automation? Sixth, is the issue of *feedback*, comparing human and automated intervention? The final issue is that of *driver stress* and its implications for vehicle automation. All of these factors are well established in the psychological literature but have yet to be fully explored with respect to vehicle automation.

4. CONCLUSIONS

Whilst these latter arguments are rather speculative at present, and there is clearly much more research effort required, we do hope to indicate pertinent issues for predicting and explaining the results of the research. We feel that the studies presented show that the simulator environment is ideal for this research, as it provides an environment which is both safe and repeatable.

We also feel that the data from the moving-base simulator look comparable to that from the fixed-base simulator. This suggests that the data from fixed-base simulator may be equally valid. The studies do suggest that there are some effects of automation which impact upon driving performance that need exploring further. Finally, we would like to suggest a framework for psychological research into vehicle automation that comprises the issues raised in the final section. Rather than link each issue to every other issue, or make no links at all between the issues, we have attempted to identify the obvious links. This is offered as a starting point for our investigations, rather than a fully validated model.

REFERENCES

NILSSON, L., 1995, Safety effects of adaptive cruise control in critical traffic situations. *Proceedings of the Second World Congress on Intelligent Transport Systems: 'Steps Forward',* Volume III: 1254–9 (VERTIS).

PARKES, A. M. and FRANZEN, S., 1993, *Driving Future Vehicles* (London: Taylor and Francis).

PEACOCK, B. and KARWOWSKI, W., 1993, *Automotive Ergonomics* (London: Taylor and Francis).

STANTON, N. A. and MARSDEN, P., 1996, From fly-by-wire to drive-by-wire: safety implications of vehicle automation. *Safety Science*, 24(1): 35–49.

STANTON, N. A. and YOUNG, M., 1998, Vehicle automation and driving performance. *Ergonomics* 41(11): 1014–28.

STANTON, N. A., YOUNG, M., and McCAULDER, B., 1997, Drive-by-wire: the case of mental workload and the ability of the driver to reclaim control. *Safety Science*. 27(2–3): 149–59.

YOUNG, M. and STANTON, N. A., 1997, Automobile automation. *International Journal of Cognitive Ergonomics* 1(4): 325–36.

Compatibility

C. D. Wickens

University of Illinois, Institute of Aviation, Aviation Research Lab, 1 Airport Road, Savoy, IL 61874, USA

1. OVERVIEW

Compatibility refers to "the capability of … performing … in harmonious … or congenial combination with another or others" (*American Heritage Dictionary*). As applied to the psychology of information processing, the concept thus refers to the congenial or harmonious combination of information representation at a set of entities (i.e., more than one). In the context of human factors (human interaction with systems), five such entities concern us: (1) information represented within the system; (2) information displayed by the system; (3) information perceived by the human; (4) information understood by the human using their cognitive mental model; and (5) information associated with the response. From the perspective of compatibility, the representation of the display and within perception can be combined into a single entity. In psychological research, the term "stimulus" usually refers to perception. In human factors applications, the concern is typically with the display, which is perceived. Since compatibility defines the harmonious relationship or *mapping* between sets of entities, there are several types of compatibility to consider here. High compatibility between the display and either the mental model or the response is a critically important feature for effective systems design (Wickens 1992, Proctor and Van Zandt 1994).

2. STIMULUS RESPONSE COMPATIBILITY

Historically the compatibility concept was first articulated in terms of the compatibility of *spatial* relations between a set of stimuli and a set of responses (Fitts and Seeger 1953). Compatible relations between stimuli and responses lead to shorter response times and fewer errors in responding to the stimuli. These relations can be described either in terms of location or in terms of movement.

2.1. Location Compatibility

Location compatibility is revealed by the rapid response when the location of the response is close to, or involves movement toward, the location of the stimulus. For example, the finger touches an illuminated light, or touches a key that is closer to that light than to any other light (solid arrows of Figure 1(a)). This defines a mapping of "colocation", which supports effective design. However, when stimuli and responses (or displays and controls) cannot be colocated, stimulus-response (SR) location compatibility can still be achieved by creating an array of stimuli that is *congruent* with the array of controls (Figure 1(b)). For example, controls on the left of a linear control array should be mapped to displays on the left of the display array, even if the two arrays themselves are quite distant from each other. A circular array of controls will be compatibly mapped to a circular array of displays, etc.

Good location compatibility, defined by either colocation or congruence, is the basis for effective design of control-display interfaces (such as stovetop controls for stove burners), because
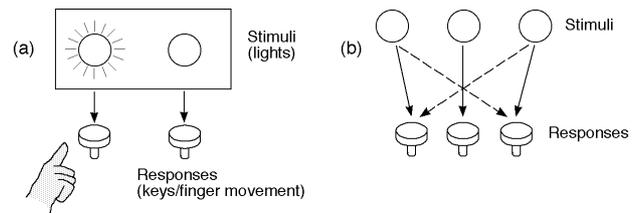
Figure 1. (a) Compatible (colocated) mapping between stimulus lights (top) and response (below), which can either be a movement toward the source or depressing a key or button. (b) Compatible (congruent) mapping between an array of displays and an array of controls, as shown by the solid arrows. Note that the stimuli are not themselves close to the responses. The dashed arrows represent an incompatible mapping.

the mapping is visually apparent and does not require any learning or training to use (Norman 1988).

2.2. Movement Compatibility

When controls involve some degree of motion, as when manipulating a joystick, mouse, slider or rotary dial, then compatibility is preserved when the direction of motion of the control is congruent with the direction of movement of a display, which is to be either tracked by the movement, or presents feedback of system elements controlled by the movement. For example, a circular (steam gauge) display is movement compatible with a rotary control but not with a linear control; a screen symbol should move upward (not downward or sideways) to reflect the action of a vertical control activity.

2.3. Noncompatible Mappings

Mappings that violate either location or movement compatibility may do so in one of two qualitatively different forms. *Incompatible* mappings are those in which the stimulus directly signals an inappropriate *but available* response. For example, considering the dashed arrows of Figure 1(b), a display on the left is mapped to a control on the right. Such mappings offer a direct invitation for error, because the appearance of the stimulus may directly activate the inappropriate response. *Acompatible* mappings are those when the response that might be signaled by the stimulus is not available. For example, a rotary dial mapped to a linear indicator, or a vertical array of lights mapped to a horizontal array of indicators, or the appearance of the digit "1" mapped to the spoken response "A". An acompatible mapping will not necessarily lead to greater errors, but will generally produce longer response times. Acompatible or noncompatible mappings result whenever there are added mental transformations between the stimulus and the response. For example, the mapping $1 \rightarrow A$, $2 \rightarrow B$, $3 \rightarrow C$ imposes such a transformation, and is thus less compatible than the mapping $a \rightarrow A$, $b \rightarrow B$, $c \rightarrow C$.

2.4. Modality-based Compatibility

Auditory stimuli appear to be more compatible with voice responses, and visuospatial stimuli more compatible with manual responses (Teichner and Krebs 1974, Wang and Proctor 1996). The basis for this effect is said to lie in the nature of the feedback provided by the responses. When the feedback itself is congruent with the appropriate stimulus, this "ideomotor" compatibility is

attained (Greenwald 1970). This will be the case if the correctly mapped vocal response is to repeat the sound of the auditory stimulus, or if the hand moves in the same direction as an illuminated light. Spatial compatibility effects appear to be most pronounced when modality compatibility is high (Wang and Proctor 1996).

2.5.   Expectancies and Population Stereotypes

Although many aspects of SR compatibility can be accounted for by the direct spatial or modality relationships between stimuli and responses (or displays and associated controls), many others are mediated by the *expectancies* of the operator regarding these mappings. Because these expectancies are generally learned through experience with past mappings, and because that experience is typically shared by a group of people, the expectancies are typically called *population stereotypes* (Smith 1981). For example, it is a population stereotype in North America to expect that movement of a switch upward will activate power; whereas in Europe people expect it to remove power. Population stereotypes often mediate acompatible relationships. For example, it is expected that rotating dials clockwise will increase a controlled variable, and it is expected that increase will be signaled by an upward or rightward movement on a linear display. It is also expected that the digits 1, 2, 3 will be mapped to the letters A, B, C, because both scales are mapped according to an experienced learned ordering. Because these population stereotypes are embodied in expectancies, based upon knowledge and learning, they form the basis of *cognitive compatibility*.

3.   COGNITIVE COMPATIBILITY

Cognitive compatibility expands upon SR compatibility in two respects:

- It accounts for behavior that is not stimulus driven, but characterizes instead that of the operator who *intends* to perform a particular action (a cognitive plan, such as "increase a variable"), responds in such a way as to execute the intention, and then evaluates system feedback of the action.

- It accounts for cognition and understanding in many environments in which human responding itself is relatively infrequent or impoverished, compared to the amount of displayed information. Thus, in such circumstances, it is important that the displayed representation of the system is compatible with the operator's expectancies of how the system behaves, in order to support better understanding of system changes.

These operator expectancies have often been embodied in the psychological concept of the *mental model* of complex systems (Norman 1988). The mental model represents the operator's conception of the physical and causal relations of the real system under control or supervision. A mental model can be accurate or inaccurate. When discussing cognitive compatibility it is important to preserve compatibility between three entities: the system, the display, and the mental model. Compatibility between the physical system and the mental model is achieved through appropriate training (or learning) of the system. If these two are compatible (i.e., the operator correctly understands the system dynamics and relations) then compatibility between the display and both the mental model and the true system, will support effective system monitoring. In this regard it is important to con-

sider the synergism between display and learning. That is, a display which is compatible with the physical system dynamics will serve as an effective training aid to form a mental model which is compatible with those dynamics.

For relatively simple systems, with few dynamic components, cognitive compatibility can be relatively easy to achieve. For example, the *principle of the moving part* (Roscoe 1968) dictates that moving elements on a display should move in the same direction and pattern of the operator's expectation of that motion. The *principle of pictorial realism* (Roscoe 1968) dictates that the configuration of elements on a display should be congruent with (and therefore a pictorial representation of) their counterparts in the real system. However, in designing compatible displays for more complex systems, like those involved in the supervision of nuclear power plants or other energy conversion systems, where there may be hundreds of time-varying elements, it becomes far more challenging to achieve cognitive compatibility.

There must be careful analysis of the constraints between various system elements, and careful analysis of the relationship between these constraints and the goals of the designer and system monitor/supervisor. This analysis is embodied in the design of what are called *ecological interfaces* (Vicente and Rasmussen 1992). Thus, while one aspect of cognitive compatibility may be achieved by configuring the location of and connections between displayed elements, in a way that represents the location and connections in the real system (the so-called mimic diagram), a second aspect must depict constraints and relationships between more abstract system variables in a way that is compatible with the well-trained operator's mental model. These more abstract variables may not have a particular physical location within the system. For example, such a display could show the conservation of mass and energy

Figure 2. Ecological interface display showing the relationship between flap settings, bank, and airspeed in an aircraft, as these are configured to depict the potential for an aircraft stall. As each of these variables changes value, it will affect the area of the rectangle, which is the stall margin. If this area reduces to zero, the aircraft will stall (Reprinted from Wickens and Andre, 1990, with permission of the Human Factors and Ergonomics Society).

in a thermodynamic process, or the relations between airspeed and angle of attack in an aircraft to represent the stall margin of an aircraft. It would present these variables in a way that is compatible with how a well-trained supervisor understands their relationship (Figure 2).

4. PROXIMITY COMPATIBILITY

Underlying the configuration of any display of multielement systems, which must therefore include multielement displays, is the concept of proximity or nearness. Furthermore, the concept of proximity can be defined across all four entities of the compatibility representation: system, display/perception, cognition, response. Two display elements can be brought closer together or moved further apart; two response devices can be varied in the same manner. Furthermore, cognitive or processing proximity can be defined by the extent to which two variables need to be related (compared or otherwise integrated) in performing the task at hand, and system proximity refers to the proximity of elements in the system itself. The *proximity compatibility principle* (Wickens and Carswell 1995) dictates that compatible relations will exist to the extent that proximity relations are preserved across the three different information processing stages. For example, if the task processing (cognition) requires close proximity of information channels (information integration or comparison), the displayed elements corresponding to those channels should also be close. The converse proximity relation applies as well. If two channels of information should be processed independently (lower task proximity), their processing will be relatively better served by display channels that are not in as close proximity.

The concept of proximity at various stages can be elaborated considerably. For example, task (or cognitive) proximity between two information channels may include simple comparisons (Is the actual value of a quantity equal to the commanded value?) or more complex integrations (What is the total amount produced, given the rate of production and the time of production?). Furthermore, there are several different ways of achieving display proximity. Moving two elements close together is the simplest way. But perceptual proximity can also be achieved by displaying two elements in the same color, by connecting them with a line, by using similar representations (e.g., both digital or both analog signals), or by *configuring* the elements such that an emergent feature arises from their combination. Two relevant examples are shown in Figure 3. In Figure 3(a) command and actual values are displayed by vertical bar graphs on the same scale and aligned with the same baseline. Hence the integrated variable — their equivalence — will be directly signaled by the emergent feature, the flat line perceived across their tops. In Figure 3(b) production rate and time are displayed as the height and width of a rectangle respectively. Hence the integrated variable — total amount — will be directly signaled by the emergent feature, the area of the rectangle.

5. THEORIES AND PRACTICE OF COMPATIBILITY

There is no single uniform theory of compatibility; there are several different theories that underlie different aspects. For example, the Kornblum *et al.* (1990) theory of dimensional mapping, in which certain features of a perceptual array automatically activate response tendencies, provides a strong theoretical account for differences between compatible, incom-
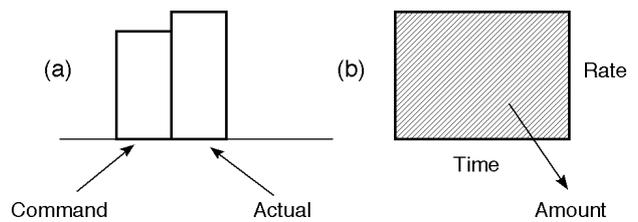

Figure 3. Configural displays illustrating the proximity compatibility principle. (a) Display of command and actual values of a system: the emergent feature signaling 0 error is the continuous line across the top of the two bar graphs. (b) Display of rate × time: the emergent feature of amount (rate × time) is signaled by the area (height × width) of the rectangle.

patible, and acompatible mappings. Various cognitive theories of mental model formation and representation can form the underlying theoretical base for cognitive compatibility in general, and some aspects of ecological interface design in particular (Vicente and Rasmussen 1992). The proximity compatibility principle has as its underpinnings theoretical conceptions of attention, and the distinction between focused attention (best served by low proximity displays) and divided attention (best served by close proximity displays), as well as the theories of perceptual similarity.

Above all, it is best to describe the concept of compatibility, as applied to good system design, as based upon the careful linking of engineering and psychology: human factors engineering must provide the *task analysis* regarding what information is available and relevant, what actions are required, and what cognitive (i.e., internal) states must be monitored and understood in order to achieve the goals of effective system operation. The psychology of information processing must be consulted to understand the natural or preferred representations at the different stages (e.g., "on" is conceived as "up"; common color is perceived as related), or what dominant population stereotypes are relevant for the design space at hand; and the craft of the engineering psychologist must then be applied to join the two domains in configuring compatible designs in systems as diverse as stove controls, calculators, airplanes, or nuclear power plants.

REFERENCES

FITTS, P.M. and SEEGER, C.M., 1953, S–R compatibility: spatial characteristics of stimulus and response codes. *Journal of Experimental Psychology*, 46, 199–210.

GREENWALD, A.G., 1970, Sensory feedback mechanisms in performance control: with special reference to ideo-motor mechanism. *Psychological Review*, 77, 73–99.

KORNBLUM, S., HASBROUCQ, T., and OSMAN, A., 1990, Dimensional overlap: cognitive basis for stimulus-response compatibility — a model and taxonomy. *Psychological Review*, 97, 253–70.

NORMAN, D., 1988, *The Psychology of Everyday Things* (New York: Harper and Row).

PROCTOR, R.W. and VAN ZANDT, T., 1994, *Human Factors in Simple and Complex Systems* (New York: Allyn and Bacon).

ROSCOE, S.N., 1968, Airborne displays for flight and navigation. *Human Factors*, 10, 321–32.

SMITH, S., 1981, Exploring compatibility with words and pictures. *Human Factors*, 23, 305–16.

TEICHNER, W. and KREBS, M., 1974, Laws of visual choice reaction time. *Psychological Review*, 81, 75–98.

VICENTE, K.J. and RASMUSSEN, J., 1992, Ecological interface design: theoretical foundations. *IEEE Transactions on Systems, Man, and Cybernetics*, 22, 589–606.

WANG, H. and PROCTOR, R.W., 1996, Stimulus-response compatibility as a function of stimulus code and response modality. *Journal of Experimental Psychology: Human Perception and Performance*, 22(5), 1201–17.

WICKENS, C.D., 1992, *Engineering Psychology and Human Performance* (2nd ed.) (New York: HarperCollins).

WICKENS, C.D. and ANDRE, A.D., 1990, Proximity compatibility and information display: effects of color, space, and objectness of information integration. *Human Factors*, 32, 61–77.

WICKENS, C.D. and CARSWELL, C.M., 1995, The proximity compatibility principle: its psychological foundation and relevance to display design. *Human Factors*, 37, 473–94.

Computer Systems Design for Psychophysical Safety of Human Operations

A.Yu. Burov

Centre of Ecology, Ergonomics and Safety, ELAN Ltd, Kyiv, Ukraine

1. INTRODUCTION

The analysis of accidents in power plants shows that one of the main causes of the increase in their number is a discrepancy of operators' psychophysiological capabilities and professional demands. The economic losses are stipulated not only for accidents, but also for the unefficient operation of equipment. Besides, the operators who do not meet professional requirements get sick and receive work trauma more often. There is a necessity to implement a monitoring of operator's psychophysiological professionally important qualities (PPIQ) during all stages of their professional career (initial professional selection, periodical and daily check). Such an approach allows not only reduced economic damage because of unefficient and erroneous human actions, but also prevents premature deterioration of his/her health and a premature quitting of the job.

Ergonomic approaches to the evaluation of operator's fitness to work have typically assumed that the functional characteristics of the operator (e.g. knowledge about task, skill level, information processing capacity, etc.) are invariant. They are assumed not to be influenced by such variables as operator mood condition, sleep disorders, time spent on task, etc. Research efforts aimed at enhancing the safety and reliability of the complex process plant operations have focused, as a result, on the improvement of the technical skills of the operator, as well as on reducing operator requirements imposed by automating processes. In complex technological systems human involvement in causing the accidents is estimated to be in the range of 40–60% (power plants) and 80% (aircraft accidents). This testifies to the inadequate evaluation of the operator's efficiency by existing techniques. The monitoring of an operator's physiological parameters during their professional career is known to be related to effective skill utilization and can be used to enhance system performance. While the requirements for professional knowledge and skills may be defined for various kinds of physical and mental work, the evaluation of their functional state over time suffers from the failure to take into consideration not only the levels of psychophysiological parameters, but also the changes of the operator's states, especially those occurring over different periods (year to year, day to day variations, within work schedule variations, variations produced by changing shifts, etc.). The direct or indirect measurement of psychophysiological changes allows for the evaluation of the structure (or moment-to-moment state of the operator) of the functional state of the operator, allowing one to predict the operator's individual fitness and reliability for effective work.

The objective was the development of a methodology and applied systems of psychophysiological maintenance of an operator's fitness to work during all stages of their professional career in order to reduce human errors and to raise his/her efficiency.

2. SYSTEMS FOR EXPERIMENTAL PSYCHOPHYSIOLOGICAL RESEARCH

To make psychophysiological research methods valid to the industrial implementation, it is necessary to integrate scientific research into industrial systems when using new developments in the field of psychophysiological initial professional selection, e.g. of operators, periodical and day-by-day check of human fitness to work as applications for the industry.

To solve this problem, a computer system for psychophysiological research was developed that could be simply converted to the industry systems for the initial psychophysiological professional selection, periodical and day-by-day check of operators' fitness to work (Burov and Chetvernya 1996). The system is intended for psychophysiological research and provides:

- the opportunity of running a complexity, rate, sequence of presentation, colorful and dimensioned characteristics of test tasks;
- registration of characteristics of each test task performance;
- registration of physiological parameters;
- synchronization between test tasks and physiological parameters;
- database control; and
- time series analysis and statistical modeling.

To record physiological parameters, a hardware–software complex is used, which consists of a set of sensors, a multichannel analog–digital converter, a personal computer and software. The software realizes the functions of control process of registration, processing and database control of physiological signals realization. Depending on the tasks, analysis of the data received during experimental research can be conducted in different directions, e.g.:

- assessment of the level of the individual's psychophysiological parameters as regard a surveyed group;
- dynamics of psychophysiological parameters during psychological tests performance;
- connection of physiological parameters and mental activity parameters; and
- study of psychophysiological reaction of a human organism to external factors (industrial, household, medical).

The researcher has flexible possibilities to control the experiment (Figure 1). The developed interactive display systems (IDS) of psychophysiological security allow one to solve research and assessment problems on the base of following principles:

- adaptation of used tests to functions that are being investigated;
- ergonomic information display on the screen;
- minimal touch;
- taking into account the time factor;
- taking into account the dynamics of the operator's work;
- specialized database management system (DBSM) as the core of the system; and
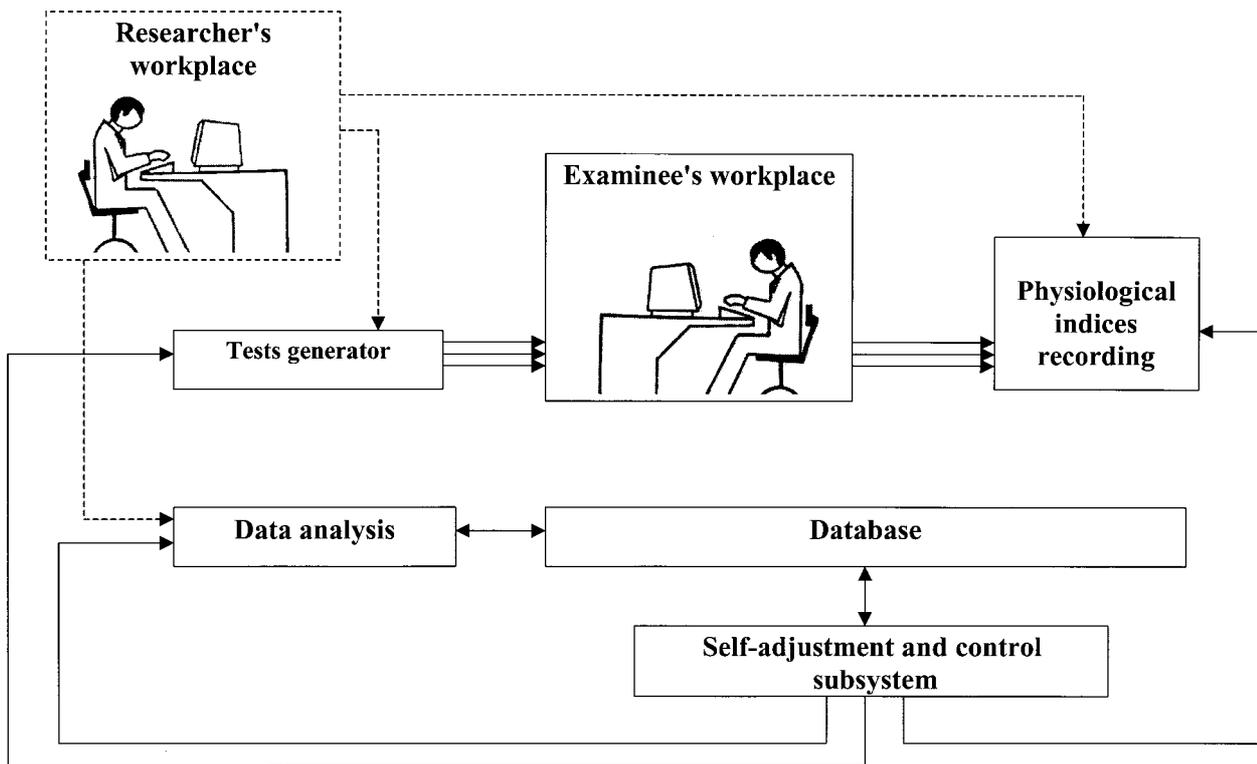- using of external criterion.

Figure 1. System structure for psychophysiological researches.

3. SYSTEMS FOR INDUSTRIAL USE

Observation of the mentioned principles permits one to create psychodiagnostical systems for industrial purposes on the basis of the system of psychophysiological research as the specialized variant for the solution of concrete psychodiagnostical tasks — a system for initial professional selection, and daily and periodic psychophysiological check. Thus, the architecture of the system is transformed to account for the following:

- exception of subsystem for registration of physiological parameters;
- transmission of test control from researcher to the DBSM; and
- store in the database only the resulting data and, as a consequence, more economical use of computer resources.

Such a structure provides the availability of the managing program (monitor); subsystems, ensuring of functions of training, testing, data processing, construction of evaluations and of predictive models. The implementation of all these functions, as well as functions of database control, is executed by specialized DBSM, which guarantees the integrity of internal database and data exchange with an external program environment within the system.

The IDS allow one to bring unified to effect on method and to form keeping of information, got on all control stages of functional state (FS) and operator fitness to work (OFW):

- Psychophysiological professional selection (with recommendations on training individualization).
- Periodic psychophysiological monitoring check of FS and operator OFW including professional aging.
- Daily (pre-shift) check of professional OFW of operator in "regular" routines.
- Daily (inter-shift) check of professional OFW of operator in extreme routines.

- Means of functional rehabilitation (psychophysiological correction of functional operator's state).
- Training process support included in teaching aids.
- PPIQ training with specialized psychological playing simulators.

All IDS intend to work in real-life conditions and can be combined with automated systems, both instructive and simulated, and they do not need additional personnel of medical and psychological specialists.

The systems of initial psychophysiological selection are aimed at estimating unspecific professionally important qualities that prevent from mastery good by selected profession or to work in the profession. The means and estimation methods must be sufficient to make a decision concerned a potential professional fitness.

The periodical check systems allow the estimation of the permanent organism descriptions and personality, their evolution and can be used periodically (e.g. once a year). Time of inspection can be considerable (> 1 h), but the inferences of status must be sufficient to make a decision concerned with the renewal options of the professional activity.

Systems for initial psychophysiological selection and periodic check can use an objective medical-biological and psychological information concerning human state — psychophysiological parameters registered by special apparatuses in laboratory conditions.

Daily pre- and inter-shift check. The pre-shift psychological check must be directed at revealing the unfitness to professional activity affected by functional (non-constant) changes in physiological or psychological status of the individual.

An inter-shift psychophysiological check is directed to evaluate the psychophysiological activity's cost and human
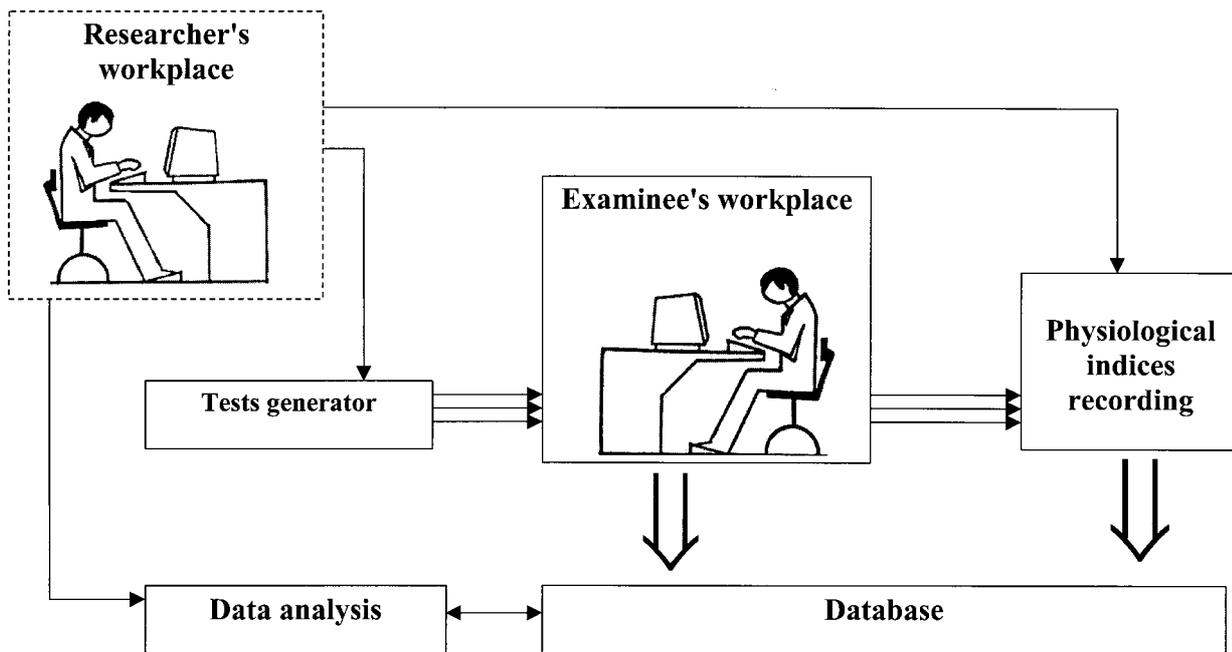
Figure 2. General structure of an industrial system for the psychophysiological security.

reserve. The means and methods of verification of this type exert a minimal influence on the professional activity, but are sufficient for detention of possibility for its continuation on a proper level.

The means and methods of functional rehabilitation restore the functional abilities, which decreased as a result of professional activity. Such methods can include diverse exercises and measures on renewal of health, and also have a full arsenal of physio- and psychotherapeutic means.

At present, to realize this approach the systems of industrial application were developed and used:

- for psychophysiological initial professional selection;
- for periodical check; and
- for daily check.

The systems permit the evaluation of the reliability and efficiency of an operator's work, as well the construction of the prognosis of changes of these parameters during his professional career. The results of valuation and prognosis permit one to make arrangements, not to admit the unreliable operator to work or to increase his reliability and fitness for work to make preventive measures for accident precautions.

Psychophysiological initial professional selection (Burov and Chetvernya 1995a, b). The efficiency of operator's work depends not only on his/her professional knowledge and working conditions, but also on conformity of psychophysiological features of a person to the requirements of the trade. As the results of preliminary researches have shown, the following structure of tests for the determination of professional fitness group of operator is optimum: structure of personality, structure of intelligence, an individual's psychodynamic features, and a bent for certain kinds of mental activity. The most informative parameters of efficiency of tests performance are included in the model of a "standard operator" enabling one to conduct the psychophysiological prediction of the group of the professional fitness of a candidate for operator.

Periodical check. The periodical check is intended for the

evaluation of changes of PPIQ slowly varying that permits one to evaluate the moment when the operator needs to do some rehabilitation steps or when it is necessary for him to be prepared to leave his trade if the irreversible age-related changes came. With this purpose the system of valuation of an operators' professional aging rate is used (Burov and Chetvernya 1996). The results of valuation of staff's professional age are presented as an integrated age value which is calculated by the chosen model, as well as an "age profile," which is a vectorial diagram of the main parameters determining biological and professional human age.

Daily check (Burov and Chetvernya 1997) used the system of valuation of an operator's functional state and prediction of his professional fitness to work in the workplace which is a method of providing feedback to the operator so allowing him to make necessary changes in case of deviation from his individual norm. The following problems are solved in the system:

- Development of objective methods of evaluation of a humans' psychophysiological condition in a relation to the norm.
- An individual approach to evaluation of a human's serviceability, that is a construction of individual psychophysiological norm instead of the group's one.
- Development of organization-technical aspects of evaluation's use.

The system allows:

- Economic running of the technological object (power industry, aviation).
- Economy of energy resources.
- An increase of the operators' professional "longevity."
- One to make the system self-adjustable. It forms automatically the individual "norm" of operator's working ability.
- The achievement of a high precision of prognosis (85–90%).

The use of the systems allows one to monitor the operator's psychophysiological fitness for work on all stages of his/her

professional biography to secure the necessary level of the human–machine system's effectiveness.

REFERENCES

BUROV A. and CHETVERNYA YU.V., 1995a, The reliability and productivity of operators of various age on enterprises of energetics. In A.C. Bittner and P.C. Champney (eds), *Advances in Industrial Ergonomics and Safety VII* (London: Taylor & Francis).

BUROV A.YU. and CHETVERNYA YU.V., 1995b, Structure of psychophysiological indexes of operators' reliability. In *ErgoCon'95 Conference Proceedings. SVEI Conference & Exposition*, San Jose, Hilton & Towers, 21–24 May.

BUROV A. and CHETVERNYA YU., 1996, Methodological principles of psychophysiological forecasting of operators' reliability. In *Advances in Occupational Ergonomics and Safety I. Proceedings of the XIth Annual International Occupational Ergonomics and Safety Conference*, 8–11 July, Zurich.

BUROV A. and CHETVERNYA YU., 1997, Development and implementation of information technology for the psychophysiological valuation and forecasting of the efficiency and safety of operators' work. In *Proceedings of the XIIth Annual International Occupational Ergonomics and Safety Conference*, Washington.

Creating Pleasurable Products

P. W. Jordan

Philips Design, Building W, Damsterdiep 267, 9700 AE Groningen, The Netherlands

1. INTRODUCTION: NEW HUMAN FACTORS

Usability-based approaches to product design tend to view people as *users,* while products are seen as *tools* with which these users complete tasks. Because of this usability-based approaches to user requirements specifications can be limited, tending to emphasize the practical aspects of interacting with products, while paying little attention to emotional or hedonic aspects of interaction.

Over the past 4–5 years much of industry-based human factors has moved away from usability-based approaches and towards pleasure-based approaches to defining user requirements. Pleasure-based approaches look at people holistically — as rational, emotional and hedonistic beings. They look at products as *living-objects* with which people have *relationships*. Such approaches have been described as "New Human Factors" (Fulton 1993).

2. HIERARCHY OF USER NEEDS

Pleasure can be thought of as being at the top of a hierarchy of user needs (Figure 1).

2.1. Level 1 — functionality

Clearly a product will be useless to the user if it does not contain appropriate functionality. A product cannot be usable if it does not contain the functions necessary to perform the tasks for which it is intended. If a product does not have the right functionality it will dissatisfy the user. To be able to fulfil user needs on this level, the human factors specialist must have an understanding of what the product will be used for and the context and environment in which it will be used.

2.2. Level 2 — usability

Once users have got used to having appropriate functionality, they will then want products that are easy to use. This seems to represent the situation at the moment in many product areas — people are used to well functioning products, now they expect usability too. Having appropriate functionality is a prerequisite of usability, but it does not guarantee usability. To assure usability
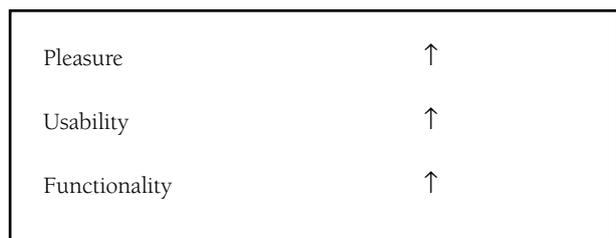
Pleasure	↑
Usability	↑
Functionality	↑

Figure 1. A hierarchy of user needs.

the human factors specialist must have an understanding of the design principles associated with usability and an understanding of how to address usability issues in the product creation process.

2.3. Level 3 — pleasure

Having got used to usable products, it seems inevitable that users will soon want something more. Products that offer something extra. Products that are not merely tools, but which are "living objects" which people can relate to. Products that bring not only functional benefits but also emotional benefits. To achieve product pleasurability is the new challenge for human factors. It is a challenge that requires an understanding of people — not just as physical and cognitive processors — but as rational and emotional beings with values, tastes, hopes and fears. It is a challenge that requires an understanding of how people relate to products. What are the properties of a product that elicit particular emotional responses in a person, how does a product design convey a particular set of values? Finally, it is a challenge that requires capturing the ephemeral — devising methods and metrics for investigating and quantifying emotional responses.

3. PLEASURE WITH PRODUCTS: THE FOUR PLEASURES

Pleasure with products has been defined as: "the emotional and hedonic benefits associated with product use" (Jordan 1997). Within this definition, four conceptually separate components of pleasure have been identified (based on Tiger 1992). These provide a framework for considering the emotional, hedonic and practical user requirements within the product creation process.

3.1. Physiological Pleasure

This is to do with the body — pleasures derived from the sensory organs. This covers, for example, tactile and olfactory properties. Tactile pleasures concern holding and touching a product during interaction. This might be relevant, for example, in the context of a telephone handset or a remote control. Olfactory pleasures concern the smell of the new product. For example, the smell inside a new car may be a factor that effects how pleasurable it is for the owner.

3.2. Social Pleasure

This is the enjoyment derived from relationships with others. Products can facilitate social interaction in a number of ways. For example, a coffee-maker provides a service which can act as a focal point for a little social gathering—a "coffee morning." Part of the pleasure of hosting a coffee morning may come from the efficient provision of well-brewed coffee to the guests.

Other products may facilitate social interaction by being talking points in themselves. For example, a special piece of jewelry may attract comment, as may an interesting household product, such as an unusually styled TV set. Association with other types of products may indicate belonging in a social group — Porsches for "Yuppies," Dr Martin's boots for skinheads. Here, the person's relationship with the product forms a part of their social identity.

3.3. Psychological Pleasure

This relates to how the user's mood is affected by interaction with a product — for example, the sense of satisfaction gained

from accomplishing a task. It might be expected, for example, that a word processor that facilitated quick and easy accomplishment of, say, formatting tasks would provide a higher level of psychological pleasure than one with which the user was likely to make many errors.

3.4. Ideological Pleasure

Ideological pleasure refers to the pleasures derived from "theoretical" entities such as books, music and art. In the context of products it would relate to, for example, the aesthetics of a product and the values that a product embodies. For example, a product made from biodegradable materials might be seen as embodying the value of environmental responsibility. This, then, would be a potential source of ideological pleasure to those who are particularly concerned about environmental issues. Ideological pleasure would also cover the idea of products as art forms. For example, the video cassette player that someone has in the home, is not only a functional item, but something that the owner and others will see every time that they enter the room. The level of pleasure given by the VCR may, then, be highly dependent on how it affects its environment aesthetically.

4. TECHNIQUES FOR DESIGNING PLEASURABLE PRODUCTS

Having considered and specified the emotional, hedonic and practical benefits that a product should bring to its users, the next step is to determine the design properties through which the product can deliver these benefits. Four techniques — previously reported in the human factors literature — for linking benefits to design properties are outlined below. Two of the methods — Kansei Engineering and SEQUAM — are based on quantitative statistical analyses. The other two are qualitatively based.

4.1. Kansei Engineering

Kansei Engineering (Nagamachi 1995) is an empirical technique aimed at linking the design characteristics of a product to users' responses to the product. The technique involves manipulating individual aspects of a product's design to test the effect of the alteration on users' overall response to the product. This technique has been used to assist in the design of a diverse range of products. Nagamachi (1997) describes examples that range from automobiles through camcorders to brassieres. To demonstrate how the technique works, a case study — reported by Ishihara et al. (1997) — is summarized below demonstrating how Kansei Engineering was applied to the design of cans for coffee powder.

4.1.1. Example of the application of Kansei engineering

In a study aimed at supporting the design of coffee packaging, 72 alternative designs of coffee can were presented to a panel of 10 subjects. The 72 designs represented permutations of various design variables, including, for example, color, font styles, graphics and form. Each member of the panel was asked to rate each of the designs according to how they fitted with a series of descriptor adjectives. These adjectives are known as "elements." There were 86 elements generated from brainstorms with consumers and marketing experts. Examples of these elements were: showy, calm, masculine, feminine, soft, individual, high-

grade, sweet and milky. Ishihara et al. (1997) report that each participant rated each of the 72 designs according to each of the 86 elements by marking five-point Lickert Scales. Thus, each participant made a total of 6192 responses on Lickert Scales!

A cluster analysis was then carried out to determine how panelists tended to group the designs according to their elements. A number of clusters emerged which Ishihara et al. (1997) identified as being linked to particular design features. For example, one cluster of cans emerged that were regarded by the panelists as being "milky," "soft" and "sweet" — this cluster of cans was characterized mainly by the use of beige coloring for majority of their surface. Another cluster was seen as being "masculine," "adult" and "strong" — these elements were all associated with having a large logo on the cans. A third example of a cluster was one that was seen as being "unique," "sporty" and "individual" — this was related to the use of blue and white coloring in the designs.

4.2. Sensorial Quality Assessment (SEQUAM)

This technique (Bandini-Buti et al. 1997) involves analyzing the link between the physical properties of a product and users' responses to tactile contact with the product. Users are presented with models of product components (e.g. car door handles) that are mocked up in combinations of materials and finishes exhibiting a range of tactile qualities. Users are then asked to handle these mock-ups and to comment on their sensorial qualities. SEQUAM has been applied in the automotive sector by FIAT.

4.3. Case Studies

MacDonald (1998) has analyzed a series of products and has suggested links between particular elements of their design and particular pleasures of use. The table below summarizes the outcomes of a study in which the benefits associated with a number of products were analyzed within the context of the "four pleasure framework" (for details of the study, see Jordan and MacDonald 1998) (Table 1).

4.4. Semi-structured Interviews

Jordan and Servaes (1995) carried out a series of interviews asking users about their most pleasurable products. They recorded users'

Table 1. Four pleasure analysis of the benefits associated with a sample of products from MacDonald's case studies.

Product	Associated pleasures
Karrimor's Rucsac Buckle	*physio:* positive click on closing
Global Knives	*physio:* comfortable to hold, weight balance *psycho:* reassuringly hygienic
NovoPen™ (for the injection of insulin)	*physio:* pleasant to hold *socio:* plays down negative associations (from others) *psycho:* easy and reassuring to use *ideo:* plays down negative associations (from user)
Samsonite Epsilon Suitcase	*physio:* comfortable, lifting, tilting and trailing *psycho:* responsive movement — suitcase 'obeys' user
Mazda Car Exhaust	*socio:* status symbol; *ideo:* positive, youthful self-image

reported reactions to pleasurable products and the product properties associated with these reactions. Reactions to pleasurable products included, for example, feelings of security and comfort, pride, excitement, freedom and nostalgia. Examples of associated properties included features and functionality, usability, aesthetics, performance, reliability and associated status.

5. NOTES ON THE TECHNIQUES

Kansei Engineering represents a thorough, formal approach to the linkage of product properties with user responses. Like SEQUAM it relies on statistical analyses to make these links. Such techniques appear to offer the most reliable and valid means of making product property/user response links. However, when many property dimensions are involved, as in the coffee can example, such methods can become unwieldy and time consuming. Designing 72 concepts and asking respondents to mark > 6000 scales each, simply to give input to the design of a coffee can, seems excessive!

Another possible criticism of quantitative techniques is that, in analyzing the effect of individual design elements, they are implicitly based on the assumption that a design is the sum of its parts. The merit of such an assumption is a matter for debate. It could be, for example, that a "Gestalt" model is more appropriate. Gestalt theories assert that entities, such as designs, must be considered holistically — that is to suggest that the overall response to a design could amount to either more or less than the sum of reactions to each of its constituent parts.

From the Gestalt point of view more qualitative approaches — such as case studies and interviews — may be more appropriate, as they may be better suited to looking at the product as a whole. However, it might be argued that, unless design elements are separated, it is difficult to give designers any meaningful advice about how to create a product that will elicit particular responses.

6. CONCLUSIONS

The trend in industry-based human factors is away from "traditional" usability-based approaches towards "new" pleasure-based approaches. Although the study and application of such approaches is still in its infancy, it is developing rapidly and there are now a variety of frameworks and techniques for holistic user requirements setting and making the associated design decisions. In this paper, a definition of pleasure with products has been given, a framework for considering person–product relationships has been outlined, and techniques for linking product benefits to design decisions have been discussed.

Products have the potential to bring a wide range of emotional, hedonic and practical benefits. Ensuring that these are delivered requires looking both at and beyond usability to create useful, usable and pleasurable products that will delight the customer.

REFERENCES

BANDINI-BUTI, L., BONAPACE, L. and TARZIA, A., 1997, Sensorial Quality Assessment: a method to incorporate perceived user sensations in product design. Applications in the field of automobiles. In *IEA '97 Proceedings* (Helsinki: Finnish Institute of Occupational Health), 186–9.

FULTON, J., 1993, Physiology and design new human factors. *American Center for Design Journal*, 7.

ISHIHARA, S., ISHIHARA, K., TSUCHIYA, T., NAGAMACHI, M. and MATSUBARA, Y., 1997. Neural networks approach to Kansei analysis on canned coffee design. In *IEA '97 Proceedings* (Helsinki: Finnish Institute of Occupational Health), 211–13.

JORDAN, P.W., 1997, Putting the pleasure into products. *IEE Review*, November, 249–52.

JORDAN, P.W. and MACDONALD, A.S., 1998. Pleasure and product semantics. In M.A. Hanson (ed.) *Contemporary Ergonomics 1998* (London: Taylor & Francis).

JORDAN, P.W. and SERVAES, M., 1995, Pleasure in product use: beyond usability. In S. Robertson (ed.) *Contemporary Ergonomics 1995* (London: Taylor & Francis), 341–6.

MACDONALD, A.S., 1998, Developing a qualitative sense. In N. Stanton (ed.) *Human Factors in Consumer Products* (London: Taylor & Francis), 175–91.

NAGAMACHI, M., 1995, *The Story of Kansei Engineering* (Tokyo: Kaibundo).

NAGAMACHI, M., 1997, Requirement identification of consumers' needs in product design. In *IEA '97 Proceedings* (Helsinki: Finnish Institute of Occupational Health), 231–3.

TIGER, L., 1992, *The Pursuit of Pleasure* (Boston: Little, Brown), 52–60.

Engineering Principles of Ergonomics

A. Stork

Ergonomics & HCI Unit, University College, 26 Bedford Way, London WC1H
0AP, UK

1. INTRODUCTION

Suggestions that the disciplines of Ergonomics, Human Factors (HF), and Human–Computer Interaction become more like engineering disciplines have resulted in the proposal of a discipline of Cognitive Engineering (CE). The explicit engineering knowledge of this new discipline of CE is termed "engineering principles" (EP). CE and the development of EP are current research topics. The proposed discipline of CE, and the relationship of engineering with pure and applied science, are discussed here under "Cognitive Engineering", to be read along with this entry, which briefly characterizes the current state in the explicit development of EP in CE.

2. ORIGINS

Norman (1986) first terms the explicit knowledge of CE as 'engineering principles', which are "principles that get the design to a pretty good state the first time around". Dowell and Long (1998) identify a requirement for CE to develop "design principles" and distinguish them from craft and scientific knowledge:

> Craft disciplines give way to engineering disciplines: personal experiential knowledge is replaced by design principles; "invent and test" practices (that is to say, trial-and-error) are replaced by "specify then implement" practices. Critically, design principles appear not to be acquired by direct translation of scientific theories. Rather, they are developed through the validation of knowledge about design problems and how to solve those problems.

> Therefore, EP offer a guarantee that their application will support some effective design. The basis of that guarantee is their validation for design.

3. EXPLICIT PROGRESS TOWARDS ENGINEERING PRINCIPLES

There has been little explicit progress towards EP. Dowell and Long (1998) offer a conception of the general design problem of CE to define the scope of EP. They propose that the scope of EP is the design of cognitive human–machine systems to achieve a desired level of effectiveness. Effectiveness is conceptualized as the task quality of the work performed in a domain by the human–machine system, and the structural and behavioral costs to the human–machine system of performing that work. Dowell and Long also distinguish substantive EP, which prescribe the properties of effective designs, from methodological EP, which prescribe the process to prescribe those properties.

A few researchers have moved towards EP based on Dowell and Long's research. They have suggested alternative conceptions of the general design problem, conceptions of EP, strategies for their acquisition, operationalizations of design problems, solutions and putative engineering principles. However, there are, as yet, no explicit validated engineering principles for CE.

4. SOME OBSTACLES TO ENGINEERING PRINCIPLES

This research attracts two main types of criticism. First, that sufficient rigor to deliver EP cannot be applied to human behavior and goals: human behavior is too complex and human goals too nebulous to be adequately formalized or measured. The counter claim is that design is currently based on an, often implicit, understanding of human behavior and goals, and that there has not been a satisfactory attempt at their adequate formalization or measurement. Second, that the conception of the general design problem of CE is insufficient: for example, that social or marketing aspects need to be addressed. The counter claims are, for example, that (1) cognitive structures and behaviors support social structures and behaviors, and (2) that CE needs to limit its scope at some point, and marketing is appropriately outside that scope.

5. CONCLUSIONS

EP cannot be precisely defined at this stage. Further attempts at their acquisition along with further reflection on what actually constitutes engineering knowledge more generally should enable more precise definition in the future.

REFERENCES

DOWELL, J. and LONG, J., 1998, Conception of the cognitive engineering design problem. *Ergonomics*, 41, 126–39. [This is a peer-reviewed paper based on an earlier paper published in 1989. Dowell and Long propose the acquisition of engineering principles with a guarantee and the scope of those engineering principles.]

NORMAN, D., 1986, Cognitive engineering. In D. Norman and S. Draper (eds), *User Centered System Design* (Hillsdale: Erlbaum), 31–61. [Norman was the first to identify a requirement for engineering principles for ergonomics and HF.]

Ergodynamics

V. F. Venda† and N. I. Venda§

†University of Manitoba, Canada

§The Venda Ergonomic Advantages, Inc., USA

1. INTRODUCTION

Ergodynamics was developed as theory and methods of ergonomic design, improvement, and studies of dynamic work in a dynamic environment (Venda and Venda 1995). Ergodynamics allows the study of work dynamics when the *functional structure* of the work (skills, strategies, methods, technologies) changes. In a more narrow sense, ergodynamics is an application of the *transformation dynamics* theory in ergonomics. A foundation of ergodynamics comprises of the following three laws of ergonomics: mutual adaptation; plurality of work functional structures; and transformations and interactions. Ergodynamics is widely applicable in ergonomic design of work places, work stations and control rooms, in mutual adaptation the in human–machine–environment, in professional training, control systems safety, prediction, planning and optimizing efficiency dynamics during implementation of ergonomic and engineering projects, transfer of new technologies, environments, work functions and skills, management structures, software, products, changing economic strategy, or when an otherwise sudden decrease in productivity, quality, or safety causes unpredicted losses.

2. DEFINITIONS IN ERGODYNAMICS

Ergodynamics is based on the following definitions:

1. Work efficiency is any positive outcome of work ergonomist wishes to maximize. The examples are work safety, health, satisfaction, quality, productivity.
2. Work complexity is any negative outcome of work ergonomist wishes to minimize. The examples are risk of injures, errors, complaints, time spent on one task or product.
3. Work factor is any parameter of interaction and mutual adaptation between human and work environment, including machines, tools, partners, managers, etc. The examples are desk height, workspace, work posture, number of emergency signals perceived by operator, noise and light as they influence concrete worker, etc. Work factor as a factor of human–environment mutual adaptation does not belong to the human or environment alone. It is a parameter of interaction and mutual influence in the course of training, work, overexertion, etc.
4. Mutual adaptation is a process of synthesis and development of human–machine–environment system structure which is a basis for human and system work process and strategy. Ergodynamics studies work structure, strategy, efficiency and complexity in the processes of human–machine–environment mutual adaptation, particularly when human work changes its structure and strategy.
5. Work structure reflects a regularity of internal mutual adaptation processes in human, machine, any other system. Strategy reflects a regularity of external, interactive mutual adaptation processes of human or machine with environment based on the structure.

3. THE FIRST LAW OF ERGODYNAMICS

The First Law of Ergodynamics is the Law of Mutual Adaptation: "Work efficiency is a bell-shaped function and work complexity is a U-shaped function of any factor of mutual adaptation in human–machine–environment system."

As an example demonstrating the First Law, experiments were conducted on typing a text on a laptop computer installed at different heights. Subjects typed the text in a sitting position. Work productivity, Q, measured as a mean number of characters typed in 1 min, appeared to be a bell-shaped function of desk height, F. Work complexity measured as time to type 1000 characters, appeared to be a U-shaped function (inverted bell-shaped function).

4. THE SECOND LAW OF ERGODYNAMICS

The experiment on typing at the laptop computer was extended so the subjects were typing in either a sitting or standing position, given varying desk heights. Two different functional structures for every subject were studied: the processes of reading texts and perception of control board information. A variety of functional structures and strategies was found. For example, the same text may be read by separate letters or by syllables. Then there are two different $Q(F)$ curves for these two strategies.

These studies allowed the wording of the Second Law of Ergodynamics, which is the Law of Work Structures Plurality: "Every work can be done using different structures (strategies) presented by different respective bell-shaped curves for efficiency, $Q_i(F)$, and U-shaped curves for complexity of work."

A visual image of the Second Law is shown in Figure 1, where S_{sub} and S_{op} are two different functional strategies used by laboratory subjects (S_{sub}) and operators (S_{op}) in the same work.

The Second Law explains the plurality of human reactions on the same signal or certain value of an environment factor, F. For example, the same desk height leads to different typing productivities in sitting and in standing positions. The Second Law helps to solve many various ergonomic tasks. For example, it may help predict a human operator's decision-making efficiency when the operator perceiving the same information from the same video terminal but uses different cognitive strategies.

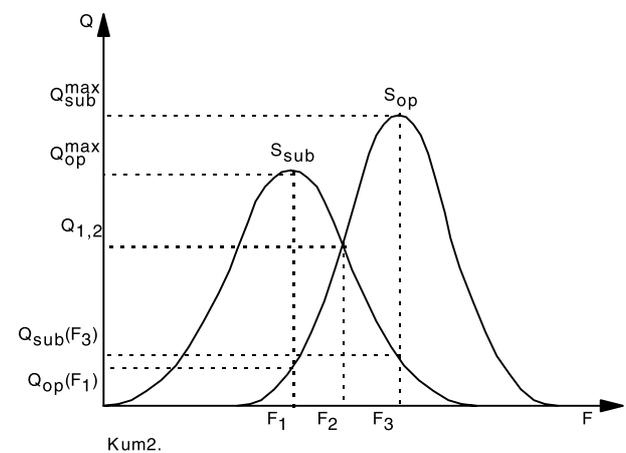The Second Law explains why ergonomics laboratory results

Figure 1. Bell-shaped curves for subject's (S_{sub}) and operator's (S_{op}) strategies. Q, efficiency; F, work factor.

obtained with subjects using strategy S_{sub} in many cases cannot be used to design and optimize work of operators using different strategy, S_{op}. Figure 1 shows that in the interval of work factor F_1 Æ F_3, the decrease of F recommended based on data for S_{sub} would lead to the inverse effect for the operator: in this interval S_{sub} increases its efficiency and S_{op} decreases its efficiency. This task is discussed in more details in *theoretical ergonomics*.

5. UNIVERSAL AND SPECIALIZED STRATEGIES: A HISTORICAL PROSPECTIVE

A brief outline of the history of work strategies over centuries based on ergodynamics methodology will be given. At the dawn of industry the craftsman was the main production force. Every craftsman performed all the operations needed to produce a whole product. Therefore, the craftsman accomplished tasks in a wide range of work factors, F_{min}/F_{max}, but work efficiency measured as productivity was very low (Figure 2).

There was not often the necessity for the craftsmen to transform work strategies, skills and methods. First, factories implemented some of distribution of functions between industrial workers. The range of tasks for each worker got narrower, and the strategy was more specialized and effective in comparison with craftsmen. At the next stage more detailed functions were distributed between production workers and, thus, work efficiency increased (Figure 2). Assembly-line workers performed
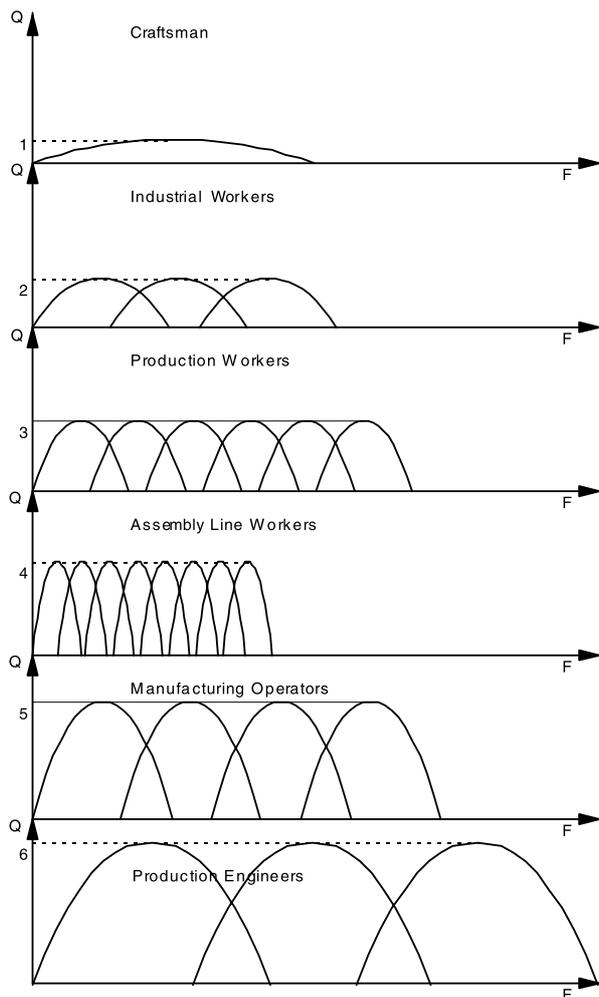

Figure 2. Ergodynamics history of work strategies.

small sets of operations with very high productivity, but this repetitive work led to frequent injures and low job satisfaction. Now a tendency in industrial development takes a reverse direction. Using various semi-automatic tools manufacturing operators can reach a higher productivity while expanding their range of tasks. Production engineers (Figure 2, stage 6) continue this tendency further, thanks to wider use and assistance of CAD, CIM, robotics, computers, expert systems, etc.

6. DISCOVERY OF WAVE–LIKE WORK EFFICIENCY DYNAMICS

The Second Law states the plurality of the work functional structures, but only one functional structure can be used by the individual at one time. If several different structures are available, and only one is used at the time, then a process of transformations between the structures should be specially studied. Traditional monotonic exponential models of human development (dynamics of efficiency) have been in exceptional use around the globe for more than a century (since Ebbinghaus 1885). These models, however, run contrary to the many practical and experimental data.

The following training experiment was conducted. Subjects (12 students at the Engineering Faculty) accomplished a compensatory tracking of dynamic signals presented simultaneously on several (1–6) measurement instruments, with the subject controlling an equal number of switches (1–6).

A distinctly non-monotonic character to the curves was found when $n = 2, 4, 5$ and 6. The data show that increased experience may sometimes result in an actual worsening of performance. It was found that in those periods the subjects were changing a functional structure of their performance of perceiving signals: (1) separately one-by-one; (2) grouping them by two into two-dimensional coordinates of a dot track; (3) by three, as 3-D coordinates of a dot in space that should be moved to the initial point with the coordinates 0,0,0.

When the learning curves were averaged over the number of signals, the result was strictly monotonic. Such averaging is, unfortunately, a traditional method of statistical processing of experimental ergonomic data. It, however, disguises the real dynamics of learning processes, which depend upon the cognitive strategies used and the number of signals being monitored and grouped into different information chunks.

We avoided traditional statistical averaging and analyzed a dependence of the signal tracking efficiency on different cognitive strategies: perception of the signals by one, two and three as an information chunk.

7. THE THIRD LAW OF ERGODYNAMICS

While analyzing all previous experimental data, Venda found that during transformations of any two work functional structures (S_i and S_{i+1}), work efficiency drops to the level corresponding to the intersect point of the curves $Q_i(F)$ and $Q_{i+1}(F)$. Thus, he discovered a transformation state that is a common and equal state for both structures. He suggested the law of transformations for any kind of complex systems (Venda and Venda 1995) and especially for ergodynamics.

The Third Law of ergodynamics was worded by Venda in 1989 as The Law of Transformations and Interactions: "Transformations and interactions between two different work
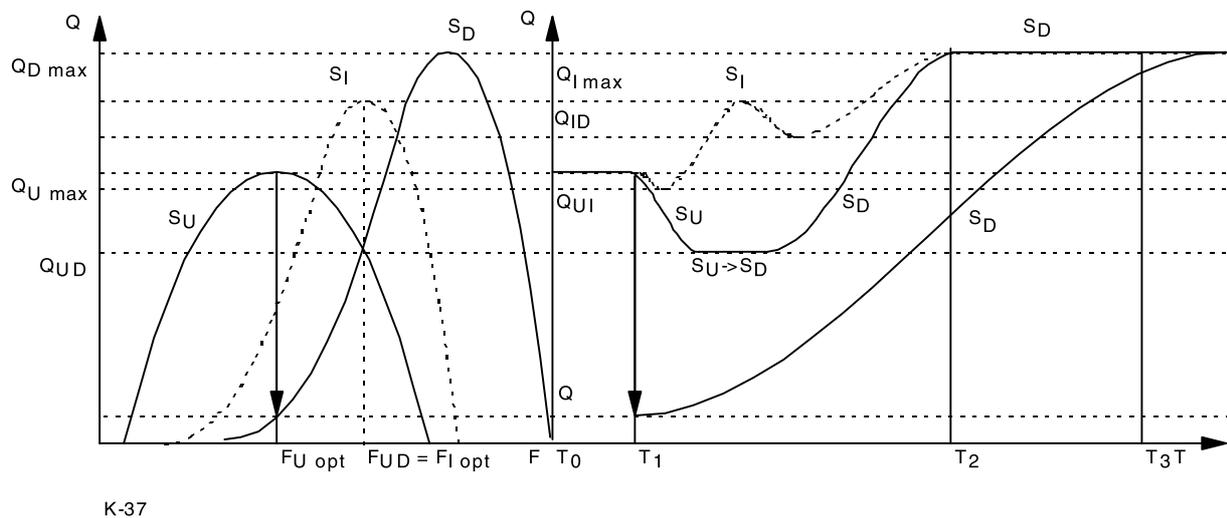
Figure 3. Work strategies S_U, S_I and S_D, and transformations between them. Q, work efficiency; F, work factor; T, time.

strategies are maximally effective if they go through a state common and equal for the structures presented as an intersect point of bell-shaped curves $Q_i(F)$ or U-shaped curves $C_i(F)$ of the two strategies."

The common and equal state for two strategies, S_U and S_D, is presented in Figure 3 as the intersect point of the respective $Q(F)$ curves of the strategies: $F = F_{UD}$ and $Q_U(F_{UD}) = Q_D(F_{UD}) = Q_{UD}$.

Work strategies may belong to the same individual and may be used sequentially, transforming one to another. Work structures may belong to different individuals (or human and machine) interacting in the complete work process.

8. ERGODYNAMICS IN IMPLEMENTATION OF ERGONOMICS PROJECTS

Every essential, innovative ergonomics project leads to some kind of transformation in the work process designed or re-engineered.

Let us assume there is an existing user strategy, S_U, that allows maximal efficiency $Q_{U\,max}$. Designer-ergonomist suggests a new work strategy, S_D, that promises maximal work efficiency $Q_{D\,max}$ if work factor (task complexity for example) increases to $F_{D\,opt}$.

There are several trajectories available to transform S_U into S_D:

1. To implement S_D instead of S_U right the way. Then efficiency of S_D while $F = F_{U\,opt}$ will be much lower than $Q_U(F_{U\,opt}) = Q_{U\,max} \gg Q_D(F_{U\,opt})$ (see arrow in Figure 3).
2. Implementation of S_D may be done as transformation of S_U into S_D along with gradual increase of task complexity F from $F_{U\,opt}$ to $F_{D\,opt}$. This trajectory leads to relatively deep decrease of efficiency to the level Q_{UD} with following increase to Q_D max by the time T_2.
3. If the decrease of efficiency to the level Q_{UD} is too deep and unacceptable for the company, the transformation may be done through intermediate strategy S_I, in two stages: $S_U \: Æ \: S_I \: Æ \: S_D$, with much less decrease of efficiency.

Choice of the best transformation trajectory depends on relative integral gains and losses.

9. TRANSFORMATION DYNAMICS IN INDIVIDUALS, COMPANIES AND NATIONS

Transformation means a smooth transition of one system structure (a human being, team, manufacturing facility, company) to another. A "pure" transformation occurs with constant system components and, consequently, its energy and material resources.

If a businessman decides to change a product and technology at his facility, s/he can use two main different politics: transformation or replacement. Transformation means fast and smooth changing of the old technology into a new one, using the same human resources.

Transformations in science, technology, trade and education help a nation or society to save and use a big part of the previous knowledge, skills, equipment, human and business relations. We had applied a transformation dynamics theory to predict the dynamics of a social-economical transformation in the former USSR, and explained transformation difficulties there well in advance (Venda 1989).

Using the Third Law and analyzing characteristic curves of existing and new desirable technological or management strategies, every company may predict, plan and optimally execute transformation of the existing into the new strategy with minimal losses in efficiency.

10. ERGODYNAMICS OF STRESS

There is one special way in fast transformations which are important for industrial companies but even much more important for operators of power plants and air pilots when situations turns from normal to emergency and time for transformation is very limited. Fast and reliable, accurate transformation to successful strategy may save lives and systems. Slow and inaccurate transformations lead to the catastrophes.

Ergodynamics is the first and only existing methodology analyzing this problem. We found one of effective solutions of this problem among many inventions by biological evolution. We mean abilities of living system to turn their states into stress condition.

The stress condition looks like partial destruction of the organism structure with low temporary efficiency. System
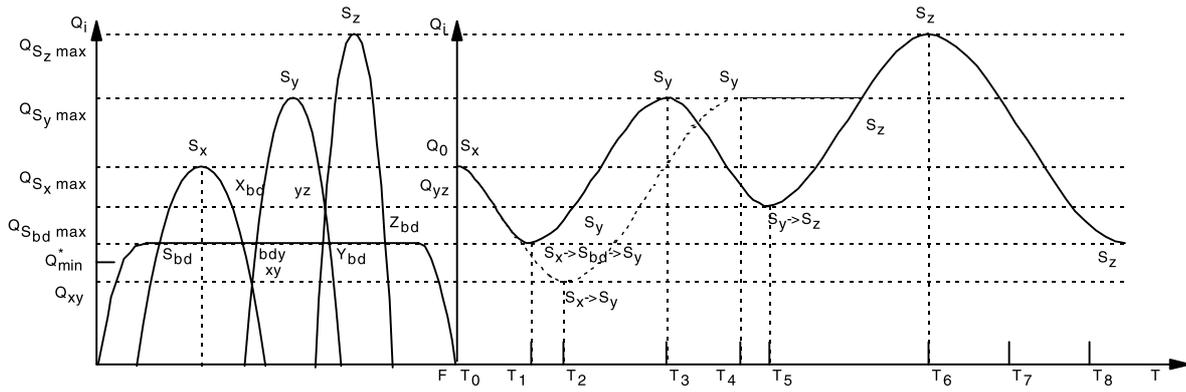
Figure 4. Three specialized structures, S_x, S_y and S_z, and a basic diversified structure, S_{bd}, and transformations between them. Q, work efficiency; F, work factor; T, time.

structures may be narrowly specialized that allows to reach high maximal efficiency in a narrow range of conditions (DF). High maximal efficiency depends on convergence and synchronization of the system components.

The structure may be more universal, with less maximal efficiency, but very wide range of acceptable conditions. Importance of the stress structure is that it is very universal as a result of divergence of the system components.

Figure 4 presents three specialized structures, S_x, S_y and S_z, and one more universal, basic diversified structure, S_{bd}.

Let us assume system cannot survive if its efficiency (safety) is lower than some minimal level Q^*_{min}. If system has currently structure S_x, it cannot transform to any other structure, S_y or S_z, when F changes and goes outside of range acceptable for S_x, because common and equal state for S_x and S_y, xy, lies lower than Q^*_{min}.

The system can survive if it has a basic diversified structure S_{bd}, then through common states between S_x and S_{bd}, between S_y and S_{bd}, and between S_z and S_{bd}, the system has flexibility to survive in a very wide range of environmental conditions.

This great invention by nature may be effectively used by the industrial companies to maintain flexibility in a dynamic market.

This principle is even much more important for the operators, air pilots and other workers who may be in emergency situation. S_{bd} may be considered as a specially trained standard structure that could be easily obtained from any non-standard structure and then may be quickly transformed into narrowly specialized highly effective structure (strategy) needed to use in the emergency.

11. PERFORMANCE PREDICTION AT TRANSFORMATION STAGES

Speed and reliability of the transformation between two strategies, S_{a1} and S_{a2} (Figure 5), depends on the following factors:

1. Difference in efficiency between current strategy and transformation state to another strategy, D$Q = Q_{a1\ max} - Q_{a1a2}$ (Figure 5); the greater the difference the more difficult, slow and unreliable is transformation.
2. Frequency of transformations between S_{a1} and S_{a2}; if system was trained or accustomed to the transformations between these strategies, the transformation will go fast and correct.
3. How long time the system was in a current condition. The longer the period, the more conservative is the system, the slower and less reliable is transformation.
4. How fast F changes. The faster is change, the more difficult inertial system may keep up with the transformation.
5. Whether the system passes through the points it used to transform to other structures is also important. The search to find a new transformation point and to transfer to a new structure is also important. This search includes trial runs to predict the consequences involved in a change.

So, as we study the probable and actual system structure dynamics, we should examine the current structure, the newly found ones, prehistory and the earlier structures of the system. Suppose that over the range F the structure S_{a2} gives a gain of efficiency DQ. Here the system seeks to pass from S_{a1} to S_{a2}. If, on the other hand, Q decreases for a change in F, the system resists
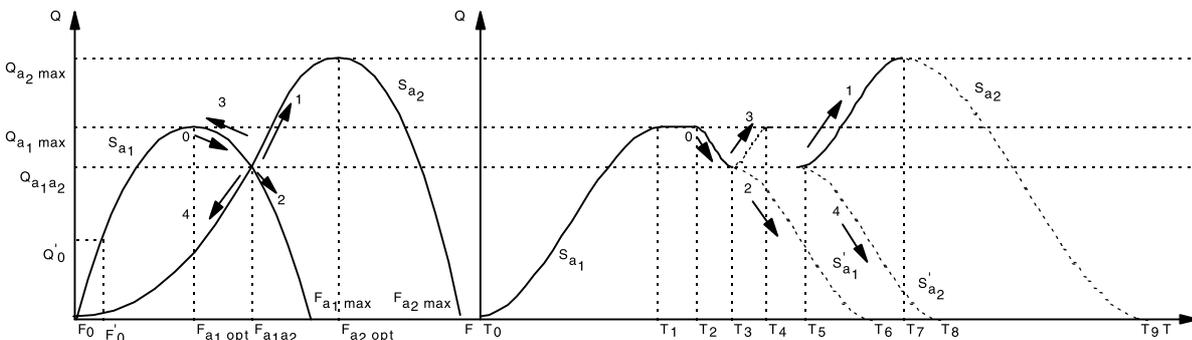

Figure 5. Performance prediction at transformation stages. Q, work efficiency; F, work factor; T, time.

the effect of the changed F. It musters its components' mutual adaptation potential to redress the change.

Even when the transformation state is reached, three types of "losing" pathways can mask real progress. Figure 5 shows that movement from maximal efficiency of the structure S_{a1} at $Q_{a1\ max}$ and $F_{a1\ opt}$, through the transformation point (Q_{a1a2}, F_{a1a2}) may have the following variations:

1. *Successful transformations.* If a system reaches transformation state (Q_{a1a2}, F_{a1a2}), and stays there long enough for the structure to transform from S_{a1} into S_{a2}, and if F increases from F_{a1a2} to $F_{a2\ opt}$, the system will then go to $Q_{a2\ max}$. This behavior model is called *progressive*. Any successful restructuring involves this kind of transformation.

2. *Collapse of the system as a result of ignoring its inertia for transformations.* Let us assume the system is in initial state 0 ($Q_{a1\ max}$; $F_{a1\ opt}$). If the system goes from $F_{a1\ opt}$ to F_{a2} (through F_{a1a2}) too fast, the system's Q will change from $Q_{a\ max}$ to zero (track 0 Æ 2). It will skip the transformation point and fail, having structure S_{a1} when factor $F = F_{a2\ opt}$, because $Q_{a1}(F_{a2\ opt}) = 0$. This behavior pattern is called *passive system fallback to zero efficiency*. This means that the system is deadlocked.

3. *Return to initial state just in time.* If a system left initial state ($Q_{a1\ max}$, $F_{a1\ opt}$) changed environment from $F_{a1\ opt}$ to F_{a1a2}, but then it does not remain at transformation state (F_{a1a2}, Q_{a1a2}) long enough for the S_{a1} Æ S_{a2} transition and F reverts to its $F_{a1\ opt}$, the system efficiency will be again $Q_{a1\ max}$. This is a standstill policy. It leads to a *relapse*. This throwback will result in track 3 in Figure 5.

4. Track 4 means "Transformation is already in progress, it is too late to go back." If a system stays at the transformation state (F_{a1a2}, Q_{a1a2}) long enough for the S_{a1} Æ S_{a2} transition, and then goes from F_{a1a2} to a direction 4 with lower (previous) values of F the system efficiency will drop to zero. This is the fatal *half measures and retreat* policy: after an extended effort to make the S_{a1} Æ S_{a2} transition, one makes a futile attempt to revert to the past. That was what Gorbachev did when he tried to restore Communist party rule, return to the "socialist ideals," retain the privileges of the party "nomenclatura" and his personal role as the General Secretary of the Communist party. By that time the people of the former USSR has already changed. Gorbachev's attempt to move back a country with new-born strong elements of social and psychological structure and free-minded people led to the crash of the superpower, the USSR and Gorbachev himself in 1991. There was a book published in 1989, *The Waves of Progress*, by V. Venda, predicting and explaining the transformation process in that country well in advance.

This historical lesson should convince everybody that if a functional structure of a firm, technological facility, design team, work skill, country society has been changed then reverse transformation should be planned and organized to avoid catastrophe of the system.

Thus, of the four possible transformation point paths, only the first is successful. The probability of a successful dynamic progress is thus only one in four if the proper transformation concept is taken in account. If there is no proper transformation concept, probability of success in changing structure and strategy may be equal to zero.

So, one should be able to predict and control transformation process to get a success. This, in turn, calls for great courage and determination when the system fails to show the proper efficiency level after a long wait.

The *transformation state (or point)* is defined as a system's critical point, when its future development route is decided. It is the point where two or more structure-strategies have the same efficiency. Any departure from this point, to the left or right, causes a rapid change in efficiency.

Thus, if the main search tendency (for instance, in scientific research) aims at increasing factor values, it must form the embryo of the next structure well before the change. The lower efficiency of the current structure during this phase will also have to be considered. The next structure could thus take over with enough ease and promptness.

12. CONCLUSION

Ergodynamics is helpful at least in dealing with two major challenges of contemporary ergonomics: (1) creating a theory that could accumulate, combine and generalize the ergonomic experience and act as a common foundation for easing ergonomics teaching, studying and practicing; and (2) improving methods of ergonomic design, improvement and studies of dynamic work in dynamic environment.

Ergodynamics is an application of the transformation dynamics theory in ergonomics. Three Laws of ergodynamics are being suggested: (1) The Law of Mutual Adaptation; (2) The Law of Plurality of Work Functional Structures; and (3) The Law of Transformations.

Ergodynamics is an operational methodology of qualitative and quantitative analysis and design of dynamic work functional structures and planning their transformations.

13. RECOMMENDATIONS

1. Ergonomic recommendation for practice may be based on laboratory data only if laboratory subjects' strategies are identical to those of industrial operators and workers.

2. Interactions between human and machine, communications between people must be organized as a processes of mutual adaptation. Maximal efficiency of the interactions may be found as an intersect point of the two interacting strategies' characteristic curves. The same principle based on the Third Law must be used for transformations between strategies of the same person, company or nation.

3. If work task is very well defined in a narrow range of work factor value, specialized strategies allow higher work productivity in a short run, but monotony of work may cause big troubles in a long run.

4. Avoid statistical averaging of data collected while different work strategies were used. Analyze data on each strategy separately.

5. Do not promise clients that implementation of innovative ergonomic project, new hardware or software will immediately lead to increase of efficiency. Temporary dip in efficiency is inevitable, it may be minimized in time duration and depth using laws and principles of ergodynamics.

6. Operators must be trained not only strategies effective in normal and emergency situations but also special skills to transform one strategy into another.

7. While organizing transformations in the company chose the

best transformation trajectory considering relative integral gains and losses.

8. Do not rush to eliminate stress in individuals and company, it is natural, necessary and very helpful in well-planned and organized transformations.

9. When planning transformations analyze current and desired structures and system's prehistory.

10. Do not use monotonic linear or quasi-exponential models to predict performance dynamics if transformations of strategies may occur.

11. Do not use ergostatics in analysis and design of dynamic work. Use ergodynamics.

REFERENCES

KARWOWSKI, W., 1991, Complexity, fuzziness and ergonomic incompatibility issues in the control of dynamic work environments. *Ergonomics*, 34, 671–86.

KARWOWSKI, W. and MITAL, A., 1986, *Applications of Fuzzy Set Theory in Human Factors* (Amsterdam: Elsevier).

ROWE, A.L., FRENCH, J., NEVILLE, K.J. and EDDY, D.R., 1992, The prediction of cognitive performance degradations during sustained operations. In *Proceedings of the Human Factors Society 36th Annual Meeting*, Atlanta, GA: HFS, 111–15.

VENDA, V.F., 1989, *The Waves of Progress* (Moscow: Znanie).

VENDA, V.F., 1990, *Hybrid Intelligence Systems: Evolution, Psychology, Ergonomics* (Moscow: Mashinostroenie).

VENDA, V.F. and VENDA, Y.V., 1995, *Dynamics in Ergonomics, Psychology, and Decisions: Introduction to Ergodynamics* (Norwood: Ablex).

The Ergonomic Qualities of Products

L. Bandini Buti, L. Bonapace, G. Caggiano, E. Fubini, F. Marcolin, P. Molinari, A. Tarzia and G. Tomei

Società di Ergonomia Applicata, Milan, Italy

1. INTRODUCTION

Is a sharp, pointed knife with a 30-cm blade a safe product? One may be tempted to say not, and it is not the kind of object one would not give to child. Yet a butcher cannot be expected to cut and prepare meat with a round-ended blade!

Thus, we see that it is impossible to give a simple "yes" or "no" answer to the question of a knife's safety, as a product is safe or otherwise not only in terms of its intrinsic nature, but also, and above all, in relation to those who use it, the tasks the user must perform, their specific training and the instructions they have received. It is possible to think of examples other than the knife in the domestic sphere: products range from ovens to pressure cookers, and from stepladders to irons. Arguments made in the context of safety do not change very much if applied to that of comfort. Seating provided on the underground is comfortable, but certainly could not be adopted for use in an office where one sits not for minutes but for hours at a time.

From such observations it would seem that it is not easy to evaluate the quality of a product or system, as reference must be made not only to the specific product, but also to the specific user, the specific environment of use and the characteristics of use. That which might be considered a positive quality by one person may not be by another, or indeed by the same individual in differing circumstances and for other purposes. Such reflections propose a very complex framework for the evaluation of the qualities of a system or product. They require the consideration of the potential qualities of the product, the characteristics of the user and above all the interface between these.

Markets that have become extremely competitive at national, international and even global levels, and users who demand high levels of performance, require products and services to be ever more responsive to the actual needs of the user, and this has tended to bring about an increase in the number of options available, causing the greater part of products, from those in the workplace to cars, and from electro-domestic appliances to entertainment systems, to become increasingly complex and difficult to understand for those who use them.

The concepts and methods of ergonomics are being used more and more in the evaluation and design of products, and are consistent both with existing European regulations and directives and those currently being drafted. In recent years, industry, stimulated by a market more sensitive and awake to the concept of health and well-being, has begun to recognize an ergonomic approach to industrial design as a planning method capable of surveying, revealing and predicting the needs brought about by the complex system of inter-relations between man and machine and between these and the environment. We are becoming ever more conscious of the fact that safety and the safeguarding of health can only be confronted in the design phase and that ergonomics is a design technique that is capable of dealing effectively with design quality.

2. ERGONOMICS AND QUALITY

The analysis of individual behavior in the use of a product or system (i.e. of factors linked to highly variable factors in the user, such as perceptive capabilities, psycho-physical characteristics at the moment of use, habits, knowledge of and familiarity with the product and relevant procedures, etc.) covers a high level of complexity and must be confronted with approaches that have the use of solid and proven methodological foundations to inform the process of making choices with objective arguments (i.e. arguments that may be evaluated, compared, measured and transmitted.

Ergonomics has developed theories and tested criteria and methods capable of highlighting the wishes and needs of users and evaluating the usability of products and systems, and which have shown themselves to be effective in the conception of products and systems, proving successful at intervening in the design phase with innovative proposals.

Studies of product quality highlight the user's perception of the qualities of the product or system. Perception is the first, indispensable step on the road to correct use; in fact, if a quality, even if present, is not perceived, it will have a reduced, or indeed, no effectiveness.

A product's qualitative characteristics may be defined as common, perceivable, non-perceivable, self-explanatory and induced.

2.1. Common Qualities

These are qualities that must be considered common and generalizable to all potential users; they are those linked closely to safety and security, which may in no case be left out of consideration.

2.2. Perceivable Qualities

These may be perceived through the normal senses of the individual (sight, touch, hearing, feeling a weight in one's hand, smell, taste). They are linked to experience (like the mechanical characteristics of wood which are expressed by its configuration, its weight which is perceived as a guarantee of its robustness, etc.) but also to stereotypes (such as considering wood a noble and reliable material, but plastic, on the other hand, artificial and as such not noble and perhaps toxic, etc.). Perceivable qualities change over time with the enriching of individual experience (such as the weight of a product being perceived as a guarantee of a large technological content, a perception that has been changed by the advent of miniaturization).

2.3. Non-perceivable Dualities

These may not be perceived by ordinary senses of the individual. They are either hidden (such as the quality of the structure of a panel covered by an outer finish, etc.) or, to be quantified, require the execution of analyses and tests (such as the durability of paint over time, safety requirements, toxicity of the product, etc.). Such qualities need to be studied, analyzed and guaranteed by specialist bodies, laboratories or other experts.

2.4. Self-explanatory Qualities

These are present when the product itself informs the user of its qualities and use, through its appearance. Above all, they are present in products of a mechanical nature, products whose form is necessarily conditioned by their mechanical nature (such as typewriters, scissors, corkscrews, etc.). With electronics, products become increasingly less self-explanatory and thus require ever-greater intervention in the design phase in order to manipulate the product's image to render it more self-explanatory.

2.5. Induced Qualities

These are proposed by various means of advertising as possible or necessary for the user. For example, car safety campaigns have induced safety as a problem and robustness as a quality in the public eye; an increasing ecological awareness has led to respect for the environment on the part of producers in production, use and disposal of goods, becoming a required quality.

The qualities that users ask for and expect also change when innovative products or systems become widely available, so that that which was optional or an accessory yesterday becomes a minimum requirement today. In the case of motor vehicles there are many such examples of products realized as optional but which have become qualitative standards of reference: electric windows, power-assisted steering, central locking, motorized wing-mirrors. One should also note that many options created only to facilitate a particular operation become transformed through use. Again with cars — the case of the motorized right (left in some countries) wing mirror which, today, is often used to check the distance between the car wheel and the sidewalk during parking.

3. THE USER

Ergonomic design proposes that the characteristics of a product respond to the actual needs of the end-user and that they may be used with a minimum physical and psychological "cost." Ergonomic design requires, therefore, the needs analysis of the user and the "costs" related to use.

The generic definition of "user" or that used in the common language of clients, consumers, indeed users themselves, is very limiting and misleading compared to the objectives of ergonomics which intends all those who enter into contact, even sporadically or casually, with the product.

3.1. User Classification

For a correct evaluation of ergonomic quality it is necessary, therefore, to define who will use the product, which group of users they belong to, which means they possess for the use of the product, and what their needs are.

3.1.1. Those who work with the product

These are the end-users of the product and they are:
- *unique*, when the product is destined for exclusive or private use;
- *multiple-defined*, when the product is destined for a defined an definable group of people, such as occurs with photocopiers or drinks machines in a large office;
- *multiple-generic*, when the product is destined for an undifferentiated collective use such as occurs with automatic teller machines or with photocopiers or drinks machines in public places; and
- *improper*, when the product may be used by categories differing from those for which it was intended (e.g. use on the part of children, use of professional products by hobbyists).

For some products, users might be *generic* in that anybody may be interested in using them, while for others the users may be *defined* since the products are destined for use by characteristic groups, such as military conscripts, Italians only, school children, licensed drivers, etc.

3.1.2. Those who work on the product

They are:
- generic "maintenance" people, who are often identified with those who work with the product and are not in possession of any specific training (e.g. changing ink cartridges in printers and photocopiers and re-filling them with paper, changing batteries, changing the wheel on a car, etc.);
- professional maintenance workers who are in possession of suitable and specific knowledge and tools; and
- personnel authorized for production and fitting.

3.1.3. Those who handle the product

These are retailers, warehouse workers, transport workers, installation workers, both those involved with distribution and those involved with the end-user.

those who use the product	• those who work with the product • those who work on the product • those who handle the product • those who eliminate the product	end-user maintenance and production personnel retailers, warehouse workers, transporters ecological workers
user category	• generic category • defined category	anybody, undefined by sex, age activity....
means possessed	• diffused/innate knowledge • specific knowledge • learning • instruction • specific training	language, meaning of symbols, etc. previous product experience product self-explanation instruction booklets, exemplification professional training, licences
needs expressed	• primary • of functional quality • of aesthetic quality	what one can do how one can do it pleasure in doing

Figure 1. Product users.

3.1.4. Those who eliminate, remove and destroy the product

These are ecological, recycling and disposal workers.

3.2. Knowledge Possessed by the User

Users may also be classified in relation to the means they possess for the correct use of a specific product. These means may be:

3.2.1. Innate/diffused knowledge

That is, what is considered common knowledge among large population groups (for us, those in industrialized society), whether this be innate or gained in the course of physical and intellectual development, such as knowledge of one's mother tongue, of the meaning of symbols and colors, the automatic performance of the actions of screwing and unscrewing, etc.

3.2.2. Specific knowledge relative to the product

That is, knowledge acquired through the use of the product or other similar products. For example, habitual use of computer technology renders more natural the management of symbolic systems.

3.2.3. Self-explanatory knowledge derived from the messages that the product transmits

Self-explanatory knowledge derived from the messages that the product transmits without the use of verbal means, but with the application of diffused, innate or specific knowledge which acquire a universally self-explanatory value (as with codes relative to upside-down, left-right, or hot-cold on controllers).

3.2.4. Instructions accompanying the product

Instructions accompanying the product, which regulations have in many cases rendered obligatory, and which should complete rather than replace the self-explanatory capacities of the product. They are the instructions and recommendations for use, printed instruction booklets or those on computer (the "help" facilities in programs), or the various types of "demos" for applications and video arcade games.

3.2.5. Specific Training

Specific training, when it is necessary to be in possession of a license or specific professional qualification to use a product such as with the use of electro-medical equipment on the part of sanitary personnel or the use of specialist military equipment

3.3. Needs Expressed by the User

One may define a group of users as homogenous from the point of view of needs expressed, which can be distinguished on three levels:

3.3.1. Primary needs

These regard the basic needs that affect one's motivations for buying a product. Typically, to communicate, to move around, to sit, etc.

3.3.2. Needs of a functional quality

These regard the mode in which primary needs are acquitted. They regard the sphere of ergonomics: safeguarding of health, safeguarding of physical integrity and comfort.

3.3.3. Needs of a symbolic and aesthetic nature

These regard the sphere image and the quality of the relationships this tends to establish between subject, object, environment (agreeability). Umberto Eco states (1982) "one must not think that symbolic functions are something added to the object, rather they form part of its functions and it would be a mistake to think that if 'form follows function,' form mustn't also follow symbolic function."

The three levels analyzed are never disconnected. That is, no primary need can be acquitted without the user also expressing the need for functional and formal qualities; subdivision into categories must, therefore, be considered only as a useful analytical instrument.

We have already that needs, when they are subjective, are specific to each individual depending on their anthropometric, physiological, functional and cognitive characteristics. It would, however, be reductive to think of the individual as unchanging both in the short and long term. Indeed, in relation to particular points in time (the stress, tiredness, motivation, etc. of the moment) and environment (lighting, noise, climate and microclimate, etc.) their needs can change considerably.

3.4. Typology of Needs

Various types of need can be distinguished:

3.4.1. Real needs

Those whose satisfaction allows one to perform complex, productive or leisure activities, which have an effect on quality of life and work and which go beyond the abstract pleasures of possession and control.

3.4.2. Imaginary needs

These do not derive from the desire or need to improve performance, but rather they refer to models induced by opinion leaders, the mass media or social trends. In a consumer society, desires often correspond to needs that are not in the sphere of utility but of representation. The confines between needs of a symbolic character and simply succumbing to consumerism are fragile and not easily defined. A lot of people, for example, would like to own a motorcycle not so much to be able to travel with more ease and speed, as for the intrinsic pleasure of owning it, or in the sphere of Information Technology, many people acquire the most up-to-date models of computer, equipped with lots of powerful applications, but they simply have no need for them; here pleasure is simply derived from the domination of technology.

3.4.3. Latent needs

These are real needs of the individual that are not expressed as they are felt to be unobtainable or simply because they are not recognized, but whose presence is felt. Identifying these needs is important for innovations.

3.4.4. Future needs

These are needs that will manifest themselves in the future, but are not yet expressed as the technological, economic and social conditions which will render them topical are not mature. One might predict that some important future needs will be expressed in the sphere of the use of leisure time and that they will only be fully expressed with the lengthening of life after work and the

shortening of the working day. The importance of considering future needs resides in the fact that between the idealization and the realization of a project on an industrial scale a lengthy period of time usually passes; thus it is not sufficient only to consider the present when working on a project but one must try to predict what the needs of the user will be when the product finally enters the market. It is not easy to study future needs in as much as they are usually only expressed by small signs and trends that must be interpreted. They also evolve with the offering of new levels of performance and quality which tend to render previous proposals obsolete and to create new points of reference and new standards, with the associated moving of the scale of values asked for (expected).

4. THE ERGONOMIC QUALITIES OF INDUSTRIAL PRODUCTS

In recent years the concept of "usability engineering" has been consolidated, a concept developed principally for application to products in the area of Information Technology. This area has been formalized with the creation of normative ISO 9241 (Ergonomic requirements for office work with visual display terminals — VDUs), which deals with the design of office equipment making use of video terminals. Part II of ISO 9241, "Guidance on Usability," deals with approaches to usability and provides a basis for measuring and specifying such approaches. It emphasizes the fact that products do not have an intrinsic usability, since this is determined by the characteristics of the user, of the task to be performed, and the immediate environment, as well as by the characteristics of the product itself.

This regulation indicates that it is possible to effect comparisons between planned solutions. The effectiveness, efficiency and satisfaction of two products (or two design situations) being used in the same context can thus be compared. Usability may thus be operatively defined as the effectiveness, efficiency, and satisfaction with which specific users reach specific objectives in particular environments.

- *Effectiveness* is the accuracy and completeness with which a user reaches the global objectives set for a system. Measurements of effectiveness relate the objectives and secondary objectives of use in the system to the accuracy and completeness with which such objectives may be reached.
- *Efficiency* is the accuracy and completeness of the objectives reached in relative to the resources used. Measurements of efficiency relate the level of effectiveness reached to the

spending of resources, such as physical and mental efforts, time spent and economic costs.

- *Satisfaction* is the level of comfort and acceptability in using the system, the subjective reactions of the system user. Measurements of satisfaction relate to the comfort and acceptability of the global system for those who must use it and for those affected by its use.

Normative ISO 9241 clearly states that there is no such thing as the ergonomic quality of a product, but that one may speak of the ergonomic quality of an object in a situation of determined use on the part of a determined subject in a determined environment (i.e. that ergonomic quality is not attributed to the product itself but to the use of the product). The concepts expressed in this chapter stem from the work of the "Industrial Design and Ergonomics" group of the Società Italiana di Ergonomia (SIE) and we believe it is possible to extend such affirmations, formulated for VDU, to ergonomic quality in industrial products in general, and we, therefore, believe it is possible to definitively affirm that:

- the ergonomic quality of an industrial product is an attribute of the use of the product and not of the product itself; and
- the ergonomic quality of an industrial product is refers to the user, to the task to be performed, to related equipment, to the immediate physical and organizational environment.

These affirmations would seem to exclude the possibility of defining the ergonomic qualities of various products or, further, of certifying them, which would rule out the possibility of providing clarity and certainty for the user. We, on the other hand, have come to the conclusion that it is possible to identify, evaluate and diffuse ergonomic quality in industrial products, operating on the characteristics, parameters and qualities of the product.

The ergonomic characteristics of a product are the aspects that have either a positive or negative influence on the relationship between the product and the particulars of possible users, of possible environments, and of possible uses, and they represent the first level of study.

The ergonomic parameters are criteria capable of measuring, evaluating or judging the ergonomic characteristics of the product in objective terms.

The ergonomic qualities of products are linked no more to general conditions, but to specific situations of use on the part of determined users and in determined environmental contexts. The ergonomic quality of products is determined by verifying the positive or negative effects of the ergonomic characteristics under
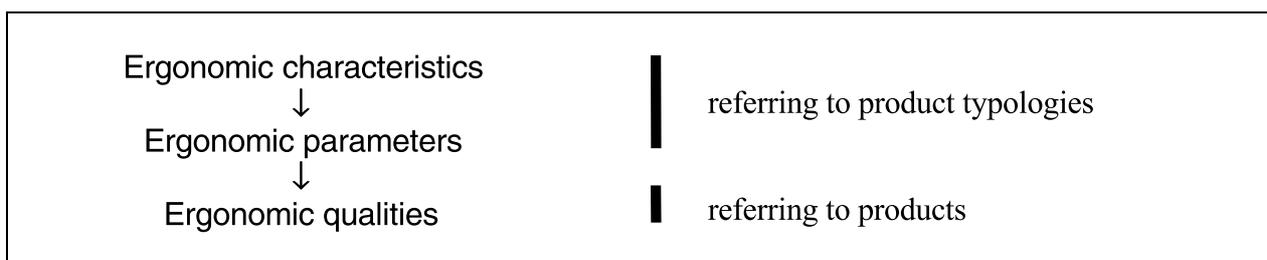

Figure 2. Diagram illustrating ergonomics qualities of products.

determined conditions of use, and evaluated by means of ergonomic parameters.

4.1. Ergonomic Characteristics

Ergonomic characteristics are those aspects which are capable of influencing the quality of the relationship between products and the particular user. Ergonomic characteristics are not, in themselves, either positive or negative, rather they are arguments capable of influencing the psychophysical well-being or the pleasure of those who use the product. All products can be embodied and expressed in terms of product typologies which possess the same ergonomic characteristics, or rather which are homogenous with each other in terms of performance offered and characteristics of the interface with the user. It is not necessarily sufficient for them to offer the same level of performance. In fact, a printed encyclopaedia and the same encyclopaedia on CD-ROM belong to two completely different ergonomic categories, in as much as, even if the content of both is identical, the characteristics of use, the operative knowledge required of the user and the constraints imposed by the work place are completely different.

A product's ergonomic characteristics must be defined for every homogenous product typology, making reference to a number of general arguments relative to the characteristics of the product–user interface. This general list of ergonomic characteristics of a product must be constantly up-dated in line with research developments in the sector, with the evolution of user expectations, and with the enrichment of technology available.

4.2. Ergonomic Parameters

Ergonomic characteristics can only be evaluated and transmitted if they are expressed by "parameters" that indicate and specify the methods, instruments and units of measurement by which they can be described in objective terms. Ergonomic parameters are, therefore, criteria for measuring, evaluating, or judging the ergonomic characteristics of a product.

Ergonomic parameters must make reference to knowledge possessed in relation to various disciplines closely linked to ergonomics, and it will only be necessary to define new parameters in the absence of satisfying parameters. Ergonomic parameters can derive from:

- Methods, instruments, and units of measurement taken from the disciplines closely linked to ergonomics. For example, intensity of sound may be measured in A' decibels, ease of handling by means of measurements made using the metric system, the equilibrium of the object in relation to the center of gravity and its distance from the center of the object, etc.

- Methods, instruments, and units of measurement chosen as the most suitable among those elaborated by disciplines closely linked to ergonomics. For example, to characterize types if noise it will be necessary to make use of the most suitable sonograms, those most representative of the phenomena; to describe forces one will have to choose representative means that place intensity and movement in relation to one another; for a usability analysis it will be necessary to choose one of the many techniques developed for use in this area, etc.

- Definitions of methods, instruments, and units of measurement developed specifically for situations in which the representation of ergonomic characteristics cannot rest on support provided by disciplines closely linked to ergonomics. For example, the definition of human body simulators of different sizes, of units of measurement and study methods for the evaluation of the comfort of mattresses.

- Specific research that it can be necessary to develop for arguments that cannot make reference to an adequate experience of ergonomic characteristics, either because of a lack of relevant literature or because of the newness of the sector or the innovation.

4.3. Ergonomic Qualities

The ergonomic qualities of a product are the responses to the ergonomic characteristics that derive from the use of the product in real life.

Ergonomic qualities, which may be either positive or negative, are, therefore, specific attributes of the product referred to the interaction of the ergonomic characteristics relative to specific users in specific environments and for specific uses. In order to define ergonomic qualities, scales of quality and parameters relative to all the ergonomic characteristics of individual products must be assigned. These scales may be the object of objective measurements or of subjective evaluation. Objective scales refer to measurable parameters that come under the various disciplines closely linked to ergonomics, and which have units of measurement that refer to the metric system.

REFERENCES

BANDINI BUTI, L., 1998, *Ergonomia e Progetto* (Milan: Maggioli).

BANDINI BUTI, L. *et al.*, 1995, Ergonomic industrial design, IEA World Conference, Rio de Janeiro.

NORMAN, D. A., 1988, *The Psychology of Everyday Things* (New York: Basic Books).

RABARDEL, P., 1997, Gli strumenti dell'uomo. *Ergonomia*, no. 9.

Evaluation of Software Usability

A. Dillon

HCI Lab, SLIS, Indiana University, Bloomington, IN 47405, USA

1. INTRODUCTION

Usability is a measure of interface quality that refers to the effectiveness, efficiency and satisfaction with which users can perform tasks with a tool. Evaluating usability is now considered an essential part of the system development process and a variety of methods and have been developed to support the human factors professional in this work.

2. CONCEPT OF USABILITY

Historically, the concept of usability has been defined in multiple ways, usually on one of the following bases:

- Semantics: in this case usability is equated to terms such as "ease of use" or "user-friendliness", without formal definition of the properties of the construct.
- Features: here usability is equated to the presence or absence of certain features in the user interface such as Windows, Icons, Menus or Pointing devices.
- Operations: where the term is defined in terms of performance and affective levels manifest by users for certain task and environmental scenarios.

The first type of definition was common in the 1970s but has proved largely useless for design purposes since it offers neither useful guidance for designers nor perspective for evaluators. The feature-based approach stems from the longstanding desire to specify in advance desirable interface attributes that enhance usability. However, this type of definition rests on an assumption that usability is an inherent part of the application. This assumption is false since one could always envisage a combination of users, with certain task demands, in a particular environment, for whom a given set of features would be suboptimal. Recognizing this, most human factors professionals now employ an operational definition in their work.

Shackel (1991) is the major developer of this operational approach. He defined usability as the artifact's capability, in human functional terms, to be used easily, effectively and satisfactorily by specific users, performing specific tasks, in specific environments. The essence of the operational definition is that it explicitly places usability at the level of the interaction between users and the artifact. This takes it beyond the typical features-based definitions common in the field. Furthermore, in setting criteria for assessing usability, this approach better supports the evaluation of any tool and the subsequent interpretation of the test results. Usability therefore refers not to a set of interface features, but to a context-dependent measure of human–computer *interaction*.

3. EVALUATING USABILITY

There exist multiple methods of evaluating usability depending on available resources (time facilities and labor), evaluator experience, ability and preference, and the stage of development of the tool under review. In broad terms it is worth making the following distinctions between evaluation methods:

- User-based: where a sample of the intended users tries to use the application.
- Expert-based: where an HCI or usability expert makes an assessment of the application.
- Model-based: where an HCI expert employs formal methods to predict one or more criteria of user performance.

3.1. User–based Methods

Testing an application with a sample of users performing a set of predetermined tasks is generally considered to yield the most reliable and valid estimate of an application's usability. Performed either in a usability test laboratory or a field site, the aim of such a test is to examine the extent to which the application supports the intended users in their work. Tightly coupled to the operational approach to usability definition, the user-based approach draws heavily on the experimental design tradition of human factors psychology in employing task analysis, predetermined dependent variables and, usually, quantitative analysis of performance supplemented with qualitative methods.

In a typical user-based evaluation, test subjects are asked to perform a set of tasks with the technology. Depending on the primary focus of the evaluator, the users' success at completing the tasks and their speed of performance may be recorded. After the tasks are completed, users are often asked to provide data on likes and dislikes through a survey or interview, or may be asked to view with the evaluator part of their own performance on video and to describe in more detail their performance and perceptions of the application. In this way, measures of effectiveness, efficiency and satisfaction can be derived, problems can be identified and re-design advice can be determined. In certain situations, concurrent verbal protocols might be solicited to shed light on users' thought processes while interacting with the tool so that issues of comprehension and user cognition can be addressed. In a usability lab, the complete interaction is normally video recorded for subsequent analysis of transactions, navigation, problem handling, etc. However, more informal approaches are also possible. Some user-based tests are unstructured, involving the user and the evaluator jointly interacting with the system to gain agreement on what works and what is problematic with the design. Such participative approaches can be very useful for exploring interface options in the early stages of design where formal quantitative assessments might be premature.

In an ideal world user testing with a large sample of the intended user population would routinely occur, however due to resource limitations, user-based tests are often constrained. As a result, there is considerable interest among HCI professionals in determining how to gain the most information from the smallest sample of users. While popular myths exist about being able to determine a majority of problems with only two or three users, Lewis (1994) has shown that the sample size requirement is largely dependent on the type of errors one seeks to identify and their relative probability of occurrence. Whereas three users might identify many problems in a new application, substantially more users will be required to tease out the remaining problems in a mature or revised product.

3.2. Expert–based Methods

Expert-based methods refers to any form of usability evaluation

Table 1.   Relative advantages and disadvantages of each usability evaluation method

Usability method	Advantages	Disadvantages
User-based	Most realistic estimate of usability	Time consuming
	Can give clear record of important problems	Costly for large sample of users
		Requires prototype to occur
Expert-based	Cheap	Expert-variability unduly affects outcome
	Fast	May overestimate true number of problems
Model-based	Provides rigorous estimate of usability criterion	Measures only one component of usability
	Can be performed on interface specification	Limited task applicability

which involves an HCI expert examining the application and estimating its likely usability for a given user population. In such cases, users are not employed and the basis for the evaluation lies in the interpretation and judgement of the evaluator. There is considerable interest in this form of evaluation since it can produce results faster and presumably cheaper than user-based tests.

In HCI, two common expert-based usability evaluation methods are Heuristic evaluation (e.g. Nielsen 1994), and Cognitive Walkthrough (Wharton *et al.* 1994). Both methods aim to provide evaluators with a structured method for examining and reporting problems with an interface. The heuristic method provides a simple list of design guidelines that the evaluator uses to examine the interface screen-by-screen and while following a typical path through a given task. The evaluator reports violations of the guidelines as likely user problems. In the cognitive walkthrough method, the evaluator first determines the exact sequence of correct task performance, and then estimates, on a screen-by-screen basis, the likely success or failure of the user in performing such a sequence. In both methods, the expert must make an informed guess of the likely reaction of users and explain why certain interface attributes are likely to cause users difficulties.

These methods differ in their precise focus. Heuristic methods are based on design guidelines and ultimately reflect the expert's judgement of how well the interface conforms to good design practice. The cognitive walkthrough method concentrates more on the difficulties users may experience in learning to operate an application to perform a given task. In practice, usability evaluators tend to adapt and modify such methods to suit their purpose and many experts who perform such evaluations employ a hybrid form of the published methods.

3.3. Model-based Methods

Model-based approaches to usability evaluation are the least common form of evaluation but several methods have been proposed which can accurately predict certain aspects of user performance with an interface such as time to task completion or difficulty of learning a task sequence. In such cases, the evaluator determines the exact sequence of behaviors a user will exhibit through detailed task analysis, applies an analytical model to this sequence and calculates the index of usability.

The most common model-based approach to estimating usability is the GOMS method of Card *et al.* (1983), a cognitive psychology-derived framework that casts user behavior into a sequence of fundamental units (such as moving a cursor to given screen location or typing a well-practiced key sequence) which are allocated time estimates for completion based on experimental findings of human performance from psychology. In this way,

any interface design can be analyzed to give an estimate of expert users' time to complete a task. The model has shown itself to be robust over repeated applications (e.g. Gray *et al.* 1992), though it is limited to predicting time and only then, for error-free performance in tasks involving little or no decision-making.

4. COMPARISONS OF METHODS

The relative advantages and disadvantages of each broad method are summarized in Table 1. Since usability evaluators are trying to estimate the extent to which real users can employ an application effectively, efficiently and satisfactorily, properly executed user-based methods are always going to give the truest estimate. However, the usability evaluator does not always have the necessary resources to perform such evaluations and therefore other methods must be used.

HCI professionals have recently reported comparisons of evaluation methods but few firm conclusions can yet be drawn. John and Marks (1997) compared multiple evaluation methods and concluded that no one method is best and all evaluation methods are of limited value. Andre *et al.* (1999) attempted a meta-analysis of 17 comparative studies and remarked that a robust meta-analysis was impossible due to the failure of many evaluation comparisons to provide sufficient statistics. Caveats noted, several practical findings follow:

- It is generally recognized that expert-based evaluations employing the heuristic method locate more problems than other methods, including user-based tests. This may suggest that heuristic approaches label as problems many interface attributes that users do not experience as problems or are able to work around.
- The skill-level of the evaluator performing the expert-based method is important. Nielsen (1992) reports that novice evaluators identify significantly fewer problems than experienced evaluators, and both of these groups identify fewer than evaluators who are both expert in usability testing and the task domain for which the tool under review is being designed.
- Team or multiple expert evaluations produce better results than single expert evaluations.
- Finally, there are good reasons for thinking that the best approach to evaluating usability is to combine methods e.g., using the expert-based approach to identify problems and inform the design of a user-based test scenario, since the overlap between the outputs of these methods is only partial, and a user-based test normally cannot cover as much of the interface as an expert-based method. Obviously, where usability evaluation occurs throughout the design process, the deployment of various methods at different stages is both

useful and likely to lead to greater usability in the final product.

5. CONCLUSION

Usability evaluation is a core component of user-centered systems design and an essential competency for Human Factors professionals working in the software domain. Test methods vary from laboratory studies of user performance to model-based predictions based on an examination of the interface specification. Choosing among methods is largely a matter of determining what information is needed and at what stage in the development cycle is the evaluation to occur. It will be difficult for an expert examining a prototype to predict user satisfaction, for example, or for a user to reliably estimate her own efficiency from an interface specification, hence the need for more than one type of evaluation method. Clearly, the ultimate test is the behavior of real users interacting under normal working conditions and any single usability evaluation method is an attempt to predict some or all of the issues that will occur in real use. Improvements in evaluation methodology are therefore tied directly to increases in the theoretical analysis of the determinants of use.

REFERENCES

ANDRE, T., WILLIGES, R. and HARTSON, H. 1999, The effectiveness of usability evaluation methods: determining the appropriate criteria. In *Proceedings of the Human Factors and Ergonomics Society 43rd Annual Meeting* (Santa Monica: HFES), 1090–4.

CARD, S., MORAN, T. and NEWELL, A. 1983, *The Psychology of Human–Computer Interaction* (Hillsdale: LEA).

GRAY, W. JOHN, B. and ATWOOD, M. 1992, The precis of project Ernestine or an overview of a validation of GOMS. In *Proceedings of the Annual ACM SIGCHI Conference: CHI '92* (New York: ACM Press), 307–12.

JOHN, B. and MARKS, S. 1997, Tracking the effectiveness of usability evaluation methods. *Behaviour and Information Technology*, 16, 188–202.

LEWIS, J. 1994, Sample sizes for usability studies: additional considerations. *Human Factors*, 36, 368–78.

NIELSEN, J. 1992, Finding usability problems through heuristic evaluation. In *Proceedings of the ACM SIGCHI Conference: CHI '92* (New York: ACM), 373–80.

NIELSEN, J. 1994, Heuristic evaluation. In J. Nielsen and R. Mack (eds), *Usability Inspection Methods* (New York: Wiley), 25–62.

SHACKEL, B. 1991, Usability — context, framework, definition, design and evaluation. In B. Shackel and S. Richardson (eds), *Human Factors for Informatics Usability* (Cambridge: Cambridge University Press), 21–38.

WHARTON, C., RIEMAN, J., LEWIS, C. and POLSON, P. 1994, The cognitive walkthrough method: a practitioner's guide. In J. Nielsen and R. Mack (eds), *Usability Inspection Methods* (New York: Wiley), 105–40.

Fuzziness, Requisite Compatibility and System Design

W. Karwowski

Center for Industrial Ergonomics, Department of Industrial Engineering, University of Louisville, KY 40292, USA

J. Grobelny

Institute of Production Engineering and Management, Technical University of Wroclaw, Poland

1. INTRODUCTION

Ergonomics, which focuses on the design of human–machine systems with respect to requisite compatibility between the users, products, systems, and environments, is often faced with the complex and uncertain relationships between human characteristics, various artifacts, and human activity requirements (Karwowski 1991). Investigation of such relationships must take into consideration the natural fuzziness in human and system performance, which can be expressed in terms of degrees of vagueness, and which formally can be modeled through gradation in categories (Zadeh 1965). Uncertainty due to fuzziness is inherent to any complex system and to the human thought and perception processes (Smithson 1982). It should be noted that fuzziness, which measures the degree to which an event occurs, and not whether it occurs, differs from randomness, which describes the uncertainty of event occurrence. (Kosko 1992). Furthermore, fuzziness, i.e. a type of deterministic uncertainty, also models ambiguity as a property of physical phenomena.

In order to study complex human–machine systems it is necessary to use modeling approaches that are approximate in nature (Zadeh 1973; Karwowski 1983). Fuzzy systems methodologies allow accounting for human and human–artifact fuzziness, and, therefore, provide the necessary framework for successful modeling efforts in the ergonomics (Karwowski 1991; Karwowski and Mital 1986). As fuzziness occurs at all levels of human interactions with the outside environments, ranging from physical to cognitive tasks, it can be used as the model for human–artifact interactions. It can also help to explain the complex phenomena of human sensation, information processing, decision-making and communication and functioning (Karwowski 1991, 1992; Karwowski and Salvendy 1992).

2. HUMAN–MACHINE FUZZINESS

Contemporary models used in ergonomics utilize formal structures to represent the human–artifact–environment reality as perceived by the human operator (Karwowski *et al.* 1999). However, the process of human perception is intrinsically governed by the natural fuzziness due to the human (human fuzziness), and by complexity and related uncertainty of the system under observation (human–machine fuzziness). Fuzziness is a useful model for human language and categorizing processes, and fuzzy mathematics can be used to augment conventional statistical techniques in the analysis of fuzzy data, reliability analysis and regressions, and structurally oriented methods such as hierarchical clustering and multidimensional scaling (Smithson 1982). In view of the above, fuzzy systems provide a useful framework for modeling variety of complex tasks, situations, systems, artifacts and environments, and their interactions with people.

For example, an assessment of resource demand or mental workload imposed on human operators in human–machine systems is of great importance to assure acceptable levels of performance (Wickens 1987). The measures of mental workload, such as those of perceptual–cognitive effort or response loading, are fuzzy phenomena, by the very nature of the underlying processes (Karwowski 1992). Therefore, the assessment methods must exhibit fuzzy properties of the investigated systems. As fuzziness occurs at many levels of cognitive processes, people have the unique ability to understand and apply vague and uncertain concepts (Hersh and Caramazza 1976). Research in semantic memory showed that people perceive natural categories as sets with no clear boundaries separating members from non-members (McCloskey and Glucksberg 1978). People are also able to comprehend vague concepts of the natural language, as if fuzzy labels represented them, and manipulate them according to rules of fuzzy logic (Oden 1977).

2.1. Human Performance

Human functioning can be described in terms of perception, information processing and decision-making, memory, attention, feedback, and human-response processes. Any task requires processing of information that is gathered based on perceived and interpreted relationships between the system elements. According to the model of human information processing (Wickens 1984), the sensory information coming from the sensor organs is compared to some internal representation of the recognized object stored in the permanent (long-term) memory. After the stimulus is identified, a decision can be made regarding the appropriate course of action. The decision process may require some transformations of the available information in the working memory, information storing in the long-term memory, and generating a response. The response can be realized through a process of coordinated muscular control actions, which effects are perceived through the feedback system. An inherent fuzziness of these natural processes is discussed below.

2.2. Perception and Memory

The purpose of the human perceptual system is to provide sufficient conditions for adaptive behavior (Foley and Moray 1987). Perception is constructed by the observer rather than determined by parameters of physical signals received from the receptors. The process of conscious perception is a fuzzy one, with uncertainties as to the physiological thresholds for physical stimuli, pattern recognition, and judgment requirements. For example, a question of "At what frequency will an intermittent light be perceived as just flickering?" is an ambiguous one, and hence the eye flicker frequency is a fuzzy, not random, phenomenon, as it refers to the degree of noticeable flickering rather than the question of whether it occurs (Karwowski 1992).

Many of the psychophysical laws that relate the magnitude of change in physical stimuli (DI) that will just be noticed by an observer, also exhibit requisite fuzziness. In the Weber-Fetchner Law, the magnitude of change in a physical stimulus that will be just noticed by an observer is a constant proportion of the

stimulus. Referring to the magnitude of DI that will result in a judgment of a difference in the levels of physical stimuli 50% of the time (so called just noticeable difference or jnd), this law is not a probabilistic statement of whether the perception of change will occur. Rather, it concerns the question at what level (degree) of the difference in physical stimuli such a change will be observed.

Human ability to process information from different stimulus objects at one time is limited. However, several dimensions of a single object can be processed in parallel (Wickens 1987). The degree to which many sources of object the human brain can process stimuli and different dimensions of the same objects exemplifies the fuzzy nature of human perceptual and central processing processes (Karwowski 1992). In the human–computer system, information displayed on the screen can be arranged along a continuum that defines the degree to which that information is spatial-analog in nature (i.e. information about relative location, transformation or continuous motion), linguistic-symbolic or verbal. The two end points of this verbal–spatial continuum distinguish the verbal and spatial memory systems (Wickens 1987). The border between these two systems, however, is subject to fuzzy interpretation. For example, the capacity of working memory is limited to 7 (+/-) 2 of the unrelated items (Miller 1956). The important question is to what degree one can recall and operate on a given number of items. The degree to which the memory is short or long is also a fuzzy category (Karwowski 1992).

2.3. Attention and Decision–making

Attention is the ability to process information in parallel or time-sharing fashion (Wickens 1987), which can be measured by the graded (fuzzy) categories of success. The selective allocation of human attention is determined by the human operator's internal model of the statistical properties of the environment (Moray 1984). Decision-making and diagnosis in human–machine systems are also of uncertain nature. Decisions are typically followed by responses, many of which are intrinsically fuzzy. It should be noted here that the prevalent definition of the relationship between choice reaction time and degree of choice (the Hick-Hyman law) is based upon the information content of a stimulus (S) in bits as follows: $RT = a + b\,S$. The presence of human fuzziness cannot be overlooked in the paradigm of stimulus—response compatibility (Wickens 1987), which described the physical relationship between a set of stimuli and the speed of human response. The spatial relations in arrangements of signals and response devices with respect to direction of movement and logic of adjustments, often exhibit high levels of uncertainty regarding the effects of intended and unintended control actions (Karwowski 1992).

2.4. Human Sensations

Traditional methods for quantification of the relationships between human sensations and physical characteristics affecting them are based on multivariate analysis techniques such as multiple regression analysis and quantification theory. With higher order data, however, it is more difficult to find an adequate formula that would represent the non-linearity factor. In addition, conventional methods have typically excluded the ambiguities that can arise in the process of recognizing and making subjective

evaluations of the physical characteristics. Shimizu and Jindo (1995) developed a model accounting for non-linearity property of information processing. The benefits of the fuzzy approach is that uncertainties and non-linearity of human sensations can be accounted for and quantified in order to derive correlations with the considered physical characteristics of the product.

2.5. Mental Workload

Mental workload is the amount of mental effort necessary to perform a given task. Mental workload assessment can be used to maintain a workload level allowing acceptable performance (Proctor and Zandt 1994). Subjective assessment techniques are based on empirical methods that measure workload directly in an operational system or a simulated environment through the use of human operators' judgments. The operators are asked to rate the perceived mental effort, time load, and stress load of particular tasks. Therefore, workload is not as a scalar quantity, but rather a vector quantity associated with multiple dimensions (Moray 1982). The subjective measures of mental workload utilize the scale method, which calls for the subjects to express their feelings through rating scales or questionnaires. Unfortunately, a precise scale often cannot represent the human feelings adequately. Using the linguistic quantifiers can help to overcome this difficulty, as evidenced by recent applications of fuzzy methodologies to risk analysis with mental workload assessment (Liou and Wang 1994; Chen 1996).

2.6. Human Cognitive Processes in Industrial Tasks

Ukita *et al.* (1996) described a fuzzy-based system to model the decision-making process of the human operators involved in tuning of microwave circuits in a real-time environment. The process was automated through the application of a fuzzy knowledge-based system. In a complex tuning process, multiple circuit specification criteria have to be simultaneously satisfied by several trimmers. To reduce the number of trial-and-error steps required to meet specific circuit tuning criteria, and, consequently, the tuning process time, the order of subsequent trimmer adjustments in the tuning process, as well as the extent of each individual trimmer tuning magnitude, must be chosen very carefully. In the past, experienced workers, who would skillfully adjust a set of trimmers by hand, performed the circuit tuning process. The quality of the adjustment process was primarily based on the human operator's mental model of the circuit tuning behavior formulated from the long practice and requires specific cognitive skills of the human operator. Ukita *et al.* (1994) proposed to account for the effect of each trimmer on each of the tuning criterion by a grade of fuzzy membership related to the circuit output. The overall effect of each trimmer on the circuit tuning performance was modeled by an aggregation of the fuzzy grades used for trimmer selection. The model simulation showed that the geometrical average operator was the best method for evidence aggregation of fuzzy evidence in modeling of the human cognitive processes underlying manual circuit tuning tasks.

3. HUMAN–MACHINE RELIABILITY

Human reliability is typically defined as a probability that a person correctly performs some system-related activity in a required time period, or as the probability of a successful performance of a

task. However, given the natural fuzziness of human interactions with the outside world, both probabilistic and possibilistic approaches to human reliability are needed (Terano et al. 1983). Since the human operator characteristics which affect the reliability include such fuzzy processes as sensory perception, motor control, cognitive functions (information processing and decision-making), and meta-cognitive functions (intuition and abstract processing), human reliability models must allow for qualitative reasoning through human error possibility measures, and should account for natural vagueness of higher cognitive functions. Onisawa (1988) proposed the concept of human error based on the possibility measure that is affected by several human performance-shaping factors. Pedrycz (1990) also discussed cognitive aspects of information processing and proposed a fuzzy framework for development of human perception perspective as opposed to machine perspective.

4. FUZZINESS IN HUMAN—COMPUTER INTERACTION

The interactions between people and computers often reflect the cognitive fuzziness of the data, as well as users' uncertainty exhibited in perception of the computing environment. The field of human–computer interaction (HCI) focuses on developing effective models of such inter-relationships, including the fuzzy-based communication tools. Karwowski et al. (1990) discussed the problems of fuzziness due to high complexity of human–computer systems and the nature of user's perception and information processing.

4.1. Fuzzy GOMS Model

According to GOMS model (Card et al. 1983), the user's cognitive structure consists of the following four components: (1) a set of Goals, (2) a set of Operators, (3) a set of Methods for achieving the goals, (4) a set of Selection Rules for choosing among competing methods for goals. Karwowski et al. (1990) proposed extensions to GOMS to account for uncertainty within selection rules and to generalize the Goals, Operators and Methods components as either precise or fuzzy. The Selection Rules were also expressed in either probabilistic or fuzzy variables. A variation of the manuscript editing experiment by Card et al. (1983), with the subject verbalizing five different cursor placement rules, was used. The fuzzy GOMS model utilized the sets of Goals and Operators that were precisely defined, while the (predicted) Methods used by subjects, as well as specific Selection Rules applied to accomplish the editing task included application of linguistic descriptors, fuzzy logic, and possibilistic measures of uncertainty. The rules, methods, and corresponding membership functions were elicited from the subjects, and theory of possibility was used to model the subject's rule selection process. Each of the potential rules was assigned a possibility measure equal to the membership value(s) associated with it during the elicitation phase of experiment. The non-fuzzy GOMS model successfully predicted 58.7% of the responses, while the fuzzy GOMS model predicted significantly more correct responses, i.e. 82.3% of all decisions (Karwowski et al. 1990).

4.2. Computer Screen Design

Effectiveness of the human interactions with computer systems depends to a large extent on computer screen design (Tullis 1981).

Among well-defined relationships between screen formats, there are many rules of thumb, which are based on subjective views and anecdotal knowledge (Grobelny *et al.* 1995). Limited empirical data and lack of consistent measures and quantitative criteria for assessment of screen quality makes the evaluation of system efficiency and comparison of different screen designs difficult. Fuzzy-based linguistic patterns for assessment of the computer screen design quality have been proposed. The linguistic patterns are based on intuitive expressions closely related to natural language and truth-values. The proposed system of concepts, relations and definitions, includes: (1) the implication and definition of linguistic variables, (2) a degree of truth of an implication, (3) intensity levels of implication variables, (4) degree of truth of the consistency of two expressions, (5) definitions of linguistic relationships, and (6) definitions of modifiers for linguistic expressions and connectors.

For example, searching times for desired information can be reduced if the information presented on the screen is organized in groups of closely related items. Furthermore, the amount of presented information greatly influences the searching times. If groups are significantly larger than optimal, the mean group size becomes the main factor that can be used to predict searching times. The study demonstrated that it is possible to achieve rational and relatively easy to interpret assessment of different screen designs in the form of the degrees of truth (Grobelny *et al.* 1994).

5. PHYSICAL WORKLOAD

5.1. Modeling of Manual Lifting Tasks

Application of the psychophysical approach to setting limits in manual lifting tasks requires the subject to adjust the weight of load according to his or her perception of effort in order to minimize the potential for over-exertion or excessive fatigue. Karwowski (1983) and Karwowski and Ayoub (1984) applied fuzzy sets to model and assess the acceptability of stresses involved in manual lifting task. They hypothesized that a combination of acceptability measures of the biomechanical and physiological stresses leads to an overall (psychophysical) measure of lifting task acceptability.

Fuzzy modeling was also used to investigate the relationships between physical weight, its perceived heaviness, and size of load. Luczak and Ge (1989) noted that it is not clear why the handling of a small box with a certain weight is sometimes perceived heavier than handling a bigger box with the same weight (size-weight illusion). They asked the subjects to express the relationships between physical weight and its perceived heaviness. Load heaviness levels were expressed using fuzzy sets, including: "very light," "light, moderate," "heavy" and "very heavy." The derived relationships illustrate how fuzzy measures can be used to quantify natural fuzziness underlying human cognitive processes. Recently, Yang et al. (1998) extended the study by Luczak and Ge (1989) to Asian population of workers, focusing on load heaviness in relation to perceived weight lifted.

Following the 1982 study, Karwowski and his co-workers performed variety of research projects focusing on understanding of human perception of load heaviness, and fuzzy modeling of acceptable and safe weights for lifting tasks (see Yang *et al.* 1998; Karwowski *et al.* 1999). Recently, Karwowski *et al.* (1999) demonstrated experimentally that perception of load heaviness

is subject to fuzziness of human cognitive processes, and concluded that this phenomenon should be taken into account when setting limits in manual lifting tasks. Examples of other important studies in this area include development of the model of ergonomic workload stress index (Chen *et al.* (1994), and fuzzy knowledge-based decision support system for recommending the maximum acceptable weights of lift, based on data generated from the job severity index (Ngo *et al.* (1996).

5.2. WORK–RELATED MUSCULOSKELETAL DISORDERS

Work-related musculoskeletal disorders (WRMDs), such as carpal tunnel disorders, tendinitis, chronic muscle strain, and degenerative joint diseases, are a major occupational health problem (Karwowski and Marras 1997). These cumulative trauma disorders (CTDs) may occur when a force is applied repeatedly over a prolonged period to the same muscle group, joints, or tendons, and are linked to jobs that exhibit repeated or awkward postures. In order to implement adequate prevention and health program, it is necessary to document the relationships between work exposure (prevalence of job related risk factors in a production environment) and specific musculoskeletal disorders. From the perspective of CTDs prevention, a desirable situation is to predict the possibility of occurrence of CTDs in a given occupational setting. A fuzzy system methodology can be used to determine the possibility of occurrence of CTDs, given the available relationships between CTDs, risk factors and their severity level. It should also be noted that in many cases, the human-expert is the only available "measurement device" for observing, coding and classifying interesting data. Therefore, the linguistic expressions are more adequate and reliable tools for human analyses than precise numerical measurements in many real life situations. Zadeh (1978) introduced a concept of the linguistic variable, which enables an application of linguistic, expert-defined expressions. The main point of this approach is that natural language expressions describing level or intensity of a given value can be defined in appropriate spaces as fuzzy sets.

5.3. Fuzzy Modeling of CTDs

Theoretical framework for fuzzy modeling of the risk of cumulative trauma disorders (CTDs) was first proposed by Grobelny and Karwowski (1992). The developed conceptual model for quantification of risk for work-related musculoskeletal disorders was based on fuzzy logic, linguistic variables, and knowledge provided by the human experts. For example, the following linguistic proposition was defined with respect to work-related risks of CTD:

PROPOSITION:
B: IF Required force is BIG THEN Exposure time is SHORT
C: IF Deviation from neutral posture is BIG AND Required force is BIG THEN Exposure time is VERY SHORT

The logical expressions (A AND B AND C) were defined using the "General Desired State Patterns" (GDSP), i.e. "natural" descriptions that constitute a basis for a formal assessment of working environment. The GDSP can be compared with risk factors associated with any industrial job to generate the logical truth-values: FALSE or TRUE, respectively, for jobs fulfilling or not fulfilling the above pattern. This was done according to the principle of implication (truth value determination).

For example, using the categorical risk descriptors (HIF — high force, HIR — high repetition), a general pattern of "desired conditions for repetitive work of the hand" was defined as follows:

PROPOSITION:
A: IF Force is HIF THEN Exposure time is SHORT, AND
B: IF Repetition is HIR THEN Exposure time is SHORT, AND
C: IF Force is HIF AND Repetition is HIR THEN Exposure time is VERY SHORT

Assuming the values of measured risk factors for CTD to be: Force = 8 kg, Repetition = 0.8/s and Exposure time = 10%, one can set the values: HIF(8) = 0.8, HIR(0.8) = 0.8, SHORT(10) = 0.5, and VERY SHORT (10) = 0.25.

The following formula proposed in multivalued Lukasiewicz logic (Grobelny 1988) can then be used:

If p and q are both truth values of P and Q statements, respectively, then truth value of implication

P => Q can be calculated according to: T (P => Q) = min (1, 1 - p + q)

in order to assess the values of this PROPOSITION as follows:
A: min (1, 1 - 0.8 + 0.5) = 0.7
B: min (1, 1 - 0.8 + 0.5) = 0.7

Also, the "min" operator defined in the Cartesian product of relationship space was used to represent the relationship "AND". According to this rule, the left side of the sub-pattern C for exemplary data (Force = 8 kg, Repetition = 0.8/sek) can be evaluated as follows:

min(HIF(8 kg), HIR (0.8/sek)) = min (0.8, 0.8) = 0.8, and consequently: C: min (1, 1 - 0.8 + 0.25) = 0.45

Using the "min" operation for evaluation the whole pattern: (A AND B AND C):

T (PROPOSITION) = min (0.7, 0.7, 0.45) = 0.45.

The "possibility measure" (Zadeh 1978) allows to widen the proposed approach to situations when the input data to PROPOSITION is fuzzy. Assuming that it is possible to obtain a reasonable system of linguistic representations for problems modeled by PROPOSITION, the following expert's opinions could be possible:

Force is ABOUT 4, Repetition is HIGH, Exposure time is MEDIUM. Having all parameters defined and implemented on a computer model, the user can derive the general "pattern satisfaction index", as well as the sub-pattern truth-values, expressing quantitatively the relative CTD risk at the workplace.

Recently, Bell and Crumpton (1997) also proposed a fuzzy linguistic model for predicting the risk of carpal tunnel syndrome (CTS). The model utilizes fuzzy sets to quantify the risk associated with development of this neuropathy. The first set of membership functions involved utilizing the linguistic risk level obtained by the expert knowledge acquisition. The second set of membership functions was derived to rate the possibility of the hazard associated with a particular linguistic variable. The membership functions of the two variables were intuitively determined based on graphical representation of the physical data.

6. CONCLUSIONS

Fuzziness represents natural vagueness and uncertainty inherent to any human–artifact–environment systems. Fuzzy systems

methodologies allow accounting for the human and system fuzziness, and provide a useful framework for investigating the requisite human—system compatibility, and for modeling in the human factors and ergonomics discipline in general. Designers should treat the fuzziness in human performance as natural system requirement.

REFERENCES

BELL, P.M. and CRUMPTON, L., 1997, A fuzzy linguistic model for the prediction of carpal tunnel syndrome risks in an occupational environment, *Ergonomics*, 40, 790–9.

CARD, S.K., MORAN, T.P., and NEWELL, A. (1983). *The Psychology of Human-Computer Interaction* (London: Lawrence Erlbaum).

CHEN, J.-G., JUNG, H.S., and PEACOCK, B.J., 1994, A fuzzy sets modeling approach for ergonomic workload stress analysis, *International Journal of Industrial Ergonomics*, 13, 189–216.

CHEN, S.-M., 1996, New methods for subjective mental workload assessment and fuzzy risk analysis, *International Journal of Cybernetics and Systems*, 27, 449–72.

FOLEY, P. and MORAY, N., 1987, Sensations, perception and systems design, G. Salvendy (ed.), *Handbook of Human Factors* (New York: Wiley), 45–71.

GROBELNY, J., 1987, The fuzzy approach to facilities layout problems, *Fuzzy Sets and Systems*, 23, 175–90.

GROBELNY, J., 1988, The 'Linguistic Pattern' method for a workstation layout analysis, *International Journal of Production Research*, 26, 1779–98.

GROBELNY, J. and KARWOWSKI, W., 1992, Quantifying risk estimations of a work-related musculoskeletal disorders according to an expert knowledge: the methodological framework, unpublished report, Technical University of Wroclaw, Wroclaw, Poland.

GROBELNY, J., KARWOWSKI, W. and ZURADA, J., 1995, Applications of fuzzy-based linguistic patterns for the assessment of computer screen design quality, *International Journal of Human-Computer Interaction*, 7, 193–212.

HERSH, M.M. and CARAMAZZA, A., 1976, A fuzzy set approach to modifiers and vagueness in natural language, *Journal of Experimental Psychology: General*, 105, 254–76.

KARWOWSKI, W., 1983, A pilot study on the interaction between physiological, biomechanical and the psychophysical stresses involved in manual lifting tasks, K. Coombe (ed.), *Proceedings of Ergonomics Society Conference*, (London: Taylor & Francis), 95–100.

KARWOWSKI, W., 1991, Complexity, fuzziness and ergonomic incompatibility issues in the control of dynamic work environments, *Ergonomics*, 34, 671–86.

KARWOWSKI, W., 1992, The human world of fuzziness, human entropy, and the need for general fuzzy systems theory. *Journal of Japan Society for Fuzzy Theory and Systems*, 4(5), 591–609.

KARWOWSKI, W. and AYOUB, M.M., 1984, Fuzzy modeling of stresses in manual lifting tasks, *Ergonomics*, 27, 641–9.

KARWOWSKI, W. and MARRAS, W. S., 1997, Cumulative trauma disorders, in G. Salvendy (ed.), *Handbook of Human Factors & Ergonomics*, (New York: Wiley), 1124–73.

KARWOWSKI, W. and MITAL, A., 1986, *Applications of Fuzzy Set Theory in Human Factors*, (Amsterdam: Elsevier).

KARWOWSKI, W. and SALVENDY G., 1992, Fuzzy-set-theoretic applications in modeling of man-machine interactions. In R.R. Yager and L.A. Zadeh (eds), *An Introduction to Fuzzy Logic Applications in Intelligent Systems,* (Boston: Kluwer), 201–220.

KARWOWSKI, W., MAREK, T., and NOWOROL, C., 1988, Theoretical basis of the science of ergonomics, *Proceedings of 10th Congress of International Ergonomics Association*, Sydney, Australia (London: Taylor & Francis), 756–8.

KARWOWSKI, W., MAREK, T., and NOWOROL, C., 1994, The complexity incompatibility principle in the science of ergonomics. In F. Aghazadeh (ed.), *Advances in Industrial Ergonomics and Safety,* vol. VI, (London: Taylor & Francis), 37–40.

KARWOWSKI, W., AYOUB, M.M., ALLEY, L.R., and SMITH, T.L., 1984, Fuzzy approach in psychophysical modeling of human operator-manual lifting system, *Fuzzy Sets and Systems*, 14, 65–76.

KARWOWSKI, W., KOSIBA, E., BENABDALLAH, S., and SALVENDY, G., 1990, A framework for development of fuzzy gomes model for human-computer interaction, *International Journal of Human-Computer Interaction*, 2, 287–305.

KARWOWSKI, W., LEE, W.G, JAMALDIN, B., GADDIE, P. and R. JANG, 1999), Beyond psychophysics: a need for cognitive modeling approach to setting limits in manual lifting tasks, *Ergonomics*, 42(1): 40–60.

KOSKO, B., 1992, *Neural Networks and Fuzzy Systems: A Dynamical Systems Approach to Machine Intelligence* (Englewood Cliffs NJ: Prentice Hall).

LIOU, T.-S. and WANG, M.-J.J., 1994, Subjective assessment of mental workload: a fuzzy linguistic multi-criteria approach, *Fuzzy Sets and Systems*, 62, 155–65.

LUCZAK, H. and GE, S., 1989, Fuzzy modeling of relations between physical weight and perceived heaviness: the effect of size-weight illusion in industrial lifting tasks, *Ergonomics*, 32, 823–38.

MCCLOSKEY, M.E. and GLUCKSBERG, S. (1978). Natural Categories: Well-Defined or Fuzzy Sets?, *Memory and Cognition*, 6, 462–472.

MERRITT, T.W. and GOPALAKRISHNAN, S., 1994, An application of fuzzy set theory to the prediction of cumulative trauma disorders of the upper extremity, *International Journal of Industrial Ergonomics*, 13, 95—105.

MILLER, G.A., 1956, The magical number seven plus or minus two: some limits on our capacity for processing information, *Psychological Review*, 63, 81–97.

MORAY, N., 1982, Subjective mental workload, *Human Factors*, 24, 25–40.

MORAY, N., 1984, Attention to dynamic visual displays in man-machine systems. In R. Parasuraman and R. Davies (eds), *Varieties of Attention*, (New York: Academic Press), 485–512.

NGO, A.D., MALINOWSKI, P., KAJI, S., and BALAZINSKI, M., 1996, Fuzzy logic approach to manual material handling tasks. In R.J. Koubek and W. Karwowski (eds), *Manufacturing Agility And Hybrid Automation-I*, (Iea Press), 588–91.

ODEN, G.C., 1977, Human perception and performance, *Journal of Experimental Psychology*, 3, 565–75.

ONISAWA, T., 1988, An approach to human reliability in man machine systems using error possibility. *Fuzzy Sets And Systems*, 27, 87–103.

PEDRYCZ, W., 1990, Fuzzy set framework for development of a perception perspective. *Fuzzy Sets and Systems*, 37, 123–37.

PROCTOR R.W. and ZANDT, T.V., 1994, *Human Factors: In Simple and Complex Systems*, (London: Allyn & Bacon).

SHIMIZU, Y. and JINDO, T., 1995, A fuzzy logic analysis method for evaluating human sensitivities. *International Journal of Industrial Ergonomics*, 15, 39–47.

SMITHSON, M., 1982, Applications of fuzzy set concepts to behavioral sciences. *Mathematical Social Sciences*, 2, 257–74.

TERANO, T., MURAYAMA, Y., and AKIJAMA, N., 1983, Human reliability and safety evaluation of man machine systems. *Automatica*, 19, 719–22.

TULLIS, T.S., 1981, An evaluation of alphanumeric, graphic and color information display. *Human Factors*, 23, 541–50.

UKITA, A., KARWOWSKI, W., and SALVENDY, G., 1994, Aggregation of evidence in a fuzzy knowledge-based method for automated tuning of microwave electric circuits. *Journal of Intelligent and Fuzzy Systems*, 2, 299–313.

WICKENS, C.D., 1984, *Engineering Psychology and Human Performance* (Columbus, OH: Charles Merrill).

WICKENS, C.D., 1987, *Information Processing, Decision-Making, and Cognition*. In G. Salvendy (ed.), *Handbook of Human Factors* (New York: Wiley), 72–107.

YANG, Y.N, KARWOWSKI, W. and LEE Y.H., 1998, Load heaviness and perceived weight lifted: implications of human cognition for setting design limits in manual lifting tasks. *Occupational Ergonomics*, 1(4), 291–303.

ZADEH, L.A., 1965, Fuzzy sets. *Information and Control*, 8, 338–53.

ZADEH, L.A., 1973, Outline of a new approach to the analysis of complex systems and decision processes. *IEEE Transactions on Systems, Man and Cybernetics*, SMC-3, 28–44.

ZADEH, L.A., 1978, Fuzzy sets as a basis for a theory of possibility. *Fuzzy Sets and Systems*, 1, 3–28.

ZURADA, J., KARWOWSKI, W. and WRIGHT, A.L., 1996, A fuzzy logic-based decision unit for a robot safety system. In R.J. Koubek and W. Karwowski (eds), *Manufacturing Agility and Hybrid Automation–I* (Louisville, Kentucky: IEA Press), 580–3.

Human Factors Testing and Evaluation

S. G. Charlton
Department of Psychology, University of Waikato, Private Bag 3105,
Hamilton, New Zealand

1. INTRODUCTION

It can be argued that human factors testing and evaluation is among the most applied of endeavors within an avowedly applied discipline. Testing and evaluation occur near the end of a long human factors chain of basic research, applied research, and product design. Yet, for a variety of reasons, the procedures and findings of tests are not widely disseminated or read, and human factors testing and evaluation remains a somewhat insular discipline. This article explores some of the similarities and differences between human factors testing and the more familiar practice of basic human factors research, in the hopes of making testing more accessible to the wider human factors community.

At first blush, basic human factors research and human factors testing and evaluation would appear to have much in common. Laboratory researchers and testers often share a common background and use similar terminology and techniques. The differences between the two disciplines only become apparent when we survey their goals, work environments, practices, and products.

2. THEORETICAL INTEREST VERSUS PRAGMATIC UTILITY

It has been argued that human factors fieldwork in general and testing in particular too often focus on test-specific measures, failing to address important theoretical issues and the advancement of general human factors knowledge (Mackie 1984). In counterpoint, human factors testers have decried the topics and methods chosen by laboratory researchers as being overly simplistic and of little practical application (Meister 1986).

Academics and students new to the practice of human factors testing are often disconcerted by the expediency and pragmatism required by the testing environment. Experienced testers, on the other hand, may have little tolerance for the details of a key theoretical conundrum, preferring instead to scan the methods section in the hopes of finding useful procedures or forms for field use. These contrasting outlooks, theoretical interest versus pragmatic utility, underlie many of the differences between human factors testers and human factors researchers.

3. WORK ENVIRONMENT

Testers' predilections for pragmatism over theory arise in part from the people involved in testing and who they work with. Whereas a doctorate or other advanced degree is often a prerequisite for academia or research laboratories, these qualifications are seldom needed by testers. In the world of test and evaluation a PhD may actually mark an individual as potentially abstruse or difficult until proven otherwise. One of the defining features of the work environment enjoyed by academics and laboratory researchers is the collegial company of other members of their discipline. In contrast, someone employed in test and evaluation is typically surrounded by individuals with diverse backgrounds, training, and practical experiences.

This day-to-day isolation from one's peers has two effects on the human factors tester. First, human factors testers typically become staunch advocates for human factors as the only spokesperson for their discipline. Second, and somewhat paradoxically, it has the effect of making the tester more acutely aware of other perspectives, those of engineers, designers, and users. In fact, some of the most effective testing occurs when the individuals involved set aside the parochial issues of their own disciplines and coordinate their efforts towards the test program.

4. COVERAGE OF ISSUES

The contrasting outlooks of researchers and testers can best be seen in how they structure the objects of their work, laboratory experiments and tests. Whereas laboratory experiments are focused on one or even a few human factors issues, a test program will cover a potentially vast range of issues. Human factors tests often favor a "scattergun" approach, collecting data across many aspects simultaneously. The careful control of the laboratory is typically sacrificed for the more realistic, and more unpredictable, conditions of the field test. The measures and metrics employed are selected to fit the available test conditions and data management resources rather than out of allegiance to any theoretical perspective.

This "unprincipled" approach has led some laboratory-based researchers to argue that there should be a greater reliance on psychological theory in the selection of human factors test measures. These authors hold that a strong theoretical orientation provides a means of clarifying one's view of the world in designing a test and interpreting the data. In other words, human factors testing, when conducted from a particular theoretical model, would enable the tester to more readily identify the important system issues, focus test resources on those issues, generate predictions, and help to fill in where data is lacking (Kantowitz 1992).

These arguments have not been uniformly well received by the test and evaluation community. Their pragmatic sensibilities tend to reject any approach that carries with it a danger they will be drawn beyond the data in forming conclusions about the results of a test. Although the role of theory in guiding laboratory research is unchallenged, some testers have noted that theories can be seductive in their apparent ability to provide precise predictions about what will happen when a variable is changed in a new way. The danger of theoretical reification, treating the theoretical constructs as if they were real processes and mechanisms, makes many testers chary about relying on them to any great degree in human factors testing and evaluation (O'Brien and Charlton 1996).

5. OBJECTIVE AND SUBJECTIVE MEASURES

Another difference lies in the routine use of questionnaires to collect a variety of human factors test and evaluation data. Although objective performance data is collected for many tests, the cost of instrumentation and impracticality of factorial designs for an invariably short test schedule force testers to rely on test participants as data integrators and to report on many aspects of

performance. This reliance on subjective measures is an anathema to many laboratory researchers trained in a scientific method which worships at the altar of objectivity.

Even the content of test and evaluation questionnaires and interviews is viewed askance by many laboratory researchers. Instead of molecular questions designed to extract the greatest diagnostic value possible on a psychological state or operator task, test and evaluation questions are fairly high-order, asking test participants to integrate multiple aspects of usability and performance. Given the large number of issues to address, a test and evaluation questionnaire must adopt this high-order approach or run the risk of overloading the respondents with overly long questionnaires and compromising the operational fidelity of the test (Andre and Schopper 1997).

6. SAMPLE SIZES AND STATISTICS

Researchers and testers also differ in the statistical treatment of their data. Only rarely will a classical hypothesis testing model comparing control and experimental groups fit a test and evaluation programme. Because of the expense involved in conducting test and evaluation, sample sizes are typically rather small and the data contains a high proportion of ordinal and categorical measures from questionnaires. For these reasons, testers are more likely to use nonparametric statistics than their laboratory-based counterparts.

Testers also make extensive use of descriptive analyses and circumstantial relationships between variables rather than the cause and effect level of proof expected of a well-designed laboratory experiment. Regression analyses, for example, can be used to identify situations and operator tasks that are "at risk" or highly correlated with human error and poor system performance. This approach is analogous to the medical epidemiological approach specifying risk factors associated with a disease in the absence of causal data specifying the physiological mechanism of the disease (Charlton 1992).

When inferential statistics are used in test and evaluation it is to compare the test results to a system requirement or standard. Instead of comparing an experimental treatment to a control group, the comparison is to a specific criterion to achieve a "met" or "did not meet" evaluation of the system under test. The availability of clear criteria for human factors measures, however, is the exception rather than the rule. Although design standards do exist for the layout and anthropometry of an operator's workstation, they are seldom the limiting factors in larger systems. The more important issues in human factors testing involve presenting information and designing tasks so that system operators can correctly understand and respond to the information. Few criteria exist for evaluating the way system information is presented and for determining whether it is adequate to support a diverse and changing repertoire of operator tasks.

7. REPORTS AND PUBLICATIONS

Finally, given the differences between researchers and testers, it should come as no surprise that the product of their work is different as well. The results of human factors tests are documented in technical reports of various sorts. Because of their limited availability and nontraditional format, technical reports are regarded as nonarchival in many academic circles. Because technical reports do not go through the traditional peer review process (although they are subjected to considerable review and editorial scrutiny) they are viewed as of lesser merit than a journal publication. Because of its differences from traditional laboratory research, the subject matter of test and evaluation is often seen as not worthy of publication in journals and books. All of these conditions have conspired to keep some important findings out of the mainstream of human factors.

Although they frequently appear in the reference sections of journal articles, technical reports are difficult to find and obtain. Learning about the existence of an important technical paper is haphazard at best and obtaining a copy depends on the goodwill of the authors. Sometimes a report is not available because a government has restricted its distribution or because a company thinks it contains commercially sensitive information. Some reports do not fit the traditional format of literature review, method, results, and discussion. But it would be unreasonable to change them so they more closely resemble journal articles; they are written for a different audience and for quite different reasons.

Rather than review the literature or describe competing hypotheses, technical reports tell a human factors story to an audience more concerned with the system than with psychological principles. Most laboratory researchers would be ill-prepared to report human factors results in a way that makes sense to decision makers outside the human factors profession. Testers must focus on developing a compelling human factors story that places the findings in the perspective of whether the system passed or failed the test, and why.

In spite of all the differences between technical reports and mainstream books and journals, there is a wealth of information that human factors testers and test results can offer the laboratory researcher. Researchers looking for applied topics to pursue or even fresh ideas on existing areas of inquiry could readily find them in test and evaluation reports. Technical reports are also an underutilized means of verifying the external validity of generalizations made in the laboratory. Until they become more available and more palatable to academic audiences or until there is another conduit for reporting test and evaluation, their usefulness to other researchers will remain thwarted.

8. CONCLUSION

Human factors testing and evaluation is by necessity quite different than laboratory-based human factors research. Tests are dedicated to the evaluation of particular products and processes rather than explicating a theoretical question or general principle of human engineering. The methods of test and evaluation are more sensitive to practical expediency and resource constraints than laboratory research. Their extensive use of subjective measures and nonparametric statistics is unfamiliar and unsettling to many laboratory researchers.

The requirement to select issues and employ methods that are uniquely suited to the testing environment has created a situation where test results are seen as contributing little to the advancement of general knowledge about human factors. By the same token, laboratory research is often viewed as being out of touch with the "real-world" considerations of test and evaluation. Perhaps this article has shown that human factors testing and evaluation is no less worthwhile than traditional research, and ultimately it may serve the same goal: making the products

and processes used by humans safe, effective, and satisfying to use.

REFERENCES

ANDRE, T.S. and SCHOPPER, A.W., 1997, *Human Factors Engineering in System Design* (Dayton OH: CSERIAC).

CHARLTON, S.G., 1992, Establishing human factors criteria for space control systems. *Human Factors*, 34, 485–501.

KANTOWITZ, B.H., 1992, Selecting measures for human factors research. *Human Factors*, 34, 387–98.

MACKIE, R.R., 1984, Research relevance and the information glut. In Muckler (ed.) *Human Factors Review 1984* (Santa Monica CA: Human Factors Society), pp. 1–12.

MEISTER, D., 1986, *Human Factors Testing and Evaluation*. (Amsterdam: Elsevier).

O'BRIEN, T.O. and CHARLTON, S.G., 1996, *Handbook of Human Factors Testing and Evaluation* (Hillsdale NJ: Lawrence Erlbaum).

Human Factors and Ergonomics Testing

S. G. Charlton

Department of Psychology, University of Waikato, Private Bag 3105, Hamilton, New Zealand

1. INTRODUCTION

Testing brings out the best and the worst aspects of human factors and ergonomics. At its best, it takes knowledge and methods out of the laboratory and makes them available to the general population in the form of tools, processes, and systems that are effective, efficient, intuitive, and easy to use. At its worst, it can represent a battleground where the competing interests of developers and users, scientists and engineers, designers and financial managers, are locked in conflict over the cost, implementation, and even the purpose of system designs.

Human factors and ergonomics testing is conducted to verify the efficiency, safety, and user acceptance of products and processes. This is accomplished through the systematic measurement and analysis of a range of system characteristics, performance variables, and user opinions. There are numerous stakeholders in the test and evaluation process, each bringing different perspectives and motivations to the test proceedings and doing much to shape the how, what, and when of the testing, even the ultimate interpretation of the results. This attention exists not because of the intrinsic interest of human factors issues, but rather because the outcome of human factors testing often has profound implications for production costs, product sales, training, maintainability, and liability protection. Thus, human factors testing often has a significant impact on the ultimate design of new products, yet the issues considered during testing are invisible to the public.

Human factors testing occurs in a variety of forms, and for a variety of reasons, and as a consequence defies a simple characterization based on steps to be accomplished or features to be included. We can, however, consider some of the defining moments encountered in the life of a human factors test, and in doing so illustrate many of the issues and decisions that occur as a test moves from the earliest stages of planning through to execution and evaluation.

2. REQUIREMENTS DEFINITION AND SYSTEM DESIGN

The first defining moment for human factors testing occurs during the early stages of concept exploration and system design, the definition of system requirements and design specifications. The human factors characteristics of the design (e.g., physical ergonomics, software usability, operability, design for maintainability, training, and staffing requirements) must be defined in such a way that they are verifiable during test and evaluation. It is self-evident that the requirements will direct the design effort, but they will also determine the character and accomplishment of the test program. Yet the statement of human factors requirements is as difficult as it is important. The source of this difficulty lies in defining human factors requirements in testable terms with measurable criteria. All too often, consideration of what, if any, human factors requirements a new system or product should have is limited to the simple (and untestable) statement that it should be "user friendly."

Human factors and ergonomics requirements can be stated either in terms of specific design characteristics or in terms of their impact on human and system performance. Requirements based on design characteristics might refer to relevant anthropometric and ergonomic standards for physical sizing of equipment, locations of controls, labeling and display of information, and so forth. Defining human factors requirements in terms of their impact on human and system performance is practised less frequently and requires more analytical effort.

Typical performance-based requirements might state that the system should be designed to enable a novice user to operate it proficiently within 1 hour, prevent operator errors from causing any critical system failures, enable a trained operator to complete an average of 10 transactions within 30 minutes, or some other performance-based measure. Performance requirements adopt a functional or system perspective and individual system components (including operator interfaces, training programs, etc.) are designed to achieve the required performance levels (Andre and Schopper 1997).

3. TEST PLANNING

A second defining moment for human factors test and evaluation occurs in the development of the test plan. The test plan documents the structure of the test, the test participants, the data collection methods, and the analyses to be performed. The written plan may have important contractual and resource functions but it is not nearly as important as the act of test planning. The planning exercise confronts the tester with a number of important trade-off decisions between the degree of test control and the test fidelity, the expertise of test participants and their representativeness, and the comprehensiveness of measures employed and the expense of the test (O'Brien and Charlton 1996).

The first set of decisions confronting the test planner is the overall design and structure of the test. Some aspects of the test design may be determined by the type of human factors requirements that have been defined (if any). Consider the test implications of the two types of human factors requirements described earlier. Characteristics-based requirements will encourage a testing program directed at verifying whether human engineering design standards are met. Checklist evaluations of the anthropometry and even user preferences can be collected early on, using equipment mock-ups and prototype components.

In contrast, performance-based requirements treat the human user as an integral component of the system being designed. The inputs and outputs of the user in relation to the other system components become explicit factors in the evaluation of the design. Because the test must focus on human performance in the context of other system components, the test conditions are more constrained, often requiring a nearly complete system or at least a fully functional prototype to conduct the test.

The next consideration in test design is selection of the test venue and the test participants. The test venues range from the carefully controlled environs of the laboratory to relatively unstructured field trials. The degree of experimental control during a test is a trade-off with the operational realism of the test

conditions. By employing a controlled test scenario in a laboratory or simulator environment the researcher runs the risk of decreasing the representativeness of the test, hence the generalizability of the results produced. Laboratory situations and equipment configurations may differ substantially from the eventual products and conditions of use.

Similarly, while field tests may provide a more realistic portrayal of the ultimate use of the product or process, the multitude of factors outside the evaluator's control (bad weather, delays, data loss, etc.) and the expense of field testing may restrict the quality and quantity of the data produced. Nonetheless, there is an achievable balance between test control and test fidelity, and the system issues and design requirements will have much to do with the final selection of the test venue (Meister 1986).

A related issue concerns the selection of the test participants. Ideally the human factors tester would seek participants who are representative of the intended user population in terms of training and experience, and who are happy and willing to participate in the test, and in sufficient numbers for statistical reliability. But sometimes it may be difficult or inappropriate to obtain truly representative participants for testing an unproven system or product. In the classic case of aircraft design, highly experienced test pilots are used instead of "typical" aviators; this is for safety and because they add expertise to the design effort.

In other cases, time constraints may militate against the use of actual members of the intended user population. This is often the case in the design of new training systems. Rather than test a new training system by subjecting actual students to a lengthy and unaccredited course of instruction, subject matter experts and training professionals can be enlisted to thoroughly exercise and evaluate the potential training effectiveness of the system. The use of subject matter experts as participants carries with it the danger that the test may not capture problems relevant to the ultimate user population. For example, even highly experienced instructors may not detect a significant design flaw until they use the system to teach real students.

Furthermore, when participants are selected on the basis of their expertise, it may be difficult to obtain them in numbers sufficient for statistical reliability. In choosing between representative users and experts, the human factors tester must balance the needs of the test design against the safety, comfort, and availability of the test participants.

Test planning also involves the selection of measurement methodologies. For any aspect of human factors and ergonomics design there are always multiple metrics available to the tester. The measures used in a test are often selected out of familiarity and convenience and they very much depend on the expertise and capabilities of each tester. One test planner may favor a psychological research perspective whereas another may have an industrial engineering orientation, hence the test metrics they employ may differ quite substantially.

A more significant concern lies in the failure to select a comprehensive set of measures. Selection of measures that are insensitive or irrelevant to the design issues at hand may provide false or misleading human factors data. There is no standard suite of test measures, just as there is no standard system to be tested, so for each system a tester must select a set of measures that are relevant to the requirements and design issues, as well as being

comprehensive and appropriate to the test structure (Bittner 1992, Kantowitz 1992).

4. DATA COLLECTION

The next defining moment occurs as the test plans are implemented and data collection begins. Depending on when they occur, human factors tests can be quite dissimilar in their structure, the resources employed, and the conclusions that can be drawn from them. Early on, human factors and ergonomics testing can occur as part of the design and engineering work. Computer design tools, simulations, and physical mock-ups are often used by human factors testers to evaluate ergonomic and safety issues in what is sometimes called developmental test and evaluation. Component-level testing can assess the usability of prototype controls and displays and consider their implications for user performance. Testing at this stage is conducted to reveal and correct problems through a process of design, test, and redesign.

When one or more fully functional prototypes have been developed, testing can begin to focus on full-scale or end-to-end tests with representatives of the intended user population. The testing performed at this stage is somewhat independent of development and design activities. Sometimes called operational testing or acceptance testing, the goal is to determine whether the new prototype can meet the design specifications and requirements. It is typically conducted on complete production-representative articles under realistic conditions and sometimes involves comparison between competing products or systems.

Note that the degree of fidelity inherent (and necessary) is different for tests at various stages of system development. Tests occurring early in development are typically limited to laboratory-based components, whereas the production-representative prototypes needed for full-scale tests will only be available later in the development process. In this context it is important to view testing as a means of minimizing or mitigating the risk inherent in developing and producing a new product. As system development progresses towards production, changes to the design become more difficult and much more expensive to accomplish. The earlier a test can be conducted, the greater the potential savings in terms of costly redesign or engineering changes. Here again the tester is confronted with a trade-off decision: early test results may save a lot of money but only if they produce meaningful and representative results.

5. ANALYSIS AND EVALUATION

The final defining moment we will consider comes in the analysis and interpretation of the test data. A common difficulty with human factors measures used in test and evaluation, particularly those involving cognitive issues such as workload, situation awareness, and decision making, is that they do not possess definite criteria or "red line" limits. For these sorts of measure, testers will typically report the results in a narrative without making an explicit pass/fail judgment on the human factors measure in question. Alternatively, a tester may try to establish the measure's impact on user and system performance post hoc, either logically or through quantitative statistical methods such as multiple regression analysis.

Another evaluation issue can arise regarding the statistical methods used to determine whether a human factors criterion

was in fact met. There are many stakeholders in a system's development, each bringing quite different assumptions to the interpretation of test results. For example, application of data analysis methods used in academia would lead the tester to attempt a statistical proof that the system meets the human factors requirements (against the null hypothesis that the system doesn't meet the requirements). In order to demonstrate with some statistical certainty that the system can meet the requirements when fielded, the system must actually perform somewhat better than the requirement during the test (how much better will be a function of the type II error or alpha level required and the sample size used for the test).

Others involved in system development might reasonably adopt a quite different perspective on the data analysis; they might consider that the burden of proof is on the test to prove that the system doesn't work. Here the component or prototype under test would have to perform significantly worse that the requirement before receiving a failing grade. With large sample sizes and results that are clearly more or less than the requirement, these two positions are not in contention. But with small samples and results slightly above or below the criterion, the evaluator and developer can come into conflict over the conclusions to be drawn from the test data.

A final evaluation issue concerns the logical interpretation of test data. It is a common finding that human error rates are higher than the error rates for system hardware and software. Consequently, the operator is often portrayed as the bottleneck or limiting factor in system throughput, a problem to be solved by additional training. But in many of these cases it can be argued that design flaws are the cause of poor human performance. Furthermore, users will routinely attempt workarounds to make the system function effectively, often with the effect of masking true hardware and software reliability. This disagreement over the cause of poor system performanceæuser error versus designer erroræoften comes down to differences in design philosophy and is not easily settled.

6. CONCLUSION

This article has attempted to illustrate the human factors and engineering test process by describing four defining moments in testing: requirements definition and system design, test planning, data collection, and the analysis and evaluation of results. Each of these defining moments contains fundamental issues and trade-off decisions that must be confronted by the tester. The solutions are as diverse as the systems undergoing test, but they will have significant ramifications for ensuring the products and processes we develop are safe, effective, and satisfying to use.

REFERENCES

ANDRE, T.S. and SCHOPPER, A.W., 1997, *Human Factors Engineering in System Design* (Dayton OH: CSERIAC).

BITTNER, A.C., 1992, Robust testing and evaluation of systems: framework, approaches, and illustrative tools. *Human Factors*, 34, 477–84.

KANTOWITZ, B.H., 1992, Selecting measures for human factors research. *Human Factors*, 34, 387–98.

MEISTER, D., 1986, *Human Factors Testing and Evaluation* (Amsterdam: Elsevier).

O'BRIEN, T.G. and CHARLTON, S.G., 1996, *Handbook of Human Factors Testing and Evaluation* (Hillsdale NJ: Erlbaum).

Human Factors System Design

D. Meister

Consultant, 1111 Wilbur Avenue, San Diego, California 92109, USA

1. INTRODUCTION

The human factors engineering (HFE) orientation toward systems, equipment and products (SEP) design is that each new design (as well as an update or design revision) represents a problem (sometimes a complex one) to be solved by the design team. This is because there is no single design configuration which will satisfy what the customer wants for the SEP. There are always several alternative ways of designing the SEP, some of which are inadequate, others are marginally adequate, and others are optimal or nearly optimal. The problem is in finding the most effective.

The function of the behavioral specialist in design is to evaluate the effect of the design on the likely performance of the human, and to avoid undesirable effects by counseling other members of the design team who are usually engineers with varying backgrounds, but without behavioral training.

The behavioral specialist may also have direct responsibility for the design of the human–machine interface (HMI), the arrangement of those controls and displays that direct the SEP in its operations. The object of the specialist's concern has always been the human, but in the last quarter-century some HFE theorists have emphasized what they call "user-centered" or "human-centered" design (Norman and Draper 1986). Some believe that, in addition to or as opposed to designing an SEP to perform desired mission functions, design should enable SEP personnel to derive pleasure from their work. This is a rather far out point of view because any design geared to functionality (performing mission operations) will also attend to personnel factors, because any diminution of personnel efficiency will also reduce SEP efficiency. Nevertheless, the concept of human-centered design (however exaggerated) generally works to direct more attention to the human in the design process, and this is good.

The paradigm that guides behavioral analysis in design is SOR, which stands for stimulus-organism-response, because the human is a reactive animal. All actions are taken in response to external stimuli (events, phenomena, displays) or to internal ones (self-directed behavior). Design analysis methods usually involve *decomposition* to SOR.

Design in behavioral terms has a cyclical history (initial design analysis, detailed design, design testing, and production). Design is also viewed as an information processing problem. Successful design depends on answering a number of questions which are listed in Table 1. These questions, to which there are no automatically correct answers, will have to be answered again and again as the design becomes more detailed, because changes in design detail require changed answers. The answers to subsequent questions may also require a change in the answers to previous questions; much design involves this kind of *feedback*. For this reason one cannot think of SEP design as a linear process.

The remainder of the discussion will be oriented around how answers to the questions in table 1 are secured. Answering these

Table 1: Questions posed by the design problem

Initial Design Analysis

1. What is the scope of the design problem (is it 'new' design, update, or revision?) How difficult is the problem?
2. What type of SEP is involved (e.g. aircraft, automobile, word-processing computer, general purpose commercial product); and what parameters are involved as a consequence (e.g. visual factors, strength limits, cognitive decision making?
3. What questions about the details of the design problem must be answered?
4. Are all SEP requirements fully described in the design specification?
5. What behavioral implications (e.g. human thresholds) can be drawn from these requirements; and what further information is needed that stems from these implications?
6. What technology and information can be transferred from predecessor SEPs?
7. What characteristics does the human in the SEP have (e.g. specific constraints, like physical handicaps, or needs and desires to be satisfied by design)? How important is the human role in the SEP? because this determines how much emphasis the human will receive.
8. What is the human in the SEP required to do; what is it anticipated that s/he will do; what personnel requirements, if any, have been specified in the SEP requirements; how will the human's performance affect SEP performance?
9. What factors, technological and non-technological (e.g. money, time) will affect personnel and SEP functions?
10. What design tools, analytical or evaluational, should be used to assist in answering the preceding questions?

Detailed Design
(All the preceding questions may/will be asked as design progresses.)

11. What additional information must be collected and analyzed as a result of increasingly detailed design; and where can this information be found?
12. What are the various ways in which the SEP components (e.g. HMI) can be configured? How adequate is the HMI design?
13. What criteria should be used to decide among the various design alternatives? Which alternative is the best, considering all factors? What compromises must be made among competing performance criteria and design characteristics?

Design Verification
(All prior questions may still be asked, in addition to the following):

14. How should the evaluation of the final, total SEP be performed, including behavioral criteria and test methodology?
15. What last minutes fixes, if any, must be made to the SEP production model?

questions will involve the application of behavioral design methods listed in Table 2 (Meister 1985).

The first task of the design team, of which the behavioral specialist is a member, is to scope the nature of the design problem. This is an effort to determine how difficult the problem may be, what questions will be raised by the nature of the SEP under design, and (for the specialist particularly) what behavioral implications can be derived from SEP design requirements.

The scope of the problem is limited in part by whether the problem involves "new" design, an updated design, or a design revision. Unless the SEP is a complete novelty, which is very rare (e.g., the Apollo module), it will have predecessor SEPs which will probably have technology and information that can be

Table 2: Behavioral design methods

1. Training, experience and logic.
2. Function analysis: determination of functions to be performed from description to activity.
   Function allocation: assignment of task responsibility to human, machine, or both.
3. Task description: description of activity in SOR terms.
   Task analysis: determination of factors influencing task performance (includes graphic tools such as operational sequence diagrams and function flow diagrams).
4. Design evaluation methods: checklists, walk-throughs (verbal performance simulation), and comparison of design alternatives.
5. Rapid prototyping (user analysis and exercise of initial design versions).
6. Psychological research methods: performance measurement, interviews, questionnaires, scales.
7. Special measurement methods: workload estimation, situational awareness, human performance models, error and human reliability prediction, concept mapping.

transferred without change or only slight modification to the new SEP. For example, in designing a new aircraft, a highly successful navigation subsystem will be taken directly from an earlier model and installed in the new aircraft. If the SEP is an update, new capabilities must be designed, but everything else in the SEP will be undisturbed. If deficiencies must be revised, the nature of these deficiencies will serve as indicators to what must be changed. The components of the new system may have been tested in an earlier test phase, and if these tests have involved humans, the test data may be behaviorally useful.

Beyond all this, the nature of the SEP under design will alert the specialist to certain areas of potential concern: if the SEP is a device that will be operated in a complex environment, e.g., an aircraft, the nature of the HMI may well produce special problems; if the device is a general-purpose commercial product, special attention will have to be paid to simplicity and to determining user desires.

SEP requirements, as described in the customer's design specification, are another source of information. These specifications may or may not be highly detailed, but in any event they must be analyzed.

In all the preceding analyses, what the specialist will be attempting to determine are the behavioral *implications* of any physical requirement or SEP feature, and what the operator or user of the SEP will be required to do by the SEP.

The term "implications" means effects on the human of a design requirement or the nature of a physical structure of the SEP. For example, if the design requirement specifies that the SEP will be operated underwater, it is apparent that attention must be paid to the development of the human's protective equipment, vision constraints caused by water, etc. Implications are important. Technology eliminates manual operations to which the human has adjusted, and it replaces them with technology's own imperatives, which are not at all concerned with human predispositions. If one cannot keep up with technology, as some aged or handicapped cannot, so much the worse and one will suffer in the end. HFE is opposed to creating a technological world in which the human may be lost.

The design methods listed in table 2 help the specialist recognize potentially negative technological effects. These methods are of two types: (1) general-purpose logic and deduc-

tion, the training provided the specialist, and his or her experience; (2) methods specifically created for HFE design, e.g., task description and analysis (Luczak 1997). The general-purpose methods underlie the more specific ones. Only brief descriptions of the specific methods can be given here, but the references do contain more detail.

The overall goal of the initial design analyses is to determine the behavioral implications of the SEP's physical structure. This is a matter of training, logic, and experience. Fundamental to the discovery of these implications is the specification of what the SEP's mission requires the human to do and what the human is likely to do. Function analysis, task description, and task analysis are the primary methods the specialist uses to answer these questions. Logically, certain human activities require certain human functions. For example, if the human is supposed to receive information about system status, like altitude or bearing in an aircraft, a monitoring function is involved; if an automobile must be driven, this will involve tracking the dimensions of the highway.

Functions are important to the specialist because they may suggest alternative ways of implementing those functions. When pilots carry out their monitoring, stimuli may be presented in the form of individual scales or summary indicators, and displays may combine multiple items of information; the cockpit can be more or less computerized. It is fair to say that analysis of possible effects on the operator will proceed in parallel with determination of the alternative ways in which a function can be designed. If there are *N* possible ways of presenting stimuli, these must be winnowed down to a single one or at most a few alternatives. One way of winnowing them down is to apply various *criteria* (i.e., value judgments): technological adequacy (Can a function be implemented technologically?); cost (Will the cost for a particular alternative be exorbitant?); reliability (Will performance quality remain constant over time?). One of these criteria is behavioral — operability — meaning that if the function is implemented, will it impose too great a burden on the operators? For example, if the function is to be performed manually, will its performance exceed threshold limits? In that case the design alternative must be rejected. Threshold limitations may be physical (e.g., strength required to move a weight) or cognitive (e.g., excessive memory requirements); physical thresholds are easier to recognize than cognitive thresholds. That is because physical thresholds can be demonstrated in action, but cognitive thresholds are often covert. I believe that humans have certain preferred modes of cognitive functioning (stereotypes) and that when a design requires some violation of these stereotypes, human performance suffers.

A great deal has been said in the behavioral literature about user-centered or human-centered design. This is a version of the requirement to make work more pleasurable or gratifying to the worker. What it means is that much greater attention must be paid to user needs and desires.

In contrast to traditional design, which is largely determined by what the SEP is required to do (its mission), user-centered design is presumably driven by behavioral imperatives. HFE system design has always endeavored to prevent egregious violations of human limitations, but for various reasons that concern for the human has been limited (one reason is our limited knowledge of human capabilities and limitations). Now it is proposed

to somehow elevate behavioral considerations to the same importance as that of the mission. How this can be accomplished is unclear because of our lack of knowledge about how to translate desires into physical equivalents. If the SEP does not have a mission to perform, it is hard to see what the human in the SEP will be doing. Our own point of view is that the human in the SEP deserves the greatest consideration commensurate with our knowledge, but the SEP has no purpose other than that provided by the mission.

We have somewhat arbitrarily divided SEPs into special-purpose and general-purpose devices. A commercial aircraft is a special-purpose device because it cannot be operated by everyone; an automobile is a general-purpose device. Much greater attention to the user must be paid in general-purpose devices, because in many cases the user is also the *consumer* who will or will not purchase the device. The ignominious failure of the Edsel Ford in the 1950s is an example of a device that will fail (whatever its technological qualities are) if consumers will not purchase it.

How the designer includes user needs and desires in design is difficult to say. On a molecular level, designers have altered the position of ashtrays or seat belt locations to accommodate presumed user desires, but whether more molar desires can be accommodated is not clear. One method of ascertaining user preferences is called *rapid prototyping*, which came into fashion with the advent of the computer. Rapid prototyping is the development of one or several versions of a design product and then giving them to a sample of users to exercise, use, comment on, etc. Not only opinion data can be gathered in this way (I like device X because . . .), but also performance data (How well can the user use the device?). If it is feasible from a technological, cost, and reliability standpoint, the user's responses to the prototype will be used to modify the prototype and thus to come closer to the user's desires. This procedure was more difficult to perform when hardware (e.g., in the form of wood, plastic, etc.) was involved, but software has facilitated what is essentially design, not by rational methods, but by tryout.

3. DETAILED DESIGN

All design is iterative; all design involves both analysis and evaluation. All design becomes progressively more detailed, and so questions like the behavioral implications of design and their effect on the human are asked at successively later, more detailed stages. All the questions of initial design analysis reappear in detailed design, because the changed level of detail provides a new context for the earlier questions and may therefore lead to new answers.

The greater detail of design also leads to *feedback* of information affecting prior design decisions. We see this most clearly in rapid prototyping, because the information gained by exposing preliminary design versions to users leads to changes in these preliminary designs and perhaps even modification of more basic design concepts which were, until then, considered firm.

Rapid prototyping is the clearest illustration of the role of testing in detailed design. Testing always involves exposure of the design to a human subject, but *evaluation* is also involved. Evaluation is a judgmental process which accompanies testing (e.g. What do the test results actually mean?) but it may also be used outside of testing, as a form of checking the design alterna-

tive. One of the functions of HFE research is to derive what can be termed *guidelines* or principlesæstandards of acceptable design practice. These are often incorporated in checklist form, and published checklists are used to compare the features of a specific design with the guidelines to ensure that no design feature violates one of the checklist principles. Should this happen, presumably the designer would modify the design accordingly.

Checklist evaluation is a form of testing, but it is not conducted by exposing the design to subjects; it is conducted by the specialist alone and is a means of developing confidence in the design characteristics the specialist has created.

4. DESIGN VERIFICATION

During detailed design any prototype testing has been partial, to check on one or more of the design features of special interest to the designer. Such tests are often performed by providing users with simplified versions of a design, emphasizing features of interest and leaving others in skeletal form. There is no attempt to replicate realistic working conditions in the test, nor does the test verify the overall adequacy of the final system. This is the function of the *operational system test* (OST).

The OST endeavors to test the final preproduction SEP in a manner that replicates as much as possible the actual operational conditions under which the SEP will be used. The complete SEP is involved; subjects for the test (those who will exercise the SEP) are to be as similar as possible to eventual users (if they are not such users themselves).

A full-scale OST is a major test activity and will almost always involve extensive planning. One would hope that rapid prototyping would also involve such planning, but the procedure seems to encourage a certain conceptual sloppiness. The planning requires the development of an OST test plan which includes the following sections:

- *Purpose* of the test (Will the SEP perform its mission successfully; will test personnel encounter any inordinate difficulties; what if anything needs to be revised in the design?). The OST has overall general purposes, but there may also be more detailed, specific purposes (to test such and such a feature); all of them should be specified, so that none will be overlooked in gathering data.
- *Description of the system*. This is largely pro forma and becomes important only if certain subsystems are not to be included in the test.
- *Geographic location* of the test, test instrumentation, etc. Also largely pro forma, but desirable if others besides test personnel will read the final report. All these OST sections will find their way into that final report, so writing these things into the test plan gives one a leg up on that report.
- *Performance criteria*. These are all-important. How is one to determine whether the test actually verifies the adequacy of the SEP? Criteria which may be somewhat general (e.g., mission completion) must be decomposed into their component individual measures (e.g., follow a ground track at 500 feet altitude without deviation of more than ±100 feet to reach a specified destination 1500 miles from takeoff). The example is that of a highly sophisticated terrain-following bomber. For a product/equipment like an automobile, the mission performance measure might be in terms of handling qualities, e.g., handles well in curving around obstacles. Measurements in-

volve both objective performance criteria (e.g., tracking, speed) and subjective criteria (e.g., rating the adequacy of the design by the user, design features particularly liked or disliked). Regardless of the type of SEP, it is desirable to have test personnel express their opinion about design features; to have them do this, however, it is necessary to specify the features of interest and to provide measurement instruments (e.g., checklists, scales).

- *Test instruments* must be described in detail, unless they are in common use. All instruments for securing subjective data must be described in detail, because they are almost always developed for a specific test. If the instrument is in common measurement use (e.g., a photometer), one need only refer to the instrument manufacturer. If automatic measurement instruments are part of the operational equipment, this needs to be indicated.
- *Test procedures* describe how the OST will be conducted. If certain test functions must be performed at certain times, these must be listed. Test procedures should also include any conditions that must be compared (e.g., night/day operation, performance under varying sea states or traffic conditions). Ordinarily the OST is not a research study, where conditions are compared to determine the importance of variables. OST performance will be compared under varying conditions only if those conditions are important in routine use of the SEP.
- *Anticipated data.* It is advisable to list the categories of data that will emerge from the testing, e.g. number and frequency of specific errors and response speed. This is necessary because a plan must be developed in advance of the test data to permit a statistical analysis of those data; that analysis will in part depend on the nature of the data anticipated.
- *Statistical analysis* of the test data. The anticipated statistical analysis is critical to the OST, because unless it is specified in advance, the data collected may not be appropriate for a particular analysis.

The OST resembles a formal experimental study, but there are other ways of testing a new design to verify it. Usability testing can be performed in special laboratory facilities (in which case it would be much like OST). In one form of *usability testing* a new SEP is given over to actual users (companies), and the SEP is used in a routine way (Weimer 1995). There is less control over conditions of performance, but if the specialist imposes controls over that test, very realistic data can be obtained.

Although the OST and the usability test are the final situations in SEP design, they have consequences for the design process. If all goes swimmingly, production management can proceed to distribute its products. In many cases, however, the OST and the usability test will point to design features that must be corrected.

These final tests are unlikely to lead to complete rejection of the SEP, because it cannot perform its assigned mission; previous tests would have indicated something wrong that had to be fixed, and it would have been fixed. More common, however, is the revelation of a number of detailed design faults that need fixing; and among these faults may be behavioral deficiencies (e.g., inaccuracies in operating procedures, the placement of controls and displays vexatious to test subjects). Fixes for such inadequacies will be made immediately, if they do not involve "bending hardware" or changing software code. Other fixes that require some modification of the final design will require debate; engineering management prefers to rely on additional training of personnel, warnings in operator manuals, etc., to any but the most minor changes in the basic design. Behavioral specialists may have to argue long and loudly if they are to gain their way—if they do.

System design may appear to be over with completion of final tests, but this is an illusion; as was indicated previously, systems are updated to include features that were not in the original design (e.g., provision of multivariate displays to replace discrete indicators). Despite the best efforts of all involved, the final production design may eventually reveal inadequacies that absolutely require redesign. There is life after the original design has been completed. Like a crossword puzzle, the design problem may be completed, but since designs have only a finite existence, there will be other designs and other problems.

What has been presented is merely an outline, a skeleton, of what happens in every design process. The number of individual variations is very great, especially in a design process that is pursued over long periods. This is one of the things that is most fascinating about the design process — its almost infinite variation.

REFERENCES

LUCZAK, H., 1997, Task analysis. In Salvendy (ed.) *Handbook of Human Factors and Ergonomics* (New York: Wiley), pp. 340–416.

MEISTER, D., 1985, *Behavioral Analysis and Measurement Methods* (New York: Wiley).

NORMAN, D.A. and DRAPER, S.W. (eds), 1986, *User Centered Design: New Perspectives on Human-Computer Interaction*. (Hillsdale NJ: Lawrence Erlbaum).

SALVENDY, G., 1987, *Handbook of Human Factors* (New York: Wiley).

WEIMER, J. (ed.), 1995, *Research Techniques in Human Engineering* (Englewood Cliffs NJ: Prentice Hall).

Integration of Quality, Ergonomics, and Safety Management Systems

J. S. Dzissah[†], W. Karwowski[†] and Y.–N. Yang[‡]

†Center for Industrial Ergonomics, Lutz Hall, University of Louisville, Louisville, KY 40292, USA

‡Department of Industrial Engineering, Texas Tech University, Lubbock, TX 79409, USA

1. INTRODUCTION

Quality improvement is concerned with improving products, services, quality of work, design, manufacturing and decision making within an organization in order to satisfy both the internal and external customers. Such philosophy forms a basis for the quality management system (QMS), which aims to achieve the highest levels of business quality and productivity. Such a goal can be realized through different management methods, such as process reengineering, effective product design, employee involvement and teamwork, training in statistical techniques, and customer focus. As a result, a myriad of different quality standards, methods and awards systems have evolved in the recent past (ISO 9000, BS8800, Malcom Baldrige National Quality Award (MBNQA), European Quality Award (EQA), ISO 14000, ISO TC/159, Occupational Safety and Health Association (OSHA), American National Standards Institute (ANSI) Z-365, etc.), all of which focus on the continuous improvement in different areas of the organization.

There also many obstacles to achieving global quality, including the effects of rapid technological changes, increasing complexity of products and processes, rising production costs, customers resistance to price increase, increasing demands for employee well-being, and natural environmental protection. Recently, there has been increased emphasis on the continuous company-wide improvements that link quality, ergonomics, occupational health and safety, and (natural) environmental health.

Traditionally, the QMS solutions did not go far enough in addressing the complex varieties of organizational life (quality, ergonomic, safety, marketing, purchasing, human resource management, etc.), often resulting from the complex and conflicting needs and expectation of all stakeholders involved in the organization. Because of the strong ties between the product's quality defects, working conditions defects and operational defects, it is important to develop and implement a comprehensive model for the management system that integrates quality, ergonomics and safety (occupational and health and natural environmental) aspects of organizational performance.

2. QUALITY MANAGEMENT, ERGONOMICS AND SAFETY

2.1. QMS Theory and Practice

Market, capital, management, employees, motivation, materials, machines, information and product or service requirements directly influence quality of products and services. QMS is a management philosophy aimed at improving competitiveness, effectiveness and flexibility through planning, organizing, understanding each activity, and involving each individual at each level of business activities. This concept was adopted by the Japanese and renamed as the company wide quality control (CWQC). The main characteristics of the CQWC as practiced in Japan are:

1. Quality emphasis extends through market analysis, design, and customer service rather than only the production stages of making a product.
2. Quality emphasis is directed towards operations in every department from executives to clerical personnel.
3. Quality is the responsibility of each individual and work group, not some other group such as inspection.
4. There are two types of quality characteristics as viewed by the customer those that satisfy and those that motivate. Only the latter are strongly related to repeat sales and quality image.
5. The first customer for a part or piece of information is usually the next department in the production process.

The works of Deming, Juran and Cosby have extended concept of quality from purely a technical issue to the management issue. The concept of total quality management (TQM) established foundations for implementation of the ISO 9000, which was developed in an effort to address the consistent quality needs for growing international trade.

2.2. Ergonomics

Ergonomics seeks to design tool and tasks to be compatible with human capabilities and limitations with the purpose of providing work conditions that assures safety, health, well-being and efficiency. Many studies in industrial settings aim to reduce effort, improve process, quality and occupational safety. These applications have evolved over the years from the micro-ergonomics (design of individual workstation) to a macro-ergonomics approach focusing on design of the entire corporation. The importance of ergonomics to the improvement of quality of life, product and services in the industrial settings, has prompted some organizations to develop standards or guidelines at the following levels:

- Basic standards related to fundamental properties of humans.
- Functional standards related to ergonomics characteristics of equipment, processes, product and safety.
- Environmental standards related to the effects of physical factors of the environment on human performance in the range between comfort and health hazards.
- Standards for ergonomics test procedures and for processing ergonomics data to be applied either in development of standards in the above categories or in assessing conformity to already accepted standards.

In 1975, the ISO Technical Committee 159 on 'Ergonomics' was formed to develop standards in the ergonomics area. There are also other standards that were not produced by ISO. For example, the Occupational Safety and Health Association (OSHA) in the USA has recently drafted an ergonomics regulations that focus mostly on repetitive strain injuries, like lower back pain and carpal tunnel syndrome (CTS). The American Standards Institute (ANSI) Z-365 developed a procedure for managing prevention of the cumulative trauma disorders (CTD) of upper extremities.

2.3. Occupational Health and Safety Management Systems

Although working environment has improved considerably during recent decades, occupational accidents still occur as a result of inadequate workplace design. Such accidents constitute a major problem facing industry today. Occupational health and safety management focusing on employees is the fundamental responsibility of any organization. Safety can be defined as the state for which the risk (or probability) of a hazard-related incidents is judged to be acceptable. Traditionally, safety focuses on the control of loss, actual or potential, while safety management is the application of management principles to the issues of safety.

Safety management aims to prevent repetition of accidents that have occurred in the past. However, to create and maintain the safe working environment, management must depend on the proactive rather than reactive prevention efforts. To address this issue, many organizations (British Standards Institute, National Institute for Occupational Safety and Health (NIOSH), ANSI, etc.) have developed codes of practice, procedures, and management systems. etc. to augment management's effort in improving the working environment. The BS8800 — Occupational Health and Safety Management Systems developed in the UK is an example of such a system.

2.4. Natural Environmental Safety

The natural environment is defined as the surroundings in which an organization operates, and it extends from within the organization to the global system. Many organizations strive to achieve and demonstrate sound environmental performance by controlling an impact of their activities, products and services on the natural environment. This is partly due to the public and management concerns over the safety of products, human comfort, destruction of natural environment, and the high cost of compensation or clean-up expenses incurred as a result of accidents. The universally accepted ISO14000 Natural Environmental Management System can be utilized as the subsystem for integration purposes.

3. INTEGRATION OF QUALITY, ERGONOMICS AND SAFETY

The integration of quality management system with ergonomics or safety issues has received some attention in the past. Rahimi (1995) and Weinstein (1996) proposed approaches for integrating safety, health and environment into TQM, by advocating simultaneous and continuous allocation of resources to top-down and bottom-up engineering improvement, through well-organized self managed teams and parallel training systems with flattened organization structures. They attempt to bridge the gap between TQM and its applications to safety by outlining a process for developing and implementing a safety management system that incorporates TQM concepts with the ISO 9000 quality requirements and hazard specific technical mandates.

Eklund (1997) proposed an approach for ergonomics and quality improvement by integrating the ergonomics and TQM knowledge. The field of quality would gain by incorporating ergonomic knowledge especially in the areas of work design and human capability, since these factors are decisive for improving the human performance, while the quality of work is strongly influenced by operator. Luczak *et al*. (1998) pointed out that the

objectives of ergonomics and TQM have a close relationship through the MBNQA evaluation criteria, which covers product
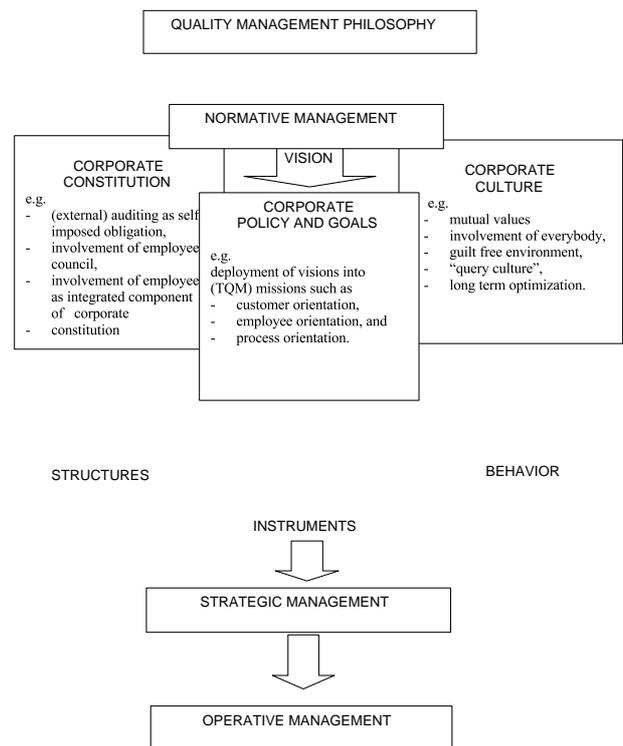

Figure 1. Quality issues integrated in the dimension of normative management (modified from Zink 1999).
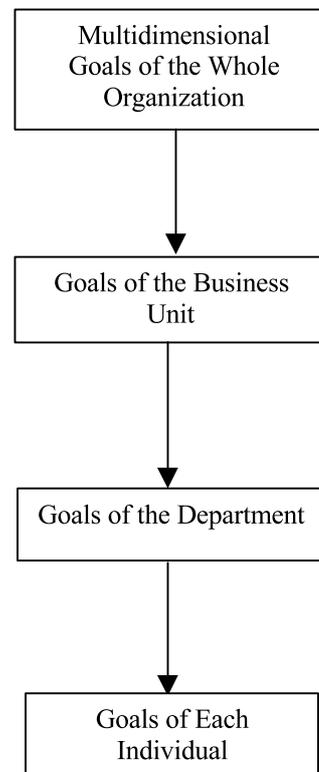

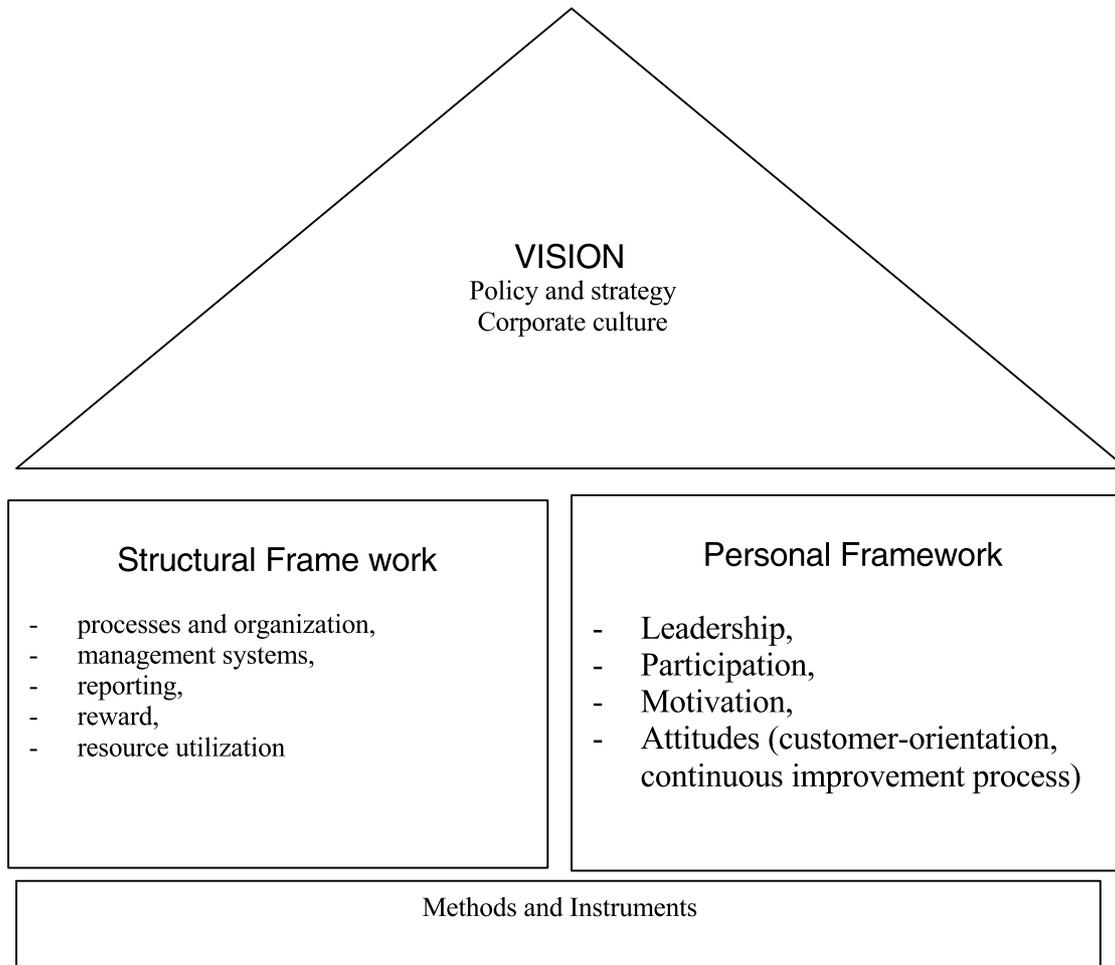Figure 2. Policy deployment process (modified from Zink 1999).

VISION
Policy and strategy
Corporate culture

Structural Frame work

- processes and organization,
- management systems,
- reporting,
- reward,
- resource utilization

Personal Framework

- Leadership,
- Participation,
- Motivation,
- Attitudes (customer-orientation, continuous improvement process)

Methods and Instruments

Figure 3. Exemplary general framework for the introduction of a TQM concept (modified from Zink 1999).

quality, process quality and quality of working life. Recently, Zink (1999) proposed a model that shows how to integrate the quality aspects into a normative management frame. The proposed comprehensive change management process involves having a

vision for a set of commonly agreed goals. Such process is summarized in Figures 1–3.

Smith (1999) proposed an approach that applies the Juran's servomechanism model to illustrate the interaction of ergonomics
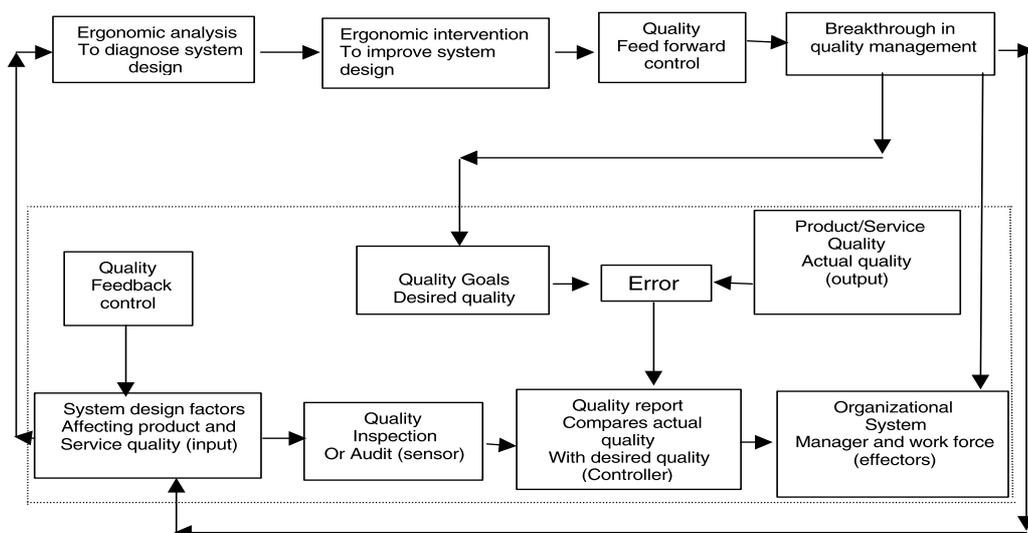

Figure 4. Model of breakthrough in quality management as a feed forward control process with ergonomic analysis and intervention specified as key contributors to feed forward control (modified from Smith 1999).

and quality. The servomechanism assumes that there is an unvarying sequence of events that occur in breakthrough from one level of performance control to the new (improved) level. This is mediated by comparing feedback from the actual system performance with performance targets. This approach is shown in Figure 4.

Carayon *et al.* (1999) presented a macro-ergonomic model of work design that takes into account not only the products and process quality, but also the quality of working environment. According to the model, TQM can influence different aspect s of work design and quality of working life (QWL). The proposed approach is shown in Figures 5 and 6.

Tables 1 and 2 show the major elements of quality, ergonomics, and safety under the traditional approach, and within-the-organization elements that interfere with the ergonomics principles. These elements influence the corporate strategy with respect to planning for quality and safety.

In many cases the production problems are not only related to ergonomics, quality or combination of these alone, but also to safety (occupational health, natural environment), as these problems are interwoven and strongly related. As a result, they must be managed simultaneously, with an integrated management system.

The analysis of the concepts of quality management, ergonomics, and safety, shows several identical elements, including leadership, evaluation and continuous improvement,
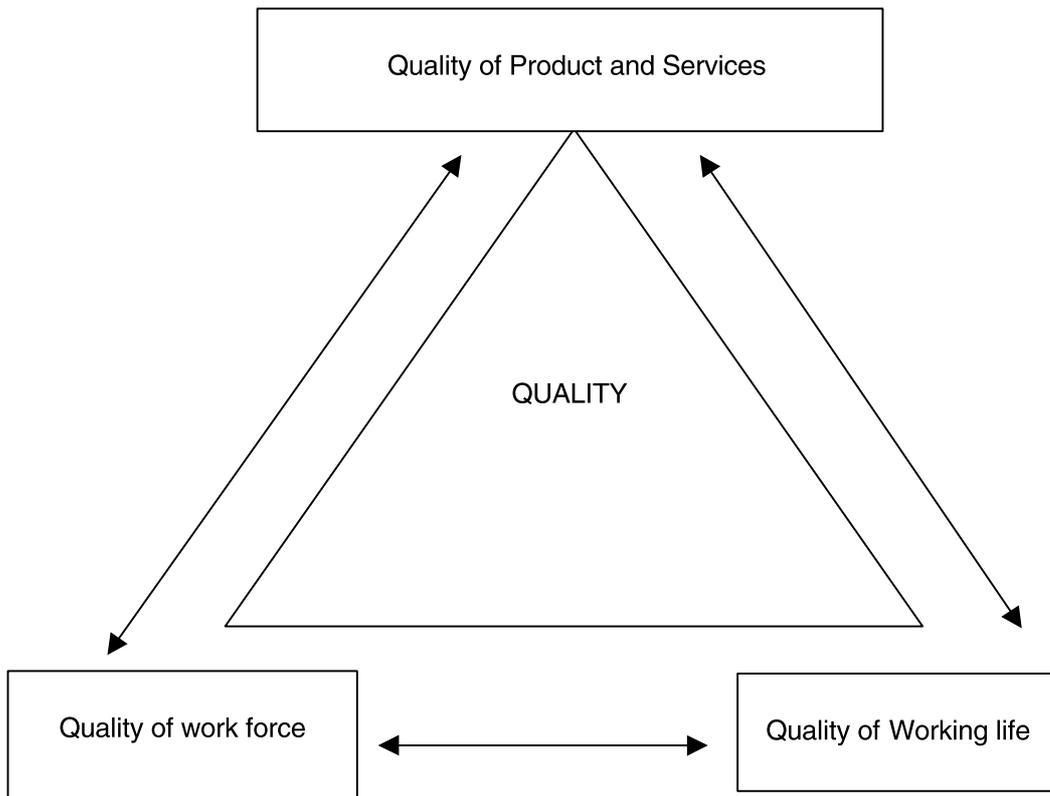

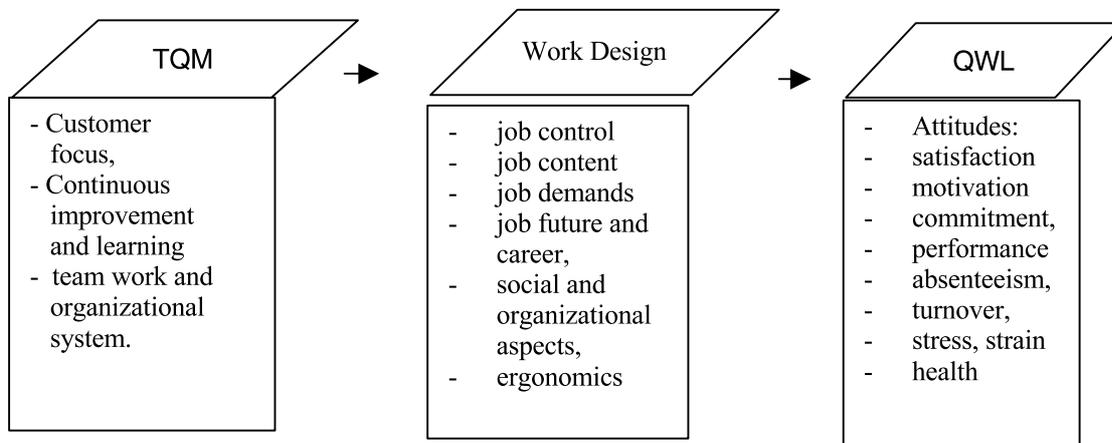Figure 5. Three interrelated aspects of quality (modified from Carayon *et al.* 1999).

Figure 6. Relationships between TQM, work design and QWL (modified from Carayon *et al.* 1999).

Table 1. Traditional approach to quality, safety and ergonomics.

Safety	Quality	Ergonomic
Zero accidents	Zero defects	Zero work place defects
Incident analysis	Event analysis	Work site analysis
Written policies procedure and guidelines	Documented policies procedures and work instructions	Written policies and procedure
Safety committee	Quality circles, employee involvement teams	Ergonomic committee
Employee participation	Employee empowerment	Employee involvement
statistical analysis	Control charts, statistical process control etc.	Experts methods including some statistical tools
All accidents are preventable	All non conformers are preventable	All work place defects are preventable
Continuous improvement	Continuous improvement	Continuous improvement

Table 2. Elements within an organization.

Material environment:
–Lighting, ventilation, temperature, volume of noises, degree of dustiness etc.
Natural environment:
–Natural environmental protection, protection against noise of urban life, litter utilization, legal resolutions with regard to environmental protection, international cooperation in shaping and protecting environment
Technical environment:
–Workstations, parameters of machines, appliances, tools, desks, seats, technical equipment, construction and building
Management environment:
–Management system, participation and decision making structure
Organizational environment:
–Work process, production, technology, work improvement, work division work space, methods, monotony, rotations, break arrangement of elements of work stand
Living and social environment:
–Building of residence areas etc.
Social environment:
–Structure of work groups, methods of supervision, instructions, adaptation and motivation work culture, technical culture, and ergonomic inventions
Economic environment:
–Working capacity and effectiveness, defective production etc.
Computer Environment:
–Interactions human computer systems etc.
Safety in work environment:
–Potential impendence for life and health, Physiopatological consequences of work (tiredness, diseases, injuries) etc.

human resource management, data collection and management, and participation and teamwork, etc. These elements constitute the main building blocks for quality management, whose fundamental core principles are focusing on the customer, participation and teamwork, effective measurement and continuous improvement. For example, TQM can be used as a suitable management framework for the integration of quality, ergonomic and safety. Thus merging ergonomics and safety efforts (occupational health and natural environment) with quality management will provide a single and coherent system that identifies all elements in relation to quality, ergonomics and safety, and helps to fulfill simultaneously different requirements related to standards, rules and regulations, guidelines (MBNQA), etc.

4. FRAMEWORK OF INTEGRATED QUALITY MANAGEMENT SYSTEM, ERGONOMICS AND SAFETY MODEL

Working conditions influence product quality, while productivity improvement is dependent on quality, macro-ergonomic and safety (occupational and health, natural environmental) management activities. System performance can be improved by minimizing productivity losses, injuries, accidents, environmental disasters and destruction, general well-being of all employee, and minimizing cost and time use for improvement activities.

Table 3 and Figure 7 illustrate the elements and relationships for integrating ergonomics, safety, and QMS. A framework of integration of quality management, ergonomics, and safety is given in Figure 8. This model is essentially an adaptation of the open system model for understanding the interactions between technological changes and organizational processes. Based on the management leadership and proper planning from the variety of perspectives (financial, safety, human resource, etc.), an organization utilizes all resources (information, material, people, etc.) to design

Table 3. Elements of the integrated set.

A
–Inspection activities (Sampling methods)
–Variability reduction methods
–Marketing and delivery strategies

B
–Ergonomic experts analytical methods, Task analysis, occupational safety

C
–Body dimension and structure in relation to work site (Antrhropology)
–Manual lifting, Postures etc.
–Physical pressure that can be endured
–Physical capabilities and limitations
–Accidents and hazards analysis in relation to work

D
–Hazards analysis in relation to the work and natural environment (Chemical, Physical, Biological)
–Machine guards, shield, warning signs
–Disposals (Toxic materials etc.)
–Emergency systems, Protective Clothing etc.

E
–External factors such as Technology (CAD, CIM etc.)

Common point

–Leadership, commitment, policies, and values in relation to quality, ergonomics and safety
–Vision priority and direction
–Provision of adequate resources
–Review performance set goals (quality, ergonomic and safety), auditing
–Employee involvement and empowerment
–Effective human resource management
–Continuous improvement activities
–Training and education (quality safety and ergonomic)
–Auditing and documentation
–Effective maintenance and installation practices
–Product and process development
–Equipment and material purchase
–Calibration and metrology services
–Statistical methods
–Defects and problem investigations
–Work instructions in relation to quality, ergonomic and safety
–Task analysis in relation to quality ergonomics and safety
–Cost benefit analysis
–Customer complaints and return product analysis
–Customer needs and expectations
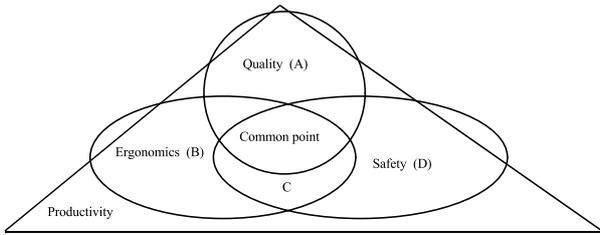
Figure 7. Relationship between quality, ergonomic, safety and productivity.

the process and produce the product in a desired way. Organization also uses all of the internal (ergonomics surveillance, process defects identification etc.) and external (ISO9000, government regulation, etc.) information to perform continuous improvement activities.

Figure 9 presents the inter-relationships between the corporate growth and integrated ergonomics, safety, and quality management. This integrated model is made up of the inputs, planning, processes, continuous improvement activities, interventions, auditing, and outputs. The overall goal of the model is the combination of the goals of Quality, Ergonomic and Safety subsystems. From the early studies and observations, it is noted that these goals interact in several areas. Consequently, the quality, ergonomics and safety related problems must be solved simultaneously to attain these goals by adjusting certain variables or changing some conditions.

Figure 8. Framework of the integrated ergonomics, safety and quality management.

A. Quality deficiency analysis B. Ergonomic (work condition) defect analysis C. Safety (occupational and health and natural environmental) defect analysis

*OHSM (Occupational Health and Safety Management)
**EMS (Environmental Management System)

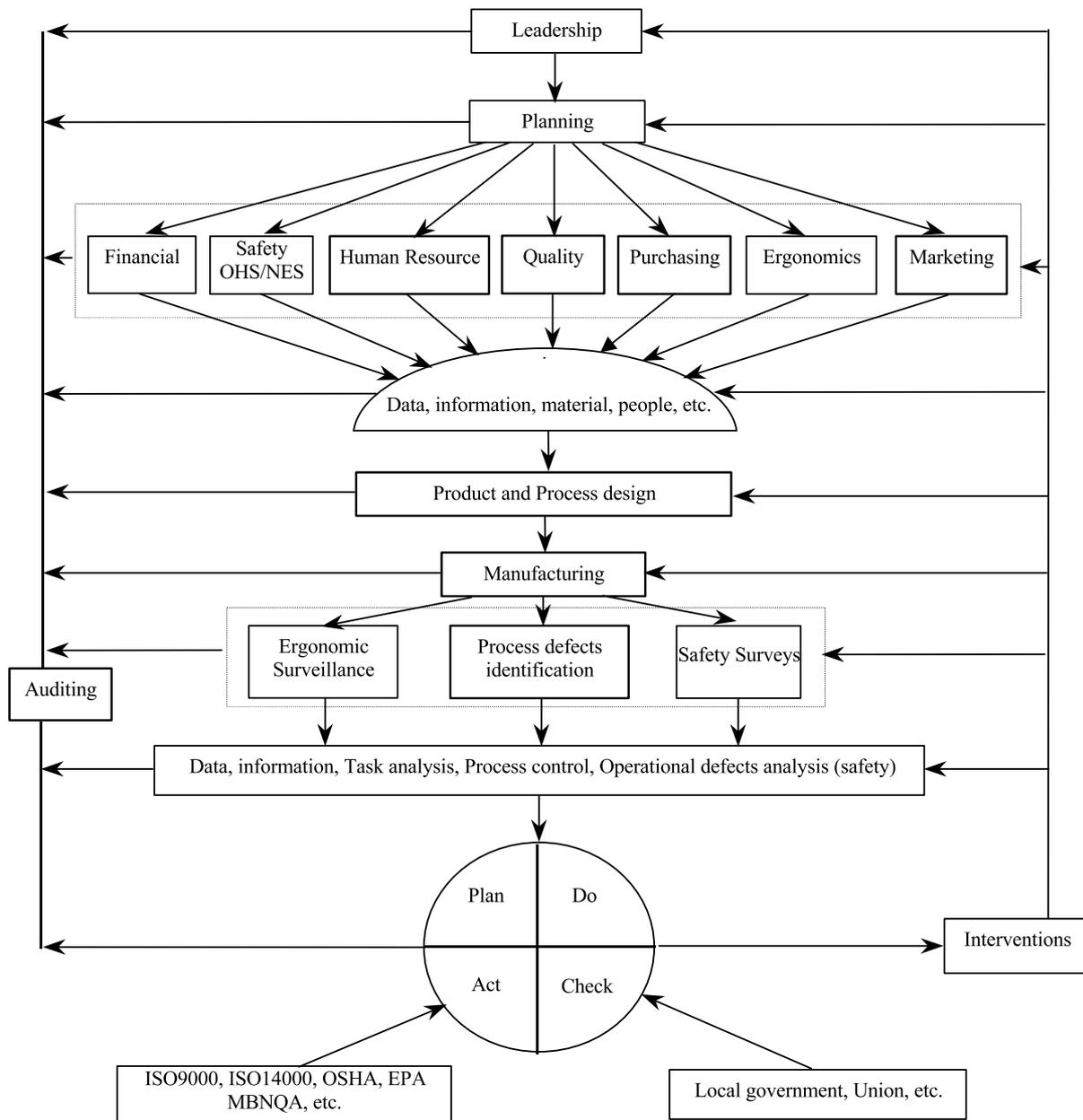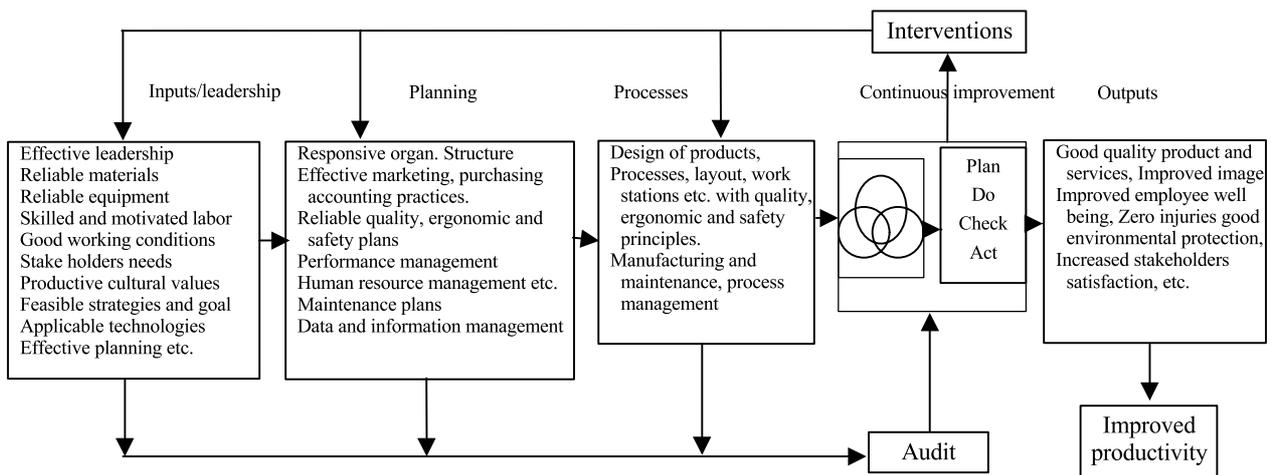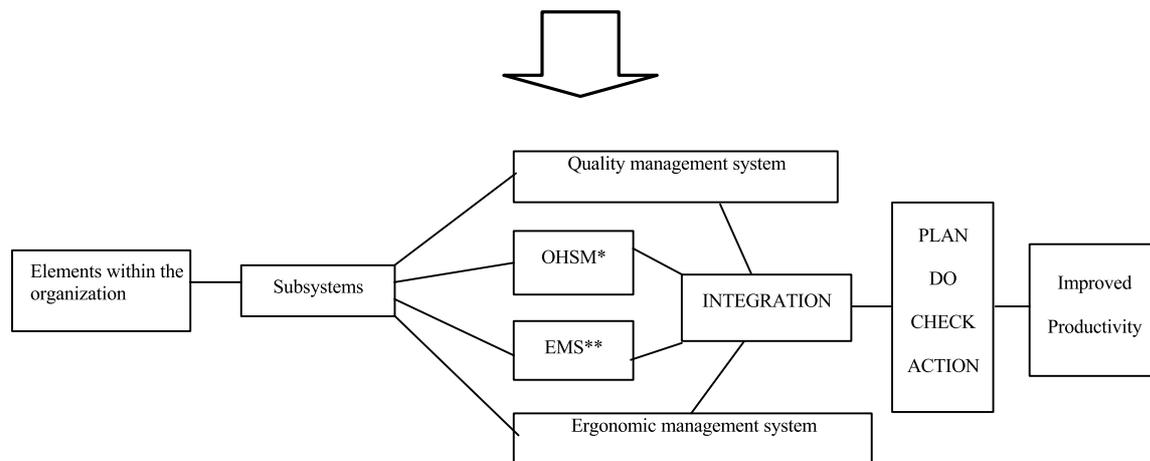**Figure 9. Integration of ergonomics, safety and quality management in organizational development.**

In view of the above, the integrated continuous improvement system can be described as:

- Attaching equal importance to quality, ergonomic and safety related problems.
- Identifying the relationship between quality, ergonomic and safety related characteristics for simultaneous improvement.
- Minimizing system procedures by performing common functions and solving common problems simultaneously.
- Combining quality, ergonomics and safety activities in all system operations

5. CONCLUSIONS

The concepts of quality, ergonomic and safety (natural environmental, occupational and health) reveal several dimensions of corporate performance that interact with each other, thereby suggesting a need for the integrated system focused on the continuous system improvements.

REFERENCES

CARAYON, P., SAINTFORT, F. and SMITH, M.J., 1999, Macroergonomics and TQM: how to improve quality of working life, *International Journal of Occupational Safety and Ergonomics*, 5, 303–34.

EKLUND, J.A.E., 1997, Ergonomic, quality and continuous improvement conceptual and empirical relationships in an industrial context, *Ergonomics*, 40, 982–1001.

GORSKA, E. and LEWANDOWSKI, J., 1998, Quality, environment and safety management in ergonomic education. In L.M. Pacholski and J.S. Marcinkowski (eds), *Proceedings of the 15th International Seminar of Ergonomic Teachers*, 289–97.

LUCZAK, H., KRUIGS K., GRYGLEWSKI, S. and STAWOWY, G., 1998, Ergonomic and TQM. In W. Karwowski and G. Salvendy (eds), *Ergonomic in Manufacturing*, 505–31.

RAHIMI, M., 1995, Merging safety, health and environment into TQM, *International Journal of Industrial Ergonomics*, 16, 83–94.

SMITH, T. J., 1999, Synergism of ergonomics, safety, and quality — a behavioral cybernetic analysis, *International Journal of Occupational Safety and Ergonomics*, 5, 247–78.

WEINSTEIN, M., 1996, Total quality approach to safety management, *Professional Safety*, July, 18–22.

ZINK, K.J., 1999, Safety and quality issues as part of a holistic (i.e., sociotechnological) approach, *International Journal of Occupational Safety and Ergonomics*, 5, 279–90.

Integration of Risk Management into Complex Management Systems

J. Sinay

Department of Transport Machines and Logistics, Technical University of Kosice, Kosice, Slovakia

1. INTRODUCTION

A high level of complexity is a typical feature of new technologies and new machine design. New machines and new machine technologies are more and more complicated. Their influence on the environment, ergonomics and human factors must be taken into consideration.

New development trends in the area of engineering also demands new technologies, new machine design, reliable control systems and new materials. Safety in the man–machine–environment system is a very important decision criterion for the correct exploitation of new technological systems. It is necessary to create conditions for the safe operation already in the phase of design of new machines or new machine technologies. These aspects of machine safety are not only important questions for designers but also they are integrated into the legislation of many countries throughout the world. It is the expression of national priorities in the area of work safety and health protection in the working process.

2. RELATION BETWEEN SAFETY AND TECHNICAL RISK

Technical safety is a very important function, i.e. the subject (machine, technology, operation) must perform its functions without any hazardous situations to persons or the environment. A hazardous situation is one in which a person is exposed to a hazard. A hazard is a potential source of harm. Harm is physical injury or damage to the health of people either directly or indirectly as a result of damage to property or to the environment.

All used methods for the analysis of global safety must take into account not only technical aspects of the safety of machines and devices, but also the protection of health and the environment.

Technical risk is defined as a combination of the probability of occurrence of harm and the severity of that harm (weight factor). For the computation of risk level, one must know real values for the probability of occurrence of harm and the values of severity of harm. If the final computed risk is greater then an acceptable level, the necessary measures for risk reduction or elimination must be performed. This process is risk management (RM), being one subsystem of global management of work safety. Safety here is a freedom from unacceptable risk.

RM (or "control of risk") is based on the three points: identification analysis, reduction and elimination of the risk. Europe has a new legislation framework for RM via the Directions of Council of European Union (EU), i.e. 89/391/EU, 89/392/EU, 93/44/EU and 93/68/EU on machine safety and health protection in the working environment.

89/391/EU is concerned with safety and protection of employees during the working process. According to it, an employer is obligated to evaluate the technical and human risks and to perform all necessary protective measures. To prevent risks, working conditions, social relations, environment, working organization and technical influences must be taken into account.

The most important Direction for machine safety is 89/392/EU (along with its supplements 91/368/EU, 93/44/EU and 93/68/EU). It legislates member countries about the safety of machines and machine devices; and its main purpose is to determine the requirements for safety and health protection during machine operation and also the creation of conditions for the free movement of commodities in the framework of the common European market.

An integrated part of International Standard ISO 14001 is the area of RM. Chapter 4.4.7 from this standard is devoted to the problematic of accident situations.

It can be stated that, according to ISO 9001, Part I, the employer has full responsibility for the activities of all managers at all management levels in the organization. After 89/392/EU the employer must integrate such activities like identification, analysis and evaluation of risks into one complex. (Everybody is also subjected to many risks at work and at home.) For this reason the employer, designer, operator, safety engineer and quality expert are obliged to identify, analyze, evaluate risks and perform the necessary measures for the reduction or elimination of them, i.e. to perform all activities within the framework of RM.

It is usually known that if there are often failures in the system, there will also be conflicts between safety and quality. Therefore, so-called "weak points" must be eliminated from complex systems
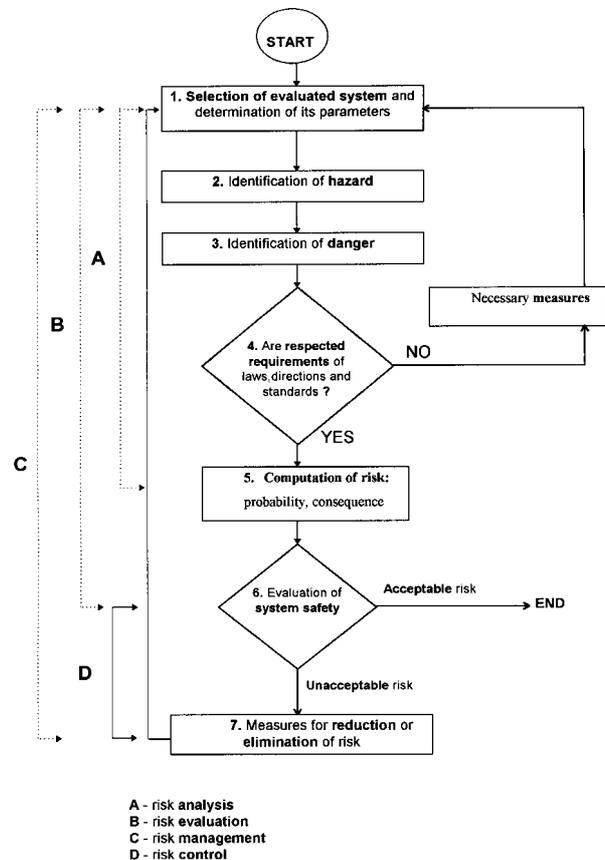

Figure 1. Activities of risk management.

to increase reliability. Accordingly, an increase of reliability means also the elimination of these weak points, and the elimination of the weak points means the minimization of risk.

The final efficiency and success of the working systems depends on the interaction of the quality of various components: technical means, organizational activities, people and social aspects.

The most important task is the control of the risk level in man–machine–environment systems, i.e. to reduce the final risk level and to obtain the so-called acceptable risk level. It is necessary from this to analyze and classify risks for the next minimization and for the realization of the whole complex of RM activities (figure 1) (Sinay 1997).

3. INTEGRATED MANAGEMENT SYSTEMS

ISO 8402 explains "quality" as a complex of properties of the subject that fulfills all determined and supposed requirements. Safety is one of very important components of quality. There are many common aspects and connections among the systems of quality management (QM), technical safety management and work safety, i.e. among the components of RM, which also includes the area of environment (Pischon and Liesegang 1997). Therefore, it is possible – like the first step – to integrate components of QM and the environment management system (EMS) into one complex system of QM by means of the total quality management (TQM) methods, and – like the second step – also to integrate the components of safety and health protection, e.g. in accordance with BS 8800/1996 (Guide to Occupational Health and Safety Management Systems). All three components of TQM are defined separately in the individual regulations, e.g. in ISO 9001, ISO 14001, EU 89/391 or EU/392. At present, there is no international or national standard that can define all three TQM components – quality, environment and safety, i.e. QM, EMS and RM – like one common base of properties of the final product.

QM, safety of technical systems, safety of work and environmental measures were separated from the organizational point of view in the past. There are many new international rules and

regulations for the technical safety and health protection that are integrated into the internal laws of countries. But it is not possible (or not so easy) to elaborate one common international standard for application in all industrial areas.

3.1. Purposes of the Integrated Management System

The main purpose of the integrated management system (figure 2) is the analysis of all partial goals in the company, the minimization of redundances and the exploitation of synergetic effect (Sinay 1997). This integrated system has many advantages, for example:

- A high level of quality.
- An increase insafety for technical devices and health protection.
- A better social acceptance of company.
- A better acceptance of environmental and safety legislation.
- The possibility of avoiding conflicts between companies and offices of state inspection.

3.2. Activities of the Integrated Management System

After creation of the integrated management system the main question is if all required aims have been achieved. There are three main categories of aims in connection with the newly created integrated management system: quality of production, safety of work and environment.

For the first category – quality – it is necessary to determine the criterion for the evaluation of quality of the final product. For the second category – safety – one must create the possibility of reducing the risk level and, above all, to define the residual risks. For the third category – environment – the question of used materials, storage systems and recycling processes is important.

Processes and technologies in man–machine–environment systems and also the final products are the subjects of an analysis in the framework of the RM (Leist 1997).

3.3. An Example of Management System Activities

An example of the integration of QM, RM and EMS activities into one complex management system (from the area of the pipeline transport systems) is shown.

During the transport of various material mediums in the pipeline system, the quality of the transported medium/material must be secured in accordance with the given specifications and requirements, i.e. this is the area of QM.

Next, a very important question is the possibility of the material leakage into the environment. From this point of view are defined the conditions for proper dimensioning of the pipeline, for the right choice of the pipe material, for protection of the pipe surface and for the surface protection paints. In such a way technical safety and the environment are joined.

Both the assembly and maintenance works demand performance of activities of the RM and environment management.

There is described a distribution of the individual activities in the framework of the integrated management system in the relation to several functions of organization in table 1 (Leist 1997).
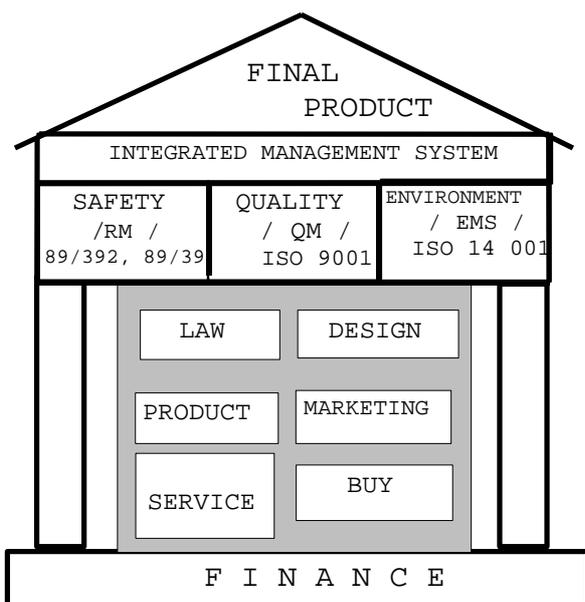
Figure 2. Integrated management system.

Table 1. Distribution of risk management functions.

Function	Quality management tasks	Risk management tasks	Management of the environment tasks
Research and development of devices	Research, development of new conception of quality management	Research, development of new methods of risk evaluation	Research, development of environmentally friendly products
Projection, planning	Conditions for quality requirements of products	Information about required safety	Determination of environmental requirements
Production, means, maintenance	Maintenance for high level of quality and durability	Identification of failures and elimination of them New safety devices	Usage of environmentally friendly production means
Material flows	To ensure the quality of articles during transport of material	To ensure the safety during transport of material	The choice of environmentally friendly transport methods

REFERENCES

LEIST, R., 1997, *Qualitätsmanagement* (Augsburg: WEKA Fachverlag für technische Führungskräfte GmbH).

PISCHON, A. and LIESEGANG, D.-G., 1997, *Arbeitssicherheitsmanagement als Bestandteil umfassenden Managementssystems – VDSI* (Heidelberg: Curt Haefner).

SINAY, J., 1997, Mana~érstvo rizika, risk-mana~ment. In *Conference Kvalita-slu~ba exportu,* }ilina, Slovakia, November, p. 429.

SINAY, J., MAJER, I., ORAVEC, M., PACAIOVA, H. and SLOBODA, A. 1997, *Riziká technickych zariadení-mana~ment rizika,* (Koaice: OTA a.s).

Intelligent Transportation Systems

P.A. Hancock

Human Factors Research Laboratory, 141 Mariucci Arena, 1901 Fourth St, SE, University of Minnesota, Minneapolis, MN 55455

1. INTRODUCTION

In many countries of the world, the problems of crashes and congestion associated with ground transportation have reached unsupportable levels. In the traditionally "developed" nations of the world, there is no longer the space or land to continually build new roadways to deal with the congestion problem. Similar strictures are placed upon developing nations who are also experiencing an explosive growth in the number of vehicles on their roads. If information superhighways are important to sustained growth and development, how much more so are physical highways that carry people, goods, and vital services? The result of these trends is that we have overcrowded and dangerous highways upon which vehicles travel at progressively slower speeds as the transportation infrastructure spirals downward in decline. What can be done about this problem? The answer in many regions of the world is to turn to innovative and advanced technologies to solve the problems of congestion and safety. Across the globe this endeavor has had several labels: in Europe it has been referred to as Transportation Telematics; in Japan Advanced Transportation; but the phrase used here is the consensus term used in the United States — Intelligent Transportation Systems (ITS). The National ITS Program Plan (1995) refers to its function as follows:

> ITS applies a broad range of advanced and emerging technologies to the needs of our surface transportation system, drawing from such fields as information processing, communications, control, and electronics. Effectively integrated and deployed ITS technologies could lead to significant improvements in safety. Mobility, accessibility, and productivity.

The reasoning behind ITS is clear. If technology can find ever more efficient ways to pack an increasing number of vehicles safely on to existing roadways, then politicians and administrators can avoid the unwelcome and unworkable alternative of ever more road-building programs. ITS seeks to accomplish this goal through the identification of user services which have been bundled according to specific domains (see Table 1). Many of these services are building blocks that can be combined for deployment in a number of different ways. For example, ITS provides an excellent opportunity to improve the dissemination of real-time weather information (RWIS) to the travelling public but also to decision-makers routing aircraft, ground commercial transport, or emergency management. Such services can be expected to change over time as different supportive technologies mature and different forms of inter-modal linkage are developed.

In addition to these user services, if innovative technologies, such as telecommuting, can replace the need for physical travel, it is possible that the problem of congestion may be solved. In

Table 1. Identified bundles and associate user services

Bundle	User Services
1. Travel and transportation management	1. En-route driver information 2. Route guidance 3. Traveler services information 4. Traffic control 5. Incident management 6. Emissions test and mitigation
2. Travel demand management	1. Demand management and operation 2. Pre-trip travel information 3. Ride matching and reservations
3. Public transportation operations	1. Public transportation management 2. En-route transit information 3. Personalized public transit 4. Public travel safety
4. Electronic payment	1. Electronic payment services
5. Commercial vehicle operations	1. Commercial vehicle electronic clearance 2. Automated roadside safety inspection 3. On-board safety monitoring 4. Vehicle administrative processes 5. Hazardous material incident reports 6. Freight mobility
6. Emergency management	1. Emergency notification and security 2. Emergency vehicle management
7. Advanced vehicle control and safety systems	1. Longitudinal collision avoidance 2. Lateral collision avoidance 3. Intersection collision avoidance 4. Vision enhancement and crash avoidance 5. Safety readiness 6. Pre-crash restraint deployment 7. Automated highway system

From the National ITS Program Plan (1995)

the near term this is unlikely. In the United States in the two decades between 1977 and 1997, the amount of drivable road increased by 2% while the number of registered vehicles increased by 50% and the number of journeys increased by an incredible 70%. Clearly, the demand is increasing and there is at present no obvious indication of any decline. Not unnaturally, administrators turn to those who created the roadways to solve the contemporary problems and since these individuals have largely been trained in the engineering sciences, it is again unsurprising that they themselves turn to their parent discipline for solutions.

2. ENGINEERING SOLUTIONS TO ADVANCED TRANSPORTATION

The marriage of engineering solutions with advanced technologies could not seem a more natural one. Since the problem is essentially one of controlling a large number of vehicles on a relatively small amount of roadway, the obvious solution is to concentrate a greater number of vehicles into a smaller space. However, there is a barrier to this which is the unpredictability associated with human drivers. Therefore, from an engineering perspective, it would appear best to circumvent this unpredictable

component and to generate automatic control over all vehicles. In the US, this effort was labeled the Automated Highway System or AHS. AHS was certainly well-intentioned and in some ways, there were signs that such a system could be successful. After all, when a complete system can be regulated, automated vehicles do quite well. For example, there are many shuttles which run at airports in the United States with no driver aboard and they perform to a tolerable level of success. What defeated the general idea of fully automated vehicles was the same stumbling block which has retarded progress in the area of Artificial Intelligence (AI), namely the conundrum of *context*. Where the problem of concern can be bounded, then computer-mediated systems do very well. For example, in the airport shuttle case, all stops and their locations are known. It is certain that there will be no other traffic on the line, etc. Similarly, in chess, AI programs have been successful since they are able to survey a bounded space of possibilities (despite the apparent enormity of that space of potential moves to us humans as individual chess players). Unfortunately, real-world problems are not so easily bounded in that same formal way. Thus, to be successful, automated vehicle systems would have to possess a wealth of "world" knowledge which appears so facile to human performers yet so difficult for their machine surrogate.

If we are unable to have a completely automated system the engineering argument ran, then let us develop a limited "world" in which some automated vehicles could operate. This was the idea behind levels of automation that could use "automated" lanes as used in the automated highway demonstration. To gain entry into this privileged world, the vehicle would have to undergo an electronic "check" to ensure that it had sufficient "intelligence" to operate within such a world. This conception seemed feasible, especially since many congested urban areas had already begun to construct or designate high occupancy vehicle (HOV) lanes in initial attempts to increase passenger-to-vehicle ratios. In order to maximize the effect of such automated lanes, the concept of "platooning" was adopted in which a group of vehicles, all travelling at the same uniform velocity, joined together in close proximity (essentially 1–2 feet apart from bumper to bumper). The technical feasibility of platooning was shown in several impressive demonstrations. However, some major problems have beset the platooning version of automated transportation. First, there is the problem of entry into automated lanes when large platoons can suddenly appear and the choice of entering the lane is at the discretion of the individual driver. Second, there is the problem of assembling platoons as they travel along well in excess of 60–70 mph. Third, there is the problem of egress from platoons or platoon dispersal as all cars reach their downtown destination at one time. This is not to say that such problems are insuperable, and indeed much progress was made on these issues culminating in a series of demonstrations of automated vehicle control in California in 1997, as had been mandated in the original Intermodal Surface Transportation Efficiency Act (ISTEA) of 1991.

Despite these successes, however, it became apparent that fully automated vehicles were not the way that Intelligent Transportation Systems (ITS) were heading. Rather than one single reason, there were a collection of influences which militated against such developments. There were questions of infrastructure. For example, if vehicle were to be guided by magnetic strips in the road, who was responsible for laying and maintaining these strips? What would happen in the event of a crash if these failed? Would car manufacturers construct vehicles that needed such external forms of support to operate efficiently and would the costs of these added features be supported by buyers in the market? In addition, how would automated and non-automated vehicles be integrated on the roadway and how would such dedicated lanes serve all road users? This is not to say that such problems of legislation, infrastructure, and responsibility — like those of a more technical nature — could not have been overcome. Rather, it is that ITS chose a different direction for progress and the vital conception in this new thrust is human-centered transportation.

3. HUMAN–CENTERED TRANSPORTATION

In actuality, human-centered transportation is an initiative of the US Government that extends well beyond ground transport to include aviation, marine, and space operations. Human-centered approaches to ground transportation can vary widely, depending upon what the specific institution or agency means by the term. In ITS, human-centered approaches have been founded upon user-services of which 29 were indicated in the 1995 ITS plan (see table 1), while some others have since begun to emerge. The purpose of ITS is to deliver these services. This is not what we mean by human-centered approaches in human factors and ergonomics (see Barfield and Dingus 1998). Our vision concerns itself with user-centered interactions with technical systems in which the human provides both the intention (either at the stage of design or at the stage of operation) and the control. Thus users in the ITS realm are largely passive consumers to be satisfied whereas in the human factors realm they are self-intention beings to be designed for. Despite this difference, I shall deal with users in the latter sense, and thus will focus largely on users as drivers of vehicles that are envisaged as including many advanced technical systems in the near future. A first question is how to structure a brief review of such a vast and growing area of research and applications? Here, I have decided to use response time as the defining characteristic. In so doing, I shall focus on some forms of current technology, mostly those in the vehicle, while necessarily giving less attention to other innovations, such as those being implemented in advanced traffic management centers. In general, the three categories I shall deal with are (1) communications, (2) navigation, and (3) collision-avoidance.

3.1. Communications Technologies

The first forms of advanced in-vehicle systems to penetrate the mass market have essentially to do with communication. In this, the cellular phone is pre-eminent. Cellular phones first appeared in the US in 1983 and were installed in vehicles closely thereafter. Although still somewhat rare in rural areas, it is difficult to drive down a major road of any urban region of the US without seeing someone telephoning from their car. Although there may be some users for whom a phone is a vital piece of operational equipment (e.g. emergency services), most car phones are discretionary technology which it may be useful to have, but not impossible to live without. A natural question that has surfaced is whether it is safe to drive and phone at the same time. It turns out that this question is not simple to answer (Goodman, Tijerina, Bents, and Wierwille 1999; Hancock and Scallen 1999) and the National Highway Transportation Safety Administration (NHTSA) has

recently conducted an extensive study of just this issue. In Japan, a breakdown of accidents involving car phone use reveals that 42% occurred when the driver was answering the phone, 31% took place when the driver was dialing and 16% resulted when the driver was simply conversing on the phone. However, as Hancock and Scallen (1999) noted, the simple act of phoning and driving together may, in themselves, not be dangerous, rather it is the context in which such actions are performed that cause them to be more or less safe activities. In view of this contextual dependency, the present safety effect of cellular phone use on driving is still largely to be determined.

Now that we have a communications channel into the vehicle there is, in principle, nothing to stop the whole of the electronic world being introduced into the driving environment. Already it is possible to receive fax and email. There are also plans for an auto-PC in which the car can become a travelling office accessing the World Wide Web and unlimited electronic information, all while travelling at 70 mph down a crowded freeway. There are already great concerns over drivers reading newspapers, applying make-up, shaving, and the like while driving. This concern must surely grow if distractions such as email, TV, and video games are placed in everyday vehicles. Many such commercial products are already being installed and operated. How and where limitations are put upon these in-vehicle systems appears, at this time of writing, yet to be determined. Of course, we can all think of ways of locking drivers out from use while the vehicle is in motion; however, for some services such communication capabilities are vital. Cellular communications have already given rise to mayday and automatic collision notification systems (e.g. OnStar, RESCU, etc.) and promise several other critical advances such as emergency notification upon airbag deployment, route assistance, real-time traffic data, and stolen vehicle recovery. What can and what should be available for operation while the vehicle is in motion is clearly a human factors question and one which many of our community is now investigating.

3.2. Navigation Technologies

If mobile phones are among the first advanced electronic systems to enter the vehicle, map navigation systems are fast following. Global positioning technologies (e.g. GPS) enable the immediate and accurate positioning of the vehicle, which can be immediately displayed on an electronic map. Some professionals, such as taxi drivers and delivery personnel, are in constant need of direction from their origin to a destination with which they are not familiar, and there is an advantage for emergency services to have information upon locations to which they are being summoned. However, is there such a strong demand from the ordinary driver? We know that the vast majority of driven miles occur on routes with which the driver is intimately familiar. Further, we know that with the increasing use of cars for commuting there is a slow but clear decrease in car use as a source of recreation. Therefore, we must ask, what advantage is it to the regular driver to possess such a capability as in-car, dynamic map-navigation systems? The answer to this seems to be bound up in the question as to how much the average driver would be willing to pay for this as a vehicle option. As of the present the answer does not seem to be a lot.

Manufacturers of such systems point to some forthcoming additional advantages. For example, when linked to information broadcast from central traffic management centers (TMCs), map systems can also display congestion. With this information drivers can choose to re-route themselves or request that the navigation system does this for them and provide a more efficient route. In saving the driver time, such systems may save the driver more money than their cost and so prove sound investments. However, this assumes that there are always viable alternative routes, but in many urban areas this simply is not so. Often, there may be only one or two routes possible between, say, downtown and a given suburban region and, if one way is blocked, virtually all drivers on the road are aware of the alternative. The optimistic advertising scenario of the powerful vehicle roaring though empty urban backstreets is likely to be misleading and counter-productive. Further, since many current map systems are downloaded from large-scale military or civilian sources, such as the USGA, they provide birds' eye views of areas from a perspective vertically above the area of interest. In reality, however, individuals do not navigate well using this perspective, but are much better at landmark recognition from the viewpoint of the human eye. Thus, turn-by-turn advisory systems are more closely allied to actual driver navigation behavior, and a number of these have begun to come to market. Perhaps integration of such understanding of human preferences and behavior will facilitate market penetration.

From a safety standpoint, navigation systems have the potential for causing significant driver distraction or increased workload. In Japan, where 12% of all new passenger vehicles come equipped with such systems, they resulted in one fatality and 93 injuries in the first half of 1998. The majority of these accidents (73%) occurred while drivers were looking at the navigation systems. This is not surprising given that driver inattention is a primary or contributing factor in a majority of all crashes in the US — even without such in-vehicle navigation systems. Thus, the challenge for designers and manufacturers is to make these devices safe and easy to operate. For example, the seemingly simple issue of deciding when to present turn instructions depends not only on the timing of the route instruction, but also on the traffic conditions and road geometry. On a positive note, standards and design guidelines are being developed for limiting access to input information to navigation systems while the vehicle is moving, as well as ensuring message uniformity, prioritization of functions for in-vehicle messages.

3.3. Collision-warning and Collision-avoidance Systems

While congestion is certainly a significant economic problem, it is the improvement of safety that stands out as the main goal of ITS. In the US alone, vehicle crashes take the lives of 42 000 people per year and millions more are injured. Adding together both the direct and indirect costs of vehicle accidents, the financial detriment alone runs into the hundreds of billions of dollars. However, road traffic accidents do more than this. Figures show that young people are far more likely to die as a result of road traffic crashes than from any other cause. Thus, road traffic accidents rob individuals, families, and society of more useful years of life than any other source of societal harm. Given this, there is a strong moral and financial impetus for us to seek ways in which technology can successfully reduce the accident toll. ITS collision warning and avoidance services

promise this increased safety and efficiency through computer-based decision and automation in the form of driver assistance or warning systems. These devices are expected to reduce both the occurrence and the severity of crashes, as well as property damage losses and crash-caused traffic delays that lead to lost work, wages, or productivity. The National Highway Traffic Safety Administration estimates that over 1 178 000 crashes could be avoided if only three systems — rear-end, lane change/merge, and road departure systems — are implemented (NHTSA 1996). To help achieve commercialization of effective collision warning and avoidance systems, NHTSA has undertaken a program of research to develop safety-based performance specifications for a number of these systems, including those for preventing rear-end, road departure, lane change and merge, and backing crashes. The automotive industry envisions such collision warning/avoidance safety services will be available within the next five to ten years.

Although collision-warning and collision-avoidance are terms that are used synonymously, there is an important distinction. Collision-warning systems are used to focus the driver's attention upon a source of threat and thus provide advanced notice of an impending collision. Collision-avoidance systems initiate some form of active intervention by the system to circumvent imminent collision.

We deal first with collision-avoidance systems. Drivers have finite capabilities and among these are restrictions upon the time it takes them to perceive and respond to stimulation in the environment. For changes that can be anticipated, humans respond relatively quickly, but their responses can be quite slow for unexpected changes. In addition, there are some events which occur with such a short latency that even the most attentive and skillful individual is unable to respond, and these are the conditions in which engineering solutions must step to the forefront. For example, an incursion into the path of progress of the driver some 250 milliseconds into the future (the equivalent of a child stepping out, from between two parked cars, 11 feet in front of a vehicle travelling at 30mph), is not amenable to significant driver response. In this case, a purely automated system could be responsible for engaging in evasive action such as swerving and/or immediate application of full braking. More complicated questions are raised when the driver and the automated system have to work together to achieve the goal of collision-avoidance. Many human factors questions are immediately evident (see Hancock, Parasuraman, and Byrne 1996). Who has control? How and when is control passed from one entity (the driver) to the other (the automation)? How and when is that control returned? Should automation ever usurp the driver's right of control? Under what circumstances might this be envisaged?

Collision-warning systems, unlike collision-avoidance systems, provide information further in advance of incipient collision. One obvious question is how far? The further off in time and space that warning about a potential collision is given, the greater the propensity for false alarms. Even with virtually flawless detection systems, there is a significant false alarm problem anyway since the probability of collision on the road is actually so low (Parasuraman, Hancock, and Olofinboba 1997). In addition, how are we to warn drivers? Is a four-dimensional auditory warning, necessarily competing for visual attentional resources, the best form of localization? How are we to calibrate different systems to different drivers' styles? Another concern is that drivers may compensate for each of the safety improvements afforded by collision-warning and avoidance technologies to gain what they perceive as the best advantage for themselves. Over-reliance on these systems may therefore lead drivers to assume additional risks and compromise any gains in safety. Drivers with forward collision-warning or night-vision systems, for example, may have a tendency to drive faster or exercise less caution than they otherwise would under degraded conditions. Product standardization and consistency is also paramount among safety critical warning systems. All these and a multitude of other questions have yet to be resolved as we endeavor to assist drivers in a task which, we must recognize, they do very well on a day-by-day basis anyway. These are part of the future challenges of ITS.

4. SUMMARY

Intelligent transportation systems using human-centered views are currently being built and implemented. However, progress in this area is moving so rapidly that it is likely that by the time this work is published some, if not all, of the developments described in this article will be either outdated or obsolete. While there have been significant and sincere attempts to bring human factors and ergonomics professionals into the development of standards, guidelines, and many specific designs, it is clear that not all such burgeoning technologies are likely to have benefited from human factors input; unfortunately, in a litigation-laden society, there will be many opportunities for forensic human factors professionals to comment and speculate upon the impact of designs of varying utility. If, however, we have contributed to the reduction of crash frequency and can demonstrate that human-centered approaches are effective in the protection of pedestrians, drivers, and passengers, we will have made a contribution that our students and followers can take into the future.

ACKNOWLEDGMENT

REFERENCES

BARFIELD, W. and DINGUS, T.A. (eds), 1998, *Human Factors in Intelligent Transportation Systems* (Mahwah, NJ: Erlbaum).

GOODMAN, M.J., TIJERINA, L., BENTS, F., and WIERWILLE, W.W., 1999, Using cellular telephones in vehicles: Safe or unsafe? *Transportation Human Factors*, 1, 3–42.

HANCOCK, P.A. and PARASURAMAN, R., 1992, Human factors and safety in the design of intelligent vehicle-highway systems. *Journal of Safety Research*, 23, 181–98.

HANCOCK, P.A., and SCALLEN, S.F., 1998, The driving question. *Transportation Human Factors*, 1, 47–55.

HANCOCK, P.A., PARASURAMAN, R., and BYRNE, E., 1996, Driver-centered issues in advanced automation for motor vehicles. In R. Parasuraman and M. Mouloua (eds) *Automation and human performance* (Mahwah, NJ: Erlbaum), 337–64.

National Highway Traffic Safety Administration (NHTSA), 1996, *Preliminary assessment of crash avoidance system benefits.* NHTSA Benefits Working Group, Washington, DC, October.

PARASURAMAN, R., HANCOCK, P.A., and OLOFINBOBA, O., 1997, Alarm effectiveness in driver-centered collision-warning systems. *Ergonomics*, 40, 390–9.

Operator Testing and Non-destructive Evaluation Performance

D. H. Harris

Anacapa Sciences, Inc., 301 E Carrilo St, PO Box 519, Santa Barbara, CA 93102, USA

1. INTRODUCTION

Non-destructive evaluation (NDE) is a critical activity in assuring the structural and functional integrity of components in power plants, production facilities, aircraft, ships, submarines and numerous other types of vehicles and facilities. NDE systems employ ultrasonic, radiographic, electromagnetic (eddy-current) and other techniques to detect and characterize flaws in materials. In these techniques probes return signals from the materials examined and operators interpret the signals (within the signal context) to assess the condition of the material. Since these techniques depend on operator proficiency for their effectiveness, operator testing can contribute to the effectiveness of system performance through: the early identification and selection of personnel who are most likely to benefit from training and experience in the application of NDE techniques; and the identification and qualification of personnel whose capabilities meet the standards required for specific NDE applications in the field.

Potential contributions of aptitude testing to the selection of candidates for NDE jobs is illustrated by the development and validation of the Dynamic Inspection Aptitude Test (DIAT). It was designed to predict success in the application of NDE techniques (such as ultrasonic and eddy-current) that require the combined aptitudes of general cognitive ability, abstract reasoning and spatial visualization. Validation studies showed that the DIAT was highly reliable, correlated with other measures of these key aptitudes and was predictive of the job performance of NDE operators.

Performance testing can provide valid, reliable, practical methods for the identification and qualification of personnel whose capabilities meet the standards required for specific NDE applications. A notable example is the performance testing that serves as the principal basis for qualifying operators for the eddy-current examination of steam generator tubes in nuclear power plants. However, while providing a valid basis for the qualification of operators, performance test results may not necessarily predict performance in the field. Various performance-shaping factors (such as alertness, work habits, environmental conditions, job preparation and motivation) might operate to degrade performance under field conditions.

2. APTITUDE TESTING

Attempts to predict job performance from scores on aptitude tests are based on the assumption that individuals who have certain innate human capacities (aptitudes) will be able to develop and apply job skills requiring those aptitudes more effectively than those who do not. An additional assumption is that higher-aptitude individuals will be able to learn the required job knowledge and skills more quickly and will be more likely to pass performance tests required by operator qualification programs. Under these assumptions the use of aptitude testing is appealing because the investments in testing are typically very small relative to the costs of operator training and qualification and to the costs incurred as the result of poor operator performance on the job. As a consequence, substantial returns on investments might be anticipated.

A computer-based test was designed to measure the dynamic combination of aptitudes required for the effective application of NDE. By measuring the requisite combination of aptitudes simultaneously, the approach employed for this test differed from the more traditional and more commonly used aptitude-testing methods in which aptitudes are measured one at a time by static tests administered either in paper-and-pencil form or by computer.

A computer-based test was developed in which scoring well on the test involved the simultaneous application of multiple aptitudes in performing a series of inspection-like tasks. It was expected that an individual's test score would reflect both the individual's aptitude levels and the dynamic interaction among these aptitudes. That is, the combined effect on task performance of the interaction among aptitudes of one individual could be greater than the combined effect of the interaction among the aptitudes of another, even when the levels of individual aptitudes were about the same.

2.1. Task Analyses

The specific job selected for the development and validation of the test was the ultrasonic detection of intergranular stress-corrosion cracking in the piping of nuclear power plants. This specific task was selected because, due to its scope and complexity, it encompassed the components of a wide range of ultrasonic NDE tasks performed in nuclear power plants and elsewhere. Task analyses were completed to identify the specific aptitudes required for the performance of this job. The detailed results of these analyses were reported previously (Harris 1990, Harris and McCloskey 1990). The aptitudes are defined briefly:

- General cognitive ability — the general ability to learn, to understand instructions and to solve problems.
- Abstract reasoning — the specific ability to determine an underlying principle, or logical rule, and apply it in understanding relationships.
- Spatial visualization — the specific ability to visualize operations conducted in three-dimensional space.

2.2. Description of the Test

The DIAT consisted of 36 items, each administered and automatically scored by computer. The test taker had 36 min to complete the items. Test files automatically maintained records of test scores and individual responses to each test item. Access to the test and to testing records was controlled by a password. In each of the 36 test items the objective was to locate a target box in an array of 20 boxes. One point was scored for each item in which the target box was correctly reported, providing for a maximum test score of 36. A two-step approach to locating the target box was required to score well on the test. Step 1 consisted of narrowing the number of possibilities to just a few boxes by identifying the characteristics of the target box through the use of analytical reasoning — determining the characteristics of the

last box in a five-box series. Step 2 consisted of applying three tests to the remaining boxes to determine which box responded with the prescribed pattern of audio signals. The item was completed by reporting the box selected as the target.

2.3. Assessments of Test Reliability

Two separate studies demonstrated that the DIAT had a very high level of reliability. Reliability is the extent to which variability among test scores reflects true differences among the aptitudes of the individuals tested rather error introduced by the testing process. The reliability of the test calculated from test scores of 46 students at Piedmont Virginia Community College was 0.94. The reliability of the test calculated from test scores of a sample of 32 experienced ultrasonic NDE operators currently employee by utilities and inspection service organizations in the nuclear power industry was 0.91.

2.4. Assessments of Test Validity

The DIAT demonstrated both content and predictive validity. As a consequence, the test is likely to be useful for the selection of personnel for jobs requiring the application of ultrasonic NDE methods. Content validity refers to the linkage between what a test measures and the tasks, behaviors, knowledge and skills required by a particular job. Predictive validity refers to the correlation between scores on a test and measures of job performance.

2.4.1. Content validity

The DIAT was shown to correlate significantly with widely used measures of the three aptitudes, as discussed earlier, that were linked to ultrasonic NDE tasks. A sample of 46 community college students completed the DIAT, the Wonderlic Personnel Test (a test of general cognitive ability) published by Wonderlic Personnel Test, Inc. (1992) and the Abstract Reasoning and Space Relations tests from the Differential Aptitude Tests (DAT) battery published by The Psychological Corp. (Bennett *et al.* 1991). These particular tests were selected for the study because of the extensive validity data that existed for them in a wide variety of jobs and tasks. Significant ($p < 0.05$) positive correlations, ranging from 0.36 to 0.65, were found between the DIAT and these other aptitude measures. These results indicated that the DIAT scores reflected each of these aptitudes.

2.4.2. Predictive validity

The predictive validity of the DIAT was assessed by correlating test scores with measures of ultrasonic NDE performance. The percentage of flaws correctly reported was calculated for each of 32 ultrasonic operators from their individual examinations of a sample of 45 piping specimens, both flawed and unflawed, which were representative of the flaw types, materials, examination limitations and component sizes encountered in nuclear plants.

Each operator required about 2 work weeks (80 h) to examine all specimens and to report the conditions found. These measures were obtained as part of an ongoing performance–demonstration process used to qualify operators to perform ultrasonic NDE tasks in nuclear power plants. A flaw detection rate of 80% was the established standard for operator qualification.

The correlation coefficient between the DIAT scores and the ultrasonic NDE performance measures obtained from the sample

DIAT Score	Percentage qualified for ultrasonic NDE tasks (demonstrated flaw detection • 80%)
• 25	92
Š24	25

Figure 1. Relationship between DIAT score and ultrasonic NDE performance.

of 32 operators was 0.51. The practical significance of this degree of correlation is illustrated in the expectancy chart of Figure 1, which relates DIAT test score to the percentage of operators who qualified for NDE jobs. As shown, 92% of those with DIAT scores of ≥ 25 qualified while only 25% of those with DIAT scores of £ 24 qualified.

2.5. Test Utility

Use of the test can result in cost savings from reductions in the amounts of training and experience required by selected candidates to meet standards of satisfactory ultrasonic NDE performance, and/or from the greater qualification rates of participants in ultrasonic NDE qualification programs. Currently < 60% of candidates qualify to perform ultrasonic NDE jobs in nuclear power plants. At this rate, it would take 167 candidates entering the qualification process (including those who choose to repeat the process) to produce 100 qualified ultrasonic NDE operators. At a cost of ~$6000 per candidate for salary, travel, lodging and subsistence, these costs alone for qualifying 100 operators would be $1 002 000 (i.e. 167 operators at $6000).

Based on the validation results, using a DIAT cut-off score of 25 to select candidates, 92% of those selected would be expected to qualify. Thus, among those scoring ≥ 25 on the DIAT, it would take only 109 candidates to produce 100 qualified operators, at a total cost of $654 000. The estimated cost savings, per 100 qualified operators, would be $348 000.

3. PERFORMANCE TESTING

Performance testing can provide valid, reliable, practical methods for measuring the capability of NDE techniques and personnel. The principal advantage of this approach is that measures can be obtained under controlled conditions that serve to eliminate potential sources of bias, subjectivity, and measurement error. Moreover, the extent of any error introduced by the measurement process can, itself, be measured and employed in establishing lower-bound confidence limits for the performance measures obtained. The standard error of measurement statistic computed for this purpose is a function of the reliability of the measurement process and the variability among the obtained measures.

Assessment of NDE capabilities is illustrated here by measures obtained from the performance of the first 286 eddy-current data-analyst candidates in the industry's qualification program (Harris 1996). On average, each candidate analyzed > 2000 tube locations containing ~400 reportable indications. Percentages of indications correctly reported (and associated 90% confidence limits) were calculated for indications attributable to the different possible damage mechanisms — thinning, wear, impingement, cracking, pitting. Thinning is the reduction in wall thickness due to the corrosive effect of water and steam over time; wear is the loss of
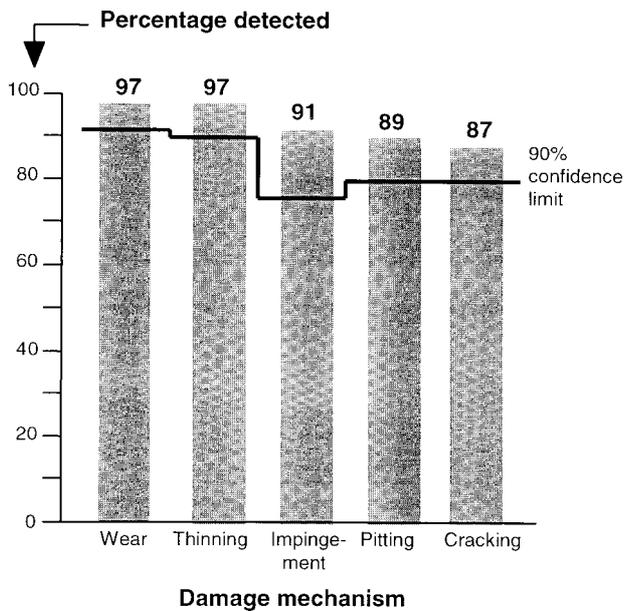
Figure 2. Percentages detected of indications attributable to wear, thinning, impingement, pitting and cracking.

tube material from the abrasive effect of another object (such as a support bar); impingement is damage caused by another object striking the tube; cracking is tube-wall reduction from a combination of stress and corrosion; and pitting is the loss of tube material due to corrosive effects of chemicals in the water. These results are shown in Figure 2.

As shown from these results, performance may differ as a function of the conditions under which an examination is conducted. For example, Figure 2 indicates differences in capabilities to analyze indications attributable to different damage mechanisms. Such findings suggest that the use of generic performance measures (those which have been averaged over various examination conditions) may lead to inadequacies in planning and conducting NDE. Other conditions which have been shown to influence operator capability for the detection of flaws of this type include: material, structural configuration, location, sizes of discontinuities and associated signal strengths, guidelines employed and specific NDE techniques employed.

3.1. Performance-shaping Factors

Measures of NDE capabilities obtained from performance testing may not necessarily be totally predictive of the levels of performance to be expected under field conditions. They would be highly predictive if field examinations were conducted under conditions equivalent to those of performance testing. However, there may be various performance-shaping factors that operate under field conditions to influence performance levels. Thus, the potential influence of performance-shaping factors must be addressed in arriving at the performance parameters to be employed in inspection planning and in assessing structural integrity based on NDE results.

There are numerous factors on the job that might degrade NDE performance from levels at which a specific technique–operator combination is capable of performing. Some factors might affect the hardware and software employed for the acquisition and analysis of data, such as less than optimal

calibration of the system, probe deterioration, variations in the probe-tube interface and so on. Other factors, such as the following, might affect the operator.

Alertness:
● Physical condition.
● Time on duty.
● Time since last break from work.
● Time since last reportable indication.
● Distractions (family, financial).

Environmental conditions:
● Temperature, ventilation, noise, lighting, interruptions.

Job preparation:
● Job-specific training and testing.

Motivation:
● Recognition, opportunities, feedback on performance, supervisory practices.

Unfortunately, relatively little is presently known about how these variables might operate, individually and in combination, to influence different aspects of NDE performance under different circumstances. As a consequence, any approach to predicting the degradation of performance by addressing each of the individual effects of specific performance-shaping factors would probably prove to be impractical.

A more reasonable alternative would be to compare measures obtained from performance testing with performance measures obtained under field conditions for the same or equivalent technique-operator systems. Data from comparisons of this type, reflecting the combined effects of the more influential factors, could then be used as the basis for developing algorithms for estimating field performance from measures obtained from performance tests.

3.3. Comparison of Performance Testing Results with Field Performance Measures

A recent study illustrated the comparison of measures made during performance testing to measures obtained under field-inspection conditions. A sample of 818 steam generator tubes was examined by 12 qualified operators under performance testing conditions and, separately, by two other qualified operators under field conditions. The sample contained 1363 reportable indications over a wide range of signal strengths (voltages). All operators, under both conditions, worked independently of other operators. A comparison of the mean percentages detected under the two conditions is provided in Figure 3. Each data point represents the average indications correctly reported within one of seven categories of signal voltages.

The results showed a very high positive correlation between performance-testing detection rates and field detection rates ($r = 0.999$), even though performance measures were obtained from two separate groups of analysts rather than from the same analysts under each of the two conditions. Thus, in this single, limited assessment, performance-shaping factors did not significantly degrade NDE capabilities under field conditions, and field performance could have been accurately predicted from measures
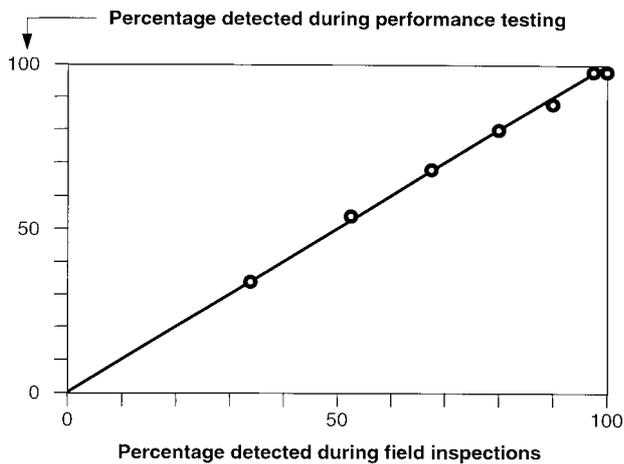
Figure 3. Relationship between testing performance and field performance.

obtained from performance testing. On the other hand, the signal voltages of the crack indications greatly influenced detection performance. Based on these results, one could not expect detection rates ≥ 90% unless signal voltages were ≥ 1.00.

4. SUMMARY AND CONCLUSIONS

Potential contributions of aptitude testing to the selection of candidates for NDE jobs was illustrated by the development and validation of the DIAT. This test was designed to predict success in the application of ultrasonic NDE techniques that require the combined aptitudes of general cognitive ability, abstract reasoning and spatial visualization. Validation studies showed that the DIAT was highly reliable, correlated with other measures of these key aptitudes, and was predictive of the job performance of NDE operators. Moreover, they demonstrated that use of the test could result in considerable cost savings in the training and qualification of NDE personnel.

Effective NDE planning requires accurate measures of the capabilities of NDE systems, the combinations of techniques and operators to be employed. To this end, performance tests have been developed and shown to be valid, reliable, practical methods for measuring the capability of NDE techniques and personnel.

However, measures of capabilities may not necessarily predict performance under field conditions, particularly if field conditions differ significantly from those under which performance testing is conducted. Numerous factors in the field environment can operate to degrade system performance. Consequently, research is required to compare the extent to which performance-shaping factors degrade NDE capabilities when they are applied under field conditions.

One recent, but limited, study indicated that capabilities for the eddy-current detection of indications in steam generator tubes was hardly degraded at all under field conditions. However, more comprehensive studies of this type are required to provide a definitive basis for the prediction of field performance from measures obtained from performance testing. The goal of the recommended research would be to develop algorithms that can be applied to performance test results to predict NDE performance in the field.

REFERENCES

BENNETT, G. K., SEASHORE, H. G. and WESMAN, A. G., 1991, *Differential Aptitude Tests for Personnel and Career Assessment, Technical Manual* (San Antonio: Psychological Corp.).

GUILFORD, J. P., 1954, *Psychometric Methods* (2nd ed.) (New York: McGraw-Hill).

HARRIS, D. H., 1990, Effect of human information processing on the ultrasonic detection of intergranular stress-corrosion cracking. *Materials Evaluation*, 48, 475–80.

HARRIS, D. H., 1996, Capabilities of eddy current data analysts to detect and characterize defects in steam generator tubes. In *Proceedings of the 15th Steam Generator NDE Workshop*, 8-1–8.

HARRIS, D. H. and McCLOSKEY, B. P., 1990, Cognitive correlates of ultrasonic inspection. Report EPRI NP-6675 (Palo Alto: Electric Power Research Institute).

MEISTER, D., 1985, *Behavioral Analysis and Measurement Methods* (New York: Wiley Interscience).

NICHOLS, R. W., DAU, G. J. and CRUTZEN, S., 1987, *Effective Non-destructive Examination for Structural Integrity* (London: Elsevier).

SCHMITT, N., BORMAN, W.C. and ASSOCIATES, 1993, *Personnel Selection in Organizations* (San Francisco: Jossey-Bass).

WONDERLIC PERSONNEL TEST, INC., 1992, *Wonderlic Personnel Test & Scholastic Level Exam, User's Manual* (Libertyville: author).

Process Control

N. Moray

DERA Professor of Applied Cognitive Psychology, Department of Psychology, University of Surrey, Guildford GU2 5XH, UK

1. INTRODUCTION

Process control systems involve industrial operations for the manufacture or transformation of energy and chemical products in a continuous stream through the interaction of mass and energy. They include, on the large scale, operations such as the manufacture of chemicals, oil refining and electrical power production, whether using fossil or nuclear energy; and on the small scale processes such as pasteurization and domestic heating and air conditioning. Processes that bear some similarities to process control but which are sufficiently different include discrete manufacturing, transportation systems and systems such as hospitals, although under certain conditions and assumptions aspects of the human factors of process control can be applied to these latter. For general descriptions of the human factors of process control see Rasmussen (1986), Woods (1986), Rasmussen *et al.* (1995) and Moray (1997a, b)

2. CHARACTERISTICS OF PROCESS CONTROL PLANTS

Process control systems have characteristics in common which cause difficulties for operators and other personnel. Typically they involve the interchange of large volumes or masses of physical substance, the feedstock, and the exchange of large quantities of energy. These exchanges are frequently on a scale which, if uncontrolled, give rise to considerable hazard, to which the events at Bhopal and Chernobyl bear witness. Control of rates of reaction, and mass–energy balances are central to process control operations. Frequently the products, by-products, or waste products include materials which are toxic, and which therefore must be controlled.

Process control operations are governed by the laws of Newtonian physics and related chemical laws. If a valve is opened, pressures are transmitted through the system almost instantaneously. Flows are proportional to pressures, and relations between pressures and temperatures obey the gas laws. Many processes can be described and modeled by mathematical equations, and transfer function descriptions of operations may be available. But while formal quantitative descriptions are often available, they are not always complete, and there is much uncertainty and risk under some operating conditions due to instabilities or imperfect quantitative models of process. The time constants of significant events in process control systems can be both very long and very short. It takes several days of complex operations involving teams of many people to start up a nuclear power plant or an oil refinery from cold shut-down; but the chain of events that destroyed Chernobyl occurred over minutes, and the final transient within seconds.

Process control plants are among the largest and most complex industrial systems constructed by mankind. They may occupy sites of many hectares, and the number of components in a plant may number hundreds of thousands. Nuclear power plants have ~50 degrees of freedom, implying that to control the plant an operator would need at all times to know the value of at least 50 variables. Older nuclear power plant control rooms have between 1000 and 2000 displays and controls, and modern ones several hundred pages of computerized displays. Furthermore, there is often tight coupling between different parts of the plant, so that events in one part may propagate rapidly to affect other parts. On the other hand, because of the inertia of the masses and energies involved, there may be lags of many hours in response to changes in controls. In general, except for abnormal transients, the bandwidth of events tends to the very low end of the spectrum, at which continuous perceptual motor control of closed loops is not performed by humans, who instead intervene from time to time to correct divergence from required set-points by discrete actions. Despite this there has been some success in modeling the operator using control theory (Baron *et al.* 1990).

Continual process control plants are bounded by physical constraints. Operators must ensure that temperatures, pressures, etc. do not exceed values defined by designers, mandated by regulators, or which would cause the failure of components or result in the release of toxic products to the environment. Operations are "real time": that is, a rapid response is required to divergences from designed operating conditions. Often there are no readily identifiable discrete events other than alarms for long periods. However, when a process has been shut down there is often a need to control the state of the system for many hours or days because of the presence of exothermic chemical reactions, radioactive decay, etc. Thus plants must be maintained in safe states even when they are not operating and when perhaps power to usual control systems has been lost.

Because of the complexity of process control systems widespread use is made of automatic control systems using closed-loop negative feedback, and increasingly of sophisticated automation. Dynamic allocation of function is therefore central. Process control systems may run under automation for many months, but operators should intervene from time to time to compensate for disturbances that cause drifts from set-points, to perform tests on equipment, and to manage any faults which may occur. Such activities require that operators know the state of the system, and this in turn may require coordination and communication with other personnel, management, and maintenance personnel both in the control room and at widely dispersed locations about the plant. Process control is essentially a team operation, even though manning levels have declined precipitously in recent decades with the development of automation and computerization.

It is easy to show from empirical data that operators cannot possibly keep track of all displayed information. Given known physiological limitations on eye movements and the dynamics of attention and the limitations of working memory, operators must construct their knowledge of system state from samples of a relatively small number of state variables. However, these, if chosen appropriately, can suffice to identify the overall system state because of the correlation (coupling) among variables within subsystems. Such sampling has been shown empirically by Iosif (1968, 1969a, b) and has been discussed theoretically by several writers (Moray 1986). It is widely believed that operators form mental models of the causal structures of the systems with which they work, and that these models are used as surrogates for a complete real time knowledge of all state variables.

To summarize, the characteristics of industrial systems which are of most relevance to human factors are system size, complexity, system dynamics (time constraints, bandwidth), system decomposition into subsystems and components, intrasystem coupling, closed loop and open loop control characteristics, degree of automation, style of automation, observability, modelability, bases for function allocation, modes of human–machine coupling, and the quality of the human–machine interface.

Typical tasks include start-up and shut down operations, maintaining variables at their set points, fault detection, diagnosis and management, and maintenance, which may occur during operations or during a previous planned system outage.

A major change from manual to supervisory control has occurred during the last thirty years due to increasing automation. Excellent summaries of the human factors of manual process control are provided by Edwards and Lees (1974a, b), who provide many hundreds of references to relevant laboratory and field studies. Such work is important, because when operators intervene during fault management they become elements in closed-loop negative feedback loops, and the early research discusses their properties in such conditions. Novice controllers in manual control behave as closed-loop controllers, lagging the plant state, but as they acquire expertise they become open-loop, predictive controllers, implying the acquisition of mental models which allow them, consciously or unconsciously, to extrapolate from present to future plant states, and thus achieve much more efficient reduction of system state error.

Automation has changed the job of operators from manual to supervisory control (Sheridan 1992, 1997). The main task is now to monitor the system, and intervene only when abnormal conditions arise, whether due to drifts from set points of due to faults. The combination of human and automated control can take many forms. For example, the automation may require confirmation from operators before carrying out actions, may carry them out unless countermanded, or may carry them out and inform the operator. Surprisingly little is known about the relative merits of different styles of human–machine cooperation. Bainbridge (1983) notes that one effect of automation is to make operators' work harder, since engineers automate what they understand well, leaving the remaining tasks to human control. Thus one effect of supervisory control is to place increasing emphasis on the role off the human as a fault manager. Various attempts have been made to use expert systems to embody the knowledge and wisdom of human operators, in effect to automate intelligence, but few of these have been successful, and currently there is little emphasis on the use of expert systems in process control.

3. PSYCHOLOGICAL CHARACTERISTICS OF PROCESS CONTROL TASKS

Sheridan (1997) defines the main tasks in supervisory control as being to monitor the system, intervene when necessary, learn about the properties of the system, reprogram the controls and automation when necessary, and to plan for future actions both in the short- and the long-term. From this list one may identify psychological functions and characteristics which are demanded of expert controllers. All require high-quality human–machine interfaces, with display controls designed in the light of human information processing and cognitive abilities.

Efficient monitoring requires an accurate appreciation of the dynamics of the components of the system, so that limited attention can be efficiently deployed. When monitoring conventional control panels eye movements (which approximate to the direction of attention) seldom exceed two per second, and when pages of computer information must be accessed, the rate of switching may fall well below this rate, although the resulting displays can be richer, and may make it easier to identify system state (Vicente 1999). Several models of attention dynamics are available (Moray 1986). It is important to realize that in a very large system, even optimal distribution of attention cannot guarantee that all abnormal signals will be detected. Efficient and reliable alarms must be provided, and their reliability should be considerably higher than the processes they monitor.

The data acquired by the observer serve to allow the latter to identify system state, and hence support diagnosis and the detection of faults. Although much research has concentrated on how operators reason when performing fault diagnosis, direct perception of patterns of displays, in particular patterns of illuminated annunciators, often provide potent diagnostic ability without reasoning (Rasmussen 1986). Hence, the design of displays is extremely important, and recent research has provided new insights into how to design displays for complex systems (Moray 1997, Vicente 1999). Monitoring and pattern perception together provide the basis in process control for system state identification, which in other domains may be called "situation awareness."

A major difference between novices and experts in process control is the extent to which they possess well-developed and accurate mental models of the plant. Having a mental model is more than merely having knowledge about the plant. It implies that the operator has learned how to choose a subset of state information which is practical to monitor and which suffices under normal conditions to deduce causal relations among plant components, and thence to make appropriate diagnoses and take appropriate actions. It also supports predictive behavior. Operator mental models can take many forms, from unconsciously embodied transfer functions in closed- and open-loop control to complex verbal or mental images of how the plant works (Moray 1997b, 1999). Training should aim to provide accurate mental models and support their use.

4. FAULT MANAGEMENT AND ALLOCATION OF FUNCTION

Although allocation of function between human and machine is not restricted to periods of fault management, it is convenient to consider them together because of the way operators decide when to intervene in plant operation. Since modern plants are usually optimized for automated control, operators should not intervene unless they believe that the automation is malfunctioning. It is therefore necessary that they should trust the automation. Characteristically, it is during fault conditions that trust fails, and operators take manual control. While this is desirable, they should be prepared to return control to the automation when conditions change. One problem is that when reasoning about faults operators tend to show certain characteristics which hinders dynamic allocation. When faced with abnormal conditions in a complex system, operators tend to choose a hypothesis to explain the fault as rapidly as possible, and are then unwilling to change

the hypothesis. This "cognitive lock up" as it has been called, results in their collecting evidence in favor of the hypothesis and neglecting alternatives. (The most well known example of this was the failure to diagnose the correct cause of the accident at Three Mile Island, where operators decided early on that they had too great an inventory of coolant, whereas in fact they had a loss of coolant.) A different kind of rigidity may be seen in what has been called "complacency" (Parasuraman and Riley 1997). If trust in automation is too great, operators may cease to monitor the automation, and hence fail to notice the need to intervene. It is not known to what extent complacency is a real problem: the mere fact that part of a plant that is highly reliable is not monitored often is not a sign of miscalibrated trust, but may be a correct strategy given the past history of high reliability. Our understanding of the psychological relations between humans and the systems they supervise is still relatively poor, and it is not clear to what extent research can be transferred between domains, for example, from studies in aviation to those in process control. (Compare, for example, Parasuraman and Riley 1997 with Lee and Moray 1994.) To understand the psychology of this kind of behavior it is most important to rely as much on field research in industrial settings as on laboratory studies. Field studies in process control have a long history, from work such as that of Bainbridge (1974) to more recent work particularly in Europe (De Keyser 1981, De Keyser *et al.* 1987, for example).

5. CONCLUSION

Process control provides an environment in which people are required to work in the most complex systems ever designed and built by humans. The transfer of knowledge from basic research in cognitive and social psychology provides an opportunity for human factors to play a major role in the development of safer and more productive socio-technical systems, and to improve the working conditions of personnel. At the same time, the remarkable skill of experienced operators in process control provides important data about human cognition, and a challenge to laboratory research to explain the abilities of humans who, despite progress in automation, it is now accepted must remain an essential component in the operation of such plants. Since the books by Edwards and Lees (1974) there have been great advances in the human factors of process control, and currently human factors can provide important insights into the role of humans in automated systems and how they should be designed for process control.

REFERENCES

BAINBRIDGE, L., 1974, Analysis of protocols from a process control task. In E. Edwards and F. Lees (eds), *The Human Operator in Process Control* (London: Taylor & Francis), 146–58.

BAINBRIDGE, L., 1983, The ironies of automation. *Automatica*, 19, 755–79.

BARON, S., KRUSER, D. and HUEY, B.M., 1990, *Quantitative Modeling of Human Performance in Complex, Dynamic Systems* (Washington, DC: National Academies Press).

DE KEYSER, V., 1981, *La fiabilité humaine dans les processus continus, les centrales thermo-éléctriques et nucléaires*. Technical Report 720-ECI-2651-C-(0) GCE-DGXII (Brussels: CERI).

DE KEYSER, V., DE CORTIS, F., HOUSIAUX, A. and VAN DAELE, A., 1987, *Les communications hommes–machines dans les systèmes complexes*. Appendice, Technical Report Contrat No. 8 (Liège: Actions Nationales de Recherche En Soutien A Fast, Université de Liège).

EDWARDS, E. and LEES, F. (eds), 1974a, *The Human Operator in Process Control* (London: Taylor & Francis).

EDWARDS, E. and LEES, F., 1974b, *Process Control* (London: Institute of Chemical Engineers).

IOSIF, G., 1968, La stratégie dans la surveillance des tableaux de commande. I. Quelques facteurs déterminants de caractère objectif. *Revue Roumanien de Science Social-Psychologique*, 12, 147–61.

IOSIF, G., 1969a, La stratégie dans la surveillance des tableaux de commande. I. Quelques facteurs déterminants de caractère subjectif. *Revue Roumanien de Science Social-Psychologique*, 13, 29–41.

IOSIF, G., 1969b, Influence de la correlation fonctionelle sur parametres technologiques. *Revue Roumanien de Science Social-Psychologique*, 13, 105–10.

MORAY, N., 1986, Monitoring behavior and supervisory control. In K.R. Boff, L. Kaufman and J.P. Thomas (eds), *Handbook of Perception and Human Performance* (New York: Wiley).

MORAY, N., 1997a, Human factors in process control. In G. Salvendy (ed.), *Handbook of Human Factors* (2nd ed.) (New York: Wiley), 1944–71.

MORAY, N., 1997b, Models of models of … mental models. In T.B. Sheridan and T. van Lunteren (eds), *Perspectives on the Human Controller* (Mahwah: Lawrence Erlbaum), 271–85.

MORAY, N., 1999, Mental models in theory and practice. In *Attention and Performance XVII* (Cambridge, MA: MIT Press).

PARASURAMAN, R. and RILEY, V., 1997, Humans and automation: use, misuse, disuse, abuse. *Human Factors*, 39, 230–53.

RASMUSSEN, J., 1986, *Information Processing and Human–Machine Interaction: An Approach to Cognitive Engineering* (Amsterdam: North-Holland).

RASMUSSEN, J., PEJDERESEN, A.-M. and GOODSTEIN, L., 1995, *Cognitive Engineering: Concepts and Applications* (New York: Wiley).

SHERIDAN, T.B., 1992, *Telerobotics, Automation, and Human Supervisory Control* (Cambridge, MA: MIT Press).

SHERIDAN, T.B., 1997, Supervisory control. In G. Salvendy (ed.), *Handbook of Human Factors* (2nd ed.) (New York: Wiley), 1295–327.

VICENTE, K.J., 1999, *Cognitive Work Analysis: Towards Safe, Productive and Healthy Computer-based Work* (Marwah: Lawrence Erlbaum).

WOODS, D.D., 1986, Human factors challenges in process control: the case of nuclear power plants. In G. Salvendy (ed.), *Handbook of Human Factors* (New York: Wiley).

Rail Transport

I. Grabarek

Faculty of Transport, Warsaw University of Technology, Poland

1. INTRODUCTION

The basic element of rail transport system is rolling stock, which can be divided into traction vehicles, passenger coaches, and freight wagons. Traction vehicles include locomotives with electric or internal-combustion engines. The rapid development of electric traction in many countries and a large number of electric locomotives concentrates the subject of this article on problems of ergonomics in that area of transportation.

From the point of view of ergonomics the most important issue is the place where the train driver works — that is, the cab. The work of a train driver is typical operator work in a moving technical object. Basic functions of a driver-operator are the processes of orientation, decision-taking, and execution. The main equipment in each cab is the control desk with steering and signaling devices. The optimization of the driver–locomotive–environment system depends on the degree in which the rules of ergonomics are applied in the phase of designing the vehicle, and then in maintaining the state during the usage.

That is why ergonomic examination is part of concept or correction ergonomics. Safety and economy of transportation enforces the designers to develop more advanced solutions of cab design. The concepts are presented in the later part of this article. In correction ergonomics, the most important factor is the method of ergonomic diagnosis, the aim of which is not only to establish the actual level of ergonomics in a locomotive cab, but first of all to formulate conclusions for its modernization, which would make driver's job more comfortable. The tools for simplifying and accelerating the diagnosis are computer-aided expert systems, whose usefulness has been proved in research.

2. CHARACTERISTICS OF DRIVER'S WORK

The function of a driver of an electric locomotive is not only to integrate received information, but also to steer the vehicle. The signals received from the environment (the track, high and low signals, obstacles, etc.) and information from indication devices in the control desk are continuously analyzed and processed. The effect of these procedures is the making of an executive decision, performed via steering devices. The operator is, first of all, the moderator of data processing. It does not change the fact that he is to a large extent a receptive element in the system, registering data from both the track and the control desk. The steering functions are accomplished by relatively small movements that do not require great precision and speed. The necessity of quick decision-taking arises in situations of possible collisions (e.g. unexpected obstacle on the track), driving at night or in bad weather conditions. It is combined with the necessity of correct judgement of the situation and correct reaction.

The structure of a driver's working place in a locomotive cab, from the point of view of its design, is subordinate to general rules of ergonomics that apply to:

i. the spatial structure;
ii. the arrangement of elements of equipment in the working field;
iii. material working environment.

The spatial structure of the place should be adjusted to:

i. dimensions of a user and his dominating body position at work;
ii. the scope of movements, visionary field;
iii. the planned equipment of the working place, providing optimum comfort and safety;
iv. projected scope of user's manipulative movements.

The ergonomics of the control desk and arrangement of steering devices and the seat, which are adjusted to the anthropometric dimensions and visionary field of a man, decide about the optimization of spatial structure of a cab.

The arrangement of elements of equipment in a cab must be subordinate to the visibility of a track and of signaling and steering devices. The angle of windscreen inclination is very important. It has to protect the driver against dazzling and disable mirror reflections while driving at night. Signaling and steering elements in a locomotive cab are arranged to make them visible from the driver's seat in natural and artificial light. Ergonomic arrangement of signaling and steering elements requires:

i. maintaining the agreement of directions of movements of steering elements with the direction of movements of indication and control elements;
ii. positioning of seldom used steering elements outside the control desk;
iii. positioning of damage control steering elements outside the control desk;
iv. grouping signaling and steering elements in the control desk in the following function blocks: basic signaling block, diagnostic block and preparatory block.

Quick and precise reception of information, being the prerequisite for safe driving, is not only dependent on the arrangement of signaling devices, but also on their shapes and types. The arrangement of signaling devices is dependent on the visual field of a man, in which we can distinguish three spheres. Sphere A includes elements requiring easy and quick perception or continuous observation; in sphere B there are elements requiring periodic control; sphere C includes elements that do not have principal meaning for the steering process.

Strict adjustment to the requirements of ergonomics has particular meaning in the positioning of the seat. It should provide the possibility of vertical and horizontal adjustment, and above all comfort (according to anthropometric requirements), that would lessen the fatigue connected with train driving.

Material work environment in the cab is created by physical and chemical factors like noise, vibrations, microclimate, lighting, air pollution, and electromagnetic field.

It is very important to keep the values of the factors above according to the requirements of appropriate norms. The values should be controlled during the usage of a vehicle and corrected if necessary.

3. NEW TENDENCIES IN ERGONOMIC SHAPING OF A WORKING PLACE

In rail vehicles designed today there are various technological developments introduced according to individual needs of long-

distance or local transport. The projects must take into account some needs connected with higher speeds of driving. From the point of view of the man–machine system, the most interesting are solutions concerning the driver's cab, the spatial structure and equipment of which provide the driver with the possibility of performing his operator actions in comfort. Modern ergonomic solutions of the cab concern first of all the spatial structure, features of signaling, and steering devices. A characteristic feature of those solutions is assisting the driver with computer systems. While driving a vehicle, the man can always evaluate and identify received information on the basis of the data he has, and then make correct decision. The higher the speed of a vehicle, the faster is the sequence of the processes described, as there is less time for their realization. A higher degree of automation of the working post simplifies the work process itself and lowers the participation of cognitive functions in completing the tasks. This is why the introduction of computer systems assisting the driver requires research and the forming of ergonomics rules that would provide:

i. increasing the speed and capacity of transmitted information as a basis for safety;

ii. early warning of the driver about approaching signals and limitations of speed;

iii. increasing the feeling of safety connected with the impossibility of passing a semaphore with a STOP signal on it;

iv. a possibility of objective evaluation of a situation in emergency situations (lessening the number of decisions made by the driver in probable situations).

One of the first systems meeting those requirements was EBICAB 700, which is used in Sweden, Portugal, France, Bulgaria, Australia, and Taiwan. It is built on the basis of a transmission device installed between rails, which sends encoded signals to a receiver on a locomotive. The new generation of the system, called EBICAB 900, which has bigger transmitting capability, enables the speed to be increased above 140 km/h. The system registers all events. It also meets the demands of speed increase and the volume of information transmission. The transmitters, located between rails, send data from signaler devices and traffic management systems as constant information to the antenna of a locomotive. The transmission lasts for several milliseconds. Quick data transmission concerning approaching signals and speed limitations enables early warning of the driver and gives him enough time for braking. The devices in the cab control the speed all the time (calculating braking curves) and immediately reacting to any deviations from the pre-programmed drive.

The modernization of the method of data transmission and reception results in some changes in the spatial structure and the whole equipment of a cab. The control desk of the new type of cab usually separates information and executive areas. The executive elements are usually placed in two consoles encircling the driver's seat. In ABB Henschel–"Eco 2000" locomotives, the console is situated inside the cab and can be half-opened to let the driver leave his seat in emergency situations. Opening of the console activates emergency braking. In the new generation of locomotives the steering of propulsion and braking is done by a deck computer, which also has control and diagnostic functions. The usage of 32-bit computers enables extension of diagnostic function (location of damage, operating instructions, data storage) and additional functions such as the energy saving drive program

or drawing the actual course characteristics. As a rule, deck microcomputers enable access to their data from a PC computer and send the data by radio to a dispatcher.

In the control desk there should be the following steering devices: a pulling power control lever, a set speed control lever, a supplementary air brake valve lever, and an automatic air brake valve lever. Other indicator devices include, first of all, a deck monitor, voltage indicators, current, pulling and braking power indicators, and indicators of pressure in the braking system. An example of an ergonomic solution is the control desk in the locomotive ABB-Re 460, from the group of ABB locomotives (in cooperation with SLM). The control desk consists of two consoles, one of which can be drawn aside. An example of a different solution is the control desk in the : "Euro-Sprinter" electric locomotive, jointly manufactured by Siemens and Krauss-Maffei. Both control desks represent the same generation of technological solutions. They perform the same functions, but differ in the ergonomic conception of the shape, equipment, and arrangement of signaling and steering devices, and finally, in color. For example, in the central and left parts of the desk we have some necessary indicators of the brake system control. As advised by the rules of ergonomics, the indicators are round. The levers of the air and regeneration brakes, which are used most often in driving, are separated from each other but installed on one panel. In the control desk there are two different colored areas: a white panel which contains the black steering devices, and the area of the upper part of the desk, containing indication devices, which is black. All the instruments have white scales, marks and pointers to simplify reading. The monitor of the deck computer is located on the right side of the desk in the central field of vision. Less important devices, and those indication and steering devices that are not used so often, are arranged in the left- and right-side fields of vision. Equipping the cab with a computer creates the necessity of taking into account many rules of ergonomics concerned with human–computer work, which is a new problem in locomotive cab design. It is worth mentioning the "Thalys" locomotive, which was designed cooperatively by French, Belgian, German, and Dutch firms, is an example of the usage of leading solutions in the construction of drivers' cabs.

4. AN EXAMPLE SOLUTION OF A CONTROL DESK IN THE DRIVER'S CAB

All new solutions of cabs in electric locomotives are usually widely tested by their manufacturers. It is the basis for the introduction of many improvements and modernization in successive versions. The research involves not only the technical efficiency of a vehicle, but also the need to increase driver's reliability and optimization of his physical and mental load. First of all, it is very important to reduce a load that results in "sharp" fatigue; this is the reaction to short bursts of effort, which causes a serious loss of efficiency and which ceases after some rest. Chronic fatigue is also important; it appears as a sum of other types of fatigue, resulting from incorrect adjustment of the cab and material working environment to a man. Negative reactions include both physical and mental fatigue. Physical fatigue affects efficiency of movement, whereas mental fatigue manifests as incorrect reaction during the reception and processing of information.

The monotony of the work process leads to boredom and sleepiness. Optimization of the driver's physical and mental loads

is a prerequisite for the optimization of the whole man–machine–environment system; it becomes a priority in the search for new ergonomic solutions which will guarantee working comfort for users from a range of populations, driving vehicles in a range of climates. Work in this direction is being conducted by European Rail Traffic Management System (ERTMS). They use the experience of European Train Control System (ETCS) in control-desk design, assuming the use of only one monitor in a cab, the operation of which should be simple for every European driver. That is why ERTMS have adopted a uniform method of presenting basic parameters to a driver, such as current maximum speed, target speed (the closest speed limit), and distance to that target. ERTMS takes into account differences in signaling in the countries that are supposed to cooperate with the ECTS system.

The control desk suggested by ETCS was created on the basis of research and experiments in simulators, which were conducted by specialists in vehicle driving and ergonomics. Having gathered usage and experimental data, they suggested the ultimate solution, which reflects drivers' and experts' opinions about the location of fields in the monitor. The monitor is divided, according to requirements of ergonomics, into six fields:

A. brake details
B. speed control
C. speed maintenance
D. planning
E. monitoring
F. driver input

The fields represent main tasks in driving a train:

A.   Brake details:
the field displays time to brake activation, distance to an area of lowered speed, projected speed in the target in relation to applied braking intensity. From the point of view of ergonomics, the most important parameter was the choice of colors: green denotes that the train will stop before the target, gray denotes that the train will stop at the target, and orange denotes that the train will stop past the target.

B.   Speed control:
the field shows basic parameters concerning the speed when the train is moving: the speedometer shows current and target speeds and the change of colors of the information on display when the train starts braking. Colors are used to attract the driver's attention. In the case of exceeding the speed limit, the driver is warned by the change of color of the speedometer's pointer and its border to orange. Further exceeding the limit causes the color to change to red and simultaneously starts the braking.

C.   Speed maintenance:
data is displayed in graphics format. It is possible to display timetables and information about suggested and target speeds.

D.   Planning, and E. Monitoring:
these fields display information about the profile of the track and all information about traffic safety.

F.   Driver input:
in this field there are some continuously lit touch keys, enabling the driver to switch the screen to maneuver position, cancel some system warnings, input train and driver data, and conduct diagnostic tests.

From the point of view of ergonomics, a correct solution in the field of speed control is a device that does complex calculations and reduces the outcome to one figure and one gray arrow, and gives the driver ready-made information — "slow down" or "accelerate". The technology reduces the mental load, as it presents information in its most coherent form. The arrow pointing upwards shows the need to increase speed, while the arrow pointing downwards shows the need to decrease speed (the movement of the arrow matches the expected result). The length of the arrow expresses the value of difference between actual and suggested speeds.

In the "planning" field the designers facilitated projecting the progress of driving by:
i.   confronting various data on the route chart;
ii. separating information concerning safety (maximum speed limit) from current information.

The driver can decide on the size of a field. Touching the upper part of the screen extends the scale, touching the lower part contracts it. The whole screen in the driver's cab is sensitive to touch and in this way every information sphere has its equivalent touch key that enables the driver to obtain other pictures. It is also used for accepting the information received. During the driving period, the pictures in the fields "brake details", "speed control", and "driver input" are always displayed, and the driver can decide himself what type of information should be displayed at any particular moment. There is also a possibility of supplying additional information in the form of electronically generated sounds. The monitor is able to transmit a wide scope of such sounds, including artificial speech in several languages, which lessens the danger of confusing these sounds with others used in the cab.

The following are the rules for color usage:
i.   white is used as a basic color that does not require special attention or action;
ii.   white changes into yellow in order to signal the need for taking an action in the immediate future;
iii. a higher degree of urgency is indicated by orange, denoting the necessity of immediate execution of an action, otherwise the system intervenes automatically;
iv.   red indicates dangerous situations, and the system intervenes automatically.

The presented solutions of the control desk in the driver's cab prove that the scope of ergonomic actions includes the need for the unification of projects to make them comply with the demands of European users.

5. SUMMARY

This article shows the main directions of ergonomics in the cabs of electric locomotives. It also presents some results of leading European research; it underlines the specific meaning of ergonomics in rail transport, in which the integral part of the optimization of the driver–locomotive–environment system is the safety of passengers. The International Union of Railways is the coordinator of unification efforts (including ergonomics), which currently boasts a membership of 140 companies in 82 countries: 59 active members (the railways of all Europe, North Africa, the Middle East, India, and East Japan), 49 associate members (railways in Asia, Africa, America, and Australia) and 32 affiliate members (companies providing public transport, sleeping car and railway catering services, etc.).

REFERENCES:

GRABAREK, I., 1999a, Computer simulation of dynamic and ergonomic properties of the man-railway vehicle system. Proceedings of the International Conference on Computer-Aided Ergonomics and Safety. Barcelona, 19–21 May.

GRABAREK, I., 1999b, Za≥o¿enia metody rangowania czynników kryterialnych ergonomicznej oceny stanowisk pracy w środkach transportu (Method assumptions to value ergonomic factors in means of transports) (in Polish). Prace Wydzia≥u Transportu Politechniki Warszawskiej.

Official Exhibitors' Catalogue, 1998, *3rd World Congress on High-Speed Rail,* Berlin, 28–30 October1998.

VORDEREGGER, Ir., J., 1997, Ergonomics moulds Europe: wide cab display. *Railway Gazette International*, 12, 859–61.

WOLFRAM, T., 1995, Nowoczesne lokomotywy elektryczne du_ych mocy kolei europejskich (The modern European electric locomotives) (in Polish). *Technika Transportu Szynowego,* 8–9, 14–37.

UIC 651 OR, 1994, Layout of driver's cabs in locomotives, railcars, multiple unit trains and driving trailers.

Systems Modeling. Physical–Control– Information Approach to Decompose Systems for Modeling: a Warehouse Analysis Case Study

D. A. Bodner

School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, GA 30332–0205, USA

H. A. Ruff and S. Narayanan

Department of Biomedical and Human Factors Engineering, Wright State University, Dayton, OH 45435, USA

M. Gopalakrishnan

School of Management, Arizona State University West, Phoenix, AZ, 85069–7100, USA

1. INTRODUCTION

A model of a system is a representation, or abstraction, of the real world being studied. System models can be used to gain an understanding of real-world complexities and can be the basis for development of support systems, such as decision aids, user interfaces, and training systems. Models can help predict performance of a system and can aid in identifying the information needed for decision-making processes concerned with system operation and design. Models are critical in the system design phase, and often valuable for system management. Models can be either descriptive (e.g. simulations that predict values of performance measures for a given input specification) or prescriptive (e.g. linear programming methods that specify the values of the decision variables that provide the normative solution for a given problem).

Irrespective of the modeling methods adopted, one of the first steps an analyst must perform is to describe the system in a manner amenable to analysis. This description leads to the abstraction of the real-world problem through the use of a modeling framework. The model results can then facilitate informed decision-making. Techniques such as Entity–Relationship diagrams help an analyst in describing a system (Chen 1983). While these techniques are generic and have been found useful in applications such as database development, they are at a very high level and are difficult to operationalize for developing dynamic system models.

For modeling dynamic and complex real-world systems, a physical/control/information (PCI) decomposition provides analysts with a better means for describing system characteristics and along with traditional diagramming techniques such as flow charts, can be powerful in identifying areas of improvements during analysis. This chapter provides an overview of the PCI approach and illustrates its application in analyzing a real-world warehouse system.

2. THE PCI APPROACH

The PCI approach is based on the premise that most complex dynamic systems have three major components: a physical part, a control part, and an information part. The physical part is the actual visible manifestation of the entities in the system. The control part embodies decision-making processes that drive the physical entities in the system. The information aspect represents data or parameters used by the physical entities or control processes during the system execution. The physical-control decomposition method has been in vogue since the early control theoretic approaches, but adding the information dimension has gained popularity with the increasing interest in the object-oriented simulation community. The PCI approach has been coupled with object-oriented programming concepts in developing reusable simulation architectures for various domains, including airbase logistics (Narayanan *et al.* 1997) and manufacturing (Narayanan *et al.* 1998; Pratt *et al.* 1994).

The PCI decomposition in systems modeling is useful because it readily separates the decision-making aspects of the system from the physical and information dimensions in analyzing complex time-varying behaviors exhibited in most real-world systems. The separation of control processes, when implemented in a computer simulation, is particularly useful in conducting what-if analysis. To test alternate control policies, for example, only the control processes need to be modified. Similarly, the explicit identification of information flow and content facilitate the development of support systems for informed decision-making. Good decision-making, in turn, is fundamentally important to efficient and effective system performance. System performance, in turn, relates to cost and customer satisfaction in most industries.

The PCI approach, along with traditional diagramming techniques, can be powerful in describing an operational system and can be used for multiple purposes. These include: (1) describing and documenting current procedures (needed for quality certification), (2) identifying the critical information needs (to improve decision-making), and (3) specifying the role of decision aids and other support systems (to aid human operators). These purposes are of great benefit to the analysis of industrial systems (e.g. factories, warehouses, distribution systems). These types of systems are complex, with a variety of operations that must be performed. They require extensive decision-making in design and operation, and increasingly information availability and quality certification are becoming major issues. The PCI approach of decomposing a warehouse system and its utility in describing current procedures and support systems needed are outlined below.

3. CASE STUDY: A WAREHOUSE SYSTEMS ANALYSIS

3.1. Elements of a Generic Warehouse System

In a broad sense, a warehouse can be viewed as a system that receives shipped material from a surrounding environment, and then ships that material, or some transformation of that material, out into the surrounding environment. During the time the material is in the warehouse, it is handled in different ways. Material may come into the warehouse in bulk units, be stored in bulk units, and be shipped out in bulk units. Alternately, a

warehouse may receive bulk units, break down these units, package the smaller units, store the smaller units, pick the smaller units, and ship the smaller units. Similarly, the warehouse may receive smaller units, package several different units together, store these consolidated units, and then ship out the consolidated units. Other combinations exist, and a warehouse may perform any number of these activities simultaneously (Bartholdi and Hackman 1998). Numerous performance measures are used in warehouse systems analysis. These include the following:

- *Cycle time.* The time needed to perform an operation or a set of activities (e.g. unload a truck).
- *Lead time.* The time needed to satisfy a customer order after it has been received.
- *Fill rate.* The percentage of an order that has been filled when shipped.
- *Error rate.* The percentage of items incorrectly included or excluded in an order.
- *Utilization.* The percentage of time spent by a resource performing operations. A resource can be either equipment or labor. Other ways that resources may spend time include being idle, being in a failed state, being on break, or undergoing maintenance.

We recently modeled a real-world warehouse using the PCI approach and developed a transaction documentation system to help document current procedures for quality certification purposes and to increase warehouse efficiency, aesthetics, and effectiveness. The case study is described below.

3.2. A Specific Example

The specific case study is referred to as Warehouse ABC. Warehouse ABC is a complex, dynamic warehouse, which handles products for a large automotive manufacturing company, and sells separate automotive products to its own customers. Warehouse ABC receives shipments from suppliers in trucks in the form of individual items, or as pallets, which are cases stacked on a wooden skid, and are moved by a forklift or pallet-jack. These shipments are manually transported from the dock to the appropriate storage area. The materials are stored for some period of time until they are removed from storage to be transformed into the customer-defined product, or transported to the dock to be shipped out. An initial analysis based on real-world observation reveals three broad scenarios that describe the warehouse operations. These scenarios are: (1) unloading and storage of incoming material, (2) retrieval and shipping of outgoing material, and (3) value added transformations of material inside the warehouse (e.g. kitting, assembly, repackaging, etc.).

3.3. PCI Decomposition

The analysis methodology consists of two major components: (1) development of a warehouse model using the PCI decomposition and (2) analysis and development of decision support systems using the model. The model is represented using flowcharts of the three scenarios described above.

The modeling goals are twofold. The first goal is to identify areas for improvement, both for short-term operations and for longer-term planning. The second goal is to catalog the existing set of operations, and any proposed changes to them, in anticipation of the system documentation required for ISO certification (i.e. quality certification needed in many industries).

The PCI decomposition for the scenarios consists of abstracting the warehouse system into physical, control, and information elements, and tying these to the necessary performance measures. These elements are united by two significant concepts: the flow of material in the system and the flow of information in the system.

3.3.1 Physical activities

The physical element of decomposition involves physical activities, which consist of the operations performed in the warehouse. These include, but are not limited to, loading items, unloading items, moving items, inspecting items, assembling or repackaging items, storing items and retrieving items. Physical activities require the use of limited resources, including space, time, equipment and labor. Activities also require procedures to be performed effectively and efficiently.

3.3.2 Control (decision-making)

The control element of decomposition captures the decision-making of the warehouse operators performing the physical activities. During the operation of a warehouse, decisions have a major impact on performance. A decision may consist of selecting a procedure to perform operations. It may also consist of prescribing a sequence of operations to perform. Making "good decisions" is important not only for operational efficiency, but also for quality assurance and customer satisfaction.

3.3.3 Information

System abstraction of information elements reveals the data used to describe the warehouse system. This data includes the system state information, as well as requirements, procedures and structure information. Good decisions depend heavily on having the right information at the right time. Information consists of a variety of data and exists in a number of formats. For example, information consists of procedures used to perform an activity. It also consists of the set of orders to be filled at any point in time, or the set of people and equipment available to perform an activity. In terms of format, information can be stored in a computer database or as a set of instructions in paper form. An *information system* is a well-defined set of procedures and data that provides information for decisions. Much of it may be computerized, although computerization of all information is not necessary. It must be designed to supply the right information at the right time to support effective decision-making.

3.4. Flow Charts

The PCI decomposition and event flow can be represented through the use of flow charts. The convention used in the flow charts is shown in figure 1.

The cylindrical shape represents information entities, the diamond represents decision-making control, and the rectangle represents physical activities. The arrows represent flow of events from one entity to the next, however, the represented flow does not hold true for all real-world situations within the three scenarios.

Figure 2 shows a portion of the flow chart created for scenario 1, in which material comes into the warehouse on a truck. There is a vertical flow from top to bottom, representing the order of the physical activities and the decision-making that precedes those actions, or is required of the operator because of those
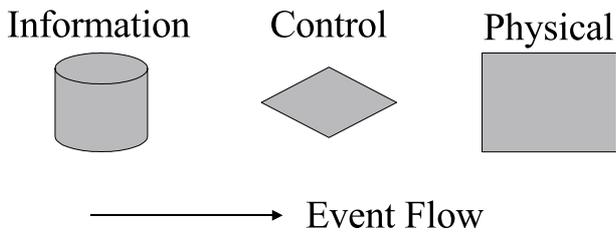
Information   Control   Physical

Event Flow

Figure 1. Notations used in flow charts.

actions or some other action or requirement. For example, the human and equipment resources are allocated before the truck can be unloaded. On the other hand, a dock can be allocated to a truck only after the truck has arrived. The flow chart also shows a horizontal flow from left to right, representing the flow of information elements of the PCI decomposition. For example, information on dock status is required to make the decision of which dock the truck will enter. Once that decision is made and the truck docks, then the dock status information is updated to reflect this. This decision will affect future dock-related decisions, and will in turn change the system state as well as the information element of the PCI decomposition. This is represented in the flow chart as a cylinder on the left representing utilized information and a cylinder on the right representing the updated information that will be utilized later. The arrows show this transition. Flow charts are similarly specified for all the operations in each scenario.

3.5. Application

The flow charts provide the basis for specific analyses and for the development of a computer-based system to document and analyze warehouse operations (i.e. a transaction documentation system). Through analysis and identification of decision points and the information and resources needed, recommendations for operational improvement were identified. These include procedures to be used for staging outgoing orders prior to shipment, procedures for scheduling weekly workload for the warehouse, and tracking of orders that need to be filled to enable effective expediting of orders that are late. Each of these recommendations focuses on improving decisions on the shop floor to improve performance in terms of efficiency (e.g. shorter operation cycle times) or effectiveness (e.g. increased number of orders shipped on time).

At the same time, the PCI approach enabled the design of a transaction documentation system for the warehouse operations. The purpose of this application is to aid an operator in tracking and documenting warehouse activities. While this is a necessary step in quality certification, it also provides a database of historical information that can be used for further analysis for improvement. Documentation of procedures is a time-consuming and tedious task. It is rarely done (or done well) by operators. A computer-based application should be designed to automate the tedious tasks associated with data entry and to reduce the time needed for data management (thus freeing the operator for other tasks in the warehouse). The application is developed using Microsoft
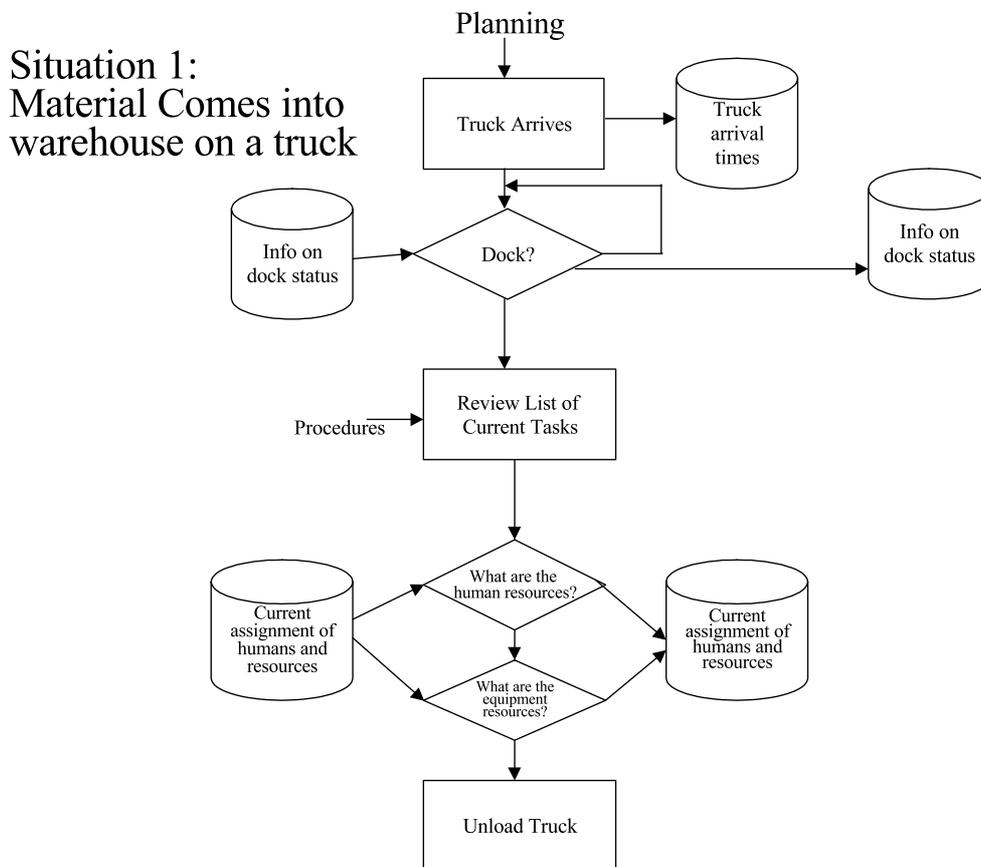

Figure 2. Flow chart for scenario 1.
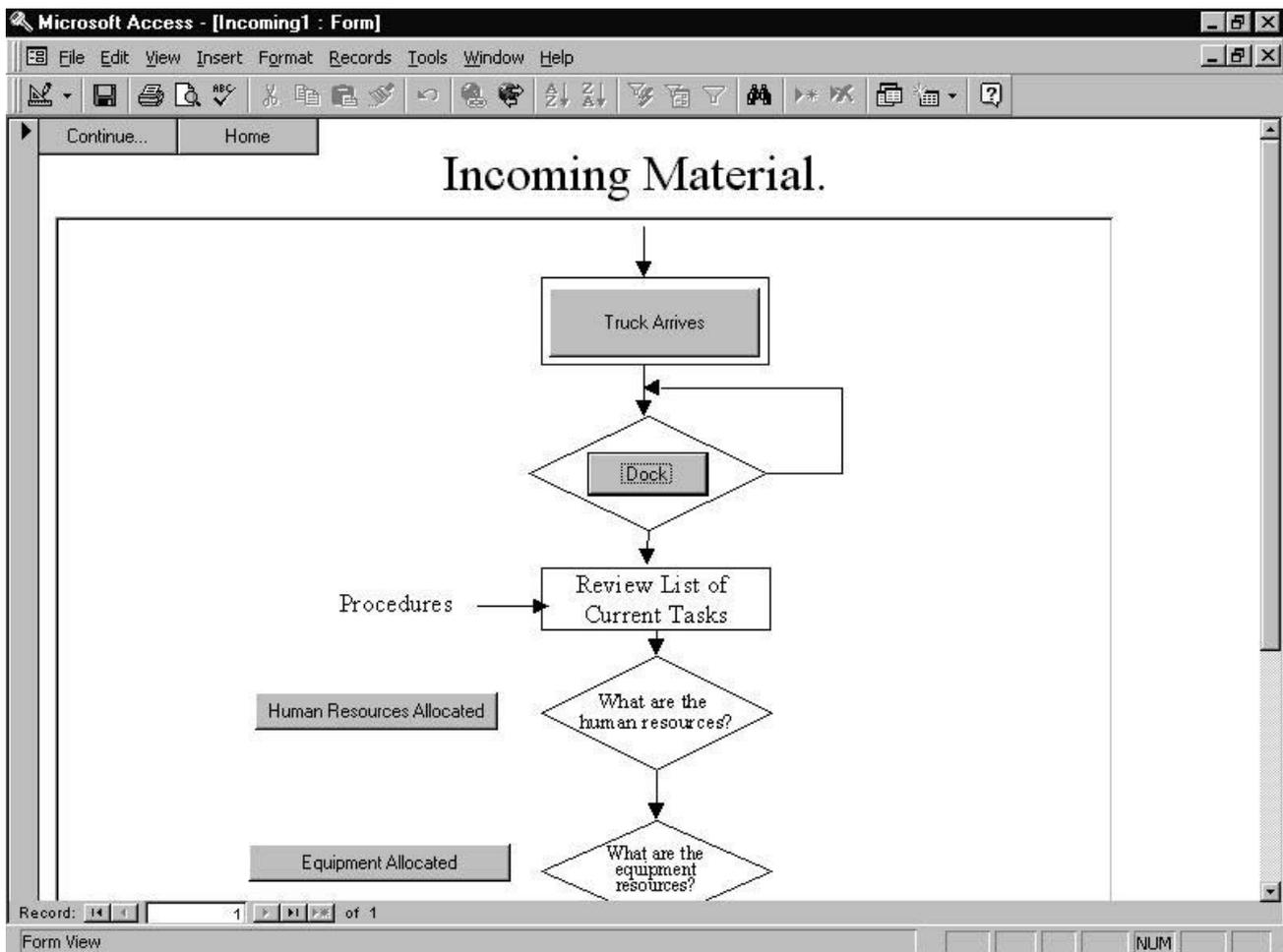
Figure 3. Screen-shot of the transaction documentation system.

Access 97, due to its database capabilities, availability, user-interface design tools, and ease of programming/re-programming. Figure 3 depicts a screen shot from the system.

The transaction documentation system allows operators to enter data into the database by following a user interface based on the flowcharts for the three scenarios. Figure 2 shows a portion of the application environment designed based on the flow chart of figure 1. Operators enter the program, select the scenario for which they will be entering data, and then point and click on the appropriate button embedded in the flow chart. For example, in figure 2, the button 'Truck Arrives' is clicked on when a truck arrives at the warehouse. The operator enters data such as the trucking company, the type of customer, the truck arrival time, and any other significant notes. The system provides a transaction number for tracking purposes.

The database is then updated, and the system analyst can review this data at a later time. For example, performance measures, such as cycle time and fill rate, can be calculated by reviewing the database. The database can also be queried to compile the procedures necessary for ISO certification and to identify potential improvements to effectiveness and efficiency of system operations. Future design improvements can be made through analyzing the databases, and interviewing warehouse operators. Regular use and analysis of the application can promote ongoing systems management, and can be a useful tool in training, decision-making, and planning for future company growth.

REFERENCES

BARTHOLDI, J.J., and HACKMAN, S.T., *Warehouse Science, release 0.01b,* May 22, 1998; revised July 2, 1998 (an unpublished work at the School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, GA).

CHEN, P. P. (1983). ER: a historical perspective and future directions. Proceedings of 3rd International Conference on Entity-Relationship Approach, pp. 71–8.

NARAYANAN, S., BODNER, D.A., SREEKANTH, U., GOVINDARAJ, T., MCGINNIS, L.F., and MITCHELL, C.M., 1998, Research in object-oriented manufacturing simulations: an assessment of the state of the art. *IIE Transactions*, 30(9), 795–810.

NARAYANAN, S., SCHNEIDER, N., PATEL, C., REDDY, N., CARRICO, T., and DIPASQUALE, J., 1997, An object-based architecture for developing interactive simulations using Java. *Simulation*, 69(3), 153–71.

PRATT, D. B., FARRINGTON, P. A., BASNET, C. B., BHUSKUTE, H. C., KAMATH, MANJUNATH, MIZE, J. H., 1994, Separation of physical, information, and control elements for facilitating reusability in simulation modeling, International Journal in Computer Simulation, 4(3), 327–42.

Usability Evaluation

L. Fu[†] and K. Schmidt[‡]

†Metaphase Technology Division, Structural Dynamics Research Corporation, 4233 Lexington Avenue N, Suite 3290, Arden Hills, MN 55126, USA

‡Cardinal Consulting, 2191 Second Street, White Bear Lake, MN 55110, USA

1. INTRODUCTION

The ISO 9241-11 definition of usability is "The effectiveness, efficiency and satisfaction with which specified users achieve specified goals in particular environments." In general, two factors should be included in a complete evaluation of human–computer interaction: the user and the computer system. Whitefield *et al.* (1991) classify the different usability evaluation methods according to how the user and the computer components of the system to be evaluated are presented in the evaluation process (Figure 1). These two components of the system can be real or representational.

For the component "users," "real" means actual users or approximations of them, such as students. The "representational user" means descriptions or models of the user. For the component "computer," "real" means the physical presence of the computer and software system or an approximation of it. The system, prototypes and simulations all count as a real computer presence in the usability evaluation. "Representational computer" refers to the notational models or the user's mental representation of the computer. In questionnaires and interviews or during code inspection, users work with their symbolic mental representations of the computer. Based on this classification, the evaluation methods are divided into four categories: analytic methods, specialist reports, user reports and observational methods.

2. ANALYTIC METHODS

When analytic usability evaluation methods are used, both the computer and user are representational. Typical methods include the Goals, Operators, Methods, Selections model (GOMS); the Task Action Grammar (TAG) method; the Systems Analysis of Integrated Networks of Tasks (SAINT); and Queuing Network

Figure 1. Classes of usability evaluation methods (Whitefield *et al.* 1991).

Models. All the methods presented here are cognitive modeling methods that can be used to make predictions about the usability of products.

The GOMS model analyzes cognitive tasks in terms of operators and methods for achieving the goals of the tasks and rules for choosing among different possible methods. The goal defines the end-state that the user is trying to achieve (e.g. saving a changed file). The operators are the basic actions available to the user for performing a task. This includes motor, perceptual, or cognitive primitives (e.g. click a mouse button, search a menu item, remember a file name). Methods are sequences of operators, or procedures, for accomplishing goal (e.g. saving a file is accomplished by moving the cursor to the File menu, clicking the left mouse button, moving the cursor to the Save menu item, and clicking the left mouse button). Selection rules are invoked when there is a choice of methods (e.g. the user could also save the file by clicking the save button on the toolbar). The GOMS model can be used to describe the methods that a user employs to carry out a set of tasks, and to estimate the time it takes to complete those tasks.

The Natural GOMS Language (NGOMSL) was developed to enable the task analysis using a GOMS-like model to be more specific. NGOMSL is an attempt to define a language that will allow GOMS models to be written down with a high degree of precision. The purpose of the language is to describe idealized, error-free behavior in terms of those four concepts (goals, operators, methods, selection rules), and to predict the time needed to execute given tasks. NGOMSL can also be used to estimate minimal learning time and to quantify different aspects of mental workload.

Task Action Grammar is a method of describing interfaces in terms of the linguistic structure of the commands. TAG requires an existing interface specification for evaluation and has little to say about non-expert performance. It is focused mainly on revealing internal inconsistency existing in the interface mechanisms rather than on evaluating user performance.

The Systems Analysis of Integrated Networks of Tasks models human interaction with the task environment as a sequence of tasks. Alternative sequences to accomplish a goal may exist and form a network. Queuing network models integrates the considerations of queuing theory and discrete network models, and supports modeling a broader range of possible mental structures that can be subjected to empirical testing. The SAINT model allows processes on the same path to be active at the same time. It supports modeling of both alternative and concurrent processes as well as sequential and parallel processes.

The principal advantages of analytic methods are that they can be used early in development (before any real computers exist), require few resources to apply, and are potentially fast. The disadvantages are that suitable modeling techniques are still under investigation and development and, consequently, the validity and reliability of the methods are uncertain.

3. SPECIALIST REPORTS

This category of usability evaluation involves one or more people who are not real users assessing a version of the real computer. The evaluators could be human factors specialists evaluating the design of a prototype version using relevant handbooks, guidelines, or their own experiences. Typical evaluation methods

are checklists, guidelines, walkthroughs or heuristic evaluations by specialists.

There are many user interface guidelines available in the literature. Some guidelines are used for a specific platform or software, such as those from Microsoft, Apple, OSF/Motif Style Guide. There are also many general guidelines designed by individual usability engineers. ISO 9241 also has guidelines for specifying and measuring usability.

The advantages of using guidelines are that they are relatively fast, use few resources, provide an integrated view and can address a wide range of behavior. The disadvantages of using guidelines are that their reliability will vary between specialists and, because their assessments are inevitably somewhat subjective, their reports are likely to be incomplete, biased and difficult to validate.

Since user interface guidelines have the disadvantage of being hard to follow (some of them have > 1000 rules), specialists proposed another method: the heuristic evaluation, which is easier to follow. Heuristic evaluation involves having a small set of evaluators examine the interface and judge its compliance with recognized usability principles such as those defined by Nielsen (1993), provide simple and natural dialogue; speak the user's language; minimize user memory load; be consist; provide feedback; clearly marked exits; provide shortcuts; provide good error messages; prevent errors; provide help; and provide documentation. Three more heuristics were extended and validated later: respect the user and her/his skills; provide a pleasurable experience with the system; and support quality work.

The advantages of heuristic evaluation include: low cost in comparison with other methods; that it is intuitive to perform; that no advance planning is required because the evaluations can be conducted by team members in isolation; and that it is suitable for use early in the development process. But because it does not involve real users, it has some degree of subjectivity. Heuristic evaluation often focuses on problems rather than solutions. Also, it requires several usability specialists to perform the evaluation, which limits its application power.

Cognitive walkthrough provides a method of analyzing designs in terms of exploratory learning. It can be applied to designs for systems that will be used by people without any prior training, perhaps in a "walk-up-and-use manner." Analysis by cognitive walkthrough involves simulating the way users explore and become familiar with interactive systems. The evaluator starts with a rough plan of what s/he wants to achieve; for example, a task to be performed. By simulating the process of a user's interaction, the evaluator asks the following questions: Will the correct action be sufficiently evident to the user? Will the user connect the correct action's description with what he or she is trying to do? Will the user interpret the system's response to the chosen action correctly; that is, will the user know if he or she has made a right or a wrong choice? The problems found by this method are focused mainly on these three areas.

The cognitive walkthrough method can help define the user's goals and assumptions and can be used by software developers. The disadvantage is that it requires task definition methodology, which is tedious for evaluators and which may miss general and recurring problems.

4. USER REPORTS

User reports involve real users and their mental representations of computers. Methods typically include the use of interviews, focus groups, and questionnaires. The methods are used to obtain data or opinions from the users on some aspect of the system.

Interviews provide a particularly rapid and congenial way of gathering data. They require less prior planning and preparation than other formal methods of analysis. Interviews may be structured or unstructured. This method can also be used in the beginning of the design process to conduct user studies. Because interview data are subjective, they are open to a number of inaccuracies.

Focus groups provide another way to assess users' needs and feelings. In a focus group, about six to nine users are brought together to discuss new concepts and identify issues. The session is completed in about two hours. Each group is run by a moderator who is responsible for maintaining the focus of the group on the issues of interest. The moderator usually needs to prepare a plan outlining issues to be discussed. Focus groups often bring out users' spontaneous reactions and ideas through the interaction between the participants.

In general, a user satisfaction questionnaire measures an individual's feeling about some aspect of a computer system. If a system meets users' needs in an easy and efficient manner, it will be used and user productivity will increase. A user satisfaction questionnaire indirectly provides information that may be used to improve the system. Ideally, a user satisfaction measure provides evaluations of the hardware components, system capabilities, screen design, the quality of the information flow, instruction manuals and technical support.

There are several questionnaires available, such as User Satisfaction Scale, Computer User Satisfaction Inventory (CUSI), Questionnaire for User Interaction Satisfaction (QUIS), Software Usability Measurement Inventory (SUMI), Purdue Usability Testing Questionnaire (PUTQ), After Scenario Questionnaire (ASQ), Post-Study System Usability Questionnaire (PSSUQ), and Computer System Usability Questionnaire (CSUQ).

The User Satisfaction Scale is also referred as the Pearson–Bailey scale, which defines satisfaction on the basis of positive and negative attitudes toward different components of the computer system. The questionnaire has 39 items that comprehensively cover the computer system, such as accuracy, error recovery, attitude of staff support, documentation and security of data. Each item is rated on six scales. Four are used for evaluation responses and include adjectives like good/bad, simple/complex, useful/useless and fast/slow. A fifth scale directly assesses the user's feelings about the item, and is anchored with satisfactory/unsatisfactory. Studies showed an acceptable reliability and validity with a modified Pearson–Bailey Scale.

The Computer User Satisfaction Inventory is a short questionnaire of 22 items. Two subscales of usability were established, called Affect (the degree to which users like the computer system) and Competence (the degree to which users felt supported by the computer system). These subscales were developed by the analysis of intercorrelations of responses to individual questions in a large initial item pool. This item pool was gathered from literature searches and discussion with end users about their reactions when carrying out their normal tasks

on their usual system. The range of systems sampled was large and heterogeneous. CUSI has an overall reliability of 0.94, with two separate scales showing individual reliability of 0.91 for Affect, and 0.88 for Competence.

The Questionnaire for User Interaction Satisfaction is a commercially available questionnaire. QUIS v.5.5 consists of one introductory section, which is an overall reaction to the software scale, and four other sections, each consisting of between four and six items. These sections are screen, terminology and system information, learning, and system capabilities. The current version has two forms: a long form with 80 items, and a short form with 27 items. The long item version was reported to have a high reliability coefficient of 0.94 for the entire scale.

The Software Usability Measurement Inventory was developed on the basis of CUSI by examining the CUSI Competence scale, expanding it, and extracting further subscales. The final version has 50 items. It provides three types of measures: an overall assessment, a usability profile and an item consensus analysis. In the usability profile, five subscales are included: affect, efficiency, helpfulness, control and learnability.

PUTQ, Purdue Usability Testing Questionnaire, was developed on the basis of information processing theory. Eight human factors are identified: compatibility, consistency, flexibility, learnability, minimal action, minimal memory load, perceptual limitation and user guidance. It was reported to have a reliability of 0.77.

The ASQ, After Scenario Questionnaire, was developed especially for scenario-based usability study. Participants use a product to do a series of realistic tasks. The ASQ can then be used to assess participant satisfaction after the completion of each scenario. The three items of the questionnaire address three important aspects of user satisfaction with system usability: ease of task completion, time to complete a task, and adequacy of support information. An average reliability of 0.93 was reported. The moderate correlation between the ASQ score and scenario failure or success provides evidence of concurrent validity to the questionnaire.

Both the Post-Study System Usability Questionnaire and Computer System Usability Questionnaire are overall satisfaction questionnaires. The PSSUQ items are appropriate for a usability testing situation, and the CSUQ items are appropriate for a field study situation. Otherwise, the questionnaires are identical, and each contains 19 items. Factor analysis identified three main factors that were measured in the questionnaires, namely, system usefulness, information quality and interface quality. High reliability, validity and sensitivity of the questionnaires were reported.

Generally speaking, questionnaires provide a means to gather enough data to perform statistical analyses. Questionnaires demand very careful design to produce reliable data and to avoid the problem of misleading questions. They may also be conducted via the Internet by using e-mails and Web pages. Relatively formal techniques for data collection and analysis are available. The methods can be applied relatively quickly, and they involve real users.

5. OBSERVATIONAL METHODS

This category involves a real system and real users interacting with a real computer. The set of such methods is very large,

ranging from informal observation of a single user to full-scale experimentation with appropriate numbers of subjects and control of variables. The observation can be conducted in the real working environment or in usability laboratories.

Usability laboratories generally have two rooms, one for the experiment, the other for observation (Figure 2). There are soundproof, one-way mirrors separating the observation room from the test room. This allows the experimenters to monitor users' actions without disturbing them. In laboratories, the evaluation methods such as user performance measurement, think aloud protocol collection, and logging actual use can be used.

For user performance measurement, usability testing can be conducted to get information on user execution time, accuracy, user's satisfaction, and cognitive workload. Generally such a test involves one test administrator and one data logger. Additional observers might also be used to gather more observational data. The test administrator is responsible for controlling the test flow and interacting with the test subject. The data logger is responsible for recording the user's comments and task performance time, sometimes with data logging software. The following performance data is helpful in identifying usability problems and can be collected during the test: failure to complete the task; task and subtask time with high variability; deviation from optimal task path as perceived by usability test administrators; long "seek" time, or the time it takes the user to find the right place in the software to accomplish a task; requests for help during the task; and verbal expression of confusion, discomfort, irritation, and so forth. From those data, usability problems could be identified. If necessary, audiotaping and videotaping can also be used to record voice and image information. Some usability practitioners also do retrospective usability testing by reviewing the videotapes with usability test subjects.

In an industrial setting, usability testing can be very costly. Based on several published budgets, Nielsen (1993) estimated that for a representative, medium-sized usability test, fixed cost was US$3000 and variable cost was $1000 per test user. The pay-off ratio between the benefits and the costs changed with various numbers of test users. The highest ratio was achieved with three test users for a typical medium-sized project. For a user number over three, the pay-off ratio began to decrease. To identify 85% of the usability problems exist in a system, about five subjects are needed for the test.

Think-Aloud evaluation can also be used in a usability test. This basically involves having a test subject use the system while continuously "thinking out loud." By verbalizing their thoughts,
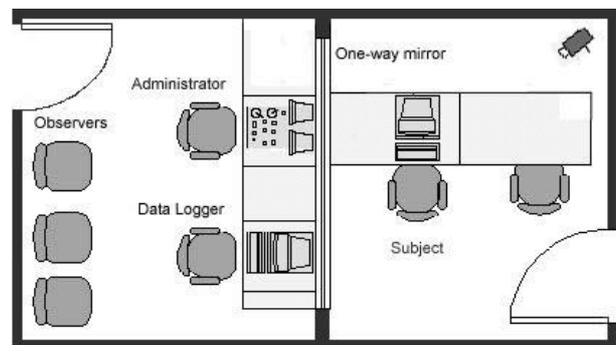
Figure 2. Usability laboratory layout sample.

test users enable an observer to understand how they view the computer system. This makes it easy to identify systems major design flaws. This method also provides process information about the human–computer interaction.

Logging actual use involves having the computer automatically collect statistics about the detailed use of the system. Normally, logging is used to collect information about field use of a system after release, but logging can also be used as a supplementary method during user testing to collect more detailed data. Typically, an interface log will contain statistics about how frequently each user has used each feature in the program and how frequently various events of interest occurs; for example, how frequently error messages are called. This method is especially useful in evaluating Web site usability.

Together, observational methods of usability evaluation have the major advantage of investigating the performance of the real system and, therefore, ought to reflect the performance more accurately than the other methods. Disadvantages of these methods are that they tend to be slow to conduct, use many resources, and require expertise in experimental design and interface issues.

6. CONCLUSION

There are many usability evaluation methods available, and each has its own advantages and disadvantages. Some scientists argue that there is no single "best" evaluation method. All of the methods examined have some disadvantages or consider only a limited number of the factors influencing an evaluation, but many of them contain useful ideas or are very appropriate for the evaluation of a specific factor. In practice, deciding which is the optimal usability evaluation method for a specific interface is a fairly complex matter, and depends on the goal of the evaluation and the resources available.

REFERENCES

NEWMAN, W.M. and LAMMING, M.G., 1995, *Interactive System Design* (Cambridge, MA: Addison-Wesley).

NIELSEN, J., 1993, *Usability Engineering* (Boston, MA: Academic Press).

RASMUSSEN, J., 1986, *Information Processing and Human–Machine Interaction: An Approach to Cognitive Engineering* (New York: Elsevier).

WHITEFIELD, A. WILSON, F. and DOWELL, J., 1991, A framework for human factors evaluation. *Behavior and Information Technology*, 10, 65–79.

Utility Analysis

N. A. Stanton and M. S. Young

Department of Psychology, University of Southampton, SO17 1BJ, UK

1. EVALUATING DESIGN

The concept of *usability* is common parlance in design, and terms such as *user friendly* and *ergonomically designed* are becoming the ad-man's by-line. Evaluation of the design of everyday objects in terms of their likely impact on human performance would seem to make good sense. Indeed, which product manufacturers would not like to make their devices easy and pleasurable to use? Methods are available which attempt to predict good or bad designs, and thus save manufacturers' time and money. Such methods are taught on every undergraduate and postgraduate program in Human Factors and Ergonomics (Sanders and McCormick 1993), and are undertaken by ergonomics groups in commercial organizations (Jordan *et al.* 1996). However, despite the proliferation of ergonomics methods in research, teaching, and industrial practice, there is little substantive empirical evidence that these methods actually work. This is most likely due to the way in which the methods have evolved and developed over this century. Interest began in the early 1900s with the examination of physical activity, looking at what people did in performing tasks and breaking these activities down into task units (Gilbreth 1911). The analysis led researchers to identify optimal task activity sequences and to prescribe performance times (Taylor 1911). Around the middle of the century, theorists began to link covert, internal, cognitive processes with observable behavior to develop theories of skilled performance (Miller *et al* 1960; Welford 1968). These theories led to the development of methods for representing both internal and external activity with the aim of assisting the development of skills and predicting performance of skilled persons in the execution of activities. From the research and development activities of many dedicated researchers spanning the past ninety years reported in substantial academic journals (e.g. *Ergonomics*; *Human Factors*) and books, modern-day techniques have begun to assume their own validity. These methods are so entrenched in the *modus operandi* of researchers and so well established in the literature that everyone, quite naturally, assumes that they are valid. Indeed, it is difficult enough to get funding to conduct research and consultancy projects without having to ask sponsors for additional funds to prove that the conclusions drawn are valid ones. This argument would be counter-productive to the initial intervention — who is going to fund the work if you have to admit that you cannot be sure that your findings have any sound basis? With this in mind, let us consider in more detail what ergonomics methods do.

2. ERGONOMICS METHODS

People use ergonomics methods with the aim to improve design by: reducing device interaction time, reducing user errors, improving user satisfaction, and generally improving device usability. Each of the methods considers different aspects of human interaction with artifacts. The methods can be broadly classified as either quantitative or qualitative. The quantitative methods predict either speed of performance (e.g. Keystroke Level Model: KLM), errors (e.g. Systematic Human Error Reduction and Prediction Approach: SHERPA), or both speed and errors (e.g. observation). The qualitative methods predict user satisfaction (e.g. questionnaires and repertory grids), device optimization (e.g. checklists, link analysis, and layout analysis), and user-device interaction (e.g. heuristics, Hierarchical Task Analysis (HTA), and interviews). Thus a fairly wide range of human activity with devices is considered using methods that have a pedigree of up to 50 years in ergonomics research.

3. COSTS AND BENEFITS OF ERGONOMICS METHODS

Claims have been made regarding the benefits of ergonomics methods in the design of devices. These normally center around cost–benefit analyses. Many researchers have made brave attempts to quantify the benefits. This usually works by calculating the cost of applying the method (in terms of person-hours, materials, etc.) and subtracting this from the estimated savings generated by the consequently improved design. The net figure is proposed as the benefit brought about by using the ergonomics methods. However, this fails to take account of the accuracy of the methods being used — potential benefits will be reduced if the method is not wholly accurate. For example, the costs associated with a major usability engineering program (Bias and Meyhew 1994) include setting up a laboratory ($20,000), conducting interviews ($4,850), task analysis ($21,900), checklists ($16,800), development of prototype 1 ($39,860), observation ($18,600), modification of prototype 1 ($5,600), observation ($18,600), modification of prototype 2 ($5,600), and final evaluation comprising questionnaires ($6,000), interviews ($7,275), and observation ($6,220). This makes the total costs of conducting the ergonomics study $171,445. The benefits arising from the improved design are expressed as profit from increased sales ($62,500), decreased training ($280,000), decreased support ($233,335), and the savings associated with making the design changes early rather than late ($16,800). This makes the total benefits $592,635. The benefits are justified on the basis that well-designed products (e.g. TVs, VCRs, software applications) are more likely to sell in greater numbers, less likely to require support, and more likely to reduce the need for training a support team, as well as reducing the need for corrections to the design. Subtracting $171,445 from $592,635 leaves $421,190. Who could fail to be impressed? However, the researchers of this work do qualify the figures by noting that the assumptions are estimated, and that they are not proven benefits as proof does not exist (Bias and Meyhew 1994). In addition, the ease of use of a device or product is not the only factor influencing its purchase; factors such as cost, functionality, aesthetics, and perceived usefulness may figure as high, or higher, in the mind of the purchaser.

Our contention with such figures is that the likely dollar benefits do not attempt to take the accuracy of the ergonomics methods into account. The application of ergonomics methods (if inaccurate) could fail to predict true problems with device design or, worse still, identify problems that do not exist. The extent to which methods do this can be calculated using Signal Detection Theory, which determines both the hit rate (the rate at which the method generates correct data, expressed as a ratio of hits to misses) and the false alarm rate (the rate at which the method generates incorrect data, expressed as a ratio of false alarms to correct rejections). Combining these two ratios will

give an overall sensitivity value for the method. We argue that the degree to which the financial benefits will be realized will depend upon the accuracy of the methods used.

4. UTILITY ANALYSIS

In order to consider utility of ergonomics methods, we must factor reliability and validity into the cost–benefit equation. A study was conducted to investigate the reliability and validity of these ergonomics methods to test their stability across analysts (reliability) and their ability to predict behavior (validity). The reliability and validity data are presented together because the two concepts are interrelated: while a method might be reliable, it might not be valid; however, if a method is not reliable it cannot be valid. The data for reliability and validity for each of the ergonomics methods, based on our studies, are presented in table 1. As shown in Table 1, five of the methods achieved acceptable levels of inter-rater reliability (i.e. the reliability statistic above 0.5: this means that there was a reasonable degree of agreement between analysts) and eight of the methods reach acceptable levels of validity (i.e. the validity statistic is above 0.5: this means that the sensitivity of the methods in discriminating between hits and false alarms was satisfactory). Only five methods (indicated with asterisk) performed at an acceptable level for both criteria. We recommend that all of the other methods are treated with caution until further studies have established their reliability and validity.

For the purpose of critiquing conventional cost–benefit wisdom, we have applied our data to the figures taken from the analysis conducted in our previous example. For this analysis we have assumed that the amount of potential benefit to be reaped from the analysis will be comparable to the effort expended on the method (based on person-hours which is expressed in terms of dollars — this is a fairly arbitrary approximation, but perhaps the fairest way of treating the data). Assessing the impact of accuracy of the methods on the potential dollar benefits is worked out using the formula:

(reliability $\times$ validity $\times$ potential benefit) $-$ cost of method $=$ net benefit

A reworking of the cost–benefit analysis for each method first assumes that the maximum potential benefit will be $592,635. From this is subtracted the cost of setting up a laboratory ($20,000), development of prototype 1 ($39,860), modification of prototype 1 ($5,600) and modification of prototype 2 ($5,600). This leaves $521,575, percentages of which are allocated as

potential dollar benefits to interviews (12%), task analysis (22%), checklists (17%), observation (43%), and questionnaires (6%). The formula for each of the methods showing credit (or debit) are illustrated below:

Interview = (0.449 x 0.446 x $62,589) - $12,125 = $409

Task Analysis = (0.226 x 0.591 x $114,746) - $21,920 = (-$6,594)

Checklists = (0.307 x 0.589 x $88,668) - $16,800 = (-$767)

Observation = (0.89 x 0.729 x $224,278) - $43,540 = $101,974

Questionnaires = (0.578 x 0.615 x $31,394) - $6,000 = $5,160

Therefore, the overall net benefit is more likely to be in the region of $100,182 (the sum of the credits and debits for the five methods above), which is less than 25% of the estimated benefit before the reliability and validity of the methods was taken into account. The utility analysis shows that some methods are likely to have greater payback than others. The take-home message of this analysis is that ergonomists need to be realistic in showing the value of their approach, and make credible claims about the payoffs. The analysis can also be used to assess the relative (financial) merits of using one method over another.

5. ENHANCING ARTIFACT DESIGN

In conclusion, we have begun to show that it is possible to collect evidence to strengthen, or weaken, the confidence that may be placed in ergonomics methods in the design of devices. The picture painted in our studies might represent a worst-case scenario, and further studies could lead to a more optimistic picture. However, in the absence of substantive evidence to the contrary, we would rather caution was exercised. There is still little reported evidence in the literature on reliability or validity of ergonomics methods. The use of reliable and valid methods would indeed enhance product design. This would be an attractive situation for ergonomists (who could use their methods with justified confidence), manufacturers (who could make their products easier to use and therefore more attractive to purchasers), and consumers (who would reap the benefits of better-designed products). Therefore, establishing the reliability and validity of ergonomics methods is to the general benefit of everyone. We suspect that, independently of academic debate, there is likely to be an increase in the use of ergonomics methods in the next century. This makes it even more important to establish the reliability and validity of methods as soon as possible.

REFERENCES

BIAS, R. G. and MEYHEW, D. J., 1994, *Cost-Justifying Usability* (Boston: Academic Press).

GILBRETH, F. B., 1911, *Motion Study* (Princeton, N.J.: Van Nostrand).

JORDAN, P. W., THOMAS, B., WEERDMEESTER, B. A. and MCCLELLAND, I. L., 1996, *Usability Evaluation in Industry* (London: Taylor & Francis).

MILLER, G. A., GALANTER, E. and PRIBRAM, K., 1960, *Plans and the Structure of Behavior* (New York: Holt).

SANDERS, M. S. and MCCORMICK, E. J., 1993, *Human Factors in Engineering and Design* (7th ed.) (Maidenhead, UK: McGraw-Hill).

TAYLOR, F. W., 1911, *Principles of Scientific Management* (New York: Harper & Row).

WELFORD, A. T., 1968, *Fundamentals of Skill* (London: Methuen).

Table 1. Reliability and validity of ergonomics methods.

Ergonomics method	Reliability	Validity
Checklists	0.307	0.587
Heuristics	0.471	0.476
Hierarchical task analysis	0.226	0.591
Interviews	0.449	0.446
Keystroke level model*	0.916	0.769
Layout analysis	0.121	0.070
Link analysis*	0.830	0.764
Observation*	0.890	0.729
Questionnaires*	0.578	0.615
Repertory grids*	0.562	0.533
Systematic human error prediction and reduction	0.392	0.614

Part 9

Work Design and Organization

Air Traffic Management

K. Corker

NASA Ames Research Center, Moffett Field, Ca. & San Jose State University, San Jose, CA 95192, USA

1. INTRODUCTION

Air traffic management (ATM) can be defined as the control of aircraft and airspace system assets to assure safe, efficient and flexible movement of aircraft and passengers through international airspace. The process of ATM is fundamentally a coordination of vehicle and airspace control. The process is mediated by three primary instrumentalities: (1) the air traffic service provider (air traffic control), (2) the airlines and air transportation service operations, and (3) the commercial flight deck and the general aviation cockpit. In addition to these primary constituents, there is a very large number of ancillary service organizations that provide maintenance, communications, weather information, scheduling, training and infrastructure to support ATM. It is a dynamic and distributed control. The system is managed in time frames that extend from seconds to months. The system operates in a physical range from surface movements at tens of miles per hour to *en route* operations with a range to 60 000 feet altitude and Mach 1.8 velocities. Air traffic management is, by the nature of its service, global in extent.

The human factors issues associated with so complex a system are similarly complex and diverse. Human factors research is performed in broad categories as follows: training and selection of personnel, design and development of controls and displays for human use, development of automation and optimization decision-support tools for the operators, and human performance contribution to safety, human–system error propagation, and human response to psycho-physiological stressors, situation awareness, vigilance, crew and team coordination, and system dynamics and stability. The research paradigms brought to bear to address the issues are similarly broad, extending from computational analyses and simulation through empirical studies in the laboratory and simulation to field studies, and large-scale institutional demographic and ethnographic studies. The following sections will describe the evolution of ATM, identify the human factors issues associated with current and future operations, and discuss the methodological implications of these issues.

2. AIR TRAFFIC MANAGEMENT OPERATIONS SCOPE AND HUMAN FACTORS

Research in ATM human factors has tended to be focused on training and selection of air traffic controllers and pilots, on technology for presentation of information to human operators in the system, and on technology for transmission of control input to appropriate parts of the system. However, the drive to efficiency and flexibility has provided technologies to make the process of ATM distributed and de-centralized. In response the foci in human factors, research has shifted to include considera-tion of collaborative decision-making, distributed control, and coordination of roles and responsibilities in dynamic systems.

2.1. Schedule and Planning

Airline operations schedules are established months in advance of flight operations. The schedules are based on market demographics and on the particular operations being scheduled (e.g. optimal cargo airline schedules from those of passenger/revenue operations). In addition to market demands, human factors issues associated with fatigue, vigilance and duty-cycle play a significant role in mediating operations duration and frequency, and in determining crew complement. The performance impact of circadian de-synchronization are the focus of a significant body of research in aviation/ATM human factors (Aviation Space and Environmental Medicine 1998). The issues of duty time and cycle are being extended from flight crew to maintenance and air traffic controller schedules.

2.2. Flow and Facility Optimization

The national and international airspace and the airports that serve commercial and general aviation operations are a constrained resource. Human operators are called on to manage these scarce assets to provide predictable and timely performance while maintaining a margin of safety to assure system robustness in response to environmental or other sources of operational disturbance. Within the US national airspace, 50 000 flights per day operate. The issues of large-scale optimization and decision-making come quickly to the fore. Automation and information aids are needed to assure all factors are accounted for in the decision-making process. Automated checks for information integrity and currency are critical to the air traffic service providers and to the operators. Appropriate resolution of human factors issues associated with human/automation integration and distributed decision-making are fundamental to effective operations.

2.3. Air Traffic Control

Air traffic service providers seek to ensure efficient passage of aircraft through the airspace and while assuring that safe separation is maintained among aircraft. An international collaboration in ATM, the International Civil Aviation Organization (ICAO), partitioned the type of management and decision-making required of the human operators in the system. Air traffic operations are delegated to distinct facilities and different services are provided by those facilities. For a review of the structure of ATM and its evolution, see Federal Aviation Administration (1995). The operations on the surface of airport facilities and the management of operations at gates are delegated to ramp control and airport surface tower operations. Arrival and departure traffic management is delegated to tower and approach/departure control operations. These transition airspace in departure and arrival are handled by the Terminal Radar Approach Control (TRACON). Finally cruise, *en route* operations are managed both over continental and oceanic operations by controllers in Air-Route Traffic Control Centers (ATRCC).

There is a range of specific human factors issues motivated by the kind of service provided and the types of tools provided to the controllers to manage those services. These services are provided by the delivery of clearance and advisory information from the ground to the aircraft and by the provision of opera-

tional state information from the aircraft to the ground service providers. Human/automation integration, training for operation, interface development, procedure development, coordination in strategic decision-making, and communication processes are the key areas of concern for human factors in air traffic control.

2.4. Flight Management

The final common path for ATM is the coordinated movement of aircraft through the airspace. The process of flight management seeks to assure safe timely and efficient operation of aircraft. The human factor attendant the evolution of flight control and flight management as the integration of human and machine is well documented in the Billings (1997). The management of flight has evolved from simple manual control with augmentation to a nearly fully automated operation. The development and evolution of flight deck automation has compelled significant development in the theory and practice of human factors (Weiner and Nagel 1988. This fundamental research in human automation integration in flight deck operations serves as an excellent precursor for similar automation integration issues in ground-based air traffic control.

2.5. Integrated Flight Deck and Air Traffic Management

There are crucial and far-reaching changes in the process of ATM that provide significant challenge to human factors theorists and practitioners. The world community of aviation operations is engaged in a vast, system-wide evolution to integrated airborne and ground-based ATM, NASA, the FAA and Eurocontrol have initiated programs of research and development to provide flight crew, airline operations and air traffic managers with automation aids to increase capacity in *en route* and terminal areas. The aiding technologies support exploitation of timely and dynamic information on atmospheric hazards, traffic fluctuations and airspace utilization. However, the human operators retain the authority and responsibility for safe, efficient operation. The integration of human decision and performance parameters with those of the automation-aiding systems offers a significant challenge to human factors and cognitive engineering in the extent and dynamics of information and control exchanged among the participants across a global aviation network (Kahne and Frolow 1996).

3. AIR TRAFFIC MANAGEMENT HUMAN FACTORS RESEARCH

The complexity of current ATM challenges human factors practice. The potentially revolutionary impact of operational changes in advanced ATM requires fundamental and extensive research agenda to assure: operational integration, management of automation, assurance of appropriate situation awareness, safety assurance and interface optimization (FAA Human Factors Team 1996, Wickens *et al*. 1997, 1998). In addition to these human–technology interactions, the context of operations for these technologies is international in scope. Language and culture will have a significant impact on the integration and operation of technologies for worldwide use. Research in the interaction of automation and culture is in its infancy in terms of theory and method.

3.1. Relaxation of Constraints and Crew–Team Coordination

Technologies have being developed to relax system constraints and integrate flight deck and ground-based operation. These impact human performance research in two ways. First, the decision-making process becomes distributed. This distributed decision process differs from current operations in the number of participants, in the timing of their communication, in their situational awareness and in the span of their authority which has direct impact on crew and team resource management processes. Second, the dynamic concept of operations provides a new coordination challenges for the human operators of that system. In current operations, the roles of the constituents in airspace operations are relatively fixed. However, through implementation of technologies to share information and authority in ATM relax that stability. The human operators (pilots, air traffic controllers, airline operations personnel) must monitor and predict any change in the distribution of authority or control that might result as a function of the airspace configuration, aircraft state or equipage, and other operational constraints. The operators are making decisions and sharing information not only about the management of the airspace, but also about the operating state of that airspace.

3.2. Automation Integration

Automation of flight management operations has had a 20-year development history and the human factors issues associated with increasingly sophisticated multi-modal control operation has received extensive research (Weiner 1989, Billings 1997). The issues of automation bias and over-reliance, lack of adequate feedback as to operating mode, workload, vigilance decrements and primary/secondary task inversion need to be investigated in the development of ground-based automation for air traffic control. These issues become even more critical as the link in control tightens between the ground and the air.

3.3. Workload, Situation Awareness and Vigilance Factors

In addition to workload issues associated with automation management, ATM human factors research will need to address dynamic shifts in workload distribution between air and ground. As automation and information systems for the participants increase in sophistication, situation awareness will need to be maintained among multiple participants in the system. As aircraft development provides for more fuel-efficient extended range operation, crew complement and crew fatigue factors will need to be re-investigated.

3.4. Control–Display Interface Integration

The promise of increased safety and efficiency in the international airspace is based in large part on improved automation aiding and information presentation to the multiple players in the system. These improvements rest on a research requirement in the human factors of controls and displays. The use of virtual environments, new color-coding and immersive technologies for information retrieval and communication are areas for significant new redevelopment. Integration of new display and communication process into the ATM requires a human factors focus on transition from current practice to near-term future improvements (e.g.

data link that moves transactions from an essentially verbal/auditory to a visual/spatial representation). Research is required to assure effective transition (operator training and interface design) and to assure safety (human reliability studies and system design for error–tolerance).

3.5. Communication

Improved information exchange, coordination of its use and assurance of its integrity are all key issues in human factors for the advanced ATM process. Group coordination and distributed decision-making research is beginning to explore the impact of shared data on aviation operations. Research needs to be undertaken to explore the impact of shifting roles and responsibilities in a complex dynamic system. Communications contribution to system safety and reliability in distributed operations needs to be characterized. The appropriate coordination of information services on the flight deck and on the ground also needs to be explored. The technology "push" tendency to overwhelm the operator because information is available should be countered by study of information requirement and information utility based on roles and responsibilities in the advanced ATM.

3.6. Methodological Developments

In addition to the research agenda based on the emerging issues of complex dynamic systems, there is shift in methodologies required to meet that research agenda. Prior paradigms used an incremental increase in testing complexity, fidelity and cost from empirical results, to prototype, to full mission simulation and then to field-testing. Such a process is still a valid paradigm. However, the rate of development of ATM systems, the tremendous economic pressure to implement and reap immediate benefits from technologies, and the significant complexity and cost of large-scale distributed air–ground tests suggest the development of other, more cost effective, methods of human factors research. Prominent among these is computational human performance modeling (Laughery and Corker 1997). In this paradigm the human and the system elements of interest are represented as computational entities (or agents). These agents interact as the system elements would in actual field operations and behaviors can be observed. The benefit accrued (assuming well-developed and validated models) is that system and human characteristics can be quickly varied (e.g. based on an assumed technology change) and the impact of that change identified in the full-system context. Such models help focus the expensive and complex simulation and field tests. In addition, performance at the edge of system safety can be explored in the computational human performance-modeling paradigm.

4. CONCLUSION

Human factors research in advanced ATM is a challenging mix of well-established human factors practice (e.g. display and control development and evaluation, training and selection processes) and more edge-of-the-art research (e.g. virtual and immersive environment development, automation, shared initiative, and decision-aiding in distributed control). These techniques are made more challenging by the size, dynamic complexity and safety–criticality of an evolving international system. The methodology and the theory of human factors are challenged by the scale of technology and the global extent of its implementation. Fundamental research is needed in the process of scaling-up basic human factors practice to global technology implementation for real-time control of such a dynamic and critical operation as advanced ATM.

REFERENCES

AVIATION SPACE AND ENVIRONMENTAL MEDICINE, 1998, Special Edition: *Crew Factors in Flight Operations* (Alexandria: Aerospace Medical Association).

BILLINGS, C.E., 1997, *Aviation Automation: The Search for a Human-centered Approach* (Hillsdale: Lawrence Erlbaum).

FEDERAL AVIATION ADMINISTRATION, 1995, *Air Traffic Services Plan 1995–2000 (*Washington DC: FAA*)*.

FEDERAL AVIATION ADMINISTRATION 1996, *The Interfaces Between Flightcrews and Modern Flight Deck Systems,* Human Factors Team, (Washington DC:FAA).

KAHNE, S. and FROLOW, I., 1996, Air traffic management: evolution with technology. *IEEE Control Systems*, 12–21.

LAUGHERY, K.R. and CORKER, K.M., 1997, Computer modeling and simulation. In Salvendy, G. (ed) *Handbook of Human Factors and Ergonomics* (New York: John Wiley).

WEINER, E.L. and NAGEL, D.C., 1988, *Human Factors in Aviation* (San Diego: Academic Press).

WEINER, E., 1989, Human factors of advanced technology ("glass cockpit") transport aircraft. *NASA Contractor Report 177528*.

WICKENS, C., MAVOR, A. and MCGEE, J., 1997, *Flight to the Future: Human Factors in Air Traffic Control* (Washington, DC: National Academy Press).

WICKENS, C., MAVOR, A., PARASURAMAN, R. and MCGEE, J., 1998, *The Future of Air Traffic Control: Human Operators and Automation* (Washington, DC: National Academy Press).

Air Traffic Management System Design

K. Kerns

The MITRE Corporation, Center for Advanced Aviation System Development, 1820 Dolley Madison Boulevard, McLean, VA 22102-3481, USA

1. INTRODUCTION

In the USA, the air traffic management (ATM) system, operated by the Federal Aviation Administration (FAA), is composed of two subsystems. The air traffic control (ATC) system assures safe separation of aircraft and an orderly flow of traffic. Air traffic specialists or controllers provide the ATC system's direct service to pilots, aircrews and passengers on board aircraft being controlled by the system, during takeoff, while en route and upon landing. The traffic flow management (TFM) system works to simplify and expedite traffic flows, ensuring that the operational environment is compatible with the capabilities of the controllers and pilots responsible for conducting flight operations. Traffic flow management specialists issue system-wide, regional and local restrictions on aircraft flight paths and schedules to ensure that traffic demand does not exceed the traffic handling capacity of the ATC system. Figure 1 depicts the hierarchical structure of the ATM system that has evolved as the demand for aviation services has grown.

Increases in traffic demand are projected to continue for a long time. The next evolutionary steps for the ATM system focus on the introduction of new technologies, decision-support tools and traffic management strategies better to utilize capacity and to accommodate user needs without compromising safety. Even as the ATM system evolves becoming more highly automated, the role and contribution of the human element to the system's performance remains significant. The interactions and interdependencies among the human operators, their equipment and their procedures present a broad and diverse set of human factors challenges. An examination of the changing role of controllers in the evolving ATC system and in the future ATM system illustrates many of these challenges.

2. THE ATC SYSTEM AND CONTROLLER TASKS

To receive air traffic services, pilots file a flight plan, obtain a clearance and maintain radio communication with controllers as they fly through airspace controlled by the ATC system. Controllers monitor flights within the specific airspace (sectors) that they are in charge of and issue instructions to pilots to ensure safe separation and efficient use of the airspace.

Controllers in terminal and en route facilities handle active flights operating under the control of the ATC system. Figure 2 shows the role of the terminal and en route controllers (Mundra 1989). In the terminal area, controllers working in airport traffic control towers (or towers) control the surface movement of aircraft and their landings and takeoffs. Once the aircraft takes off and before it lands, the aircraft is controlled by controllers working in terminal approach control facilities. Outside of this region,
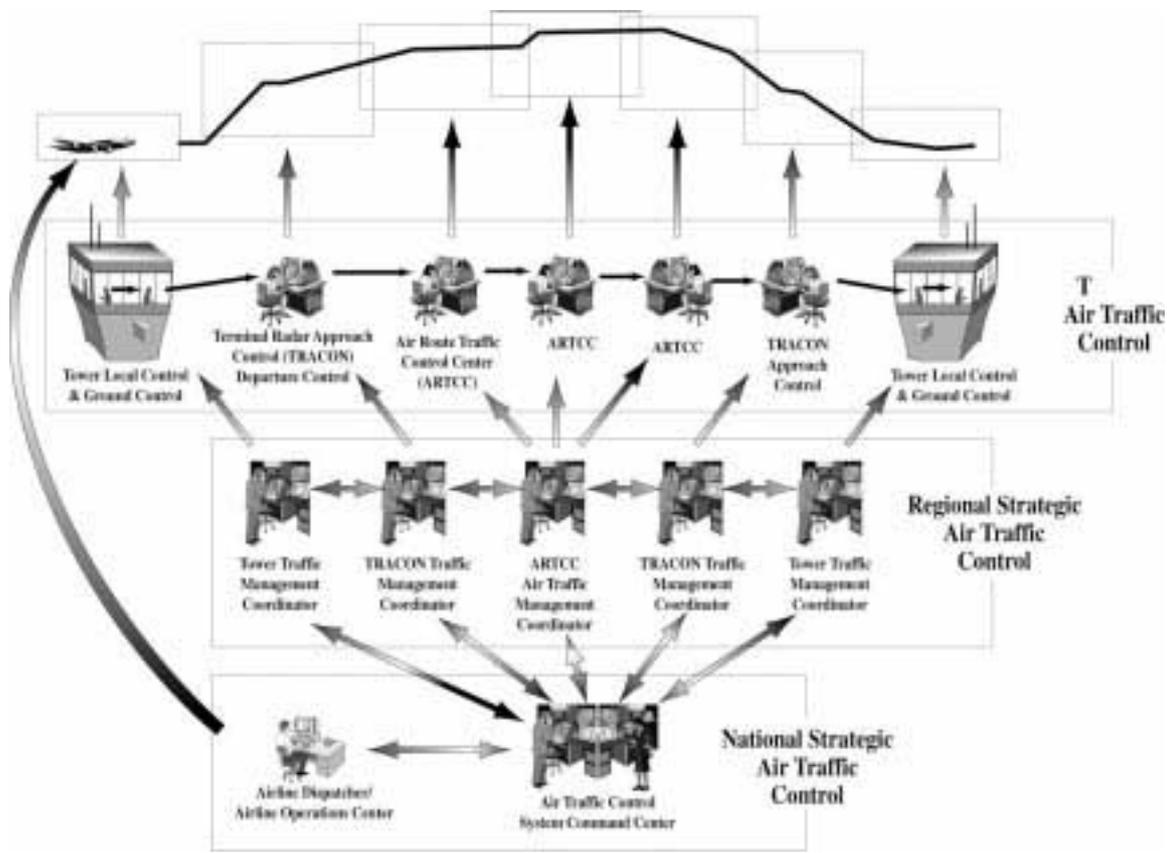

Figure 1. The air traffic management system organizational structure from Kerns *et al.* (1999), with permission.

Type of Facility	Terminal Area Facilities		En Route Facilities	
Controlling Facility	Airport Traffic Control Towers	Approach Control Facilities	Air Route Traffic Control Centers (ARTCCs)	
Type of Control	Ground Traffic Control Takeoff and Landing Control	Approach and Departure Control	ATC during Transition and Cruise	
Airspace	Airport Traffic Area Typically 5 nmi and 3,000 ft AGL	Approach Control (Tracon Area) Typically Extending Up to 40 nmi and 10,000 ft msl from the Airport	En Route Airspace Transitional Phase Typically 50-150 nmi from Airport	Cruise Phase Up to 60,000 ft
Typical Flight Time	Typical Ground Time 5-10 min	Typical FlightTime 10-20 min	Typical FlightTime 10-20 min	Typical FlightTime 20 min to Several Hours
Flight Profile				

Figure 2. Role of terminal and en route facilities in air traffic control from Mundra (1989), with permission.

controllers working in en route centers have responsibility for the flight. While en route, aircraft may pass through one or more centers, into oceanic airspace, and be transferred to an international control facility.

2.1. The Airport Traffic Control Tower

Controllers working in towers spend much of their time visually tracking aircraft movements outside the tower. The glass-enclosed tower cab is physically oriented relative to the runways to provide an unobstructed view of the airport's primary movement areas (taxiways and runways). To keep track of aircraft movements, tower controllers must move around the tower cab and interact with other controllers who are responsible for aircraft during different segments of the takeoff and landing phases of flight. Reduced visibility at night or because of weather conditions, can leave tower controllers and pilots vulnerable to errors resulting in runway incursions and other surface incidents between aircraft taking off and landing. In addition, the responsibility to monitor traffic by continuously looking out the tower cab window, makes it difficult read display information or enter data without disrupting visual scanning tasks.

A voice communication link allows the tower controller to direct aircraft movements to and from the gate and issue clearances for take off and landing. During busy periods, tower controllers must accomplish time-critical communications at a rate only marginally ahead of air traffic movements around the airport. The tempo of the operation means that controllers must speak rapidly to transmit complex messages. Moreover,

radiofrequency congestion during these busy periods often prevents pilots from verifying the communication by reading back their instructions. Dropping the readbacks or other procedural steps that are intended to assure communication can leave controllers and pilots vulnerable to communication errors.

2.2. Terminal Radar Approach Control

Once the aircraft takes off and before it lands it is monitored by the terminal radar controllers. Arrival and departure controllers rely on a radar display to keep track of the traffic situation and plan efficient traffic flows. The arriving traffic in the terminal area funnels in from higher altitudes and speeds for landing. The arrival controllers sequence and space the aircraft. The departing traffic fans out to higher altitudes and speeds for the cruise phase. Departure controllers establish the aircraft on headings to their planned routes of flight.

In this environment, there is a high frequency of aircraft maneuvering and rapid convergence of many traffic streams. Within the terminal airspace aircraft do not follow a predefined route, instead control is exercised through heading, altitude and speed instructions. Controllers must extrapolate aircraft positions to visualize the paths to be flown by the traffic and issue instructions to execute these paths. A two dimensional radar display depicts each aircraft's position, identity, altitude, and speed. The controller's performance depends on an ability to visualize the three-dimensional traffic situation and keep track of the situation and the required actions under high workload conditions.

The high rate maneuvering results in frequent communications with pilots and frequent coordination with other controllers handling nearby sectors of airspace. As in the tower, terminal controller workload and frequency congestion can result in procedural short cuts and miscommunications.

2.3. The En Route Center

Once an aircraft is established on its route of flight, en route controllers monitor the flight and ensure separation. Over the continental USA, en route ATC is predominantly based on radar information. However, the FAA also provides ATC service within a large area of non-radar airspace over the Atlantic and Pacific Oceans and the Gulf of Mexico. Oceanic ATC is based on flight plan and aircraft position information provided by pilots. This section focuses on controller tasks and procedures in the en route radar environment.

The en route center airspace is divided into sectors with various characteristics. Controllers working high altitude sectors handle departure traffic climbing to cruise altitude, en route traffic flying level and traffic descending to a lower level in preparation for arrival. Controllers working low altitude sectors handle a mixture of aircraft which fly at lower altitudes, slower speed arrival and departure traffic, and higher performance aircraft climbing to the high altitude sectors.

Characteristics of the traffic and sector airspace contribute to sector complexity and controller workload. Factors such as the type of traffic flows and the service that must be provided each aircraft account for routine tasks that are known in advance and performed for every aircraft or for a specific subset of the traffic such as a particular flow. For example, en route controllers currently have a major responsibility to routinely issue predetermined routing and altitude clearances that establish an orderly flow of traffic within the airspace. Factors such as the size and shape of the airspace delegated to the sector, the number and characteristics of aircraft within the airspace, and the nature and frequency of requests by pilots to alter their flight plans account for additional tasks that depend on the situation. For example, one factor thought to be an important determinant of complexity involves the number and pattern of potential conflicts that can occur within the sector airspace for a given period of time. The decision-making process associated with the detection and resolution of conflicts has a great impact on determining controller workload and sector capacity.

During busy periods, two or three controllers may staff en route sectors. When multiple controllers work a sector, their tasks must be closely coordinated. Typically, a radar (R) controller is in charge of the sector operation. This controller uses the radar situation display and voice communications to apply radar separation procedures and separate the aircraft from all others within the sector. A radar associate or data (D) controller is responsible for separation planning activities. The D controller uses paper flight strips which provide advance information on the aircraft's planned route and communicates with other controllers to identify potential problems and coordinate preventative control actions. When the sector is too busy for the R and D controllers to handle, a radar coordinator (tracker) or handoff controller may share the R controller's load, serving as a redundant "set of eyes and ears" to support situation monitoring and as a second pair of hands to perform data entry and coordinate actions with other sectors.

In contrast to the tower and terminal environments, the en route operations and computer system support a more structured traffic organization. Manual data entry tasks that make up a large part of the en route controller's workload can be demanding and time-consuming. The en route computer system stores maintains the ground system's flight plan data for each aircraft throughout its flight. To support ground-based flight planning, en route controllers are required to keep the computer's data up to date by entering control instructions and flight amendments as they are issued.

During busy periods, communication with pilot makes up a significant portion of the R controller's workload. Moreover, the addition of other team members does not redistribute the sector's communication workload. In fact, the R controller's heavy involvement with air–ground communication tasks tends to hinder the verbal interactions required for team direction and coordination when multiple controllers are working. As in the tower and terminal radar environments, miscommunications can occur when controller workload and frequency congestion increase.

3. HUMAN FACTORS ISSUES IN THE EVOLVING ATC SYSTEM

The ATC system is evolving to address human factors challenges that relate to the mental demands and uncertainties imposed by the ATC process and the environments in which the controllers are embedded. Fundamental human factors challenges are to manage mental demands and complexity while improving the efficiency of controller strategies. Another broad set challenges relate to the cooperative nature of the ATC activity and the need to manage communications demands and improve information transfer.

3.1. Complexity Management

Because the air traffic control system is functioning near, at or even beyond its planned maximum traffic handling capacity, there has been a long-standing interest in understanding and quantifying the controller's traffic handling capacity and in predicting when this capacity breaks down. However, despite years of research on controller workload, sector complexity, and operational errors, the measurement and management of complexity in the current system is rudimentary.

Today, a mix of procedural techniques and automated capabilities is used to anticipate and manage complexity problems. On a scheduled basis, controller positions are opened and closed and responsibility for sector airspace is divided and combined under the positions. During busy rush periods throughout the day, en route sectors are staffed with one, two, or three controllers. When traffic volume is expected to exceed a sector's capacity for an extended period, TFM is alerted by an automated function and procedures are activated to limit the volume of traffic handled by the sector. Currently, the automated function uses the predicted aircraft count to estimate sector complexity. Operational decisions on alerting thresholds and complexity management actions are made from human experience and judgment.

Operational errors and deviations have been used to study the factors that contribute to sector complexity and controller

workload giving rise to system breakdowns . Although controller workload has been analyzed extensively to derive traffic and airspace characteristics beyond traffic count that predict controller workload, this approach seems to be most applicable to measurements of the observable, motor and manual components of workload. Research on the mental and cognitive components of workload has shown that successful controllers use adaptive strategies to handle more aircraft without error or excessive workload. For example, the controller may handle an unexpected increase in traffic load by structuring the traffic so that all of the flights are in-trail and traveling at a uniform speed, thereby reducing the difficulty of monitoring for conflicts. As a result the controller spends less time on each flight during periods of high demand. In part, this adaptability may explain why studies of operational errors reveal that errors are not only uniquely associated with high complexity and workload, but also occur under low-to-moderate complexity and workload.

Because there are no normative data indicating relative frequencies of the levels of traffic volume and complexity in the operational environment, it is difficult to determine whether the reported frequencies of errors in these levels are disproportionately related to chance. A better understanding of the relationship between complexity and errors and techniques for managing complexity depends on more complete baseline data on human and system performance under normal and abnormal conditions.

Even though researchers do not fully understand the relationship between complexity, workload, and errors, there is general agreement that controllers need information on current and predicted sector complexity. Decision-support tools for conflict prediction and resolution may help to address controller needs for sector complexity information (Kerns 1999a). With these tools, an indicator of sector complexity is available in the form of display data on the number of predicted conflicts and their temporal distribution. Advance information on conflict resolution workload is expected to improve situation awareness and allow the sector team to better schedule and manage conflict resolution tasks.

3.2. Control Strategies and Efficiency

As mentioned earlier, controllers cope with increasing traffic demand and complexity by employing adaptive strategies. Research on controller strategies indicates that specific adaptations tend to lower the efficiency of individual flights and the overall traffic flow when demand is high. In general, strategies that are economical for the controller are those that preserve the primary objective of safety but take less account of secondary objectives such as flight efficiency, user-preferred paths, and fuel economy.

Another way controllers adapt their strategies is to focus on the immediate tactical situation and abandon planned strategies. To plan and execute control strategies that are more flight efficient, controllers must coordinate with each other. Under high workload, a shift from cooperative toward individual work has been observed in both the terminal and en route environments. In the terminal environment controllers may shed coordination tasks under high demand, resulting in less efficient sequencing of arrival aircraft. While in the en route environment, both sector controllers focus attention on the current tactical situation, reacting quickly and employing less efficient tactical control techniques. In many traffic situations, reactive strategies are also

communications-intensive, with the controller assuming greater responsibility for flight paths and making continuous tactical adjustments.

To preserve flight efficiency, a more proactive approach to ATC is desirable. Decision-support tools for tactical planning and selection of control strategies are being introduced in the en route and terminal environments. The preceding section mentioned the en route decision-support tools being developed to provide early detection of conflicts. This conflict probe capability will reduce the mental calculations and extrapolations involved in separation monitoring and afford a longer lead-time to plan resolution maneuvers that are less disruptive to the user's flight intent. Team performance will be aided by providing the D controller with more powerful tools for visualizing the future traffic situation and evaluating proposed maneuvers. In addition, an automated coordination aid will also allow controllers to share and approve plans electronically.

In the terminal environment, controller aids for merging flows and sequencing aircraft for approach are already in use or under test at field facilities. These tools provide advance information on the predicted sequence of arrival aircraft. One tool, the converging runway display aid (CRDA), is designed to assist the controller in conducting staggered approaches (Mundra and Levin 1990). Staggered approaches require specific separations between aircraft landing on adjacent runways as well as between in-trail aircraft. Staggered approaches are more complicated than simultaneous approaches that have only in-trail spacing requirements. The CRDA reduces the complex mental calculations and extrapolations involved in staggered approaches by projecting false targets or ghosts for aircraft arriving in one of two converging streams onto the other stream, thus allowing the controller to visualize and manage simple in-trail spacing on a single approach.

Another tool, the final approach spacing tool (FAST) provides the controller with landing sequence numbers and runway assignments to achieve an accurately spaced flow of traffic onto the final approach course (Lee and Davis 1995). Based on the displayed sequence, the controller formulates appropriate instructions for merging and spacing the arrivals. FAST advisories should improve the runway delivery precision and reduce controller workload by reducing the number of vectors issued to each aircraft and reducing the need for (verbal) coordination between controllers.

3.3. Communications and Information Transfer

There is ample evidence in the research literature that controller–pilot communications are a common and persistent problem in today's operations (Kerns 1999b). Some of the primary factors that contribute to communications problems arise from the use of spoken language to transfer information. Controllers and pilots have adopted a standardized phraseology, language conventions, and procedures which define the process for conducting the dialogue, including the cues that tell a listener when a transaction has been completed and whether a readback is required. However, despite years of refining the language and procedures, research confirms the intractable nature of many of the problems inherent in the exclusive use of spoken language for controller–pilot communications. Grayson and Billings (1981) analyzed aspects of human speech processing and conversational behavior that mediate communication performance. Their analysis found that

a tendency to fill-in information, the expectancy factor, and timing problems were implicated in many types of controller–pilot communication problems. The expectation factor contributes to misinterpretations and inaccuracies because controllers and pilots sometimes hear what they expect to hear. This generates what have been called "readback and hearback" errors in which, respectively, a pilot perceives what he expected to hear in the instruction transmitted by the controller and a controller perceives what he expected to hear in the readback transmitted by the pilot.

Congestion on the voice radiofrequency has also been implicated in communications problems. During busy periods, controllers issue longer, more complex messages in an attempt to minimize use of the radiofrequency. Analyses of routine communications (Cardosi 1993, Morrow *et al.* 1993) have highlighted the contribution of message complexity and the resulting memory burden to miscommunications. In both the terminal and en route environments, errors and procedural deviations increased as clearances increased in complexity. In response to these problems, data link communications have already been introduced as an alternate means of information transfer in the tower environment and a controller–pilot data link capability for the en route environment will be field tested soon.

In addition to reducing frequency congestion, data link technology offers several advantages for transfer of lengthy, repetitive ATC messages. The data link has capabilities for message storage and retrieval that reduce the controller's burden when preparing messages and the pilot's memory burden when receiving them. Data link also allows controllers and pilots to pace the transmission and processing of the information, thus avoiding conflicts between communications and other higher priority tasks.

Although the experience and results to date indicate that these initial applications of data link offer important operational benefits at little or no cost to the human operators, future applications of data link must be selected carefully, addressing key human factors challenges. Visual display and manual control of transmitted information may not be appropriate in environments where the controller or pilot visual and manual resources are already reaching an overload state. Delay factors associated with message composition may limit the utility of data link in rapidly changing conditions while transmission delays may limit its utility for time-critical transmissions. Finally, new procedures will be needed to maintain team performance when the communication medium is silent and may be less readily observable by multiple operators.

4. HUMAN FACTORS ISSUES IN THE DESIGN OF THE FUTURE ATM SYSTEM

Recently, the FAA and the aviation industry reached general consensus on a strategic direction for the ATM system in the USA called "Free Flight" (RTCA 1995). The goal of free flight is an ATM system based on two fundamental premises:

- The FAA retains and strengthens its safety mandate for separation of aircraft.
- The users of the system are given the flexibility to make decisions that allow them to extract the maximum economic benefit from the ATM system.

Among other characteristics, free flight assumes a shift way from the current ground-based, tactical ATC operations toward a more cooperative arrangement in which users have greater flexibility to select and manage their flight paths and to participate routinely in airspace management decisions. The new decision-support tools and data link communications technology discussed in the preceding section represent the initial phase of ATM system evolution toward free flight. As the ATM system continues to evolve, further human factors challenges will emerge in response to changes in operating philosophies, the design of new functional architectures which reallocate system functions, and the introduction of more advanced tools and technologies to support ATM operations.

4.1. New Philosophies and Procedures

Currently in ATC, the basic rule of separation is that every controller is responsible for separation of participating aircraft for the duration of time the aircraft is within the controller's sector of responsibility (Nolan 1990). In future ATC, the free flight goal is to evolve toward a more strategic operation that better accommodates user preferences. Under this philosophy, controllers may assume more responsibility for solving each other's problems as strategic predictions allow for early ATC interventions. In addition, as new technologies provide pilots with traffic information and conflict prediction capabilities commensurate with or surpassing those of the controller, they may also be called upon to participate in solving ATC problems.

A team perspective is likely to be a major focus in addressing human factors issues in future ATM. Considerable research and analysis has addressed the controller's role (often as a single operator) within an ATC domain. There is much less work that addresses the multi-operator perspective. Today team training is less formal in ATC than on the flight deck, but teamwork is likely to be a critical component of ATM for the foreseeable future. Definition, prediction and management of team performance will be an important human factors challenge in the future system.

4.2. Functional Architecture

Already, new ATM decision-support systems cut across traditional divisions of responsibility between en route and terminal controllers. While strategic planning tools such as conflict probe tend to blur the distinction between ATC and TFM responsibilities. It is also likely that the current workload concentrations in the tower environment will motivate further reallocation of tasks across ATC environments and development of capabilities for smoothing workload through greater human control over the tasks, their timescales, their scheduling, and their sequencing (Laios and Giannacourou 1995). At the same time, imbalances in air and ground system capabilities will motivate exploration of alternative allocations of functions between controllers and pilots which may increase the pilot's role in ATC. These forces, taken together with the new operating philosophies envisioned under free flight, will drive a progression toward new relationships among system participants.

In ATC, human factors can help evaluate the allocation of responsibility between controllers and pilots in all of the operational environments. And within the ATM organization, human factors can help analyze and evaluate the allocation of responsibility between controller team members and between controllers and traffic managers. For example, laboratory simulations can be used to explore the feasibility and effectiveness

of new roles, such as a planner/coordinator controller role that uses automated tools to support multiple sector controllers. Such a role might combine functions currently performed by a D controller and a traffic flow manager.

4.3. Advanced Tools and Technologies

Advanced concepts for ATC in the USA envision higher levels of automation in controller decision-support. In the terminal environment, the next version of FAST will provide automated advisories for final approach spacing while in the en route environment automated conflict resolution advisories will be provided. This level of automation will have a profound impact on the way ATC is performed. Not surprisingly, there is considerable concern in the human factors community over the impact of such automation on the controller's situation awareness and ability to provide back-up for the automation system when necessary environment (Wickens *et al.* 1998).

Automation issues will become increasingly important human factors challenges in the future ATM system. However, before questions on human roles can be addressed, more fundamental questions must be answered regarding the efficiency of the alternatives in different task environments and the nature of the back-up mechanisms. Laboratory simulations can help answer questions regarding the efficiency of controller tools under various traffic scenarios. Requirements for maintaining situation awareness and controller skills will vary according to the need for manual back-up in the future. Human factors analyses can help specify these requirements under various back-up mechanisms.

5. Conclusion

There are major challenges ahead in analyzing functional requirements and the allocation of functions between human and automated elements of the future ATM system. At this point, there is an opportunity for proactive research in human factors to gain insight and perspective on issues surrounding the evolution of ATM. Multiple research approaches, including the review and application of existing data and knowledge, will make it possible to prove the value of new concepts regarding human roles and interrelationships and their impact on the quality and efficiency of the ATM system. During the development and introduction of new ATM capabilities, human factors specialists should be involved to help define new functionality and procedures.

As the discussion above indicates, however, this is a system that has historically attempted to match the capabilities of the people with the demands of the system by distributing roles and responsibilities, and by introducing new tools and technology. Efforts to enhance efficiency through increased flexibility and shifts in the locus of control need to be based on an understanding of the complexity inherent in this system, and on a realistic assessment of the strengths and limitations of the technologies being considered to support controllers and pilots.

REFERENCES

CARDOSI, K., 1993, *An Analysis of En Route Controller–Pilot Voice Communications*. DOT/FAA/RD-93/11 (Washington, DC: US Department of Transportation, Federal Aviation Administration).

GRAYSON, R.L. and BILLINGS, C.E. (eds), 1981, Information transfer between air traffic control and aircraft: Communication problems in flight operations. In *Information Transfer Problems in the Aviation System*. NASA Technical Paper 1875 (Moffett Field: NASA Ames Research Center).

KERNS, K., 1999a, *Effects of a Strategic Conflict Probe Capability and Unstructured Traffic Conditions on ATC Performance and Flight Efficiency*. MTR98W121 (McLean: MITRE Corp.).

KERNS, K., 1999b, Human factors in ATC/flight deck integration: implications of data link simulation research. In Garland, Wise and Hopkin (eds) *Handbook of Aviation Human Factors* (Mahwah: Lawrence Erlbaum), pp. 519–46.

KERNS, K., SMITH, P.J., MCCOY, C.E. and ORASANU, J., 1999, Ergonomic issues in air traffic management. In W. Karwowski and W.S. Marras (eds) *The Occupational Ergonomics Handbook* (Boca Raton: CRC Press), pp. 1979–2006.

LAIOS, L. and GIANNACOUROU, M., 1995, Human factors issues of advanced ATC systems. In Fuller, Johnston and McDonald (eds), *Human Factors in Aviation Operations: Proceedings of the 21st Conference of he European Association for Aviation Psychology* (Cambridge: Ashgate), pp. 108–13.

LEE, K.K. and DAVIS, T.J., 1995, The development of the final approach spacing tool (FAST): a cooperative controller–engineer design approach. In *Proceedings of The International Federation of Automatic Control Conference: Automated Systems Based on Human Skill*, Berlin, Germany, pp. 219–26.

MORROW, D., LEE, A. and RODVOLD, M., 1993, Analysis of problems in routine controller–pilot communications. *International Journal of Aviation Psychology*, 3, 285–302.

MUNDRA, A.D., 1989, A *Description of Air Traffic Control in the Current Terminal Environment*. MTR88W167 (McLean: MITRE Corp.).

MUNDRA, A.D. and LEVIN, K.M., 1990, *Developing Automation for Terminal Air Traffic Control: Case Study of the Imaging Aid*. MP90W29 (McLean: MITRE Corp.).

NOLAN, M., 1990, *Fundamentals of Air Traffic Control* (Belmont: Wadsworth).

RTCA, 1995, *Final Report of RTCA Task Force 3: Free Flight Implementation* (Washington, DC: RTCA, Inc.).

WICKENS, C., MAVOR, A. and MCGEE, J., 1998, *The Future of Air Traffic Control: Human Operators and Automation* (Washington, DC: National Academy Press).

Aircraft Inspection: Computer Based Training Program

G. M. Nickles, A. K. Gramopadhye, B. Melloy, and J. Marshall*
Department of Industrial Engineering, Clemson University, USA and
*Lockheed Martin Aircraft Center, Greenville, South Carolina, USA

1. INTRODUCTION

Aircraft inspection and maintenance are an essential part of a safe, reliable air transportation system. Training has been identified as the primary intervention strategy in improving inspection performance. If training is to be successful, it is clear that we need to provide inspectors with training tools to help enhance their inspection skills. This article outlines recent efforts being pursued by the Training System Laboratory at Clemson University focused on the development of a computer-based inspection training program, automated system of self-instruction for specialized training (ASSIST). The ASSIST program was developed using a task analytic methodology.

Existing training for inspectors in the aircraft maintenance environment tends to be mostly on-the-job (OJT). Nevertheless, this may not be the best method of instruction (FAA 1991, Gordon 1994). For example, OJT feedback may be infrequent, unmethodical, and/or delayed. Moreover, in certain instances feedback is economically prohibitive or infeasible due to the nature of the task. Thus, because the benefits of feedback in training have been well documented (Weiner 1975), and for other reasons as well, alternatives to OJT are sought. Furthermore, training for improving visual inspection skills of aircraft inspectors is generally lacking at aircraft repair centers and aircraft maintenance facilities. However, the application of training knowledge to enhance visual inspection skills has been well documented in the manufacturing industry. Training improves the performance of both novice and experienced inspectors (Weiner 1975, Drury and Gramopadhye 1990). Visual inspection skills can be taught effectively using representative photographic images showing a wide range of conditions with immediate feedback on the trainee's decision (Weiner 1975). Training with realistic photographic images, in controlled practice and with appropriate feedback, appears to be superior to OJT training on its own (Latorella *et al*. 1992).

Thus, off-line training/retraining with feedback has a role to play in aircraft inspection training. One of the most viable approaches for delivering training given the many constraints and requirements imposed by the aircraft maintenance environment is computer-based training. Computer-based training offers several advantages relative to traditional training approaches; for example, computer-based training is more efficient, facilitates standardization, and supports distance learning. With computer technology becoming cheaper, the future will bring an increased application of advanced technology in training. Over the past decade, instructional technologists have offered numerous technology-based training devices with the promise of improved efficiency and effectiveness. These training devices are being applied to a variety of technical training

applications. Examples include computer-based simulation, interactive video disks, and other derivatives of computer-based applications. CD-ROMs and digital video interactive (DVI) technology will also provide us with the multimedia training systems of the future. Many of these training delivery systems, such as computer-aided instruction, computer-based multimedia training, and intelligent tutoring systems, are already being used today, ushering in a training revolution.

In the domain of visual inspection, the earliest efforts to use computers for off-line inspection training were reported by Czaja and Drury (1981). They used keyboard characters to develop a computer simulation of a visual inspection task. Similar simulations have also been used by other researchers to study inspection performance in a laboratory setting. Since these early efforts, Latorella *et al*. (1992) and Gramopadhye *et al*. (1993) have used low fidelity inspection simulators using computer-generated images to develop off-line inspection training programs for inspection tasks. Similarly, Drury and Chi (1995) studied human performance using a high fidelity computer simulation of a printed circuit board inspection. Another domain which has seen the application of advanced technology is the inspection of X-rays for medical practice.

It seems that most of the relevant work has focused on developing low fidelity simulators for running controlled studies in a laboratory environment. Thus, research efforts need to be extended in order to take full advantage of today's computer technology. Moreover, advanced technology has found limited application for inspection training in the aircraft maintenance environment. Presently, most of the applications of computer technology to training have been restricted to the defense and aviation industries for complex diagnostic tasks. The message is clear: we need more examples of advanced technology applied to training for inspection tasks, examples that draw on proven principles of training. This article describes a university/industry collaborative research effort to develop an off-line computer-based inspection training system for aircraft inspectors. The specific objective of this research was to develop an inspection training system that would help improve the visual search and decision-making skills of aircraft inspectors. ASSIST was developed in cooperation with Lockheed Martin Aircraft Center and Delta Air Lines (figure 1). Here is a brief description.

2. TRAINING FOR INSPECTION

The task analysis is the subject of another article. Along with the trainee analysis, it was used to compare the knowledge and skills required by the task with those of the inspector, in order to determine any gaps which needed to be addressed. Patrick (1992) has identified the training content, the training methods, and the trainee as the important constituents of the training program. Drury (1992) includes the training delivery system as another component of the training program. Although a considerable amount has been written about designing training systems (Patrick 1992, Gordon 1994), very little focuses directly on enhancement of visual inspection skills. Embrey (1979) states that for any training program to be effective, it should address the following three issues: attitude of the trainee at work, knowledge required to perform the job, and the specific skills required to perform the task. Specific training methods which can be used for inspection training (Drury and Gramopadhye 1990, Gramopadhye *et al*. 1997) are described below.
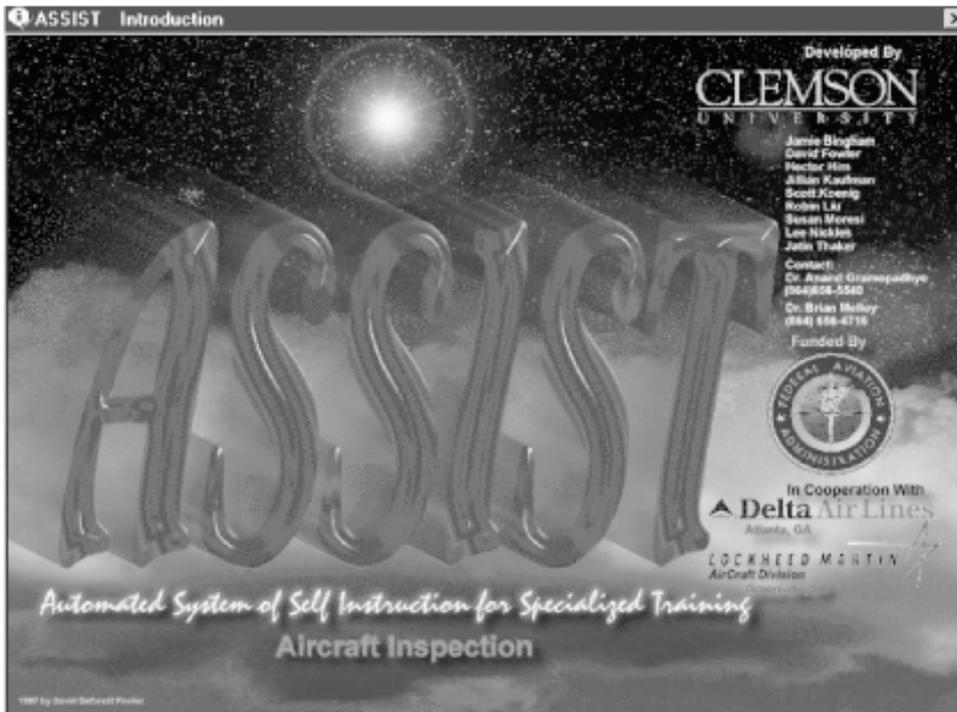
Figure 1. ASSIST title screen.

2.1 Pretraining

Pretraining provides the trainee with information concerning the objectives and scope of the training program. During pretraining, pretests can be used to measure (a) the level at which trainees are entering the program and (b) cognitive or perceptual abilities that can later be used to gauge training performance or progress. Advanced organizers or overviews have proven useful; they are designed to provide the trainee with the basics needed to start the training program. The elaboration theory of instruction proposes that training should be imparted in a top-down manner, wherein a general level is taught first before proceeding to specifics. Overviews can fulfill this objective by giving the trainee an introduction to the training program and facilitating assimilation of new material.

2.2 Feedback

A trainee needs rapid, accurate feedback in order to know whether a defect was classified correctly or a search pattern was effective. Feedback with knowledge of results, coupled with some attempt of performing the task, provides a universal method of improving task performance (Wiener 1975). This applies to learning facts, concepts, procedures, problem solving, cognitive strategies, and motor skills. The training program should start with rapid feedback, which should be gradually delayed until the "operational level" is reached. Providing regular feedback beyond the training session will help to keep the inspector calibrated. Gramopadhye *et al*. (1997) classify feedback as performance feedback and process feedback. Performance feedback on inspection typically consists of information on search times, search errors, and decision errors. Process feedback, on the other hand, informs the trainee about the search process, such as areas missed. Another type of feedback, called cognitive feedback, has emerged from the area of social judgement theory. Cognitive feedback is

the information provided to the trainee about some measure of the output of their cognitive processes. For inspection tasks, process feedback is the same as cognitive feedback.

2.3. Active Training

In order to keep the trainee involved and to help with internalizing the material, an active approach is preferred. In active training, the trainee makes an active response after each piece of new material is presented, e.g., identifying a fault type. Czaja and Drury (1981) used an active training approach and demonstrated its effectiveness for a complex inspection task.

2.4. Progressive Parts Training

Salvendy and Seymour (1973) successfully applied progressive parts training methodology to training industrial skills. In the progressive parts methodology, parts of the job are taught to criterion and then successively larger sequences of parts are taught. If a task consisted of elements E1, E2, E3, and E4, the training would be like this:

- Train E1, E2, E3, and E4 separately to criterion.
- Train E1 and E2, E3 and E4 to criterion.
- Train E1, E2, E3 to criterion and E2, E3, E4 to criterion.
- Train the entire task to criterion.

This method allows the trainee to understand each element separately as well as the links between the various elementsælinks which represent a higher level of skill. On the other hand, reviews of the literature reveal that parts task training is not always superior. The choice between parts task training and whole task training depends on "cognitive resources" imposed by task elements and the "level of interaction" between individual task elements (Gordon 1994). Thus, there could be situations in which one type of task training is more appropriate than the other. Naylor and Briggs (1963) have postulated that for tasks of

relatively high organization or complexity, whole task training should be more efficient than parts task training.

2.5. Schema Training

The trainee must be able to generalize the training to new experiences and situations. For example, it is impossible to train the inspector on every site and extent of corrosion in an airframe so they can detect and classify corrosion wherever it occurs. Thus, the inspector will need to develop a schema which will allows a correct response in novel situations. The key to developing schema is to expose the trainee to controlled variability in training.

2.6. Feedforward Training

It is often necessary to cue the trainee as to what should be perceived. When a novice inspector tries to find defects in an airframe, the indications may not be obvious. The trainee must know what to look for and where to look. Specific techniques within cueing include match-to-sample and delayed match-to-sample. Feedforward information can take different forms, such as physical guidance, demonstrations, and verbal guidance. Feedforward should provide the trainee with clear and unambiguous information which can be translated into improved performance.

3. AUTOMATED SELF-INSTRUCTION FOR AIRCRAFT INSPECTORS

3.1. System Specifications

The computer-based training program was developed using Visual C++, Visual Basic and Microsoft Access. The development work was conducted on a Pentium 120 MHz platform with a 17-inch high resolution monitor (0.28 mm dot pitch, noninterlaced), 32 MB RAM, 2 MB video RAM, ATI Mach 32 VLB advanced graphics accelerator card, 810 MB hard drive, multispeed CD drive, 210 MB Bernoulli drive and a Reveal multimedia kit. The training program uses text, graphics, animation, and audio. The inputs to the system are entered through a keyboard and a two-button mouse.

3.2. System Structure

ASSIST consists of three major modules: (1) general inspection module, (2) inspection simulation training module, and (3) instructor's utilities module. All system users interact through a user-friendly interface. The user interface capitalizes on graphical user interface technologies and human factors research on information presentation (color, formatting, layout, etc.), ease of use, and information utilization.

3.3. General Module

The objective of the general module is to provide the inspectors with a basic overview on the following topics: (1) role of the inspector, (2) safety, (3) types of aircraft, (4) factors affecting inspection performance, and (5) inspection procedure. The module incorporates multimedia (sound, graphic, text, pictures, and video) with interaction opportunities between the user and the computer. Figure 2 shows a prototypical screen of the general inspection module.

3.4. Inspection Simulation Training Module

This module of the training program provides inspection training on a simulated aircraft inspection taskæthe aft cargo bin of a Lockheed Martin L-1011 (figure 3). By manipulating the various
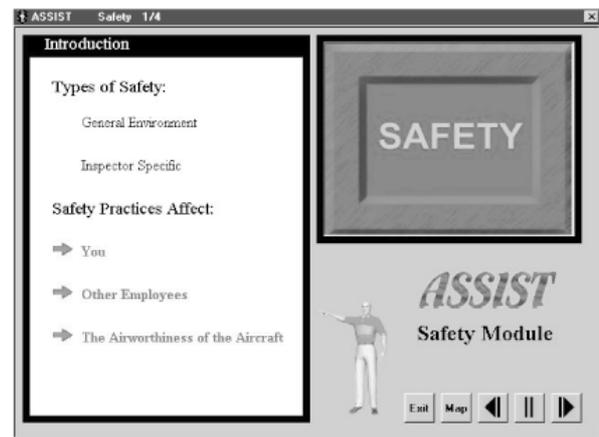
Figure 2. The safety topic of the general module.

task complexity factors, the inspector can simulate different inspection scenarios. The simulation module uses actual photographs of the airframe structure with computer-generated defects.

3.4.1 Introduction

The introduction provides the trainee with an overview of the various facets of the program, the workcard for the inspection assignment, and a graphical representation of various faults. It introduces the trainee to the search and decision-making aspects of the visual inspection task.

3.4.2 Testing

The testing module is designed to operate in two separate modes, with and without feedback. The nonfeedback mode simulates the actual visual inspection task as it would take place on a hangar floor. In either mode, the inspector first locates the defect and indicates this by clicking on the fault. Subsequently, the inspector classifies the defect. In the feedback mode, the inspector is provided with feedback about their own performance on the search and decision-making components of the inspection task. The trainee is also provided with end-of-session performance feedback. The program also features paced and unpaced modes.
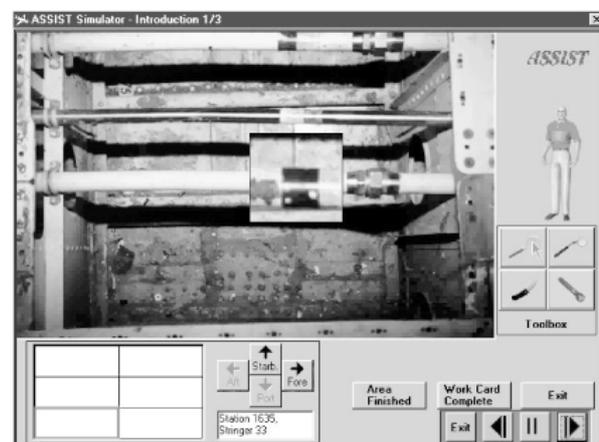
Figure 3. The inspection simulator showing the use of the magnifying glass.

Paced mode allows the inspection to continue for only a specified period of time, whereas unpaced mode allows the inspection task to be unbounded by time.

3.5. Instructor's Utilites Module

This module allows the supervisor or instructor to access the results database, the image database, and the inspection parameter modules. The module is designed as a separate stand-alone tool that is linked to the other modules of the system. The results database allows the instructors to review the performance of a trainee who has taken several training and/or testing sessions. Performance data is stored on an individual image basis and summarized over the entire session so that results can be retrieved at either level. The utility allows the instructor to print or save the results to a file. The objective of the image database module is to provide the instructor with a utility wherein a specific image along with its associated information can be viewed on the computer screen. By manipulating the inspection parameters, the instructor can create different inspection scenarios. The inspection parameter module allows the instructor to change the probability of defects, the defect mix, the complexity of the inspection task, the information provided in the workcard (thereby varying the feedforward information provided), whether the inspection will work in feedback mode or nonfeedback mode, and whether the inspection task is paced or unpaced.

3.6. Inspection Training Session

The training program was designed to use the general principles listed earlier, as derived by the task analysis. A major prerequisite was that it should be a progressive parts training scheme which enabled the inspectors to build their repertoire of knowledge and skills in an orderly manner. A typical training session proceeds as follows.

3.6.1 Initial overview

Initially the subjects use the introduction module, wherein they are introduced to the navigation map, and are familiarized with the operational aspects of the computer program.

3.6.2 General module training

In the general module the subjects are given information on the following five topics relevant to an inspector: role of the inspector, safety, aircraft review, factors affecting inspection, and inspection procedures. Using the navigation map, the subjects can either go directly to a particular topic or subtopic, or follow the default path through the topics. At the end of each topic, a brief quiz is administered to review the subject's understanding of the material. The subjects are given feedback and the correct answers. On completion of the topics in the general module, the subjects take the final test. The final test consists of questions selected from a database and covers material from each topic within the general module.

3.6.3 Simulation module

In the simulation module, subjects are initially introduced to the workings of the simulator. Then they are presented with a workcard containing the instructions for the inspection assignment. Next comes information on defect standards; this includes images of the defects, descriptions, likely locations for particular defects, and possible indicators. After this the subjects conduct inspection using representative images of airframe structures, wherein they must first search for the defect and then classify it as needing maintenance or not. The simulator allows the subject to use various inspection tools: mirror, flashlight, scraping knife, and magnifying glass.

Following inspection, the subjects complete a discrepancy report. They are then given feedback on their overall performance. Feedback is provided on the subject's search and decision-making performance (time to complete inspection, defect detection, defect classification performance, etc). The simulator can be operated in various modes (e.g., with or without feedback, paced or unpaced) and it also allows the instructor to set various inspection parameters (e.g., mix of defects, defect probability, workcard instructions), helping to create different inspection scenarios.

CONCLUSION

The high degree of control that ASSIST affords will create the opportunity to systematize the training. In addition, there are several other inherent advantages that will alleviate the problems characteristic of OJT:

- *Completeness*: Inspectors can be exposed to a wide variety of defects, with varying degrees of severity and at different locations, through the use of a library of defect images. Inspectors can also be trained on less frequently occurring critical defects.
- *Adaptability*: ASSIST can be modified to meet the needs of individual inspectors. Batch files of images can be created to train inspectors on particular aspects of the inspection task with which they have the greatest difficulty. Thus, the program can be tailored to accommodate individual differences in inspection abilities.
- *Efficiency*: Since the training will be more intensive, the trainees will be able to become more skilled within a shorter period of time.
- *Integration*: The training system will integrate different training methods (e.g., feedback training, feedforward training, and active training) into a single comprehensive training program.
- *Certification*: ASSIST can be used as part of the certification process. Since the record-keeping process can be automated, instructors can more easily monitor and track an individual's performance, initially for training and later for retraining.
- *Instruction*: ASSIST could be used by instructors in FAA-certified A&P schools for training. In this manner, aircraft maintenance technicians could gain exposure to defects on wide-bodied aircraft that they might not otherwise have acquired.

This article has described research in the area of aviation maintenance and inspection currently underway at Clemson University. Through the development and systematic application of human factors techniques, the research aims at improving the effectiveness and efficiency of aircraft visual inspection. The results of the research effort have been made available to the aviation maintenance community as deliverable products in the form of usable CD-ROMs. It is anticipated that the use of these products would lead to improved airworthiness of the US domestic aircraft fleet. Subsequent phases of this research will evaluate the utility of ASSIST in an operational setting. As to future implications, human performance models could potentially be used in

conjunction with ASSIST for a wide range of controlled studies: (1) to evaluate the effect of various task and subject factors on aircraft inspection performance, and (2) to identify specific interventions to enhance performance.

ACKNOWLEDGMENTS

REFERENCES

CZAJA, S.J. and DRURY, C.G., 1981, Training programs for inspection. *Human Factors*, 23, 473-84.

DRURY, C.G., 1992, Inspection performance. In Salvendy (ed.) *Handbook of Industrial Engineering* (New York: John Wiley), pp. 2282-2314.

DRURY, C.G. and CHI, C.-F., 1995, A test of economic models of stopping policy in visual search. *IIE Transactions*, 382-93.

DRURY, C.G. and GRAMOPADHYE, A.K., 1990, Training for visual search. *Proceedings of the 3rd FAA Meeting on Human Factors in Aircraft Maintenance and Inspection: Training Issues*.

EMBREY, D.E., 1979, Approaches to training for industrial inspection. *Applied Ergonomics*, 10, 139-44.

FEDERAL AVIATION ADMINISTRATION (FAA), 1991, *Human Factors in Aviation Maintenance æPhase One Progress Report*, DOT/FAA/AM-91/16 (Washington DC: Office of Aviation Medicine).

GORDON, S.E., 1994, *Systematic Training Program Design: Maximizing Effectiveness and Minimizing Liability* (New Jersey: Prentice Hall).

GRAMOPADHYE, A.K., DRURY, C.G., and SHARIT, J., 1993, Training for decision making in aircraft inspection. *Proceedings of the Human Factors and Ergonomics Society 37th Annual Meeting*, pp. 1267-71.

GRAMOPADHYE, A.K., DRURY, C.G., and PRABHU P.V., 1997, Training for visual inspection. *Human Factors and Ergonomics in Manufacturing*, 7(3), 171-96.

LATORELLA, K.A., GRAMOPADHYE, A.K., PRABHU, P.V., DRURY, C.G., SMITH, M.A., and SHANAHAN, D.E., 1992, Computer-simulated aircraft inspection tasks for off-line experimentation. *Proceedings of the Human Factors Society 36th Annual Meeting*, pp. 92-96.

NAYLOR, J.C. and BRIGGS, G.E., 1963, Effects of task complexity and task organization on the relative efficiency of part and whole training methods. *Journal of Experimental Psychology*, 65, 217-24.

PATRICK, J., 1992, *Training Research and Practice* (New York: Academic Press).

SALVENDY, G. and SEYMOUR, D.W., 1973, *Prediction and Development of Industrial Work Performance* (Toronto: John Wiley).

WEINER, E.L., 1975, Individual and group differences in inspection. In Drury and Fox (eds) *Human Reliability in Quality Control* (London: Taylor & Francis).

Balance Theory of Job Design

M. J. Smith and P. Carayon
University of Wisconsin-Madison, USA

1. INTRODUCTION

In 1989 Smith and Carayon-Sainfort introduced a holistic concept of job design that went beyond traditional psychological and ergonomic theories. This concept was elaborated in 1995 to address the issue of new technologies and job design (Smith and Carayon 1995). Smith and Cohen (1997) and Smith et al. (1998) further developed it. Through a general examination of prior work organization, job design and ergonomic theories, an understanding of how work can have psychosocial and psychophysiological effects on the individual was developed. These effects can directly and indirectly influence health, mood state, motivation and performance. In addition, this theory integrated the psychological concepts with physical ergonomic considerations.

In general, micro-ergonomics does not address the complexity of the organizational environment. In the "balance theory" the model of the work system takes into account the multiple aspects of work which can influence employees' performance, safety and health. From a practical point of view many theories of design seem limited because of their narrow focus and their prescriptive guidelines. They lack a "holistic" perspective and flexibility. For instance, the job characteristics theory (Hackman and Oldham 1976) specifies a set of job features necessary for good job design such as task identity, feedback, significance, variety and autonomy. The job strain model of Karasek (1979) proposes two main job characteristics that affect job stress, workload and decision latitude. Our experiences as industrial engineers have led us to question the wisdom of such prescriptive approaches. The approaches fail to address the unique characteristics of particular organizations, their unique workforce, their products and their customers. We have found that in many situations, rigid, prescriptive approaches to job design are not effective and are often not accepted by managers or employees. For instance, how can variety in work be increased when there is only a limited number of tasks to perform? How can decision latitude be increased when management is autocratic and will not share power, or when the employees do not have the knowledge of skills to manage? These real-life experiences convinced us that a job design model needs to be holistic and flexible.

2. THE BALANCE MODEL

The following is a concept about job design that integrates the psychosocial and biological aspects of employee behavior within an ergonomic framework. This model proposes that working conditions (and other environmental features) produce "loads" on the person that challenges biological resources (energy expenditure, biomechanical strain, stress responses) and psychological resources (perception, cognition, decision-making, distress). The characteristics of the load produce physiological and psychological consequences such as hormone release, muscular action, perceptions and mood states. Individual's

physical capacity, health status and motivation influence the responses to the load. The psychological responses are the product of personality, past experiences and social situation. These physiological and psychological reactions to the "loads" act as motivation for employee behavior to respond to the "loads." The response could be increased or decreased performance, or "coping" behaviors (adaptive or maladaptive).

In this theory working conditions impose the "loads" that bring about the individual's physiological and psychological reactions. The physiological reactions caused by the load produce a strain on the person if they exceed the available biological resources, such as energy resources or mechanical strength. For instance, a repetitive lifting task with heavy materials being carried out for hours without rest can exhaust energy resources and produce local muscular fatigue. This can diminish strength and affect lifting style. Or shiftwork can disrupt biological rhythms, which increases the demand on the body's energy resources. It can influence behavior such as eating and sleeping, which affect energy resources and fatigue. When both of these load are present the effects are greater and happen sooner.

Working conditions can also cause psychological reactions that have emotional, behavioral and biological consequences. These consequences are primarily determined by the individual's perception of her/his ability to meet the demands imposed, and/or her/his perception of the "acceptability" of the working conditions. In addition, the person's availability of psychological and behavioral resources, such as motivation, cognitive capacity and coping behaviors influence the consequences. However, the physiological and psychological reactions are not independent of each other. They interact and may even reinforce each other. For instance, the repetitive lifting task may cause boredom, which leads to mental fatigue. It may also produce physical fatigue by depleting energy resources. These effects reinforce each other resulting in a systemic response of "general" fatigue (Grandjean 1968).

Physical, psychological and behavioral resources are not a fixed and stable set of individual characteristics but change over time and are influenced by capacity, motivation, stress responses and the demands of the working conditions. For instance, these resources may increase because of on-the-job training or the availability of powered assistance machinery. In the repetitive lifting example the introduction of equipment to do the lifting will reduce the energy requirements and the local muscle fatigue, while training in proper lifting techniques may also reduce these same outcomes. Stress responses can influence the biological and psychological resources available to an individual. Individual personality characteristics, genetic make-up and health status all influence the physical resources available to the individual and the nature of the stress responses.

The load on the individual can be influenced by the physical demand, psychological response to the demand as mediated by perception, or both. When the load becomes too great, the person displays stress responses which are emotions, behaviors and biological reactions that are maladaptive. When these reactions occur frequently over a prolonged period they lead to health disorders. Thus, chronic exposure with cumulative reactions is a hallmark of distress. Cumulative stress responses reduce the available resources for dealing with the loads from the work environment, and a circular effect begins. This repeated circular

cycle leads to a breakdown in individual resources unless external resources are made available or the environmental load is reduced.

According to the balance theory the effects of the work system on the individual are assumed to be mediated by the stress load, which is both physical and psychological. The stress load is the resultant internal effect of the work system on the individual. The physical and psychological dimensions of the stress load are not independent. They interact with each other to produce effects. These effects have influences on the quality of working life, performance, strain and health.

Figure 1 (Smith *et al*. 1998) was developed from the concepts of Smith and Carayon-Sainfort (1989). It illustrates a model for conceptualizing the various elements of a work system and defines categories of loads that working conditions can exert on workers. At the center of this model is the individual with his/her physical characteristics, perceptions, personality and behavior. The individual has technologies available to perform specific job tasks. The capabilities of the technologies affect performance and also the required skills and knowledge needed for its effective use. The task requirements also affect the skills and knowledge needed. Both the tasks and technologies affect the content of the job and the physical demands. The tasks with their technologies are carried out in a worksetting that comprises the physical and the social environment. There is also an organizational structure that defines the nature and level of individual involvement, interaction and control.

This model can be used to establish relationships between job demands and job design factors, and loads. In this model these various elements interact to determine the way in which work is done and the effectiveness of the work in achieving individual and organizational needs and goals. This is a systems concept in that any one element will have an influence(s) on any other element(s). Demands are placed on the individual by the other four elements (environment, technology, organizational design, task) which create loads that can be healthy or harmful (motivating or stressful). Harmful loads lead to stress responses that can produce adverse health effects. Each element in this model and some examples of the potential adverse aspects are described below.

Figure 1 also displays the relationship between the work system and the employee. This can be used to examine the effects of job design factors, such as workload and task content, on employees' behavior, performance and health. In this model the various elements interact to determine the way in which work is

accomplished, and the way the work affects individual and organizational needs and goals. This is a systems concept in that any single element will have influences over any other element. Demands are placed on the individual by the environment, task requirements, the organizational structure and the technology used. These demands can match an employee's capacity and capabilities to produce a positive fit, or exceed the employee's capacity and capabilities and cause misfit. Chronic misfit leads to poor performance and/or ill health. Each element of the model and some examples of the potential effects are described below.

- Environment: various aspects of the physical environment have been implicated as job stressors. Noise is the most well-known environmental stressor that can cause increases in arousal, blood pressure and negative psychological mood. Environmental conditions, general air quality and housekeeping have been shown to affect energy expenditure, heat exchange, stress responses and sensory disruption, which make it more difficult to carry out tasks increase the level of worker stress and emotional irritation.

- Task: aspects of the task have been studied more than any other element in this model. A wide variety of influences have been studied from content considerations such as repetitiveness and meaningfulness, to workload issues such as overload and underload. Low task content, lack of control, high demand and machine-pacing lead to low motivation, stress and diminished performance.

- Technology: lack of adequate skills to use the technology leads to poor motivation, stress and diminished performance. Fear over job loss due to replacement by technology reduces motivation and increases stress. On the other hand when new technology is applied appropriately it can enhance job content and skill use which leads to increased motivation and performance with decreased stress. The physical characteristics of the tools and technology can put physiological loads on the employee. For instance, poor workstation design leads to unhealthy postures and movements and diminished performance.

- Organizational factors: the organizational context in which work tasks are carried out often has considerations that influence worker motivation, stress and performance. The way in which workers are introduced to new technology and the organizational support they receive, such as training and time to acclimate, has been related to stress and performance. The ability to grow in a job and to be promoted (career development) affects motivation and stress. Potential job loss influences motivation, performance and stress. Other organizational considerations such as work schedule (shiftwork) and overtime have been shown to have negative mental and physical health consequences, as does role conflict and role ambiguity.

- Individual: a number of personal considerations determine the physiological and psychological responses that the preceding elements of the balance model will produce. These include personality, physical health status, skills and abilities, physical conditioning, prior experiences and learning, motives, goals and needs.

The five elements of the balance model system work in concert to provide the loads and the resources for achievement of individual and organizational goals. We have described some of
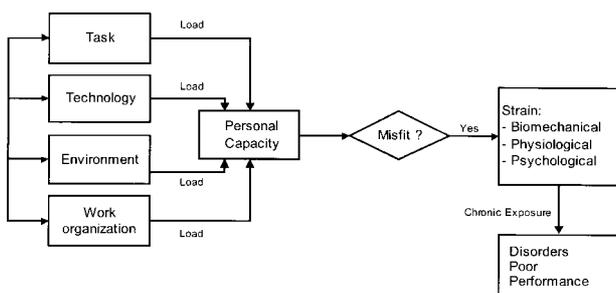

Figure 1. Balance model of misfit (from Smith *et al*. 1998).

the potential negative attributes of the elements in terms of motivation, performance and job stress, but there are also positive aspects of each that can counteract the negative influences. For instance, increased worker training can offset the negative influences of inadequate skill to use new technology. Or the adverse influences of low job content can be balanced by an organizational supervisory structure that promotes employee involvement and control over the tasks. Jobs with many negative elements are jobs that produce the most adverse impact on the employee, whereas jobs that have better balance are less stressful (Carayon 1994).

The essence of this theory is to improve motivation and performance and reduce stress and the negative health consequences by "balancing" the various elements of the work system to provide positive aspects to counter the negative ones. The "best" job design can be achieved by providing all characteristics of each element of the model that can meet recognized criteria for worker ego needs fulfillment and that set proper physiological and psychological loads to eliminate stress and strain. In reality such a perfect job is not attainable. Then, this model proposes using good elements to compensate for poor aspects in other elements to balance the "loads" to achieve reduced stress and enhanced motivation and performance.

Various theories of job design can help us define the positive and negative characteristics of the work system. For instance, theories of occupational stress have defined work stressor which are negative characteristics, such as high workload, shiftwork, low job control, high role ambiguity and role conflict (Smith 1987). Theories of job design have also specified positive characteristics such as high task variety, feedback, opportunities for learning, and autonomy (Herzberg 1966, Hackman and Oldham 1976). Ergonomic models have also defined negative characteristics or work, and their interactions with the individual. generic risk factors have been defined, such as repetitiveness of motions, forceful motions and poor postures.

3. ACHIEVING A BALANCED WORK SYSTEM

There are two aspects of "balance" that need to be addressed: (1) the balance of the total system and (2) compensatory balance. System balance is based on the idea that a workplace or process or job is more than the sum of the individual components of the system. The interplay among the various components of the system produce results that are greater (or lesser) than the additive aspects of the individual parts. It is the way in which the system components relate to each other that determines the potential for the system to produce positive results. If an organization concentrates solely on the technological component of the system, then there is an "imbalance" because the personal and psychosocial factors are neglected. Thus, job improvements must take account of and accommodate the entire work system. Our model of the work system can be used to establish relationships between job demands, job design factors and stress.

The second type of balance is "compensatory" in nature. It is seldom possible to eliminate all psychosocial factors that cause stress. This may be due to financial considerations, or it may be because it is impossible to change inherent aspects of job tasks. The essence of this "balance" is to reduce psychological stress by making changes in aspects of work that can be positively changed to help improve those negative aspects that cannot be changed.

In one strategy, proper job design can be achieved by providing all of those characteristics of each work element that meet recognized criteria for physical loads, work cycles, job content, control and socialization, and that provide for individual physiological and psychological needs. With this approach, the best designs will eliminate all sources of stress. However, such a perfect job cannot often be achieved in reality, so a second strategy is to use positive work elements to compensate for poor work elements which can balance the stress by moderating those negative factors to reduce the total demands (loads).

4. ADVANTAGES OF THE BALANCE MODEL

Our model defines the process by which working conditions at different levels (i.e. individual, task, environment, technology, organization) can produce loads that can lead to poor outcomes such as low motivation, diminished performance, increased stress and poorer health. It also proposes a system that helps balance these loads to produce better outcomes. When balance cannot be achieved through changing the negative aspects of an element, then it can be improved by enhancing the positive aspects of other elements of the job. Thus, the good aspects of work can be used to "counter-balance" the bad.

A major advantage of this model is that it does not highlight any one factor such as shiftwork, or a small set of factors such as demand and control. Rather it examines the design of jobs from a holistic perspective to emphasize the potential positive elements in a job that can be used to overcome the adverse aspects. Thus, all aspects of the job must be considered in developing a proper design. This model does not subscribe to only one approach for job design such as content enrichment or participation. Both approaches may have some positive benefits given the right circumstances. In fact, it is likely that there will be circumstances in which both approaches can be used in concert to provide less stressful work. This model is similar to an "organizational development" approach in that it uses one or more aspects or elements from many different theoretical perspectives to solve specific problems. The emphasis differs in that stress and not productivity is the outcome of interest and, therefore, may direct different interventions than an "organizational development" approach.

Another advantage of this model is that it provides a research direction that can be fruitful in defining the causes and cures of occupational stress. Most current job stress research is directed toward one or two critical job factors as the primary elements in producing worker stress. The results of such research have proven connections between specific factors and workers stress responses but in each case have only accounted for a small portion of the total effect. To us this suggests that a multifactor theory will be most fruitful in defining how job conditions influence stress. One simple way to look at multiple factors is to use an engineering analogy to the total load that leads to some level of stress which if sufficiently great can produce strain.

The balance theory is a total systems viewpoint that shows how to integrate the diverse workplace job design factors into a characterization of positive work design. We think that our concept of "load" which is a combination of physical and psychological attributes of the environment does this in a logical way. Much research has to be carried out to define how the various on and off the job factors can be fit together to determine the

"stress load" of a job. Some factors may always have a greater influence in promoting an "unhealthy load", or situational characteristics may influence the specific contribution that a factor makes to the total load. Only research and experience can answer this. But for the time being our approach offers a clear direction for job design.

REFERENCES

CARAYON, P., 1994, Stressful jobs and non-stressful jobs: a cluster analysis of office jobs. *Ergonomics*, 37, 311–23.

DAVIS, L.E., CANTER, R.R. and HOFFMAN, J., 1955, Current job design criteria. *Journal of Industrial Engineering*, 16, 1–7.

FRENCH, J.R.P. JR, 1963, The social environment and mental health. *Journal of Social Issues*, 19, 39–56.

GARDELL, B., 1977, Autonomy and participation at work. *Human Relations*, 30, 515–33.

GRANDJEAN, E., 1968, Fatigue: its physiological and psychological significance. *Ergonomics*, 11, 427–36.

HACKMAN, J.R. and OLDHAM, G.R., 1976, Motivation through the design of work: test of a theory. *Organizational Behavior and Human Performance*, 16, 250–79.

HERZBERG, F., 1966, *Work and the Nature of Man* (New York: Thomas Y. Crowell).

KARASEK, R.A., 1979, Job demands, job decision latitude, and mental strain: implications for job redesign. *Administrative Science Quarterly*, 24, 284–308.

LEVI, L., 1972, *Stress and Distress in Response to Psychosocial Stimuli* (New York: Pergamon).

MASLOW, A.H., 1970, *Motivation and Personality* (2nd ed.) (New York: Harper & Row).

MCGREGOR, D., 1960, *The Human Side of Enterprise* (New York: McGraw-Hill).

SELYE, H., 1956, *The Stress of Life* (New York: McGraw-Hill).

SMITH, M.J., 1987, Occupational stress. In G. Salvendy (ed.) *Handbook of Human Factors* (New York: Wiley), 844–60.

SMITH, M.J. and CARAYON, P.C., 1995, New technology, automation and work organization: stress problems and improved technology implementation strategies. *International Journal of Human Factors in Manufacturing*, 5, 99–116.

SMITH, M.J. and CARAYON-SAINFORT, P., 1989, A balance theory of job design for stress reduction. *International Journal of Industrial Ergonomics*, 4, 67–79.

SMITH, M.J. and COHEN, W., 1997, Design of computer terminal workstations. In G. Salvendy (ed.) *Handbook of Human Factors and Ergonomics* (New York: Wiley), 1637–88.

SMITH, M.J., KARSH, B. and MORO, F., 1998, *A Review of Research on Interventions to Control Musculoskeletal Disorders* (Washington, DC: US National Academy of Science) [www/nas/edu].

TAYLOR, F.W., 1967, *The Principles of Scientific Management* (New York: Norton & Co., 1911).

TRIST, E. and BAMFORTH, K.W., 1951, Some social and psychological consequences of the longwall method of coal-getting. *Human Relations*, 4, 6–38.

Changes in Modern Manufacturing Practices

D. Holman, C. Axtell, C. Clegg, K. Pepper, P. Waterson, I. Cantista, and M. Older–Gray

Institute of Work Psychology, University of Sheffield, UK

1. INTRODUCTION

Manufacturing organizations compete in an environment that is characterized by uncertainty, increased global competition, the fragmentation of markets, an increasing dependence on non-price competitiveness, and a high level of technological change (Bessant 1991). To survive and compete in such an environment, there is widespread agreement on the importance of an organization's ability to innovate and manage change (Burnes 1993; Wolfe 1994). Recent research would seem to suggest that UK companies are responding positively to the need for change and innovation. For example, a recent CBI/NatWest survey of innovation in UK companies found that 50% of manufacturing companies had introduced between two and five new processes in the last year, 15% had introduced between six and ten processes, and 20% had introduced over ten processes. Only 8% had introduced one new process, and 3% did not introduce any new processes (CBI/Nat West 1998). Moreover, a recent survey of 564 UK manufacturing companies across 18 sectors, has shown the widespread use of modern manufacturing practices, such as total quality management, team-based work, just-in-time, supply chain partnering, and outsourcing (Waterson *et al.*, in press).

At one level, then, it would appear that UK companies are responding well to the present competitive environment. Yet, although the use of modern manufacturing practices is fairly widespread, the expected improvements in performance have often failed to materialize (Webster and Williams 1993; Wemmerlov and Johnson 1997; Waterson *et al.,* in press). For example, a survey of IT-based practices found that approximately 80% of cases failed to meet their objectives and that 40% of systems failed outright (Clegg *et al.* 1996). There are clearly a number of reasons why an innovation will be successful or not, but it has been argued that one key factor is the effective management of the change and innovation process (Voss 1992; Tidd 1991). The change management literature suggests that there are well-established principles of change that can enable the successful introduction of manufacturing practice innovations (Jaikumar 1986; Voss 1992; Bessant 1991; Tidd 1991; Dawson 1994). They include:

- Conducting a strategic review.
- Considering a variety of objectives and not solely focusing on cost reduction or technical objectives.
- Taking the long-term view and ensuring a sustained level of interest by management.
- Ensuring resources are adequate, particularly with regard to training.
- Communicating and selling the change.
- Using project planning tools and cross-functional teams.
- Using champions of change.

- Considering the systemic effects that an innovation might have, particularly human resource issues.
- Encouraging the participation of all in change.
- Examining the expertise within the firm.
- Evaluating the change process.

However, empirical research on the implementation of modern manufacturing practices has demonstrated that many of these principles are not necessarily practiced (Bessant 1991; Voss 1992; McLoughlin and Harris 1997), and that this can contribute to a considerable reduction in the overall effectiveness of the manufacturing practice being introduced (Kearney 1989; Wemmerlov and Johnson 1997). For example, a lack of planning or project management, a failure to think systemically about the impact of the practice, especially in relation to human resource practices, and a failure to consult and involve end-users have all been shown to reduce the effectiveness of the manufacturing practice (Webster and Williams 1993; Fawcett and Scully 1995; Clegg *et al.* 1996). It is interesting to note that similar results have been found across a range of manufacturing practices, organizational sizes, and sectors.

By showing that the principles of change management may not be widely practiced, the research conducted on UK samples (e.g. Kearney 1989; Clegg *et al.* 1996, clearly calls into question the "change and innovation capabilities" of UK companies with regard to the implementation of modern manufacturing practices. Capabilities are the skills, routines, and ways of doing things that are built up and learnt over time by organizational members (Collis 1994; Bessant *et al.* 1996). Change and innovation capabilities may therefore be thought of as the skills and routines that organizational members have learnt which enable them effectively to manage the change process. While present research provides many important insights, e.g. in establishing principles of change and illuminating the relationship between the change process and the outcomes of change, it does not provide data on the nature of the change process or the change and innovation capabilities of UK companies more generally. This is primarily due to the fact that existing studies have been based on case studies of a small number of firms, and on surveys of one type of manufacturing practice. This limits, to some extent, the generalizability of the findings. Given this limitation of previous research, the main aim of this research was to examine the change and innovation capabilities of UK manufacturing companies across a range of companies in different sectors, and with regard to a variety of modern manufacturing practices.

This article is divided into three further parts: the methodology describes the rationale behind the research design, the experience of the participants, and the range of companies and practices covered. This is followed by the findings of the research and a conclusion.

2. METHOD

To achieve the aims of the research, experts drawn from research, consultancy, professional bodies, and manufacturing companies were interviewed, all of whom had extensive experience of managing or studying the change and innovation process in manufacturing companies. The "expert panel" approach was chosen as it enables the collection of data about a larger number of companies, and across a wider variety of manufacturing practices, than case study approaches, while also enabling

descriptively richer accounts of the change process to be gathered than is possible by a questionnaire survey.

Data was collected by interview using a semi-structured interview schedule that had three main sections. The first part focused on the participants' experience and expertise. In particular, interviewees were asked about the type and number of companies they had worked for or researched in, and what type of manufacturing practices they had most expertise of. Interviewees were asked to indicate from a list of twelve practices (see table 1, which type of practice they had most experience of. (Between them, interviewees had experience of all 12 practices, although less experience was evident in the area of concurrent engineering). This list was based on a detailed evaluation of previous manufacturing surveys in the UK, and on earlier research that classified the variety of modern manufacturing practices (Bolden *et al.* 1997). This list was also used successfully in a previous survey (Waterson *et al.* 1999). Once the interviewees

Table 1 List of modern manufacturing practices used in the study

Practice	Definition
Just-in-time production	Making products in direct response to internal or external customer demands (rather than building in advance to maintain stock levels).
Manufacturing cells	Organising the shop-floor so that one or a group of operators has the resources to produce a whole product (rather than routing the product through different functional areas).
Intergrated computer-based technology	Linking together computerized equipment to enable enhanced integration. (Such as CADCAM, CIM and FMS rather than stand-alone CAD or CNC machinery).
Concurrent Engineering	Simultaneously designing and manufacturing products (rather than keeping design and production as seperate functions).
Total production maintenance	Involving all operators in minor maintenance and repairs (rather than having all maintenance done by engineers).
Team-based work	Placing operators into teams with their own goals and giving the team freedom to allocate work between members (rather than dividing labour according to job title).
Decentralised decision-making (Empowerment)	Passing considerate responsibility for operational management to individuals or teams (rather than keeping all decision-making at the managerial level
Learning culture	Providing a range of development oppurtunities for employees, not merely training them for their present job
Outsourcing	Contracting out certain manufacturing processes and sub-processes to other companies (rather than making everything in-house).
Supply-chain partnering	Developing strategic alliances and long-term relationships with suppliers and customers (rather than negotiating on a short-term basis)
Total quality management	Seeking continous change to improve quality and making all staff responsible for the quality of their work. (Such practices include Kaizen and Continous Improvement).
Business process re-engineering	Radically re-designing or rationalizing production processes to eliminate unnecessary practices and procedures (rather than the more gradual process of continual improvement).

had selected the practices that they had most experience of, they were asked to concentrate on these practices for the rest of the interview. The second part of the interview schedule was based on process models of innovation. These attempt to understand the manner in which innovations (an idea, practice, or artifact, are adopted, implemented, incorporated, and evaluated within organizations (Van de Ven *et al.* 1989; Wolfe 1994). They are also concerned with very similar issues to that in the change management literature — for example, the extent and nature of planning, and the extent and nature of evaluation. Process models of innovation, therefore, provided an appropriate means with which to structure both the interview and generate data useful to the aims of this study. The main topics covered in the second part of the interview were (see Appendix 1 for complete interview schedule):

- the reasons why companies adopt particular manufacturing practices;
- the implementation process;
- the degree to which the practices are successfully incorporated into the organization;
- how companies evaluate the process of implementation.

The third part of the interview focused on the factors that interviewees thought could improve the change and innovation process in UK companies.

Each interview lasted on average 1–1.5 hr. Where possible, all interviews were tape-recorded and fully transcribed. Using NUDIST, a software package that facilitates qualitative analysis, each interview was then coded according to a pre-defined coding scheme. The coding scheme was based on the interview structure. Early analysis of the data indicated that there were variations in response between the different types of interviewee. The analysis therefore started by separating responses according to interviewee type.

Different sampling rationales were chosen for the various types of participant. The researchers, consultants, and members of professional bodies were selected as some of the leading experts in this field in the UK, and to cover the leading groups with expertise in this area. Researchers were selected to reflect various disciplinary backgrounds and expertise in researching different manufacturing practices. Consultants were selected from the leading consultancies as well as from smaller specialists in manufacturing. Professional bodies were chosen to reflect the different interest groups in manufacturing industry. The company practitioners, manufacturing managers, and directors were chosen from a database of participants in an earlier survey (Waterson *et al.* 1999). This database, derived from the Dun and Bradstreet Database of UK companies, covered 15 manufacturing sectors (see Appendix 3, and four sizes (150–249, 250–499, 500–999, 1000+ employees). While we were limited in the number of companies we could select, attempts were made to include one company from each sector, as well as to select companies of varying size.

Forty-eight people were interviewed. By occupation there were 16 researchers, 9 consultants, 17 company practitioners, and 6 representatives from professional bodies. As a group, participants had experience of work in all 15 manufacturing sectors, although we were unable to obtain company practitioners from the "clothing" and "rubber and plastic" sectors. Participants also had experience of working in all sizes of company. However,

participants, particularly the consultants, had less experience of working in small organizations, and some participants had particular expertise in working with small companies. Consultants tended to have worked with more companies than researchers and practitioners. The average number that consultants had worked in was 300. Researchers had researched or consulted in from 20 to 300 companies, the average being about 50. Company practitioners had worked in an average of seven companies. Members of professional bodies had contact with hundreds of companies. In total, interviewees were able to draw on knowledge of approximately 10,000 organizations.

3. FINDINGS

The major findings are described below, and organized into the following five parts:

3.1. Adoption

Interviewees were asked about the reasons why companies adopted new manufacturing practices. The clear message from interviewees was that companies often adopt a new manufacturing practice in reaction to a "significant event" (which some labeled a "crisis", such as declining market share, the loss of a major customer, or poor profits. Another, but less common trigger, is pressure from a major customer, particularly in relation to the introduction of total quality management and supply chain partnering. Less common still are the more proactive approaches to change, and these were thought only to occur in the "better" companies. Proactive approaches involved continuously seeking out new ways of doing things through, for example, scanning the market place, benchmarking, and evaluating the usefulness of new methods.

One researcher posited that, although the pressures for change are common to adopters and non-adopters, what differentiates the two are the forces against change that prevent the innovation process from starting in the non-adopters. All interviewees mentioned that one of the main reasons why change does not occur is that the risks are seen to be too high, especially among senior management. The key risks were generally perceived to be financial (e.g. loss of profits, or personal (e.g. loss of personal credibility). Risks were thought to be higher if the scale of change is large and the practice relatively new (e.g. company-wide BPR). Twelve interviewees with extensive experience of IT-based practices suggested that management seriously underestimated the level of risk involved in their adoption.

Other forces against change are "initiative fatigue", a history of failure and abandonment, a lack of knowledge by management about alternative working practices, and a preoccupation with operational firefighting. Twenty-two respondents commented that a significant barrier to the introduction of new manufacturing practices in the 1990s is middle management. This is especially the case when a new practice challenges middle managers' traditional job roles and calls for new forms of managerial control. Commenting on the introduction of team working one manager noted that:

> there was a perceived risk [by managers] that the more you empowered people the less control you have, and when you are used to having all the control, [losing it] was very much a fear in a lot of their eyes.

By contrast, only two respondents thought that the shop floor, whether unionized or non-unionized, was a significant and consistent inhibitor of change. This may be a reflection of unions' reduced power, but five interviewees actually stressed the positive role that unions can play in promoting change.

Whether a reactive or proactive adoption of a manufacturing practice, the following four factors were regarded as important in the adoption stage, and it can be noted that these are similar to those suggested by the change management literature:

i. *A strategic review*. A strategic review was considered to be fundamental to the success of a new manufacturing practice, although researchers and consultants were particularly skeptical about the extent to which this occurred in UK companies. For example:

What's missing is a real strategic sort of outlook in terms of what is the nature of their business, on what basis are they going to compete [and] what are the drivers of competitiveness. They don't go through that process at all (Researcher).

Concern was expressed by researchers and consultants about the general level of strategic ability in senior managers in the UK More critical comments suggested that managers may only be getting better at telling "strategic tales".

ii. *The consideration and specification of various objectives.* The best adoption practices were believed to be characterized by companies considering a variety of objectives (e.g. cost, quality, customer responsiveness, the environment). The evidence from interviewees suggests that while more and more companies are starting to consider a wider variety of objectives, management's prime objective for the introduction of many practices is cost reduction. Indeed, six interviewees suggested that there is a pervasive obsession with cost-cutting, even in the more strategically thought-through implementations. Such an obsession was seen to obscure the alternative ways in which competitive advantage might be achieved. Consultants and researchers cited many examples of strategic reviews in which they had been involved where companies realized that cost-cutting was not such an important driver of competitive advantage as they had first thought.

iii. *Participation in the strategic and operational review*. Strategic and operational reviews were seen by some fifteen respondents to be improved by the involvement of middle management and other members of the organization (e.g. unions). When middle management are involved, two main benefits were thought to accrue. First, operational issues are better considered in strategy formulation. Second, there is less chance of middle management being unsure of the strategic objectives of change, and configuring the change process to suit purely operational objectives.

iv. *A long-term view*. There was a common consensus that senior management needed to take a long-term view of the introduction of new manufacturing practices. There was much criticism about "short-termism' and the pressure for quick fixes and quick returns. In some instances, the pressure for quick returns, particularly on cost, was seen to negate the value of any strategic planning. In most cases though, short-termism was seen to be symptomatic of a lack of strategic thought and a mechanistic view of organizations in which practices could be simply plugged in and made to work almost immediately. About two-thirds thought that

manufacturing practices often took three or four years to fully realize their potential, and that this was especially the case with team work, empowerment, learning culture and total production maintenance. For example:

In Total Production Maintenance . . . you won't get much under three to five years, and people are hoping that TPM will be paying for itself within a year or two. I think in many cases it has been a forlorn hope . . . I think it's all pressure to get a pay back on whatever investment you've put into the work (Practitioner).

The desire for quick fixes was believed to lead managers to fail to realize that symbolic and material resources would be required beyond the initial adoption and implementation phases. Forty-six respondents, nearly the whole sample, noted that the pressure for quick fixes was intense in the current economic climate.

Interviewees were asked about whether they thought practices were introduced because they were the latest fad. While about one-third agreed with this statement, the rest, especially company practitioners, were skeptical about the degree to which the implementation of practices could currently be described as faddish, although they did recognize that this may have been more true in the past. What emerged from the interviews was that the problem was not so much with the fads themselves, but how they were adopted and implemented. Thus, a more positive interpretation of "faddism" is that it does bring new ideas and practices to the attention of managers. It gets new practices onto the innovation agenda, which if adopted and implemented well, can be of immense benefit. The willingness of UK companies to adopt new practices is thus an indication of their innovativeness. Organizations such as the Department of Trade and Industry, the Engineering Employers Federation, Higher Education Institutions, and consultancies play an important role in disseminating knowledge of these new innovations. "Fads" are only problematic when they are adopted without careful reflection and attention to strategy, and with the belief that they are a universal panacea which will produce quick returns.

3.2. Implementation.

In response to questions about the process by which manufacturing practices are implemented, the following factors were thought to be some of the most important in successful implementations: participation of all stakeholders; an examination of business processes; flexible project planning; champions of change; a systemic view of change; the use of tools (e.g. soft systems analysis, sociotechnical theory, that facilitate a systemic approach; and, the selling and communication of change. The fact that many of these are well-established "principles" led one-quarter of interviewees, mainly consultants and researchers, to wonder why these well-known principles of change are not widely practiced. The evidence from this study suggests that UK companies are getting better at implementing changes, that "the principles of change" are becoming more widely practiced. In particular the evidence suggests that companies are generally improving at:

- project planning, i.e. the "mechanics" of change;
- using champions of change;
- consulting end users;
- communicating and selling the reasons for change;

- recognizing that change impacts on human resource practices, particularly payment systems;
- choosing and using consultants.

The evidence from interviewees also suggests that, in general, companies could benefit from:

i. *Taking a more systemic view of change.* Interviewees were asked whether a range of topics were considered during the planning of change (see Appendix 2). The evidence from their responses indicates that change tends to be technology or technique led. For example:

Change very much seems a technical matter. So personnel, human resource issues and the personnel function tend not to be playing leading roles [and] get involved after the event to deal with "human" problems (Researcher).

As this quote illustrates, other aspects of the organization, such as human resource practices, job and work design, accounting systems and supply chains are often considered late in the change process, or only when it is clear that a problem has arisen. One result of this is that changes to these systems are rushed through with little consultation and participation. Another outcome is that the resource implications of such changes are not costed in the project plan. A consequence being that such changes are under resourced and inadequately implemented.

ii. *Understanding change as messy, iterative, political and emotional.* Companies are getting better at managing the mechanics of change (project planning, etc., but many tend to see change as linear and mechanistic. This can lead to change being simply cascaded down the organization in an autocratic manner and passed from one level to the next, and a lack of review and reflection. It was also noted that little explicit recognition is given to the fact that change is a political process involving the construction of alliances, and a highly emotional event that can cause anxiety about future work roles and job security.

iii. *Using a range of tools.* A number of interviewees, five in total, suggested that many project planning tools focus too much on the timing and sequencing of change. While this is useful, such tools can encourage mechanistic and linear approaches, and make little or no explicit provision for promoting systemic thought, or for encouraging participative methods.

iv. *Encouraging end user participation rather than just consultation.* The degree of involvement by end-users (e.g. the shop floor , is often limited to consultation towards the end of the implementation process, rather than participation throughout the whole change process. This was mentioned by sixteen respondents, ten of whom were researchers.

v. *Understanding the impact on middle management.* Thirty-three interviewees noted that many new manufacturing initiatives had a significant impact on middle managers' career paths, status, identity, job security, job roles, appraisal, remuneration, and the forms of control that they can exercise. Only a few cases were cited where this had been explicitly considered in the change process.

vi. *Attending to other practices whilst new ones are being introduced.*

My experience tells me that [manufacturing practices] fail in the long term because something else comes along and it's "Let's put that to one side and do this". That's not a sustainable culture, that's a series of projects (Practitioner).

This quote illustrates that existing manufacturing practices were often felt to be ignored while new ones are being introduced. This can lead to the possibility of older practices decaying or being abandoned.

vii. *Attending to the management of learning and expertise*. Nine participants stressed the need for companies to play closer attention to the promotion of learning during the implementation process, e.g. through training and periods of review and reflection. It was also thought that organizations need to play closer attention to the way in which change affects expertise, e.g. the possible disruptive impact of a new manufacturing practice on expertise, or, the degree to which the expertise of external experts is transferred internally.

When asked about the extent to which changes were completed on time and on budget, interviewees found it difficult to provide specific figures as deadlines and budgets were sometimes not set, or set and subsequently revised. Forty-four interviewees thought that projects were nearly always late and rarely within budget. Practitioners tended to perceive higher rates of success, suggesting that 50% of practices were on time and within budget a lot. Researchers, consultants and professional advisors thought that only 20% of practices were on time and within budget a lot.

3.3. Incorporation

The best examples of change were found to be characterized by a sustained effort by management and the shop-floor to ensure that a practice became embedded into everyday organizational life after the "big bang", i.e. when it had gone live. Thirty-five interviewees stated that, in a significant number of cases, adequate and sustained support, particularly by managers, did not occur. The lack of continued support is often compounded by a "quick-fix" mentality which expects large and exciting returns. The best companies were also thought to recognize that setbacks and failures would occur after the initial implementation. Indeed, it was noted that with practices such as TQM and JIT, if problems do not surface, then the practice is not working properly. As one researcher noted: "Our experience with JIT is that it's high risk because you want the system to break down, find out where the problems are [that] you've got to fix."

Furthermore, setbacks and failures are treated as learning events. This, as one member of a professional body posited, differs from the norm in which companies "have a crack at something, fail, then don't say we can learn from our failure. They just simply cut it off there and then."

Many interviewees, thirty-seven in all, believed manufacturing practices often failed to meet their full potential, and that this was rarely due to their inherent technical complexity. Incompatible human resource practices and lack of continued managerial effort were offered as more likely causes. With regard to human resource practices, inappropriate payment schemes were cited the most often. The "classic" example was the continuance of a payment scheme that rewarded individual performance when team work or manufacturing cells had been implemented. Such misalignments were seen to be avoidable if a more systemic approach had been taken earlier in the change process.

Three interviewees stressed that careful attention to learning processes is crucial during the post-implementation period as:

"Users can acquire the technology, they can't acquire the knowhow. And of course, it's the knowhow and the technology that make the innovation" (Researcher).

It was thought that most individual learning about a new practice occurs when employees make it workable and smooth its rough edges. End-users also discuss and disseminate this knowledge developed between them in an informal manner. It was also suggested that companies should facilitate this process so that the effectiveness of the practice may be enhanced. Training and the opportunity in formal settings to discuss problems and communicate what they have learnt were offered as examples of how this could be achieved.

Interviewees were asked about the extent to which newly implemented practices meet their objectives. The most optimistic assessments were given by practitioners. Practitioners stated that, on average, 30% of practices met their objectives entirely and that 40% met their objectives a lot. Consultant and professional advisers, and researchers in particular, were much more pessimistic about success rates. Practices were thought rarely, if ever, to meet their objectives entirely. Only 20% of practices were seen to meet their objectives a lot. The majority of practices were thought to meet their objectives moderately or only a little. Differences in the perception of success probably stem from different criteria of success being used. From the results it would seem that practitioners have lower criteria for judging success. However, and in apparent contradiction to this, many interviewees noted that extremely high objectives are often set at the beginning of the project as a political device to get funding. This either points to official and unofficial objectives being aimed for, or managers going through some form of post hoc rationalization and adjusting the criteria of success downwards. Differences may also be a result of consultants and researchers being called in after a failure; they may therefore see fewer examples of success. The practices that were thought to be the most successful were TQM, JIT, and Concurrent Engineering. The least successful practices were thought to be BPR, team working and IT-based practices.

Interviewees were asked about the extent to which practices were abandoned. Again there were differences between practitioners and other interviewees. Practitioners suggested that very few, if any, practices were totally abandoned. Other interviewees thought that 10% of practices were abandoned. The reason for this discrepancy might be that practices are not necessarily completely abandoned but simply wither away leaving behind some traces of their existence.

3.4. Evaluation

Forty-five interviewees said that nearly all companies do not evaluate new practices against initial objectives. Good practice, the continuous evaluation against well-specified initial objectives, was thought to be extremely rare. The main reasons why evaluation against initial objectives did not occur were thought to be:

- It could not be done, as initial objectives were not specified.
- Political reasons; it may not be in the interests of managers to conduct an evaluation.
- The effort and costs involved are seen to be too high.
- That companies tend to look forward rather than backwards. Evaluation, when it was done, tended to be against current

measures. These were normally quantitative cost or quality measures. Few examples were offered of companies evaluating the impact of change on their employees. Eight participants thought that the lack of evaluation and reflection on the introduction of new practices significantly reduced further capacity for successful change.

3.5. Improvements needed

Interviewees were asked about what needed to occur to improve the process of change in UK manufacturing companies. A large proportion of interviewees, thirty-nine in all, believed that one of the best ways to improve the change process in UK companies would be to disseminate knowledge and expertise more widely. One way of disseminating knowledge was thought to be through using examples of "good practice". However, it was often noted that such examples should not just reinforce the relatively well-known principles of change. Rather, examples of "good practice" should also focus on the practicalities of managing change, particularly the social and political processes involved, and include information about the tools and processes that help companies to think strategically and systemically. Three interviewees wondered whether too active a promotion of "good practice" may be problematic. For example, do examples of "good practice" lead to companies having unduly high expectations about a practice? Does it encourage the following of fads? Thirty-two interviewees thought that while knowledge dissemination was adequate, in many instances other forms of direct intervention were needed. These more direct forms of intervention included Teaching Company Schemes and the creation of learning networks which bring companies with similar problems together to share knowledge and learn from each other.

Ten respondents argued that the dissemination of knowledge and expertise should be targeted at the long tail of "average" companies, and not the small head of excellent companies. It was further suggested that organizations such as the Department of Trade and Industry, the Confederation of British Industry, the Engineering Employers Federation, Business Link, European Community Regional Development Funds and institutions of higher education should play a major and pro-active role in disseminating and delivering knowledge and expertise. A more proactive approach was thought to be needed as knowledge and expertise from the head was not thought to "trickle down" to the tail in an efficient and effective manner.

Just over half of the participants, twenty-five in all, also thought that managers needed more education in management. Yet, one perspective that emerged on this is that, although more and more managers receive some form of management training (witness the phenomenal expansion of business schools and business courses), the problem may not be so much with the extent, but the type of education they are receiving. A clear problem is that while managers may know the theory, they often think it irrelevant or difficult to apply. This implies that management education should focus more on the practicalities of managing through the use of experientially based teaching practices such as action learning (Pedler 1991). Given the lack of strategic planning, twelve of the twenty-five who mentioned management education, thought that particular attention should be given to improving the strategic skills of senior managers.

4. DISCUSSION

This survey indicates that only a small proportion of UK manufacturing companies are thought to manage consistently the change and innovation process in an effective manner. Such companies can be thought of as having well-developed change and innovation capabilities. The results also indicate that, in the majority of companies, the quality of the change process is variable and that their change and innovation capabilities are not particularly well developed. Thus, companies are generally good at managing the mechanics of change (e.g. project planning), placing more importance on selling and communicating change, consulting end-users more often, and increasingly recognizing that human resource practices often need adjusting to realize the benefits of change. Companies also appear to be less subject to fads. However, this research also shows that in most manufacturing companies:

- manufacturing practices are adopted with relatively little strategic thought;
- change remains technology and technique led;
- the systemic impact of change on employees and human resource practices is generally considered too late;
- user participation is usually limited to consultation;
- rarely evaluate the impact of new practices against initial objectives.

If read optimistically, these findings suggest that some of the lessons of change management are slowly becoming embedded within the change capabilities of UK organizations. From another perspective, the findings make for more pessimistic reading as they fit into similar patterns found by earlier studies (e.g. the lack of strategic thought (Kearney 1989); the lack of systemic thinking (Clegg *et al.* 1996); that change is often technology or technique led (McLoughlin and Harris 1997)), and demonstrate that this is occurring across a large number of organizations and with respect to many different types of modern manufacturing practice.

This study has also found that the principles of change seem to be fairly well known if not widely practiced. Companies, it would appear, experience difficulty in implementing the principles of change; the devil is quite clearly in the detail. An example of this can be seen in the fact that while managers may realize the need to think strategically or systemically, they often lack the tools that enable them to do this in a detailed and practical manner. A lack of tools, therefore, goes some way toward explaining the difficulty of using the recommended principles of change. But, as case studies of successful change have shown, change is not simply a matter of following the "correct" prescriptions or using the "right" tools. Rather, change is also a social and political process that involves the interactions of different interest groups with different aims (Mintzberg 1973; Bessant *et al.* 1996; Thomas 1996). Recognizing and attending to the social and political aspects of change is as important as attending to its technical, strategic, and systemic aspects. That the political aspects of change were not particularly well highlighted in this study points to one of the limitations of the expert panel approach.

A further interesting theme to have emerged from this study is the importance of using techniques that develop learning and expertise, and this is in keeping with concepts such as

"organizational learning" and the "learning organization" (Argyris 1996; Cohen and Sproull 1991; Pedler 1991). Techniques that are thought to promote learning and expertise include end-user participation, cross functional teams, periods of review and reflection, action learning, evaluation of the practice, and evaluation of the change process. With regard to the change and innovation process, these techniques appear to have three benefits. First, they aid the creation and dissemination of knowledge about an innovation, its systemic implications, and how to enhance its effectiveness once implemented (Nonaka 1995). (It can be noted that improving the knowledge base and expertise of organizations was implicit in the recommendations made by interviewees, see section 4.5., Second, techniques such as action learning and cross-functional teams can provide an important means with which to examine, reflect on, and manage the social, political, and technical practicalities of change (McLoughlin and Thorpe 1993). Third, evaluating, reviewing, and reflecting throughout the process of change, and at the "end" on the effectiveness of the practice and the change process itself, can help organizational members to learn from their experiences. Indeed, reflecting on experience is one of the critical tasks of learning and development (Schon 1981; Kolb 1984, What is perhaps worrying is that although evaluation and reflection appear to be so critical to the development of change and innovation capabilities, they rarely happen in any systematic manner. Overall, it can be suggested that techniques that focus on learning and expertise can contribute not only to a successful change process, but also to the development of change and innovation capabilities (Bessant 1996; Tidd *et al.* 1997).

One criticism of this study that could be made is that, due to the relatively small number of participants, the picture drawn is not representative of UK companies as a whole and it is therefore difficult to be confident in the claims made. In defense, it can be argued that although the number of participants was small, the actual sample of companies drawn on was relatively large in comparison to other examples of research in this area. Moreover, the interviewees responses indicated that the sample drawn on represented a wide range of company sizes and sectors.

Another criticism is that the interviewees may be biased in the picture they present. Company practitioners may be too positive due to their personal investment in the change process. Researchers and consultants might be too negative on account of witnessing more "failures" or higher standards of success. However, such biases are likely to be a problem for any methodology. The fact that the expert panel draws on a range of perspectives may go some way in providing a more balanced account. Further confidence can also be drawn from the fact that the results are similar to those found in other studies which we commented on earlier (e.g. Kearney 1989; Clegg *et al.* 1996).

5. CONCLUSION

The aim of this survey was to gather information, based on the experience of experts, about the change capabilities of UK manufacturing companies with respect to a range of twelve common modern manufacturing practices. It indicates that in the UK there are a small number companies that have good change capabilities, while the majority of companies have change capabilities that are more variable in quality. With regard to the latter group, there are optimistic signs for believing that companies are slowly learning the lessons of change and developing their

change capabilities. Indeed, many of the principles of change appear to be widely known and companies do seem to be getting better at aspects of change such as project planning, selling and communicating change, and consulting end-users. This survey also indicates that companies could develop their change capabilities further and that they are poor at learning from their experiences of managing change and innovation. In order to rectify this situation there are a number of factors that companies need to pay close attention to. These include: examining the strategic and systemic (e.g. human, organizational, aspects of change; using tools that help manage the practicalities of change; increasing end-user participation; using practices that promote learning and expertise; evaluating against initial objectives; and evaluating change projects. Organizations, such as governmental agencies, higher education institutions, consultancies and professional bodies can also play an important role in facilitating the development of change capabilities by providing examples of "good practice" and through schemes that promote the transfer of knowledge and expertise more directly.

Future research on this topic will need to examine and explore a number of different areas. Given the relatively small number of participants and the methodological approach, further survey work is needed to confirm the whether this survey's results have a wider validity, and examine more closely the similarities and differences between company sizes and industrial sectors. In addition to this, detailed case studies should be conducted that examine the development of change and innovation capabilities over time. Of particular importance in these studies would be an examination of the role and nature of learning and expertise, how they affect the change and innovation process, how they affect the subsequent effectiveness of the manufacturing practice, and how they affect the long-term development of change and innovation capabilities.

APPENDIX 1
MAIN BODY OF INTERVIEW SCHEDULE
1.   Adoption of innovation
1.1 When companies decide to adopt the practices you have just identified to what extent are these choices: strategically led; business-led; customer driven; because it is the latest fad; seen as a quick fix?
1.2 What are the objectives of selecting these practices?
1.3. How risky are these practices perceived to be? and what are the key risks?

2.   Management of change/planning and implementation
2.1 Can you please summarize how the implementation process is/was usually managed?
2.2 To what extent did the plan consider the impact of the new practice on other organizational systems and practices? *Respondent shown Appendix 2.*
2.3 What was the extent of end user involvement (including managers, in the redesign process?
2.4. Do companies use "champions of change" either in the form of individuals or groups?
2.5. To what extent have new practices been delivered on time and within budget?
     entirely   _____%

a lot	_____%
moderately	_____%
a little	_____%
not at all	_____%
Total	_____%

3.   Incorporation of new practices

3.1 To what extent do you estimate that these practices have met the company's objectives for which they were introduced? For example: what percentage have:

met company objectives entirely	_____%
met company objectives a lot	_____%
met company objectives moderately	_____%
met company objectives a little	_____%
not met company objectives at all	_____%
Total	_____%

3.2 What percentage of investments in these areas have been abandoned? _____%

3.3 To what extent have the practices given value for money? and/ or made an acceptable return on investment? (financial performance measures?)

3.4 Where practices have not succeeded, why has this been the case?

3.5 Did the new practice have the intended impact on other organizational practices? *Respondent shown Appendix 2.*

3.6 Did the new practice have unforeseen consequences in relation to other organizational practices? *Respondent shown Appendix 2.*

4.   Evaluation

4.1 To what extent do companies evaluate new practices against initial objectives?

4.2 How have companies done it (financial audits, surveys, benchmarking)?

4.3 Were any of the new practices evaluated against "emergent objectives"? (i.e. did a company take on a practice for one reason, but find it particularly effective in relation to another important objective).

4.4 To what extent have these new practices had an effect on the firms financial performance?

APPENDIX 2
EXAMPLES OF ORGANIZATIONAL PRACTICES

1. Human resource practices
   Payment systems
   Appraisal/development practices
   Levels and type of training needed
   Employment contracts
2. Other manufacturing practices and technologies
3. Job design/work organization
4. Organizational structures
5. Traffic across organizational boundaries (e.g. contractual relationships between firms, particularly in relation to outsourcing and supply chain partnering)
6. Culture (communication patterns, power)

7. Others?

APPENDIX 3
STANDARD INDUSTRIAL CLASSIFICATION OF MANUFACTURING ORGANIZATIONS

20/21 Food products, beverages and tobacco
22 Textiles
23 Apparel and clothing
24/25 Lumber, wood, furniture and fixtures
26 Paper and allied products
27 Printing and publishing
28/29 Chemicals and petroleum
30/31 Rubber, plastic and leather products
32 Stone, clay glass and concrete
33 Primary metal industries
34 Fabricated metal products, except machinery and equipment
35 Industrial and commercial machinery
36 Electronic/electrical equipment
37 Transportation equipment
38 Measuring and analyzing equipment

REFERENCES

ARGYRIS, C., 1996, *Organizational Learning II: theory, method and practice*, (Wokingham, Addison Wesley).

BESSANT, J., 1991*, Managing Advanced Manufacturing Technology: The challenge of the Fifth Wave*, (Oxford/Manchester: NCC/Blackwell).

BESSANT, J., CAFFYN, S. and GILBERT, J., 1996, Learning to Manage Innovation*, Technology Analysis & Strategic Management*, 8(1), 59–70.

BOLDEN, R., WATERSON, P., WARR, P., CLEGG, C. and WALL, T., 1997, A new taxonomy of modern manufacturing practices, *International Journal of Operations & Production Management*, 17(11), 112–1130.

BURNES, B., 1993, *Managing Change*, (London: Pitman).

CLEGG, C., AXTELL, C., DAMODARAN, L., FARBEY, B., HULL, R., LLOYD-JONES, R., NICHOLLS, J., SELL, R., TOMLINSON, C., AINGER, A., and STEWART, T., 1996, *The performance of Information Technology and the role of human and organizational factors*, (Sheffield: Institute of Work Psychology, University of Sheffield).

COHEN, M. and SPROULL, L. (eds), 1991, *Organizational Learning* (London: Sage).

COLLIS, D., 1994, How valuable are organizational capabilities? *Strategic Management Journal*, 15, 143–52.

Confederation of British Industry (CBI)/ National Westminster Bank, 1998, *Innovation Survey*, http://www.cbi.org.uk/innovation.

DAWSON, P., 1994, *Organizational Change: A Processual Approach* (London, Paul Chapman Publishing).

FAWCETT, S. and SCULLY, J., 1995, A contingency perspective of just-in-time purchasing: globalization, implementation and performance, *International Journal of Production Research*, 33(4), 915–31.

JAIKUMAR, R., 1986, Post-industrial manufacturing, *Harvard Business Review*, Nov–Dec, 69–76.

KEARNEY, A.T., 1989, *Computer Integrated Manufacturing: Competitive Advantage or Technological Dead-end* (London: A.T. Kearney Consultants).

KOLB, D., 1984, *Experiential Learning* (New Jersey: Prentice Hall).

McLOUGHLIN, H. and THORPE, R, 1993, Action Learning: A paradigm in emergence; the problems facing a challenge to traditional

management education and development, *British Journal of Management*, 4(1), 19–27.

McLOUGHLIN, I. and HARRIS, M., 1997, *Innovation, Organizational Change and Technology* (London: Thompson Business Press).

MINTZBERG, H., 1973, *The Nature of Managerial Work* (New York: Harper Row).

NONAKA, I., 1995, *The Knowledge-Creating Company* (Oxford: Oxford University Press).

PEDLER, M., 1991, *Action Learning in Practice* (Aldershot: Gower).

SCHON, D., 1981, *The Reflective Practitioner* (London: Maurice Temple Smith).

THOMAS, P., 1996, The devil is in the detail: Revealing the social and political processes of technology management, *Technology Analysis & Strategic Management*, 8(1), 71–84.

TIDD, J., 1991, *Flexible Manufacturing Technologies and International Competitiveness* (London: Pinter).

TIDD, J., BESSANT, J., and PAVITT, K., 1997, *Managing Innovation: Integrating Technological, Market and Organizational Change*, (Chichester: Wiley).

TOMANEY, J., 1994, A New Paradigm of Work Organization and Technology? In Amin, A. (ed.) *Post-Fordism: A Reader* (Oxford: Blackwell) pp. 157–94.

Van de VEN, A., ANGLE, H., and SCOTT POOLE, M. (eds), 1989, *Research on the Management of Innovation: The Minnesota Studies* (New York: Harper Row).

VOSS, C., 1992, Successful innovation and implementation of new processes, *Business Strategy Review*, Spring, 29–44.

WALTON, R., 1986, *Human Resource Practices for Implementing AMT: Report of Committee for Effective Implementation of AMT* (Washington: National Research Council).

WATERSON, P., CLEGG, C., BOLDEN, R., PEPPER, K., WARR, P., and WALL, T., in press, The use and effectiveness of modern manufacturing practices in the UK, *International Journal of Production Research*.

WEBSTER, J. and WILLIAMS, R , 1993, *The success and failure of computer-aided production management: the implications for corporate and public policy*, Edinburgh University, Programme on Information and Communication Technologies.

WEMMERLOV, U. and JOHNSON, D., 1997, Cellular manufacturing at 46 user plants: implementation experiences and performance improvements, *International Journal of Production Research*, 35(1), 29–49.

WOLFE, R., 1994, Organizational innovation: Review, critique and suggested research directions, *Journal of Management Studies*, 31, 405–31.

Change Management

R. J. Badham

Department of Management and BHP Institute for Steel Processing and
Products, University of Wollongong, NSW 2522, Australia

> There is nothing more difficult to execute, nore more dubi-
> ous of success, nor more dangerous to administer than to
> introduce a new system of things: for he who introduces it
> has all those who profit from the old system as his enemies,
> and he has only lukewarm allies in all those who might profit
> from the new system.
>
> (Machiavelli, *The Prince*; Bondanella and Musa 1979: 94)

1. INTRODUCTION

Interest in change and innovation dates back to antiquity, yet
what has now come to be called change management has mainly
developed as an explicit area of action and inquiry in the post-
World War II period. The dominant view of *change management*
is a prescriptive model of planned organizational innovation in
an environment that is making organizational change more rapid
and more complex yet more essential for survival. Planned
organizational change initiatives may take a number of different
forms and occur at a number of different levels. In 1974,
Friedlander and Brown (1974) made a classic distinction between
planned human processual and techno-structural changes, which
was extended into a four-fold typology in McMahan and
Woodman's (1992) survey of organizational change initiatives in
Fortune 500 industrial firms. The types of change initiatives
observed were: *human processual*, i.e. an emphasis on human
relationships, team building, work team interaction, process
consulting or conflict resolution; *techno-structural*, i.e. emphasis
on socio-technical systems, task and technology, work designs,
or organization and group structure; *strategic planning*, i.e.
emphasis on strategic business planning processes, strategic
change or visioning; primarily top management involvement; and
systemwide, i.e. emphasis on organization-wide improvement
activities; leadership, culture, quality improvement, and trans-
formation type organizational change projects. It is now
commonly argued that organizations are increasingly required
to undertake more strategic and systemwide changes in the face
of increasing rates of change in external environments. This broad
"managerial" or strategic planning model stands in sharp con-
trast to more "reflective" research on *organizational change* that
seeks to understand and describe the full complexity of change
processes.

2. CHANGE MANAGEMENT AS PLANNED INNOVATION

As a prescriptive model for managing change, change manage-
ment has traditionally been dominated by "project management"
and "organizational development" approaches that date back to
the periods of World Wars I and II.

2.1. Project Management

The project management model views all change as projects
involving people working within a fixed time frame to achieve
particular goals, balancing parameters of cost, time and
specification. The focus is on the technical content of projects,
rational planning and control. Project management control tech-
niques stem from the work of the American Henry Gantt who
designed the bar chart ("Gantt chart") as a planning and control
aid in World War I. They were further developed in the 1950s in
methods seeking to calculate and coordinate the relationship
between different sets of activities, such as the US Navy's Program
Evaluation and Review Technique (PERT) and Remington Rand
Vac's Critical Path Method (CPM). More recently, the Project
Management Institute in the US and other countries certifies
project managers who combine experience with tested compe-
tence in the Project Management Body of Knowledge in the eight
areas of: managing scope, managing time, managing money,
managing quality, managing communications, managing human
resources, managing contracts and supply, and managing risk.
Project management consultancy firms and practices within firms
normally present different stage models of projects, from initial
problem definition to implementation and improvement. These
stage models have been continued in many equally linear staged
models of the information system development lifecycle
(Buchanan and Badham 1999). This general project management
approach has been widely criticized as a change management
model for its overly simplistic view of the linear course of projects,
rationalistic assumption of clear and unitary goals to be followed,
artificial separation of formal projects from ongoing informal
processes of change (or failure to change), and failure to
adequately consider the central importance of involvement and
participation as means of orienting projects and preventing as
well as overcoming resistance.

2.2. Organizational Development

The "organizational development" model has its origins in Lewin"s
(1951) classic three stage model of change, as "unfreezing,"
"changing" and "refreezing," as well as the commitment of
organizational development practitioners to addressing the "soft"
or "human" side of change processes. Unfreezing is the initial
stage in which a recognition develops of a need for change, and
action is taken to unfreeze traditional attitudes and behavior.
Changing is the second stage as the implementation of new
patterns of behavior and modes of operation occurs. Refreezing
is the third and final stage, where positive reinforcement of the
change is provided and the internalization and institutionaliza-
tion of new attitudes and behaviors is achieved. Lewin's (1951)
classic "force field" analysis examines two types of forces operating
in a change environment, driving forces promoting change and
restraining forces that act to maintain the status quo. To bring
about change, Lewin argues, driving forces must be strengthened
and resisting forces weakened.

Many organizational development specialists that adopt this
general approach seek to provide data that would unfreeze the
existing system by reducing resisting forces rather than increas-
ing the driving forces (Dawson 1994). The classic experiments
of Coch and French (1948) in a US pajama factory, identified
participative approaches to change as a key factor in overcoming
resistance and mobilizing support for change. This has led to
organizational development practitioners advocating what
Pettigrew (1985) characterizes as the "truth, trust, love and
collaboration" approach to change management. A key orienta-
tion is to establish "ownership" of change, with change consultants

providing process expertise to the "real" owners of change processes, helping them to incrementally improve their own problem solving, communication and learning capabilities in an open and honest change process rather than providing expert solutions to be implemented.

Approaches to the project management approach vary within the organizational development approach. Some emphasize the need for a "soft" organizational development accompaniment to "hard" project management in a total project management approach, others support the idea that organizations need to create resources, structures and cultures to support a thriving "shadow organization" or "adhocracy" of project groups or "skunk works," while others criticize the rationalist attempt to divide change into goal oriented project-based chunks, preferring a more diffuse approach that uses projects as participative vehicles for organizational learning and culture change.(Buchanan and Badham 1999). More recently, however, there has been a critical backlash against long-term loosely defined culture change projects, advocating a return to more focused "results oriented" change strategies (Schaffer and Thompson 1992).

Since Lewin's classic work, there have been a multiplicity of refinements to the three stage change model, with numerous recipes developed offering more complex and refined views of stages of change and the type of activities and "change roles" that occur, or need to be played out, at the different stages (Buchanan and Badham 1999). More critically, there has been increasing skepticism about the degree to which existing, and often rapidly changing, organizational structures are usefully seen as "frozen," or the extent to which a new ideal structure should be "refrozen" into something difficult to change. Organizational development practitioners have also been criticized for offering a universalistic model of "one best way" of change, and for their sometime naive commitment to incremental participative change as a change model appropriate for all circumstances (Dunphy and Stace 1990). Two new models of change that are critical of both the project management and organizational development models, yet committed to achieving planned change are the "contingency-design" and "political-processual" models.

2.3. Contingency Design Approach

Contingency-design approaches to organizations argue that organizations need to rationally adapt their structure to "fit" environmental demands. Contingency-design approaches to change management apply the same principles to organizational change, emphasizing the need for planned change actions and strategies to be adapted to internal or external organizational contingencies (Dawson 1994). Contingency approaches to organizational change have identified a number of such contingencies ranging from the internal life cycle of organizations and the type of organizational culture that an organization possesses to the nature of the inter-organizational context within which an organization operates and the nature of the intended or required change. They have also drawn on participative management literature to classify more or less participative or authoritarian set of change management approaches. Dunphy and Stace (1990) provide a classic analysis. They explore the appropriateness of particular styles of change management (collaborative, consultative, directive or coercive) for different scales of change (fine tuning, incremental adjustment, modular transformation, corporate transformation)

in different types of circumstances (time available for change, degree of institutionalized opposition to change). They present four ideal strategies: participative evolution; charismatic transformation; forced evolution; and dictatorial transformation. These strategies center around Dunphy and Stace's argument that organizational development approaches may be more appropriate to incremental changes where there is little opposition and time is available, whereas in conditions of high conflict and time pressure and a need for radical change, a more coercive style of change may be required. The central argument of contingency theories that change management strategies should be tailored to different circumstances is a valid and important one. Contingency models may, however, degenerate into mechanistic prescriptions providing over-simplified and deterministic views of the environment and over-generalized and unhelpful views of change strategies.

2.4. Political–Processual Approach

Political approaches to change management have their roots in views of organizations as sets of competing individuals and coalitions making choices in conditions of uncertainty in line with their own interests, perceptions and bounded rationality. Drawing on "strategic choice" models of organizations, change strategies are seen as being influenced by dominant coalitions of senior managers, and mediated by middle management and other groups pursuing their own goals and interests. In contrast to simplistic recipes for organizational change, a "processual" model has been developed by Pettigrew (1985) and others viewing change as an "untidy cocktail" of rational plans, managerial visions and numerous power plays in different and changing internal and external contexts. As Buchanan and Badham (1999) argue, this view informs an approach to change management in which politics needs to be understood in all its complexity, and "managed" by change agents who recognize that it cannot be "managed away." While the processual approach has traditionally only offered limited prescriptions for planned change, Buchanan and Badham (1999) have further developed the implications of this approach for the self-understanding and training of change agents as *political entrepreneurs*. They detail the political *skills and knowledge* required by the change agent, the *behavioral repertoire* needed by change drivers that attempt to address the political dimension of change, and the manner in which effective change agents adopt a *reflexive practitioner's perspective* toward learning and improving on their ability to diagnose and act politically.

3. REFLECTIVE APPROACHES TO ORGANIZATIONAL CHANGE

In contrast to managerial models of planned organizational change, more reflective analysts of organizational change are highly critical of what they view as the simplistic change recipes of "snake oil salesmen" (Czarniawska and Sevon 1996). Influenced by ethnographic and post-modernist research traditions, many of these analyses adopt what Martin (1992) has described as a "fragmentationist" view of organizational change, i.e. they focus on the lack of certainty, inherent ambiguity and contradictory nature of cultures and change processes. In addition, many of the advocates of this general approach are critical of simple assumptions of an overlap of interest between management and workers, and point to the intended and unintended

coercive and disciplinary dimensions of many change initiatives and processes.

While managers and other actors are performing the necessary acts of simplifying and stylizing reality to assist action, the role of reflexive researchers is seen as complementing such activities by encouraging systematic reflection on the complexity of organizational life and the change processes in which practitioners are involved. While the contingency-design and political role models have helped to introduce a more complex view of change and change management, the more reflective approaches tend to go further in being highly sensitive to the danger of research being directed or used to oversimplify or reinforce a restricted managerialist view of change processes.

4. CONCLUSION

Change management literature has traditionally been prescriptive in character, as it has been intended to provide advice on how to improve the management of organizational change. In contrast to radical approaches to organizations that criticize the dominance of managerial interests in change initiatives, change management approaches have primarily provided advice to assist managers in their formal change initiatives. Despite this orientation, there is now a greater awareness of the complex, situational, and political nature of all change initiatives. This has been enhanced by the contribution of more reflective analysts of organizational change. In future, there may be a blending of the different approaches as more traditional project management and organizational development approaches attempt to improve their success rates in complex and changing organizational circumstances. In particular, it is likely that future approaches to change management will increasingly incorporate a blend of insights from contingency, political and reflective approaches to organizational change.

REFERENCES

BUCHANAN, D. and BADHAM, R., 1999, *Power, Politics and Organizational Change: Winning the Turf Game* (London: Sage).

COCH, L. and FRENCH, J.R.P., 1948, Overcoming resistance to change. *Human Relations*, 1, 512–32.

CZARNIAWSKA, B. and SEVON, G. (eds), 1996, *Translating Organizational Change* (Berlin: de Gruyter).

DAWSON, P., 1994, *Organizational Change: A Processual Approach* (London: Chapman & Hall).

DUNPHY, K.D.D. and STACE, D., 1990, *Under New Management: Australian Organizations in Transition* (Sydney: McGraw-Hill).

FRIEDLANDER, F. and BROWN, L.D., 1974, Organization development. *Annual Review of Psychology*, 25, 313–41.

LEWIN, K., 1951, *Field Theory in Social Science* (New York: Harper & Row).

MACHIAVELLI, N. *The Prince*, trans. and eds BONDANELLA, P. and MUSA, M., 1979, *The Portable Machiavelli* (Harmondsworth: Penguin), p. 94.

MARTIN, J., 1992, *Cultures in Organizations: Three Perspectives* (Oxford: Oxford University Press).

McMAHAN, G.C. and WOODMAN, R.W., 1992, The current practice of organization development within the firm. *Group and Organization Management*, 17, 117–34.

PETTIGREW, A.M., 1985, *The Awakening Giant: Continuity and Change in ICI* (Oxford: Blackwells).

SCHAFFER, R.H. and THOMSON, H.A., 1992, Successful change programs begin with results. *Harvard Business Review*, January–February, 80–9.

Collaborative Engineering: Spanning Time and Space

C. M. Harvey

School of Industrial Engineering, University of Oklahoma, 202 West Boyd, Norman, OK 73019-0631, USA

1. INTRODUCTION

A new paradigm of competition is facing businesses today. This new environment requires companies to market, design, manufacture and sell a product within shorter time periods than ever before. Likewise, companies must be able to respond to and collaborate with global competitors to compete in the world economy. Companies can not afford simply to react to the market in the development of new products. Instead, they must understand the potential future of the market. More flexible companies that embrace change are replacing mass production. Companies must create new opportunities for profit and growth and establish an entire new competitive environment.

Competitive factors are those that deal directly with a firms survival. Competition can create an environment in which a company must collaborate for survival, resource sharing or to reduce uncertainty to increase the predictability of the environment. As well, competition can force companies not to collaborate. For example, in a market of zero growth with resource scarcity, firms are less likely to collaborate. Several factors can create the need to collaborate. Included in these are rapid economic and technological change, declining productivity growth, increasing competitive pressures and global interdependence.

In this new environment, people's skills, knowledge and access to information become more important. With the emergence of information technology and the convergence of computer networking and telecommunication technologies, it is no longer a requirement for people or cooperative companies to be located in the same place to communicate. Instead, people or companies that are geographically dispersed can and are engaging in collaborative arrangements. However, with a transition from the traditional face-to-face (FTF) meetings to other forms of interaction, the way teams accomplish tasks has to be rethought from formal design meetings to the informal gatherings that occur in the hallways.

This chapter intends to provide an overview of the factors that affect people when collaborating across time and space. To begin, it will first discuss the collaboration process and how engineers collaborate. This is followed by a discussion of what technologies must provide to bridge distance and time for engineers.

2. FUNDAMENTALS OF ENGINEERING COLLABORATION

2.1. What is Collaboration?

The term *collaboration* is derived from the Latin verb *collaborare*, which means "to labor together." *Webster's Ninth New Collegiate Dictionary* (1983) first definition of collaboration is: "to work jointly with others or together esp. in an intellectual endeavor." Several definitions of collaboration exist within the literature. Wood and Gray (1991) have consolidated many of them:

- A process through which parties who see different aspects of a problem can constructively explore their differences and search for solutions that go beyond their own limited vision of what is possible.
- A process of joint decision-making among key stakeholders of a problem domain about the future of that domain.
- A group of stakeholders who work together to make joint decisions about the future of their problem domain.
- An interactive process having shared transmutational purpose and characterized by explicit voluntary membership, joint decision-making, agreed-upon rules, and a temporary structure.
- The formal or informal institutions, rules and decision-making procedures shaped by prevailing principles and norms by relevant actors about acceptable behavior in a given issue area.

Using these definitions of collaboration as a background, Wood and Gray provided a complete definition of collaboration:

- Collaboration occurs when a group of autonomous stakeholders of a problem domain engage in an interactive process, using shared rules, norms and structures, to act or decide on issues related to that domain.
- Collaboration requires members to share a common goal, participate in a social environment, plan and coordinate task-specialized activities necessitating joint decision-making based on an agreed-upon set of norms and rules, and complete a task within a given time frame. The question is raised then as to how engineers collaborate in traditional FTF settings.

2.2. How do Engineers Collaborate?

It is important to understand how engineers collaborate and some of the conditions that may influence the success of the collaboration.

2.3. Design Task

Engineers are trained from their very introduction to the profession that design problems are a well-ordered transformation from the formulation of the design problem to the selection of a solution. Engineers are taught early models of design that typically consist of three-to-seven stages by which a problem is transitioned from a problem to a solution. As they advance through their education, engineers are taught that problems do not simply follow a well-ordered process and most problems generally consist of many different and interrelated problems.

Design could then be viewed just as simply a composite of problems. However, as Malhotra *et al.* (1980) state, problem solving occurs "when moving from a problem state to a non-problem state." The design process is much more complex. In design, usually a person has no defined initial state and generally the means of accomplishing the design problem are only limited by imposed design constraints. In real world design problems, the requirements are generally fuzzy and poorly articulated. In addition, there is no exact final state.

Elements of the design task that could affect its accomplishment in the distributed environment include its scope, structurability and uncertainty (Harvey 1997). Task scope is the breadth, extent, range, reach or general size of a task. Task structurability indicates how well defined is the sequence and relationships

between subtasks. How structured the task is could affect its analyzability, number of alternative solutions and amount of coordination needed between the subtask team members. Task uncertainty is the degree of predictability or confidence associated with a task. The combination of a task's scope, structurability and uncertainty determine how simple or difficult a task's accomplishment.

2.4. Design Team Interaction

In practice, most design work is done by teams of engineers and the problems follow anything but a well-ordered process. The frequency of group interaction cycles is a key aspect of decision development with more frequent interactions increasing decision quality (Poole and Hirokawa 1986). Team member interaction may be different depending on the requirements of the work domain (Figure 1). For example, the work may be *individual work* (e.g. work assigned to a single individual based on expertise); *cooperative work* (e.g. work that is pooled based on each individuals' work); or *collective work* (e.g. people work together, pooling their resources and knowledge to solve a task) (Rasmussen 1991). This work can be both formal and informal in nature.

In today's world of virtual collaborative teams, interaction may be accomplished through many different media from conference calls to chat sessions to video conferencing. Whatever the medium, one of the fundamental elements of groups is the requirement to coordinate activities. Malone and Crowston (1994) define coordination as "managing dependencies between activities." Dependencies between groups would include such items as shared resources (money, space, time) or the transfer of the product one activity produces to another activity. Coordination can be concurrent as when members of a design team work in

parallel on different aspects of the same product or sequential as when one design team must provide the other team a work product before their task can begin. Processes have been developed to facilitate these dependencies in the coordination process. For example, in managing shared resources, heuristics such as first-come/first-served and priorities have been developed to attempt to mediate the coordination of resources needed by more than one activity (Malone and Crowston 1994). Cooperative work tools have also been developed to support the coordination process. For example, electronic mail acts as a message transport, whereas group-scheduling tools serve to support synchronization dependencies. The type of coordination necessitated may ultimately depend on the task to be accomplished. Organized sports make this point apparent. While coordination in a basketball team is driven by the mutual adjustment of the players, American football coordination is achieved through planning and hierarchical direction (Keidel 1984). Likewise, a group formed to complete a task in particle physics will likely have different forms of coordination than a design team.

Communication is one of the fundamental elements used by groups to manage the dependencies. A shared common knowledge and/or communicative vocabularies facilitate the communication process (Clark and Marshall 1981). Therefore, as groups start their development process to agree on the task at hand, they are faced with the need to communicate. In this process, they must at minimum begin to establish a common dialog by which they can communicate. One has to be cautious of any bias caused by a reduced bandwidth created by replacing traditional FTF interactions with new technologies. Filtering hypotheses and functional fixedness are two biases that serve to decrease decision quality by limiting the quantity of design alternatives
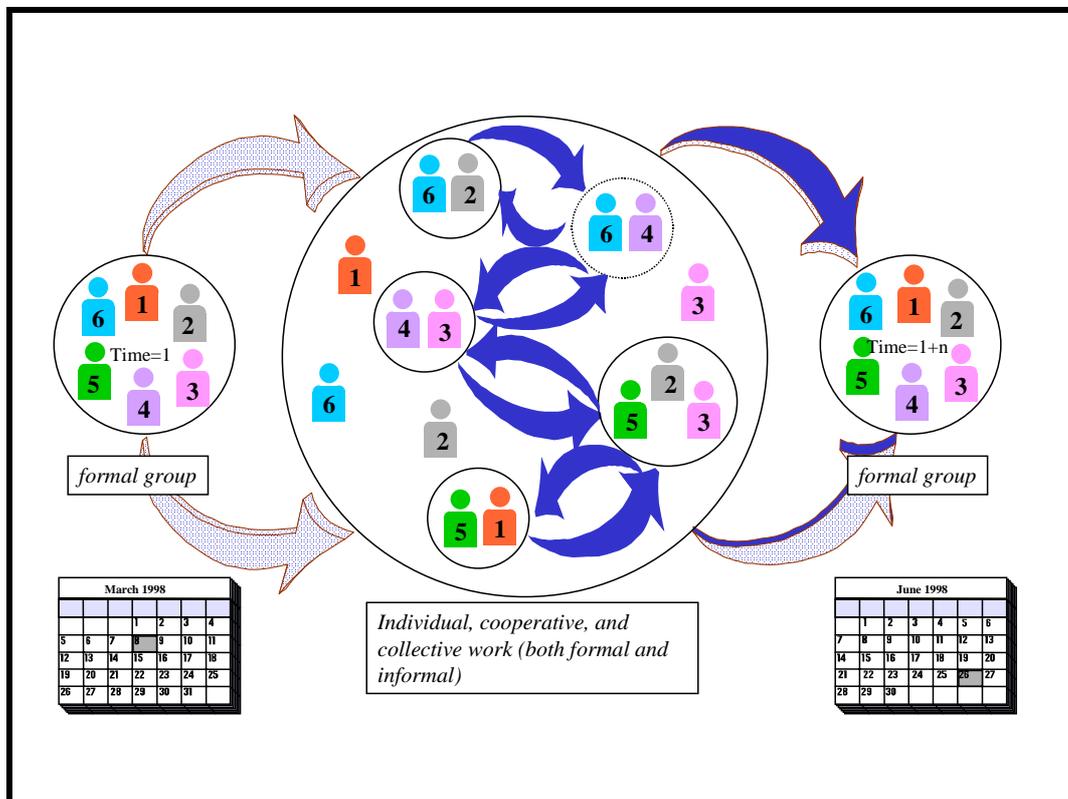

Figure 1. Engineering collaboration process.

considered by the group (Wickens 1992). The potential of these biases existing in the distributed work group could have a profound affect on the technical quality of the design, design time, schedule compliance, resource management, cost management and competition. What is unclear is whether these effects will be positive or negative. Another element that has to be considered in distributed collaboration is the other means by which engineers collaborate.

2.5. Collaborative Tools

Artifacts are one element critical to the engineering collaboration process. Artifacts are how engineers construct a representation of their ideas. Designers are notorious for not being able to think without making a drawing or sketch of rough ideas (i.e. "back-of-the-envelope-sketches") and, hence, require the ability to work within some visual medium. Ullman *et al.* (1987: 64) found six uses for the act of drawing:

- To archive the geometric form of the design.
- To communicate ideas between designers and between the designers and manufacturing personnel.
- To act as an analysis tool. Often missing dimensions and tolerances are calculated on the drawing as it is developed.
- To simulate the design.
- To serve as a completeness checker. As sketches or other drawings are being made, the details left to be designed become apparent to the designer. This in effect helps establish an agenda of design tasks left to accomplish.
- To act as an extension of the designer's short term memory. Designers often unconsciously make sketches to help remember ideas they might otherwise forget.

Dym's (1994) review of design from a broader perspective outlines six knowledge-state layers needed for design: perceived need, function, physical phenomenon, embodiment, artifact type and artifact instance. In describing these six layers, Dym outlines five representation languages as needed:

- Textual (e.g. document(s) used to articulate design projects and to communicate between design and manufacturing teams).
- Numerical-discrete (e.g. attribute values such as constraints, costs, or dimensions).
- Numerical-continuous (e.g. graphical representation of numbers which represent some element of a design).
- Graphical (e.g. visual descriptions such as sketches and CAD drawings).
- Physical (e.g. mathematical, analytical or physical models derived from the physical properties of a design that can be used to describe a function or behavior).

Dym states that "the knowledge-states and their representations are not likely to be entirely independent because the more abstract the layer, the more likely that it is rather vague knowledge expressed in text. At the other end of this spectrum, artifact types and artifact instances are increasingly specific descriptors." Hence, the designer needs different types of representation to express the design as the level of abstraction changes when moving from the initial design to the final design state.

2.6. Cognitive Workload

The distributed collaborative environment places a great strain on individual cognitive resources. Interaction within the distri-

buted collaborative engineering environment contains potentially many different input sources (e.g. video conferencing, voice communications, individual/shared computer screens) for the individual to contend with as s/he participates in the group effort.

Research has found that problem-solving is affected by the load on working memory (Sweller 1988). Likewise, some initial work looking at cognitive workload in the distributed work environment has found interesting results (Storck and Sproull 1995, Harvey 1997). In these initial findings, there appears to be an increased workload using video communication while no increase in audio communication. The video increase could be caused by the fact that social information is harder to obtain, yet there is a sense of obligation to work harder given individuals can see one another. Likewise, the non-increase in audio groups could be because it is too difficult to maintain the needed level of cognitive resources and/or the incapability between the task and the media. In either case, the distributed environment tends to impact the individuals use of cognitive resources and thereby could affect the collaboration process.

2.7. Task Cohesion

Why some groups succeed while others fail is still somewhat a mystery, yet one concept, *group cohesiveness*, is thought to impact on groups and their interaction. Cohesiveness is a complex concept to define; yet, most people can recognize whether it exists in groups in which they are participating. Three types of cohesion exist: interpersonal attraction, task cohesion and group pride. Interpersonal cohesion is viewed as the socio-emotional connection among group members (Festinger *et al.* 1950), whereas task cohesion consists of the group having a shared commitment to a task (Hackman 1976). Group pride suggests a bond between the individual and the group (Piper *et al.* 1983). Among these, task cohesion emerges as a critical component to be considered in groups of all types (Mullen and Cooper 1994).

3. SUPPORTING COLLABORATION ACROSS TIME AND SPACE

3.1. What Needs to be Supported?

A basic understanding of the engineering collaboration process allows one to start to envision the needs of engineers that are working at a distance spanning both time and space. In discussing distributive collaboration, it seems important to look at supporting technologies that support either *tele-presence* or *tele-data* (Greenberg *et al.* 1992). Technologies that support tele-presence would attempt to provide the participants participating in distributed communication with the necessary explicit and subtle dynamics that occur between participants including body language, voice cues, hand gestures, eye contact, focusing attention, etc. Tele-data is instead concerned with allowing participants at a meeting to present and access physical materials that would normally be inaccessible to distributed groups. These include notes, documents, plans and drawings, as well as a common work surface that allows each person to annotate, draw, brainstorm, record and convey ideas. These definitions provide a method of classifying technologies that support collaboration.

3.2. Supporting Technologies

Several supporting technologies exist that can be used to support distributed collaboration (Figure 2). Tools to support
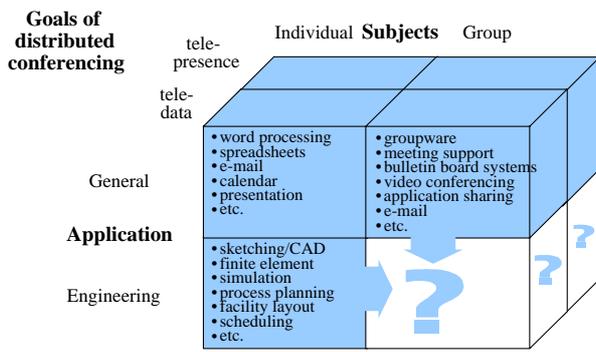
Figure 2. Engineering collaboration tools.

engineers working at a distance are still in development although research in this area is expanding. Today, engineers are taking advantage of the tools that are available to businesses for conducting distributed group work while research continues to meet specific engineering needs. Stanford's Center for Design Research (SHARE project), Carnegie Mellon's Engineering Design Research Center, Purdue's Center for Collaborative Manufacturing as well as on-going research at the University of Oklahoma's Human–Technology Interaction Center are just some of the places working on the improvement of the engineering collaboration process through the understanding of the engineering collaboration process and development of tools and technologies.

4. RECOMMENDATIONS

If an engineering group is considering the use of distributed collaboration or currently using other forms of media to communicate, the following recommendations should be considered:

- Allow team members to interact initially in a FTF environment to establish goals and work tasks before instituting distance technology. This will foster good relationships between group members.
- Ensure that the technology can support the engineering task(s) being accomplished at a distance.
- Assess design teams to ensure they are as effective at a distance as traditional FTF teams.
- Do not overuse the technology. Allow teams to reform in a FTF atmosphere from time to time.
- Train people so they can take full advantage of the collaboration tool(s).

REFERENCES

CLARK, H.H. and MARSHALL, C.R., 1981, Definite reference and mutual knowledge. In A. Joshi, J., Webber, B. L. and Sagg, I. A. (eds) *Elements of a Discourse Understanding* (New York: Cambridge University Press), pp. 10–63.

DYM, C.L., 1994, *Engineering Design: A Synthesis of Views* (New York: Cambridge University Press).

FESTINGER, L., SCHACHTER, S. and BACK, K., 1950, *Social Pressures in Informal Groups* (Stanford: Stanford University Press).

GREENBERG, S., ROSEMAN, M., WEBSTER, D. and BOHNET, R., 1992, Issues and experiences designing and implementing two group drawing tools. In *Proceedings of the '92 CSCW Conference* (New York: ACM), pp. 609–20..

HACKMAN, J.R., 1976, Group influence on individuals. In Dunnette, M.D. (ed) *Handbook of Industrial and Organizational Psychology* (New York: Wiley), pp. 1455–525.

HARVEY, C.M., 1997, Toward a model of distributed engineering collaboration. Unpublished dissertation, Purdue University.

KEIDEL, R.W., 1984, Baseball, football, and basketball: models for business. *Organizational Dynamics*, 5–18.

MALHOTRA, A., THOMAS J.C., CARROLL J.M. and MILLER, L.A., 1980, Cognitive processes in design. *International Journal of Man–Machine Studies*, 12, 119–40.

MALONE, T.W. and CROWSTON, K., 1994, The interdisciplinary study of coordination. *ACM Computing Surveys*, 26, 87–119.

MULLEN, B. and COOPER, C., 1994, The relation between group cohesiveness and performance: An integration, *Psychological Bulletin*, 115(2), 210-27.

PIPER, W.E., MARRACHE, M., LACROIX, R., RICHARDSON, A.M. and JONES, B.D., 1983, Cohesion as a basic bond in groups. *Human Relations*, 36, 93–108.

POOLE, M.S. and HIROKAWA, R.Y., 1986, Communication in-group decision-making: a critical assessment. In M.S. Poole, and R.Y. Hirokawa, (eds) *Communication and Group Decision Making* (Beverly Hills: Sage), pp. 15–31.

RASMUSSEN, J., 1991, Modelling distributed decision-making. In J. Rasmussen, B. Brehmer, and J. Leplat, (eds) *Distributed Decision Making: Cognitive Models for Cooperative Work* (New York: Wiley), pp. 111–42.

STORCK, J. and SPROULL, L., 1995, Through a glass darkly: What do people learn in videoconferences? *Human Communication Research*, 22(2), 197-219.

SWELLER, J., 1988, Cognitive load during problem solving: Effects on learning, *Cognitive Science*, 12, 257-85.

ULLMAN, D.G., WOOD, S. and CRAIG, D., 1990, The importance of drawing in the mechanical design process. *Computers and Graphics*, 14, 263–74.

WOOD, D. J. and GRAY, B., 1991, Toward a comprehensive theory of collaboration, *Journal of Applied Behavioral Science*, 27(2), 139-62.

Webster's Ninth New Collegiate Dictionary (1983), (Springfield, MA: Merriam-Webster Inc.).

Community Ergonomics: Applications

P. Carayon

Ecole des Mines de Nancy, 54042 Nancy Cedex, France and University of Wisconsin-Madison, Department of Industrial Engineering, Madison, WI 53706, USA

W. J. Cohen

Department of Management Systems, Massey University, Palmerston North, New Zealand

J.–L. Coujard

Ecole des Mines de Nancy, 54042 Nancy Cedex, France

M. C. Haims, L. Newman, F. Sainfort, J. Smith and M. J. Smith

University of Wisconsin-Madison, Department of Industrial Engineering, Madison, WI 53706, USA

1. INTRODUCTION

Community Ergonomics was originally developed as an approach and a tool for applying ergonomics to 'dysfunctional' community settings (Cohen 1996; Cohen and Smith 1994; Smith and Smith 1994; Smith *et al.* 1994). The objective of Community Ergonomics is to improve the quality in communities using ergonomic principles and methods. Community Ergonomics relies on micro- and macro-ergonomics principles and methods to design and develop communities that are healthy, safe and productive. Different levels and types of community can be examined. The community can be comprised of individuals (e.g. students (Newman 1997)), of groups (e.g. a neighborhood with different stakeholders (Cohen 1996)) or of organizations (e.g. enterprises in an industrial zone (Carayon, Coujard, and Tarbes 1997)). In this article, we discuss various applications of Community Ergonomics. See the article by William J. Cohen on the concepts and methods of Community Ergonomics.

2. COMMUNITY ERGONOMICS PROJECTS IN THE USA

2.1. The Milwaukee Project

The plan to develop the Milwaukee Community Investment Partnership (MCIP) was the framework for pursuing a strategy and specifying and presenting solutions to specific issues of concerns and aspirations of community residents. The strategy pursued (based on industrial engineering theory and techniques) was a process to create the framework to:

- establish a purpose to be achieved;
- organize a strategy to be pursued;
- specify and present solutions;
- involve people critical to planning and implementing the solution;
- develop and use information and knowledge appropriately;
- arrange for changes and improvements to the solution and the solution finding process.

This search was not some ultimate solution to the initial issues of concern, but instead was a proposal to organize a framework in which solutions develop over time and to increase the participant's ability to use the framework repeatedly for their unique needs. The search for solutions is to decide what is needed, why it is needed,

who is needed, when things are needed, and where they will be needed. MCIP consisted of local community advocacy groups, community learning centers, special homes and shelters, large local industry, business development groups, housing development agencies, religious groups, public and private banking institutions, employee unions, and local, city, state, and federal political agencies. The result was a comprehensive plan to generate economic and social results in the form of a community investment bank acting as an intermediary between traditional banking facilities and community residents who are typically rejected by those facilities.

The planning group spent several meetings delineating the issues of concern based on the planning and design matrix work of Gerald Nadler and his concept of "Breakthrough Thinking" (Nadler and Hibino 1998). The discussions of problems that generate the purposes for which solutions are needed, are framed as substantive questions, situations, phenomenon, and issues which concern members of the partnership. Such concerns, reflected in needs, uncertainty, aspirations, obstacles, desires, difficulties, or doubts are values motivating the aim for change and improvement. The specific locus of "what, where, when, and who" unique to each type of problem and issue is also defined. The purpose of this section was to define the "Something(s) That Cause Concern" within the Milwaukee inner city as defined by the Planning Committee.

The challenge was to diagnose and remedy the substantive issues in Milwaukee inner city communities. Although these inner city areas are severely affected and in need of significant attention, inner city communities are not devoid of economic or social potential. These potentials must be sought in a manner that does not promote gentrification of the area, but rather the improvement of quality of life for current residents. The potential can be identified for individuals who have an entrepreneurial attitude but no means to achieve their intentions, and for those who desire to own homes but lack both credit and capital, among others.

The purposeful activities of MCIP recognized the problems that occur during the life process of individuals or groups of individuals in terms of the need for fair and legal access to credit and capital as one way to improve community environment relations. The interactions between people and groups with institutions, agencies, other individuals, or procedures in the banking and lending domain typically do not complement personal life requirements and developmental needs for self-sufficiency and independence. The community ergonomics process helped to discern both the need for a new interface design and the opportunity for community improvement through access to systems of credit and capital. This is not to say that a community bank was the be-all-and-end-all solution to community problems. It was the process of people and resources interacting in a structured manner over a period of time that led to this one logical alternative. The partnership itself was one preliminary solution to address issues of concern, while the bank was only an alternative to do so and, as such, represents one solution in a holistic approach to inner city problems.

2.2. Improving the Educational System for Minority Students

The institution of education has been one of many institutions that has been negligent in the nurturing and development of underrepresented groups in society. Spanning decades, African

Americans, Latino Americans and Native Americans have been experiencing poor public school education. This particular research study concentrated on the under-represented population of students (African American and Latino American) currently enrolled in the University of Wisconsin-Madison College of Engineering (COE), as well as those students who were once enrolled, but who chose to leave the COE.

A qualitative research approach was used to help extract information regarding student success and/or failure. The aim was on determining those factors associated with academic success or well-being for students of under-represented populations, particularly those majoring in engineering. This research integrates principles of Total Quality Management (TQM) and systems design in an effort to identify those factors which hinder or enhance the educational performance and success of under-represented student populations pursuing higher education in engineering and science.

The findings indicated a level of misfit between the students from under-represented populations and the academic environment. Elements such as better development of advising/mentoring programs, education as a life process, pre-college preparation, more faculty–student interaction, more awareness of cultural diversity and the overall classroom environment were identified as necessary components for student success. Other elements such as alienation and isolation, miscommunication between students and faculty, administrators, as well as other students, expectations of professors, lack of financial and social support, lack of cultural diversity, lack of training, and organizational climate were identifies as factors depriving students of a rich and successful educational experience. How does one work to correct such issues? One suggestion spoke of the need to evaluate the faculty and staff on a regular basis (beyond those currently completed by students at the end of the semester). The students also indicated a need and a desire to interact with more faculty (professors, teaching assistants, etc.) of diverse backgrounds. The students indicated that the classroom environment is crucial to the model of academic learning life, because it brings with it so many other issues that link, directly, to a student's success (i.e. preparation, academic skills/foundation, classroom interaction, etc.). All of these issues set the tone of the class from the moment the student enters. They help to determine the climate — which determines a student's willingness and ability to interact or to retreat and which helps to determine whether or not students will survive.

We attempted to develop a better understanding as related to those factors contributing to student academic success. One must realize that in uncovering these negative components, one has, in fact, also stumbled upon those positive components deemed necessary to promote retention and success. The analysis performed by the community ergonomist allowed the identification of solutions aimed at improving the educational system for minority students at the COE.

2.3. Quality Improvement at the City of Madison

The objective of Community Ergonomics is to improve the quality of life in communities. Local government organizations play an important role in any community. Therefore, improving the quality of the management and organization of these organizations is an important element of Community Ergonomics. In the USA, we have worked closely with the City of Madison on their quality improvement program. In France, we have accompanied the quality management program at the City of Nancy (see below).

The City of Madison has been a pioneer among public sector organizations in adopting and implementing Total Quality Management (TQM) principles. This effort began in 1983 through collaboration between the Mayor's office and the Center for Quality and Productivity Improvement (CQPI) of the University of Wisconsin-Madison (Box et al. 1989). In its early stages, Madison's quality improvement program was strongly influenced by the Deming approach. After the initial success of applying TQM principles and methods in the City's Motor Equipment Division, the quality improvement program was spread to other divisions of the City (Sainfort and Carayon 1997). During the summer of 1994, a research team at the CQPI of the University of Wisconsin-Madison started a new partnership with the City of Madison.[1] This partnership led to an evaluation of the quality improvement program at the City of Madison, as well as active participation of the university researchers in the development of new quality improvement activities. The evaluation of the quality improvement program was based on the following methods:

- archival analysis of all documents related to quality efforts from 1984 to 1994;
- administration of an employee survey at two different times;
- in-depth assessment of quality improvement projects;
- semi-structured interviews with selected managers, union representatives and other key personnel at the City of Madison.

For more details on all of these data-collection efforts, see Sainfort et al. (1998). The data collected was used for two purposes: (1) diagnostic of quality improvement activities at the City of Madison and identification of solutions, and (2) research studies. Data was fed back to the City of Madison and identified opportunities for improvement in quality-related activities. The partnership established between the City of Madison and the university set the ground for useful contributions from the university researchers. For instance, the researchers participated in the development of quality-related activities, such as the administration of a citizen satisfaction assessment survey. Different solutions have been implemented to improve the overall quality improvement program, therefore, leading to increased quality of working life for the city employees and higher quality of services for the community members.

3. COMMUNITY ERGONOMICS PROJECTS IN FRANCE

3.1. Developing a Network of Enterprises

Community Ergonomics can also be used to improve the overall quality of life for a group of enterprises. The first French Community Ergonomics project is taking place in the industrial zone of Ludres-Fléville, located south of Nancy, France. Since its creation in 1996, the industrial zone of Ludres-Fléville (ZLF) has always tried to create optimal conditions for facilitating the implementation of new companies in the zone and for encouraging the development of companies already implanted in the zone. The ZLF is composed of about 200 small- and medium-sized enterprises.

In 1995, a group of political and economic actors of the region

decided to create an association for industrial excellence in Ludres-Fléville ('Association Excellence Industrielle Ludres-Fléville' — AEILF) in order to achieve the following objectives:

- To support and help companies of the ZLF in their own quality improvement program;
- To improve the well-being, the image and the environment and to stimulate job creation in the ZLF;
- To create a dynamic process for continuous quality improvement in the ZLF.

In practice, two types of action are put in place by the association:

- Implementation of quality improvement programs in the companies of the ZLF;
- Work groups comprised of various actors (including companies of the ZLF) on different topics related to community quality.

A total of 24 companies have started quality improvement programs. Different activities are organized by the AEILF for supporting and helping those companies: quality audits, training programs, exchange of experience and working groups. The association promotes and organizes activities of benefit to the largest number of companies. These activities foster the development of significant interactions between the companies.

In order to help establish synergies between the ZLF companies, six work groups were formed in order to study and solve problems experienced by a large number of ZLF companies: (1) traffic-transportation, (2) communication, (3) safety, (4) environment-recycling, (5) relationship with educational system, and (6) mail-multimedia-telecommunications. The work groups are comprised of ZLF companies which either have a problem related to the issue or which have a solution for the problems studied by the work group. In addition, actors outside of the ZLF also belong to the groups. For instance, representatives of the local government (cities of Ludres and Fléville) are involved in various work groups.

Quality management methods are used in the management of the association's activities and the running of the various actions. The AEILF does not implement actual solutions: it manages the entire problem-solving process, but relies on outside organizations for implementing actual solutions. The functioning of the work groups relies heavily on various problem-solving tools borrowed from the quality management literature and practice.

The community ergonomics researchers are involved in the project. They participate in some of the work groups, and provide conceptual and methodological advice to the president and the project manager of the association. Under the guidance of senior researchers, engineering students of the Ecole des Mines de Nancy, France, are regularly involved in specific projects of importance to the association.

The association has been quite effective in its various actions. Over a period of two years, the association accomplished quite a large number of projects (total of 17 realizations). The association has a total of 25 projects currently under study by the different work groups. This data shows that the association is quite effective in solving problems experienced by the ZLF companies.

3.2. Quality Management at the City of Nancy

The quality management program at the City of Nancy began in 1996 in the form of a collaboration between the mayor and the top manager (called 'secrétaire général'), on one hand, and the Ecole des Mines de Nancy, on the other hand. The quality program relies heavily on process improvement and organizational functioning in selected divisions of the City of Nancy.

The transformation process started with awareness of the top manager. Following initial contacts between the city's top manager and researchers at the Ecole des Mines de Nancy, quality management was identified as a potential solution for improving the functioning of the city organization, under tight financial conditions. In addition, quality management was seen as a useful complement to an existing project of decentralizing services throughout the city.

The implementation process involved the following phases:

- involve volunteer departments
- form an informal quality committee
- hire engineering trainees to accompany volunteer departments
- start ad-hoc training of selected employees and managers
- organize seminars on quality (examples from the private and public sectors)
- inform employees and managers
- involve more volunteer departments
- establish a small formal structure for supporting and coordinating quality efforts (quality coordinator and quality liaison persons in the volunteer departments)
- communicate on the quality management program (both internally and externally).

The quality management program at the City of Nancy in volunteer departments includes two directions: (1) clarification and improvement of the functioning of the department, and (2) process improvement. First, the volunteer departments use different methods and concepts of quality management to identify the main missions of their departments, to define the corresponding organizational structure, and to clarify the positions in their departments. A second type of actions engaged in the volunteer departments is process improvement. Key processes are first identified, then analyzed, and finally improved. Initially, the main focus of the quality management program was on internal functioning of the departments. Later on, the needs of external customers, i.e. the community members, have been taken into account and better, more appropriate, services are now offered.

4. CONCLUSION

In this chapter, we have described applications of Community Ergonomics in the USA and in France. The projects covered a range of problems and situations:

- economic problems in the Milwaukee project and in the French industrial zone;
- educational problems for minority students at an engineering college in the USA;
- quality management in local government organizations in the USA and in France.

NOTE

1. Funding for this research was provided by the 'Total Quality Organizations' program of the National Science Foundation [Grant number: SBR-9529900, PI: F. Sainfort].

REFERENCES

BOX, G. E. P., JOINER, L. W., ROHAN, S., and SENSENBRENNER, F. J., 1989, *Quality in the community: One City's experience* (Madison, WI: Center for Quality and Productivity Improvement of the University of Wisconsin-Madison).

CARAYON, P., COUJARD, J.-L., and TARBES, J., 1997, The use of quality management methods in developing network of enterprises. In P. Seppala, T. Luopajarvi, C.-H. Nygard, and M. Mattila, (eds) *Proceedings of the 13th Triennial Congress of the International Ergonomics Association. Vol. 1: Organizational Design and Management* (Helsinki, Finland: Finnish Institute of Occupational Health), pp. 150–52.

COHEN, W. J., 1996, *Community Ergonomics: Operational Requirements and Design process*. PhD dissertation. Department of Industrial Engineering, University of Wisconsin-Madison.

COHEN, W. J., and SMITH, J. H., 1994, Community Ergonomics: past approaches and future prospects toward America's urban crisis. In Human Factors and Ergonomics Society, (ed.) *Proceedings of the Human Factors and Ergonomics Society 38th Annual Meeting* (Santa Monica, CA: Human Factors and Ergonomics Society), pp. 734–8.

NADLER, G. and HIBINO, S., 1998, *Breakthrough Thinking: The Seven Principles of Creative Problem Solving* (rev. 2nd edn) (Rocklin, CA: Prima Publications).

NEWMAN, L., 1997, *Quality Improvement/Assessment of Educational System for Students of Color in the University of Wisconsin College of Engineering*. Department of Industrial Engineering, University of Wisconsin-Madison.

SAINFORT, F., and CARAYON, P., 1997, Total Quality Management in an American public sector organization, *Second International Congress International of Industrial Engineering*, 3–5 September 1997, Albi, France, http://www.enstimac.fr/dr/gind/ALBI97/LIENBD/PapFinal/index.html.

SAINFORT, F., JAMES, C., TAVEIRA, A., BISGAARD, S., CARAYON, P., HAIMS, M. C., LUND, J., MOSGALLER, T., SMITH, M. J., and VANRENSSELAER, G., 1998, Quality management in the public sector: methodology and preliminary findings from the City of Madison study. Submitted to *Quality Management Journal.*

SMITH, J. H., and SMITH, M. J., 1994, Community Ergonomics: an emerging theory and engineering practice. In Human Factors and Ergonomics Society, (ed.) *Proceedings of the Human Factors and Ergonomics Society 38th Annual Meeting* (Santa Monica, CA: Human Factors and Ergonomics Society), pp. 729–33.

SMITH, M. J., CARAYON, P., SMITH, J., COHEN, W., and UPTON, J., 1994, Community Ergonomics: a theoretical model for rebuilding the inner city. In Human Factors and Ergonomics Society, (ed.) *The Human Factors and Ergonomics Society 38th Annual Meeting* (Santa Monica, CA: Human Factors and Ergonomics Society), pp. pp.724–8.

KEY REFERENCES

COHEN, W. J., 1996, *Community Ergonomics: Operational Requirements and Design process*. PhD dissertation. Department of Industrial Engineering, University of Wisconsin-Madison.

COHEN, W. J., and SMITH, J. H., 1994, Community Ergonomics: past approaches and future prospects toward America's urban crisis. In Human Factors and Ergonomics Society, (ed.) *Proceedings of the Human Factors and Ergonomics Society 38th Annual Meeting* (Santa Monica, CA: Human Factors and Ergonomics Society), pp. 734–8.

NADLER, G., HABINO, S., and HIBINO, S., 1998, *Breakthrough Thinking: The Seven Principles of Creative Problem Solving* (Prima Publications).

SMITH, J. H., and SMITH, M. J., 1994, Community Ergonomics: an emerging theory and engineering practice. In Human Factors and Ergonomics Society, (Ed.) *Proceedings of the Human Factors and Ergonomics Society 38th Annual Meeting* (Santa Monica, CA: Human Factors and Ergonomics Society), pp. 729–33.

SMITH, M. J., CARAYON, P., SMITH, J., COHEN, W., and UPTON, J., 1994, *Community Ergonomics: a theoretical model for rebuilding the inner city. In Human Factors and Ergonomics Society, (ed.) *The Human Factors and Ergonomics Society 38th Annual Meeting* (Santa Monica, CA: Human Factors and Ergonomics Society), pp.724–8.

Cross-cultural Factors in Macroergonomics

P. Carayon[†‡] and E. Järvenpää[§]

†Ecole des Mines de Nancy, Parc de Saurupt, 54042 Nancy Cedex, France

‡University of Wisconsin-Madison, Department of Industrial Engineering, 1513 University Avenue, Madison, WI 53706, USA

§Helsinki University of Technology, Department of Industrial Engineering and Management, PO Box 9500, FIN-02015 HUT, Espoo, Finland

1. INTRODUCTION

New knowledge and understanding on cross-cultural issues in macroergonomics is necessary better to take into account cultural characteristics in the (re)design of work systems, in particular in the context of the development of global companies. Globalization of business has become a statement of success for companies in the international market. In this paper, globalization has a twofold meaning: (1) a company with international activities and operations and (2) a company with a true and deep understanding of the international business environment and the cultural characteristics of the countries where it conducts business (Järvenpää *et al.* 1997). Global companies need to understand different cultures and the effects of cultural factors on employees, and thus on organizational effectiveness and functioning. For instance, leadership styles and organizational structures suitable for a given national culture may not necessarily be applicable to the context of another national culture (Järvenpää and Immonen 1996).

In contrast with the concepts of hardware ergonomics, environmental ergonomics and software ergonomics, macroergonomics refers to an approach in which the organizational design and context are taken into account. Macroergonomics is a top-down, socio-technical systems approach to organizational and work systems design, and the design of related human–machine, human–environment and user–system interfaces (Hendrick 1991). In organizational redesign and development, organizational factors, such as organizational structure, communication, responsibilities and the functions of different operational units should be taken into account. Hendrick (1997) emphasizes the importance of cultural diversity in the macroergonomic design of organizations. This paper deals with cross-cultural issues in the management of organizations and people from the viewpoint of macroergonomics.

2. CROSS-CULTURAL MODELS

The most popular cross-cultural model of management was developed by Hofstede (1997). According to him, culture is the collective programming of the mind that distinguishes one group or category of people from another. He developed a model for understanding cultural differences at the country or national level. His original model included four dimensions of national cultures (Hofstede 1983): 1, power–distance; 2, uncertainty–avoidance; 3, individualism–collectivism; and 4, masculinity–femininity. Based on a study by Hofstede and Bond (1988) carried out in Asian countries, a fifth dimension was added later: Long- versus short-term orientation. Power–distance has been defined as the degree of inequality among people that the population of a

country considers as normal, from relatively equal (small power–distance) to extremely unequal (large power–distance). The dimension of power–distance represents the dependence on the superior, centralization of authority and autocratic leadership. Uncertainty–avoidance has been defined as the degree to which people in a country prefer structured over unstructured situations. Structured situations are those in which there are clear rules about how one should behave. In countries with high score on uncertainty–avoidance, people tend to show more nervous energy, while in countries that score low, people are more easy going. Individualism–collectivism has been defined as the degree to which people in a country prefer to act as individuals rather than as members of groups. In individualist countries, people attach importance to personal time, freedom and challenge, and relative unimportance to training, use of skills, physical conditions and benefits. The dimension of individualism–collectivism represents one's loyalty to the group and dependence on the organization. Masculinity–femininity has been defined as the degree to which tough values like assertiveness, performance, success and competition prevail over tender values like the quality of life, maintaining warm personal relationships, service, care for the weak and solidarity. Long- versus short-term orientation is a dimension that was added later in a 23-country study of the values of students (Hofstede and Bond 1988, Hofstede 1994). Long-term orientation represents values oriented towards the future, like thrift (saving) and persistence, whereas short-term orientation represents values rather oriented towards the past and present, like respect for tradition and fulfilling social obligations.

Hofstede's model was developed using data collected in a large questionnaire survey. The 1968 survey was distributed to 116 000 IBM employees in 67 countries and included 150 questions; it was also translated in 20 different language versions and was aimed at understanding how people in different countries perceive and interpret their world. Based on the survey and subsequent ones, each country is assigned a score from 0 to 100 for each of the five dimensions. The scores represent relative positions of the countries on the different dimensions. Scores for 50 countries and three regions are available in Hofstede (1997).

Hickson and Pugh (1995) have linked the five dimensions of Hofstede's model to five primary management-related features. First, power–distance is related to managing authority. Power–distance shows how removed subordinates feel from superiors in a social meaning of the word "distance". In high power–distance societies, subordinates look for directions from superiors, and usually accept those instructions without question. In low power–distance countries, subordinates are more likely to initiate discussions with their superiors and even challenge what comes from above. Second, uncertainty–avoidance is related to managing uncertainty. A way of dealing with (managing) uncertainty is through rules and regulations in organizations that give people some sense of certainty. In high uncertainty–avoidance countries, rules and regulations are plentiful. However, these countries may differ on the way rules are treated. In Germany, for instance, the rules tend to be abided by, whereas in other high uncertainty–avoidance countries, like many of the Latin countries (e.g. France and South/Central America), the rules tend to be treated lightly. In those countries, there is a conflict between a demand for rules and regulations and a tendency to go around them in practice. In low uncertainty–avoidance country, there are fewer and more

ambiguous rules. Third, individualism–collectivism is related to managing relationships. Management in an individualist country is management of individuals, whereas management in a collectivist country is management of groups. In individualist countries, employees can be moved around individually, and incentives and bonuses should be linked to an individual's performance. In collectivist countries, management decisions should take into account the groups that people belong to and feel attached to. In individualist countries, the relationship between an employer and an employee is a contract supposed to be based on mutual advantage. In collectivist countries, that relationship is perceived in moral terms, like a family link. Fourth, masculinity–femininity is related to managing oneself. In masculine countries like North America, organizations offer achievement-linked rewards, like bonuses, pay rises and promotions. In feminine countries, quality of life matters, people and environment are to be considered, and to be of service is important. Fifth, long- versus short-term orientation is related to managing time. Some cultures are more concerned with time past, others with time present or future. Short-term oriented cultures are more oriented towards the past and present, and tend to be more static. Long-term oriented cultures are more oriented towards the future (especially perseverance and thrift), and tend to be more dynamic. In short-term oriented countries, quick results are expected, whereas in long-term oriented countries, perseverance towards slow results is valued.

Hofstede *et al.* conducted the same survey twice, once in 1968 and a second time in 1972. Comparison of the 1968 and 1972 data showed no convergence of the scores obtained by the countries on the four original dimensions. Over time, there seemed to be divergence for power–distance and masculinity–femininity. There seemed to be some change in individualism–collectivism towards the more individualist side; however, the cultures shift together, so that the differences between them remain intact. Hofstede (1997) argues that national cultures have historical roots that make them highly stable and unlikely to change, at least for some centuries.

More recently, another Dutch researcher, Fons Trompenaars, has proposed a different set of dimensions for characterizing national cultures (Trompenaars and Hampden-Turner 1998):

1. Universalism–particularism: focus on rules versus focus on relationships.
2. Communitarianism–individualism: the group versus the individual.
3. Neutral–emotional: the range of feelings expressed.
4. Diffuse–specific: the range of involvement.
5. Achievement–ascription: how status is accorded.

The dimension of universalism–particularism is similar to Hofstede's power–distance. The dimension of communitarianism–individualism is similar to Hofstede's individualism–collectivism dimension. The dimension of neutral–emotional is similar to Hofstede's uncertainty–avoidance. The other two dimensions proposed by Trompenaars and Hampden-Turner do not overlap with Hofstede's model and, therefore, provides additional information on national cultures. The dimension of specific–diffuse cultures is related to the degree to which we engage others in specific areas of life or diffusely in multiple areas of our lives. In specific-oriented cultures, a manager separates out his/her task relationships with subordinates and insulates this from other dealings. The aim of management is to achieve objectives and standards with rewards attached. In diffuse-oriented cultures, management is a continuously improving process by which quality improves. The dimension of achievement–ascription specifies how we accord status: achieved status refers to doing (e.g. performance and knowledge), and ascribed status refers to being (e.g. title).

3. CROSS-CULTURAL ISSUES IN MACROERGONOMICS

Table 1 lists the cross-cultural characteristics of Hoftsede's and Trompenaars' models, and, for each cross-cultural characteristic, selected macroergonomic issues.

3.1. Participatory Ergonomics

Participation is a key method for the design and implementation of ergonomic changes and programs. Participatory ergonomics has been defined as "the involvement of people in planning and controlling a significant amount of their own work activities, with sufficient knowledge and power to influence both processes and outcomes in order to achieve desirable goals" (Wilson and Haines 1997). Participatory ergonomics has been applied in many different countries, such as Japan (Noro 1991), the USA (Imada 1994, Smith 1994, Haims and Carayon 1998), The Netherlands (Vink *et al.* 1995), Germany (Zink 1996) and Finland (Järvenpää and Eloranta 1997). These applications have provided empirical support for the use of participatory strategies for implementing ergonomic changes.

The type and methods of participation may vary considerably on different dimensions, as defined by Wilson and Haines (1998): extent/level, purpose, continuity, involvement, formality, requirement for participation, decision-making structures, and coupling. Participatory ergonomics is a complex concept within which a multiplicity of tools and methods are used. We argue that the design and implementation of participatory ergonomics should also take into account characteristics of the national cultures.

According to cross-cultural models, participation has different meanings in different cultures. Even though employees in most countries desire and prefer a participative style of management, the amount and type of employee participation seem strongly based on cultural background. In a study of right countries, it was found that Germany and The Netherlands preferred participative management style to a higher degree than the UK and Canada. The cross-cultural differences in participation can be explained by the dimensions of power–distance and individualism–collectivism of Hofstede's model (Hofstede 1997). According to Hickson and Pugh (1995), power–distance is related to managing authority, whereas individualism–collectivism is related to managing relationships. Participation is a management style that requires sharing of decision-making power and close relationships among people (e.g. project teams). Participation may be easier to implement in countries with low power–distance and high collectivism. In countries with low power–distance, management style is usually more informal and employees are used and even encouraged to express their opinions openly to their superiors. Low power–distance indicates rather equal rights for all individuals. This may facilitate participation. Every member of the organization is expected to take his/her responsibilities

Table 1. Cross-cultural characteristics and selected macroergonomic issues.

Characteristics of national culture		Selected macroergonomic issues
Hofstede's model	Trompenaars' model	
Power distance: degree of inequality among people which the population of a country considers as normal	Universalism-particularism: focus on rules versus focus on relationships	• managing authority (Hickson and Pugh 1995) • participatory ergonomics • organizational design • management style
Uncertainty avoidance: degree to which people in a country prefer structured over unstructured situations	Neutral-emotional: the range of feelings expressed	• managing uncertainty (Hickson and Pugh 1995) • design of work systems: psycho-social work factors and job stress • organizational design • management style
Individualism-collectivism: degree to which people in a country prefer to act as individuals rather than as members of groups	Communitarianism-individualism the group versus the individual	• managing relationships (Hickson and Pugh) • participatory ergonomics • organizational design • management style
Masculinity-femininity: degree to which tough values like assertiveness, performance, success and competition prevail over tender values like the quality of life, maintaining warm personal relationships, service, care for the weak and solidarity		• managing oneself (Hickson and Pugh 1995) • design of work systems: psycho-social work factors and job stress • organizational design • management style
Long-versus short-term orientation: values oriented towards the future, like thrift and persistence versus values oriented towards the past and present, like respect for tradition and fulfilling social obligations		• managing time (Hickson and Pugh 1995) • organizational design • management style • management of change
	Diffuse-specific: the degree to which we engage others in specific areas of life or diffusely in multiple areas of life	• organizational design • management style • management of change
	Achievement-ascription: how status is accorded; accorded to performance and knowledge versus accorded to one's title	• participatory ergonomics • organizational design • management of change

and participate in decision making. Because power–distance is low, employees in different levels of the organization and in different positions are able and used openly to discuss with each other. Participatory ergonomics often means that people work together in groups. Individualism–collectivism is related to managing relationships, and may therefore impact the effectiveness of participatory ergonomics in different countries. In individualist countries, it may be more difficult to implement group-based participatory ergonomics programs because the individual's performance is more important than the group's performance.

Finally, participatory ergonomics programs may challenge the existing status structure of an organization. Trompenaars' dimension of achievement–ascription should therefore be taken into account. In some cultures, status depends on what one does (i.e. performance and knowledge), whereas in other cultures, status depends on what one is (i.e. title). In cultures where status is ascribed, managers and/or supervisors may resist participatory programs because of the threat to their status, i.e. their position/title in the organization.

3.2. Design of Work Systems and Organizations

According to the balance theory of job design, a work system is comprised of the following elements: the individual, tasks, tools and technologies, and environment and organizational factors. Various factors of the work system influence the quality of working

life of employees, e.g. psychosocial work factors. Psychosocial work factors are 'perceived' characteristics of the work environment that have an emotional connotation for workers and managers, and that can result in stress and strain (Hagberg *et al*. 1995). According to cross-cultural models, the psychosocial work factors of importance to people may vary from country to country: stress and strain of employees in different countries may be influenced by different psychosocial work factors. Studies by Järvenpää and Carayon have shown that the relationship between psychosocial work factors and strain was slightly different for Finnish and American office workers. In a study of Finnish and American computer users, Järvenpää *et al*. (1993) found that high workload and low variety were related to high strain in both Finnish and American workers. However, in the American sample, high strain was also related to high job future ambiguity, frequent computer-related problems and lack of feedback; whereas in the Finnish sample, lack of participation and low co-worker social support were related to high strain. These results may be explained by the cultural differences between the USA and Finland on the characteristics of uncertainty–avoidance and masculinity–femininity. The USA has a higher score of uncertainty–avoidance than Finland, therefore psychosocial work factors related to uncertainty (e.g. job future ambiguity, frequency of computer-related problems and lack of feedback) may be more stressful for Americans than for Finnish people. Finland is a feminine country, whereas the USA is a masculine country: in a feminine country,

social support may be more important than in a masculine country, therefore Finnish people may be more stressed by lack of co-worker social support than Americans.

The design of organizations needs to fit the characteristics of the national culture. Trompenaars and Hampden-Turner (1998) proposed a mapping of national cultures onto corporate cultures. They define four types of corporate culture: (1) the family (person-oriented culture), (2) the Eiffel tower (role-oriented culture), (3) the guided missile (project-oriented culture) and (4) the incubator (fulfillment-oriented culture). They found the following patterns between corporate and national cultures:

- family companies — France and Spain
- Eiffel tower companies — Germany
- guided missile companies — USA and UK
- incubator companies — Sweden.

Other models of the fit between national cultures and corporate cultures have been proposed (e.g. Hofstede *et al.* (1990).

3.3. Management in a Cross-cultural Context

Management style, that is the way managers behave and act in leading employees and organizations, may differ across cultures. Highson and Pugh (1995) showed that in individualist countries, management is management of individuals, whereas in collectivist countries it is the management of groups. Power–distance is another cultural characteristic that affects management style. Expectations for management in countries with low power–distance differ from those in high power–distance countries, as Highson and Pugh (1995) show. These findings have implications for management in a cross-cultural context. Multinational companies need to understand the cultural context of the countries where they run some operations. The managers responsible for these operations and the employees involved need to adapt their management style to the country's culture. In product development, teams can work in several countries designing different parts of the same product. Managing cross-cultural product development teams is a big challenge, especially when the cultural characteristics of the team members' countries differ. Team members have different expectations of management. This may affect communication, knowledge sharing, responsibilities and job roles.

Management of change is another important macroergonomic issue that should take into account cultural characteristics, in particular Hofstede's short- versus long-term orientation and Trompenaars' characteristics of diffuse-specific involvement and achievement–ascription. In long-term oriented cultures, change may be easier to implement because these cultures are more dynamic and perseverance towards slow results is valued. In short-term oriented cultures, quick results are expected, therefore putting pressure on organizations and their employees to change fast. In managing change in specific-oriented cultures, specific objectives with specific deadlines should be put forward, whereas in diffuse-oriented cultures, continuous improvement approaches are more valued. Management of change should also take into account the status of those involved in the change process. For instance, project managers may be chosen according to their title in ascribed-oriented cultures and according to their performance and knowledge in achieved-oriented cultures. Korunka and Carayon (1998) confirm these cultural differences in the management of change: there were significant differences in the

way technological change was managed in Austria and in the USA. For instance, stronger negative effects on employees, such as high reduction in personnel, were found in the American cases: this finding confirms the short-orientation of US culture as compared with that in Austria, as well as the specific characteristic of the American culture as compared with the more diffuse characteristic of the Austrian culture.

4. CONCLUSION

Cross-cultural issues are extremely important today because of the globalization of businesses. Understanding and managing cross-cultural aspects in organizations not only has, therefore, become important for the competitiveness of companies, but also for the well-being of employees. This paper has presented cross-cultural models of management that define important cross-cultural characteristics that should be taken into account in macroergonomics. The importance of these cross-cultural characteristics for participatory ergonomics, work design and management has been emphasized.

REFERENCES

HAGBERG, M., SILVERSTEIN, B., WELLS, R., SMITH, M. J., HENDRICK, H. W., CARAYON, P. and PERUSSE, M., 1995, In Kuorinka, I. and Forcier, L. (Eds) *Work-related Musculoskeletal Disorders (WMSDs): a reference book for prevention* (London: Taylor & Francis).

HAIMS, M. C. and CARAYON, P., 1998, Theory and practice for the implementation of 'in-house', continuous improvement participatory ergonomic programs. *Applied Ergonomics*, 29, 461–72.

HENDRICK, H. W., 1991, Human factors in organizational design and management. *Ergonomics*, 34, 743–56.

HENDRICK, H. W., 1997, Organizational design and macroergonomics. In Salvendy, G. (Eds) *Handbook of Human Factors and Ergonomics* (New York: John Wiley), pp. 594–636.

HICKSON, D. J. and PUGH, D. S., 1995, *Management Worldwide: the impact of societal culture on organizations around the globe* (Harmondsworth: Penguin).

HOFSTEDE, G., 1983, National cultures in four dimensions. A research-based theory of cultural differences among nations. *International Studies of Management and Organization*, 13, 46–74.

HOFSTEDE, G., 1994, Management scientists are human. *Management Science*, 40, 4–13.

HOFSTEDE, G., 1997, *Cultures and Organizations: software of the mind* (New York: McGraw Hill).

HOFSTEDE, G. and BOND, M. H., 1988, The Confucius connection: from cultural roots to economic growth. *Organizational Dynamics*, 16, 4–21.

HOFSTEDE, G., NEUYEN, B., OHAYV, D. D. and SANDERS, G., 1990, Measuring organizational cultures: a qualitative and quantitative study. *Administrative Science Quarterly*, 35, 286–316.

IMADA, A. S., 1994, Self-learning organizations and participatory strategies. In *12th Triennial Congress of the International Ergonomics Association*, vol. 6 (Toronto: Human Factors Society of Canada), pp. 98–9.

JÄRVENPÄÄ, E. and ELORANTA, E., 1997, Organizational culture and organizational development: a case study in process industry. In Seppälä, P., Luopajärvi, T., Nygård, C.-H. and Mattila, M. (Eds) *From Experience to Innovation. IEA '97*, vol. 1: *Organizational Design and Management* (Helsinki: Finnish Institute of Occupational Health), pp. 76–8.

JÄRVENPÄÄ, E. and IMMONEN, S., 1996, Cross-cultural aspects in job

stress and managerial behavior. In Hacker, W., Richter, P. and Wagner, T. (Eds) *Analysis and Design of Mental Work* (Dresden: Dresden University of Technology. IV Workshop TU Helsinki — TU Dresden), December 1995, pp. 1–10.

JÄRVENPÄÄ, E., CARAYON, P., HAJNAL, C., LIM, S.-Y. and YANG, C.-L., 1993, A cross-cultural framework for the study of stress among computer users: comparison of the USA and Finland. In Smith, M. J. and Salvendy, G. (Eds) *Human–Computer Interaction: applications and case studies* (Amsterdam: Elsevier).

JÄRVENPÄÄ, E, CARAYON, P. and DERJANI-BAYEH, A., 1997, The fit between national and organizational cultures from the employee point of view (L'adéquation entre les cultures nationales et organisationnelles vue par les employés). In *Industrial Engineering in a World Without Borders*. Proceedings of the 2nd International Conference Franco-Québécois, 3–5 September 1997 (Albi: Ecole des Mines d'Albi).

KORUNKA, C. and CARAYON, P., 1998, Continuous implementation of information technology: The development of an interview guide and a cross-national comparison of Austrian and American organizations. *International Journal of Human Factors in Manufacturing* (in press).

NORO, K., 1991, Concepts, methods and people. In Noro, K. and Imada, A. S. (Eds) *Participatory Ergonomics* (London: Taylor & Francis), pp. 3–30.

SMITH, M. J., 1994, A case study of participatory ergonomics and safety program in a meat processing plant. In *12th Triennial Congress of the International Ergonomics Association*, vol. 6 (Toronto: Human Factors Association of Canada), pp. 114–16.

TROMPENAARS, F. and HAMPDEN-TURNER, C., 1998, *Riding the Waves of Culture* (New York: McGraw-Hill).

VINK, P., PEETERS, M., GRUNDEMANN, R. W. M., SMULDERS, P. G. W., KOMPIER, M. A. J. and DUL, J., 1995, A participatory ergonomics approach to reduce mental and physical workload. *International Journal of Industrial Ergonomics*, 15, 389–96.

WILSON, J. R. and HAINES, H. M., 1997, Participatory ergonomics. In Salvendy, G. (Ed.) *Handbook of Human Factors and Ergonomics* (New York: John Wiley), pp. 490–513.

WILSON, J. R. and HAINES, H. M., 1998, Participatory design in the organisational context. In Scott, P. A., Bridger, R. S. and Charteris, J. (Eds) *Global Ergonomics* (Amsterdam: Elsevier), pp. 11–19.

ZINK, K. J., 1996, Continuous improvement through employee participation: some experiences from a long-term study. In Brown, O. Jr and Hendrick, H. W. (Eds) *Human Factors in Organizational Design and Management — V* (Amsterdam: Elsevier), pp. 155–60.

Design of Shift Systems for Shiftwork

P. Knauth

Department of Ergonomics, Institute for Industrial Production, University of Karlsruhe, Hertzstraße 16, D-76187 Karlsruhe, Germany

1. INTRODUCTION

In contrast with "normal" daywork, shiftwork means either work at varying times of day (e.g. morning shift, afternoon shift and night shift) or work at constant but unusual times of day (e.g. permanent afternoon shift or permanent night shift). However, with increasing flexibility of daywork the borderline between daywork and shiftwork is difficult to define. The various features of any given shift system, such as the timing and duration of shifts, and the direction and speed of rotation of the shift system may result in various problems for the shiftworkers.

The sleep–wake cycle and body rhythms of shiftworkers may be disrupted. Furthermore, shiftworkers live at least partly disconnected from their social environment. Therefore, the individual shiftworker's problems can range from disturbed sleep to impaired well-being, health and performance, as well as to impoverished family and social life. As a consequence, their employers may be faced with increased labor turnover and absenteeism, as well as reduced safety and productivity. Therefore it is necessary to seek solutions to reduce these problems.

2. PROBLEMS OF SHIFTWORKERS

The human body is mainly 'programmed' for daytime work performance and for night-time recreation and rest. The disruption of daily variations in physiological functions caused by having to be at work at biologically unusual hours as well as by having to sleep during the day is one of the major points of stress associated with shiftwork. Many shiftworkers have the subjective impression that they feel better adjusted to nightwork after two or three night shifts. However, most authors who have studied short-term adjustment to nightwork using objective measures have found that even after 1 week of continuous nightwork the physiological rhythms of most shiftworkers are only partly adjusted.

Furthermore, a variety of workplace hazards such as noise, unfavorable climatic conditions, unfavorable lighting conditions, vibrations and a combination of these can sometimes occur more often in three-shift systems, irregular systems and night shift systems than in two-shift systems or daywork.

Although the stress of a group of shiftworkers with the same shift system is the same for each member of the group, the magnitude of individual disturbances is modified both by individual differences in age, experience, personality, capacity to cope, gender, physical fitness, etc. and by situational differences, e.g. related to domestic circumstances, environmental noise levels at home, commuting time, etc. Nevertheless, in many studies on the effects of shiftwork (Colquhoun *et al.* 1996) some general trends have been observed.

Sleeping difficulties are a major concern in the life of shiftworkers, particularly for those who have to work night shifts.

Table 1. Potential negative effects of shiftwork (Knauth 1996). Shaded areas are potential problems.

On an average day sleep after a night shift has a reduced duration (~6 h) and the quality is less good compared with normal night sleep. Therefore, in a week of night shifts an accumulation of sleep deficits has to be expected. In addition, the sleep before morning shifts is shorter, the earlier the morning shift starts. A week of morning shifts, all starting very early, results also in an accumulation of sleep deficits. Therefore, shiftworkers may be very tired at the end of a week of such morning shifts or of night shifts.

Nightwork leads to a change in the sequence and timing of meals. During the night, the body cannot adequately cope with the composition and the quantity of a typical daytime meal, because the activities of stomach, liver and kidney are very much reduced at night. Some nightworkers suffer more from disturbances of appetite than dayworkers or shiftworkers not on night shift.

In the long run, irregular food intake can lead to gastrointestinal complaints or even to disorders. However, the reasons for complex gastrointestinal symptoms are surely manifold. Some longitudinal studies have shown that those shiftworkers who transferred to daywork out of shifts because of medical problems had a higher frequency of gastrointestinal disorders than the remaining shiftworkers.

Similarly an increase in cardiovascular disorders, particular among those who transferred away from shiftwork, has been observed. Altogether, nightwork is regarded as a risk factor for health, i.e. not every nightworker will develop illnesses. However, if nightwork is combined with aggregative factors (e.g. individual and situational factors from private life) the probability for disorders may be increased.

Shiftwork may have adverse effects on family life, contacts with friends and relatives, attendance at meetings of organizations, such as sports clubs, political parties or civic groups, or participation in evening classes. Most of the activities are bound to evening hours or the weekend. If a shiftworker works in evening shifts, night shifts or at weekends, he is desynchronized from the 'normal' social environment. Women on shiftwork may have more problems with domestic duties and sleep, since household responsibilities are not equally shared by the partners.

'Headline' incidents such as the chemical/gas leak at Bhopal, the nuclear accidents both at Three Mile Island and Chernobyl, and the *Exxon Valdez* tanker wreck, all of which occurred at night, have drawn attention to the potential cost of impaired performance and safety in shift systems. It seems probable that the potential loss of efficiency and impaired safety on abnormal work

routines reflects the combined influence of a number of factors. These may include the relatively indirect influences of social factors and impaired health, as well as the more direct influences of disturbed circadian rhythms in sleepiness/alertness and performance capabilities. However, more detailed studies or 'real job' measures of safety and efficiency in different shift systems are needed before adequate recommendations can be made (for a review, see Colquhoun *et al.* 1996).

There are several possibilities for reducing the problems of shiftworkers such as special health measures (Colquhoun *et al.* 1996), guidelines for shiftworkers (Wedderburn 1991) or most effectively the design of shift systems according to ergonomic recommendations.

3. ERGONOMIC RECOMMENDATIONS FOR THE DESIGN OF SHIFT SYSTEMS

If shiftworkers are asked which shift system they feel is the best, very often – irrespective of the type of shift system – they name their roster, i.e. the one to which they are accustomed. Therefore, it seems more sensible to assess the advantages and drawbacks of a shift system on the basis of objective criteria, such as adjustment of physiological functions to nightwork, level of well-being (e.g. sleep, fatigue, appetite), health problems, disturbances in personal and social life, performance efficiency and accident rates (allowing for time of day and particular type of shift system in operation). The most important ergonomic recommendations for the design of shift systems are summarized in Table 2.

4. PERMANENT NIGHTWORK

The permanent night shift is the most disruptive of all shifts in terms of physiological adjustment, sleep and well-being. The daily variations of the physiological rhythms of most shiftworkers may require > 1 week for complete adjustment to nightwork. Any partial adjustment will be lost during the days off after consecutive night shifts. Thus, the body rhythms of permanent nightworkers are constantly in a state of disruption, whereas the least distortion of rhythms has been observed to occur after a single night shift. In

Table 2. Recommendations for the design of working time arrangements (Knauth 1993).

1. Nightwork should be reduced as much as possible. If this is not possible quickly rotating shift systems are preferable to slowly rotating ones. Permanent nightwork does not seem advisable for the majority of shiftworkers.

2. Extended workdays (9–12h) should be contemplated only if: the nature of work and the workload are suitable; the shift system is designed to minimize the accumulation of fatigue; there are adequate arrangements to assure that a complete recovery after work and a high acceptance of the working time arrangement are possible.

3. An early start for the morning shift should be avoided. In all shift systems flexible working time arrangements are realisable. The highest flexibility is possible in 'time autonomous groups'.

4. Quick changeovers (e.g. from night to afternoon shift on the same day or from afternoon to morning shift) must be avoided. The number of consecutive working days should be limited to 5–7 days. Every shift system should include some free weekends with at least 2 successive full days off.

5. The forward rotation of shifts (morning shift, evening shift, night shift) would seem to be recommendable at least in continuous shift systems.

contrast with our view that the physiological rhythms of shiftworkers should remain as near as possible to the 'normal state', i.e. should be disturbed as little as possible, some authors recommend accelerating adjustment to nightwork with the help of bright light. However, up to now not enough is known about the optimal light–work–sleep schedule for shiftworkers in terms of their ability to shift biological rhythms, improve sleep and reduce fatigue, as well as in terms of their social feasibility. Furthermore, no controlled long-term studies on the effects of such a bright light treatment have been carried out.

As detailed studies of sleep protocols of permanent nightworkers over several weeks have shown, in most cases an accumulation of sleep deficit was observed. Furthermore, former nightworkers, who transferred to daywork after many years of permanent nightwork, had more sleep problems than their colleagues, who had previously worked in alternating shift systems.

In comparison with other shift systems, permanent night shifts also have more negative effects on families, who must adapt their lifestyle to this schedule, on sexual relations and on a worker's ability to fulfil family roles.

Permanent nightwork also has some advantages. Nightworkers report a greater feeling of independence and less supervision at night. Very often nightwork is chosen because of the increase in income due to the night shift allowance. Some female nurses prefer permanent nightwork, because this represents the only way of better arranging domestic responsibilities and employment outside the home.

Although, there is insufficient knowledge about the long-term health effects of permanent nightwork and about optimal bright light–work–sleep schedules, it is known that night shifts are the most disruptive of all shifts in terms of physiological adjustment, sleep and well-being. Therefore, until results from further research are available, it will be assumed for the moment that permanent nightwork is not recommendable for the majority of shiftworkers.

5. QUICKLY ROTATING VERSUS SLOWLY ROTATING SHIFT SYSTEMS

More rapidly rotating shift systems are more advantageous compared with weekly shift rotation for the following reasons:

- Biological rhythms are not in a constant state of disruption from partial adjustment to different day and night orientations, but are kept nearer to a daytime orientation.
- The bigger accumulation of sleep deficit such as during a week of night shifts and during a week of morning shifts (starting early) are avoided.
- Shiftworkers have free evenings in every week and, thus, more regular contact with friends is possible than with weekly rotating shift systems.
- The relative risk of accidents increases during (several) consecutive night shifts.

In several longitudinal studies, the majority of shiftworkers who had experienced weekly and quicker rotating shift systems always voted in favor of quicker rotating shifts after a trial period.

6. DURATION OF SHIFTS

The potential drawbacks and advantages of 12-h shifts have been extensively discussed (Rosa 1995). According to our experience an

extended workday of 9-12 h should be contemplated only in the cases mentioned in Table 2. In particular, this means that 12-h shifts are not compatible with high physical, mental or emotional demands of a job. Furthermore, if there are many extended work shifts in a row, an accumulation of fatigue has to be expected. Similarly if there are no adequate arrangements for the cover of absentees and if overtime is added, the shiftworkers may become exhausted. As our knowledge is very limited about toxic exposure and toxic clearance during the time off work in connection with extended working hours, it is recommend that one limit toxic exposure as far as possible or not to introduce extended working hours.

7. TIMING OF SHIFTS

An early start of the morning shift seems unfavorable because it may reduce sleep duration before the morning shift, increase fatigue during the morning shift and increase the risk of errors and accidents in the morning shift. On the other hand, a late start of the morning shift means (assuming there is a constant shift length of 8 h) that the evening shift and the night shift will also start and end later.

Other factors, such as availability of public transport, opening hours of shops, main time of social contacts in the evening, meal times together with the family or particular requirements of the job, have to be taken into consideration when fixing starting and finishing times of shifts.

Flexible working time arrangements are also possible even in three-shift systems, where employees can choose their working hours in agreement with their colleagues.

8. DISTRIBUTION OF LEISURE TIME WITHIN THE SHIFT SYSTEM

When the interval of time off between two shifts is too short (e.g. 7, 8 or 10 h only) sleep is reduced, well-being impaired and fatigue increased. Too many working days in succession can lead to an accumulation of fatigue. Although it is not easy to define a limit for the maximum number of consecutive working days — because the workload, the organization of breaks, and exposure to unfavorable conditions vary from workplace to workplace — it is recommend that one limit the number of consecutive working days to between 5 and 7.

Free weekends are of particular social importance for the majority of workers. Instead of only 1 or 1.5 days off (after night shifts), every shift system should include some free weekends with at least 2 consecutive days off.

9. DIRECTION OF ROTATION

A shift system that first moves from the morning shift to the evening shift, and then to the night shift, has a 'forward rotation' (phase delay, clockwise rotation). An anti-clockwise or 'backward rotation' has a phase advance that moves from night to evening to morning shifts. The forward rotation appears preferable because it corresponds more closely to the human biological rhythm. A comparison of shiftworkers with forward and backward rotating shift systems revealed more digestive troubles, more cardiovascular complaints, more chronic fatigue and more disturbances of family and social life in the latter group. The worst case comprises backward rotating shift systems, which include several short intervals of time off between two shifts.

The only advantage of a backward rotation may be a long period off work between the end of the last morning and the start of the first night shift, in particular when this period includes a weekend.

10. EXAMPLE OF A FAVORABLE SHIFT SYSTEM

It is not always possible to adhere to all recommendations mentioned above simultaneously. However, the shorter the weekly working hours the easier it is to follow them. In Figure 1 an example of a continuous shift system is shown that corresponds quite closely to the recommendations. This shift system is operated by five teams, i.e. the basic schedule has a mean weekly working time of 33.6 h. Depending on the agreed weekly working time, additional shifts have to be worked, which may be used for continuation training, maintenance or, in a limited number of cases, to provide cover for absentees.

The shift system has short periods of morning shifts, evening shifts and night shifts, and a forward rotation of the shifts. There are longer blocks of free weekends, a regular pattern and > 48 h off between the end of the last night shift and the start of the following morning shift.

11. INTRODUCTION OF A NEW SHIFT SYSTEM

Besides the design of a favorable shift system, the implementation strategy is of particular importance. Shiftworkers who have worked for many years in one shift system have become accustomed to it and know all the drawbacks and advantages. If the employer wants to implement a new shift system this causes doubts, fear or even resistance of the shiftworkers and often also skepticism of the middle managers. To achieve a high acceptance of a new shift system, an adequate strategy is necessary which includes:

- early dissemination of this information to everybody concerned;
- participation of the shiftworkers in the planning phase, i.e. in the finding of a solution that is a good compromise between the demands of the company and the wishes of the workers;
- regular discussions during the trial period; and
- a ballot after the trial period. (If the shiftworkers know that they may vote after the trial period, their fear will be reduced and their willingness to try something new is increased.)

Week	Mo	Tu	We	Th	Fr	Sa	Su	Mo	Tu	We	Th	Fr	Sa	Su
A	M	M	E	E	N	N					M	M	E	E
B	N	N						M	M	E	E	N	N	
C			M	M	E	E	N	N					M	M
D	E	E	N	N					M	M	E	E	N	N
E					M	M	E	E	N	N				

Average weekly working hours

Men

Day off
Morning shift   33.6 h/week without additional shift
Evening shift   35.2 h/week with two additional shifts/10 weeks
Night shift   36.8 h/week with four additional shifts/10 weeks
38.4 h/week with six additional shifts/10 weeks
40.0 h/week with eight additional shifts/10 weeks

Figure 1. Continuous shift system with five teams.

REFERENCES

COLQUHOUN, W. P., COSTA, G., FOLKARD, S. and KNAUTH, P., 1996, *Shiftwork: problems and solutions* (Frankfurt a. M.: Peter Lang).

KNAUTH, P., 1993, The design of shift systems. *Ergonomics*, 36, 15-28.

ROSA, R. R., 1995, Extended workshifts and excessive fatigue. *Journal of Sleep Research*, 4(Suppl. 2), 51-6.

WEDDERBURN, A., 1991, *Guidelines for Shiftworkers*. Bulletin of European Shiftwork Topics (BEST), no. 3 (Dublin: European Foundation for the Improvement of Living and Working Conditions).

Distributed Mission Training

D. H. Andrews

United States Air Force Research Laboratory, Human Effectiveness Directorate, Warfighter Training Research Division, 6030 S. Kent, Mesa, AZ. 85212-0904.

1. MISSION TRAINING CHALLENGES FACING MILITARY PILOTS

Today's military pilots face significant challenges in their attempt to "train as they intend to fight" (Andrews 1988). A variety of constraints makes it difficult to provide realistic training environments on a consistent basis so that new skills can be learned and existing skills can be refined. Some of these constraints include:

- reductions in flying hours that can be used for training;
- training range restrictions;
- aging aircraft that makes it more difficult to use them for training hours versus operational hours;
- safety constraints such as altitude floors below which trainees cannot fly in peacetime but below which they would fly during combat; and
- security constraints that will not allow pilots to use their full range of tactics and electronic warfare gear on training ranges.

These constraints make it difficult for pilots to maintain the readiness that will be required in combat. These constraints promise to grow more challenging and, thus, will make it more difficult to conduct mission training only in actual equipment on existing training ranges (Andrews *et al.* 1995).

Mission training is aimed at teaching pilots higher-order skills necessary to conduct tactical missions. This type of training usually involves training with team members who may be on board one aircraft, and it often involves training with team members who are in other aircraft and on the ground. For example, many air forces typically have their fighter pilots fly in four ship formations when they undergo a tactical mission task. Not only will mission training involve the four pilots learning and practicing team skills together, but also it will also involve radar operators in other aircraft and on the ground.

Mission training differs from procedures training in that it is designed to train higher-order skills such as advanced problem solving in tactical settings rather than information and skills required simply to fly the aircraft safely. Examples of normal procedures training are training a pilot to land an undamaged aircraft or training a pilot to handle an emergency procedure like handling an aircraft with an engine fire. Aircraft procedures must be well learned before mission training can begin. Mission training includes such functions as suppression of enemy air defenses, defensive counter air and air–ground attack.

2. USE OF SYNTHETIC ENVIRONMENTS TO MEET TRAINING CHALLENGES

Fortunately, modeling and simulation technologies have been developed that can be of great assistance in meeting the training challenges discussed above (Hays and Singer 1989, Leplat 1989). These technologies take many different forms, and when used together allow military pilot instructors to conduct mission training on a distributed basis. Affordability is a major advantage of simulation when compared with actual live-fly training (Orlansky and Chatelier 1983). The term "distributed" means that simulators at multiple sites, often in geographically dispersed locations, are linked together via wide-area computer networks. In addition to virtual simulators, live assets such as aircraft and tanks on training ranges, as well as constructive computer models such as digital war games, can be linked into the simulated training engagement for added realism (Alluisi 1991).

This "synthetic battlefield" can consist of as few as two networked sites or many dozens or hundreds. As the networking technology matures it is feasible to think of thousands of sites representing all manner of weapons systems and military units, not just aircraft. A key engineering issue is making sure that temporal latency does not exceed tight tolerances (less than a few milliseconds) lest the pilot trainees detect non-realistic update rates in their simulators and actual weapon system. Thus far, the latency issue has not proven too great a difficulty as dozens of sites have been networked in early versions of DMT; however, it will be a concern when hundreds or thousands of sites are linked (Bell and Waag 1998). Figure 1 is a graphic representation of DMT showing the virtual (simulation), live and constructive model assets that are joined together to create the synthetic battlefield.

The following are examples of engineering technologies that have allowed militaries to provide distributed mission training to their combat pilots:

- Simulation visual systems (image generators that produce computer-generated imagery, visual displays such as domes and helmet-mounted displays, and database development tools that allow simulation developers to produce realistic visual and sensor (e.g. infrared and radar) terrain and multi-spectral environments).
- Networking technology for connecting simulated and live entities via wide area networks.
- Representation technologies that allow development realistic of computer-generated models of both friendly and enemy weapon systems and personnel.
- Instructional technologies, such as instructor–operator
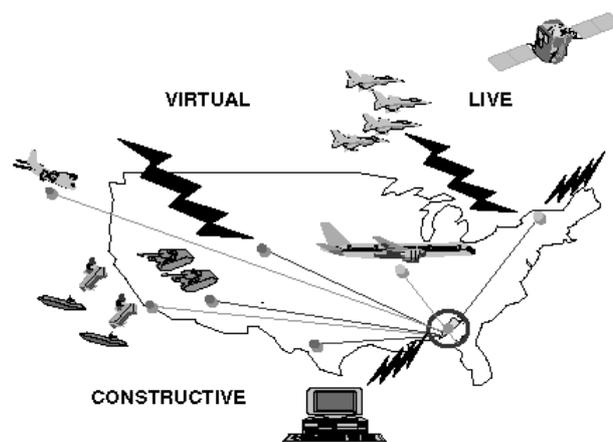
Figure 1. Graphical depiction of distributed mission training, including virtual, live and constructive assets
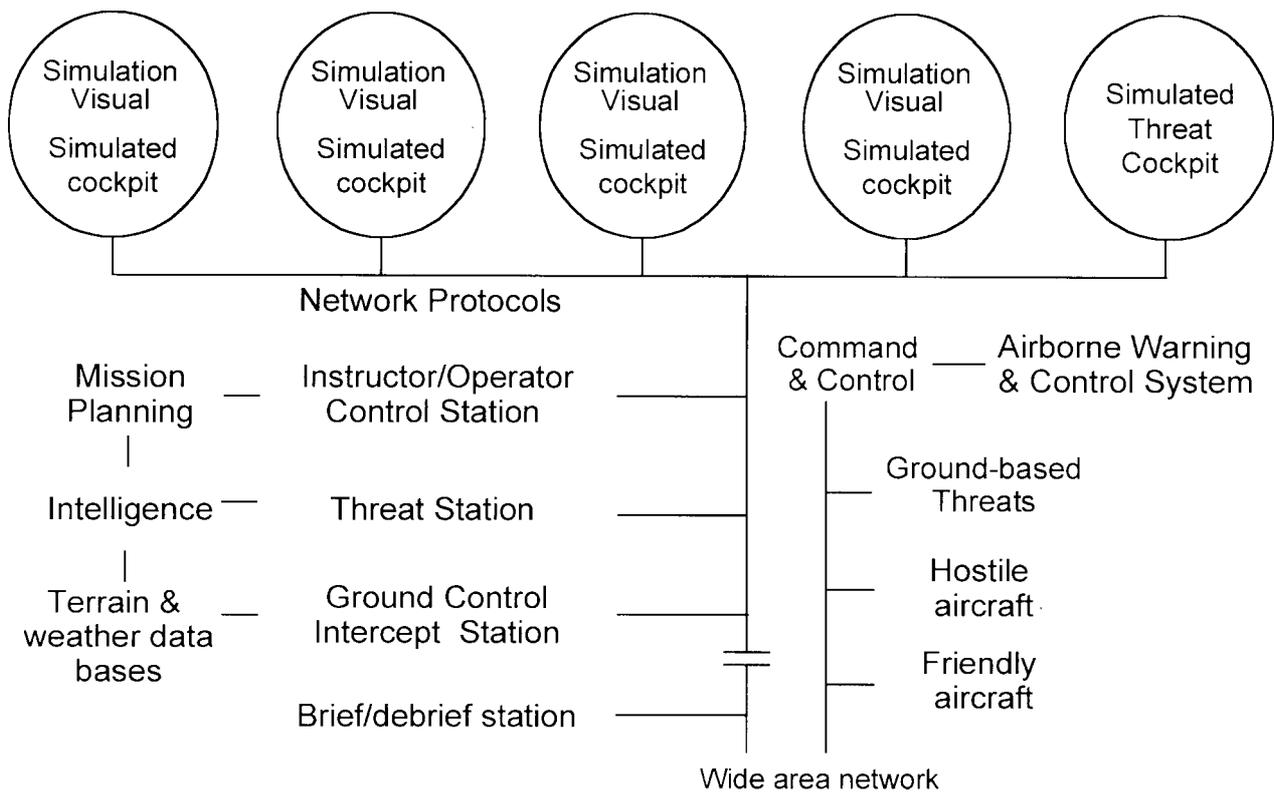
Figure 2. Typical networking schematic of distributed mission training.

stations and sophisticated debrief tools, that aid instructors in providing the necessary training help to trainees.

Figure 2 shows a typical DMT network diagram.

3. HUMAN FACTORS AND INSTRUCTIONAL ISSUES ASSOCIATED WITH DISTRIBUTED MISSION TRAINING

There are a variety of human factor and instructional issues that must be addressed if DMT training capabilities are to reach their full potential. Major examples include:

- Instructional tools for helping instructors train with trainees at distributed locations must be improved. Currently instruction, including briefs and debriefs are carried out over long-distance telephone or video conferencing. What is missing are validated tools that allow instructors at different locations automatically to see and share all the performance data collected at all the sites in a synthesized manner. Briefings (including mission planning) and debriefings are difficult enough to conduct properly when all the team members are in the same location. The task difficulty increases greatly over long distances and effective tools for the task are critical.
- Regardless of how many virtual or live training assets that are available for training, it will probably never be possible fully to represent an entire theater of battle with just human in the loop systems. Constructive models that accurately represent synthetic friends and foes are crucial if DMT is to reach its full potential. Currently constructive models either do not represent key human behaviors, or if they are represented they are not realistic. For example, current models do not represent the human trait of fear. The conse-

quence is that synthetic humans are usually fearless regardless of the situation. A human construct that is now being represented more often in wargame constructive models, but which is seldom represented very accurately is fatigue. Most model builders now recognize that warfighter effectiveness will degrade over time in battle, but the fatigue models presently in use are not based on any scientific data, but rather a degradation model that degrades at a steady rate over time. The consequence of no or poor human representation in constructive models may well be negative learning (Allesi 1988, Baker and Marshall 1989) where trainees learn to expect one behavior from a friend or foe but are surprised in the real world by a new behavior or a behavior that is different than expected.

- Visual resolution in simulator visual displays must be improved. The goal is to achieve eye-limiting (20/20) resolution. Currently, tasks such as air–ground attack mission training and formation flying training are difficult to train in DMT because of visual resolution limitations. Pilots' ability to fly in formation and to identify and attack enemy is dependent upon accurate visual representations.
- Fighter pilots must learn to distinguish key cues from a host of stimuli presented to them in the cockpit. Visual stimuli from outside the cockpit (including night vision goggle images), infrared, radar and other sensor stimuli all must be sorted and interpreted by the pilot to make proper tactical decisions. It is vital that the simulations, which present these different stimuli, be correlated so that the pilot is presented with a realistic picture. The engineering community has still not entirely solved this challenge and trainees might see

visual cues outside the cockpit that represent outside stimuli differently than a radar or forward-looking infrared sensor simulation of the same stimuli. Negative learning cannot take place if visual and sensor stimuli are not well correlated. This correlation issue is present whenever tactical aircraft flight simulators are used, but it becomes even more critical when there are multiple pilots in the loop because a synthesized picture is crucial for team tactics training.

DMT combines virtual, live and constructive assets. Experiments with DMT so far have shown that it is difficult optimally to merge this combination for maximum training effectiveness. Most of the engineering technical hurdles for combining these assets have been overcome, but the instructors who make use of these asset types often have difficulty in knowing how best to make use of the training capabilities from the other asset types. For example, constructive wargames may involve hundreds of trainees who are learning to conduct a combat operation across an entire theater. These trainees are principally concerned with large-scale command and control issues. Consequently, instructors in these settings are not very interested in particular actions of individual pilots or small teams of pilots who might be flying in the training exercises via simulators or aircraft. Therefore, the constructive instructors may ignore important training information that could be learned from the training experience of the individual pilots and small teams.

4. KEY INSTRUCTIONAL GUIDELINES FOR MAKING DISTRIBUTED MISSION TRAINING EFFECTIVE

Although DMT is relatively new and much empirical and analytical research remains to be done, a fair amount has been learned from training effectiveness studies and practical experience about how best conduct DMT training. Following are some guidelines that have emerged:

- It is critical that those who conduct DMT develop systematic instructional programs with clearly defined instructional objectives, performance measures, instructional syllabi and evaluation methods. Too often pilots are happy merely to have a synthetic battlefield environment that nearly replicates their training ranges. Once that goal is achieved, pilots have been happy merely to practice in a freeplay environment that is not well structured from an instructional standpoint. There is no doubt that learning occurs with such an approach, but it is not systematic and not very efficient. Learning to work in well-coordinated teams is a difficult process and it is best done using proven instructional systems approaches and not by mere practice.
- Owing to visual resolution limitations in current flight simulators, it is often necessary to enhance or augment visual cues so that pilots can see and react to these important cues. For example, in the real world a pilot can see the aspect angle (i.e. which direction the aircraft is pointing) of a wingman in formation out to ~15 000 feet depending upon visibility conditions. Beyond that range the pilot can see an aircraft but cannot tell the aspect angle. In current simulators a pilot can tell aspect angle only out to ~5000 feet. This limitation makes it difficult to perform many tasks that tactical teams must perform. This same limitation impacts pilots' abilities to perform tactical dogfights with simulated enemy fighters. These visual resolution limitations can be somewhat overcome by enhancing the simulated representation of friendly and enemy aircraft. For example, an aircraft's image might be made artificially larger at distances > 5000 feet. Another approach is to put artificial colored lights on the wings, fuselage and tail of a simulator aircraft better to help distinguish the aspect angle.

- Instructors who use DMT must be given clear training about how best to use the instructional capabilities that can only come from a synthetic environment learning environment. Instructional features such as freeze replay, performance measurement, and enhanced cues may be familiar to some instructors from their use of these features in traditional non-networked simulators. However, the use of these features differs in DMT, and DMT instructional designers must provide "train the trainer" sessions concerning the most effective use of these features.

5. FUTURE OF DISTRIBUTED MISSION TRAINING AND ITS APPLICATION TO NON-MILITARY ENVIRONMENTS

DMT will continue to grow in importance. New simulation technologies will allow ever more realistic training (Bolton *et al.* 1984). The affordability of synthetic environments when compared with using only expensive flying training hours will have great appeal as training budgets shrink. Currently, a simulator hour of training costs ~20% of a flying hour, and that percentage will shrink as these technologies drop in price. Already these same technologies are having an effect on mission training outside of pilot training such as: armor training, naval surface training and security police training.

Wherever large teams need to train and practice together, especially when the trainees are geographically dispersed, DMT will be of use including non-military training settings such as industrial and transportation training. An example of DMT use in industrial training might be a virtual factory training environment for industrial workers from different locations, who all work on different assemblies that are to be integrated to make the final product (e.g. an automobile that is assembled from parts made around the world). It may be that the parts production employees would have a better shared mental model of how their parts impact production workers at other locations in the assembly process if they all could assemble a virtual car together using representations of the parts they produce. Since it may well be cost prohibitive to have all the workers assemble at a central location for this training, DMT could be an efficient means for producing shared mental models.

An example of DMT training for transportation might be airline aircrews training with air traffic controllers for landing at an airport to which they have never been. Specific and peculiar conditions for landing at that airport could be depicted in the simulated visual and radar representations for the aircrew and the aircrew members could interact by voice with the air traffic controllers in that particular airport.

The key to the ultimate success of DMT is not whether the technology improves or whether it is affordable, but rather whether it fulfills training requirements. In other words, does it satisfy the basic human factors issues described in this entry by providing quality training that cannot be provided with traditional

training methods and technologies? Merely providing practice opportunities will ultimately not allow DMT to fulfill its full potential. Only when new and measurable mission skills and information are learned by trainees will DMT be a major training institution in the military.

REFERENCES

ALLESI, S.M., 1988, Fidelity in the design of instructional simulations. *Journal of Computer-based Instruction*, 15, 40–7.

ALLUISI, E.A., 1991, The development of technology for collective training: SIMNET a case history. *Human Factors*, 33, 343–62.

ANDREWS, D.H., 1988, Relationships among simulators, training devices, and learning: a behavioral view. *Educational Technology*, 28, 48–53.

ANDREWS, D.H., CARROLL, L.A. and BELL, H.H., 1995, *The Future of Selective Fidelity in Training Devices.* US Air Force Armstrong Laboratory Technical Report AL/HR-TR-1995-0195.

BAKER, S. and MARSHALL, E., 1989, Simulators for training and evaluation of operator performance. In L. D. Bainbridge and S. A. R. Quintanilla (eds) *Developing Skills with Information Technology* (New York: Wiley), pp. 293-314.

BELL, H. H. and WAAG, W.L., 1998, Evaluating the effectiveness of flight simulators for training combat skills: a review. *International Journal of Aviation Psychology*, 8, 223–42.

BOLTON, P.A., FAIGENBAUM, J.M., HOPE, A.M. and RANKIN, W.L., 1984, Simulator fidelity and training effectiveness: a comprehensive bibliography with selected annotations. Battelle Human Affairs Research Centers, Technical Report No. NUREU/CR-3726.

HAYS, R.T. and SINGER, M.J., 1989, *Simulator Fidelity in Training System Design: Bridging the Gap between Reality and Training* (New York: Springer).

LEPLAT, J., 1989, Simulation and simulators in training: some comments. In L. D. Bainbridge and S. A. R. Quintanilla (eds) *Developing Skills with Information Technology* (New York: Wiley), pp. 277–92.

ORLANSKY, J. and CHATELIER, P.R., 1983, The effectiveness and cost of simulators for training. In *Simulators*. Publication No. 226 (London: Institute of Electrical Engineers).

The Ergonomic Buddy System

L. T. Ostrom, C. A. Wilhelmsen and T. Lango
INEEL, PO Box 1525, Idaho Falls, ID 83415-3855, USA

1. INTRODUCTION

Currently, there are > 6000 employees at the Idaho National Engineering and Environmental Laboratory (INEEL). Of these, ~4000 work in an office environment to one degree or another. As of Spring 1999, ~3500 of these employees have had an ergonomic assessment of their offices and corrections made. This is quite an accomplishment given that just 3 years ago there were < 1000 employees who had an ergonomic assessment of their offices performed. Two thousand five hundred ergonomic assessments being performed in 3 years take a great effort, yet the INEEL had less than five individuals who have the training and expertise to perform a proper and adequate ergonomic assessment before 1997 and not much funding to support a program. An innovative process called the Ergonomic Buddy System (EBS) was used to accomplish this feat. This chapter will discuss that process.

2. HISTORY OF ERGONOMICS AT THE INEEL

The Human Factors unit was established at the Idaho National Engineering Laboratory (now INEEL) in the early 1980s. Its focus was to develop and test technologies designed to improve the operability of nuclear power plants and military systems. In June 1989 they hired me. My job was to work in the Human Factors Unit, but my focus was to be risk assessment and ergonomics. I immediately was also asked to head up the safety efforts for the Engineering Department. In December of that year I met Cheryl Wilhelmsen. She was in a support role at the time and was working on a degree in corporate training. I began working ergonomic issues in Spring 1990. At that time we conducted an initial survey of all the major worksites and performed initial assessments. This was under the direction of the industrial hygienist unit (Ostrom *et al*. 1991). The team set up to do this work was Gay Gilbert, Cheryl Wilhelmsen and myself. Cheryl was in a support role and helped prepare the report. As a part of the report an ergonomic checklist was developed and a rating scale for workplace design issues. In January 1991 Cheryl and I began a project to analyze the safety culture at the INEEL and found that we had similar goals for improving the working conditions of the employees at the INEEL and from that time we have been partners in crime.

From the ergonomic survey and the safety culture survey we realized that ergonomics was one of the least implemented parts of the safety and health program. We did not have support for performing ergonomics from the medical department right off the bat. In fact, the medical director did not believe there was a condition called carpal tunnel syndrome (CTS), though he claimed to experience symptoms from guiding his horses. So, during this time assessments were only performed if a person was experiencing pain and referred by medical. We were performing about one-to-two per month. It came to my attention that Cheryl was experiencing pain in her wrist and elbow. She was using a Macintosh computer and doing graphics for a book for another individual in the group. Upon reviewing the workstation I found that she was mousing a great deal and using it at a different height than the keyboard. Changes were made to her workstation and Henry Romero, a new ergonomist to the group at that time, designed a platform for Cheryl to use while doing intensive mouse work. This was installed and it helped her. Cheryl also developed a strategy to "listen to her body." Meaning, when she felt fatigued, she would take a fatigue break. This was consistent with both Grandjean's (1980) and Putz-Anderson's (1988) philosophies. Cheryl was committed not to experience an injury, as much to avoid needles as to avoid the injury itself. From this time Cheryl became committed to helping others avoid injuries because she knew what one goes through after developing a cumulative trauma disorder (CTD). Not just the pain, but the probing questions by workman's compensation insurance investigators and the medical procedures used to diagnose CTS. She was very committed not to have the CTS surgery. Cheryl also began to study ergonomics on her own so she could answer employee questions.

After this experience we began to advertise our services to the employees, in spite of the medical director. We performed numerous informational and formal training sessions on ergonomics and, as project work allowed, we performed ergonomic assessments. These were "free" assessments, meaning we had no formal budget to do them. One of the attributes of our training that was unique and helped drive the information home was Cheryl's testimony to the benefits of ergonomics and the need to listen to one's body. She would discuss her own case and how the changes to the workstation benefited her. At the end of 1993 we had performed ~200 clandestine ergonomic assessments.

In Spring 1994 Eloise Hayes asked us if we could help perform ~800 assessments at an INEEL facility run by another company. We agreed to this. This mass number of assessments set the stage for the development of the future ergonomics program. One reason being, though this facility was run by another company, they had very close ties to the one we worked for. So, employees on our side of the fence began to ask for assessments more. In addition, since we had a common medical department, medical could not deny that they were seeing benefit from the introduction of ergonomics. About this same time we began an ergonomic's committee to review office ergonomic equipment because managers were buying equipment based on sales people's recommendations or using the "looks good" principle. Again, this committee had no official sanction, but the results were widely used by managers as the basis of their purchases.

Our history of ergonomics sometimes parallel's the introduction of technology at the INEEL. In the early 1990s 1500 drop-keyboard desks were purchased. These desks had a recessed platform 21 inches long, and were non-adjustable. The desk itself was 29 inches tall and the keyboard platform was 26 inches high. These were bought because they, supposedly, were "ergonomically correct" because the keyboard height was 26 inches from the floor. However, they obviously were not designed to fit most individuals. One of the common findings of the ergonomic assessments was that a new desk was needed because the keyboard was not right and there was no room for a mouse. Management balked at the cost of the new desks, but could not deny they were needed to fix the problems. Therefore, we came up with an

idea to level the top of the desks and add an articulating keyboard platform. We estimated that our solution would save the company ~US$900,000 over the cost of new desks, not to mention the savings in human terms. We began a program to retrofit these desks. When we approached management about the fact that we could save them money as we improve the health of employees they began to buy into the ergonomic concept. This helped gain our first management champion, Derek Moore who was the Chief Engineer.

With the introduction of Windows 95 in 1996 employees began to use the mouse to a much higher degree and we began to see a much higher number of employees reporting to medical for evaluations. Managers began to see the need for ergonomic assessments due to the awareness training that had been conducted, as well as employee requests. We began to get swamped and could not keep up with the great number of assessments. We needed a scheme to help us perform more assessments because we did not have funding to hire more help. We came upon the idea of the self-assessment.

The first attempt at implementing the a self-assessment system was a behavioral-based safety project we developed where employees were asked to take a fatigue break and go to an ergonomic information kiosk and receive a Hersey's Kiss. "Take a fatigue break and get a kiss" was the obvious theme. This little program worked very well. At the kiosk there were informational flyers and a self-assessment checklist. This effort demonstrated that properly motivated employees would seek ergonomic information and would perform a self-check.

Next, we distributed self-assessment cards to employees. The self-assessment card contained questions and a diagram showing a person in a correct ergonomic position for a computer workstation (Ostrom 1994). Figure 1 shows the representation
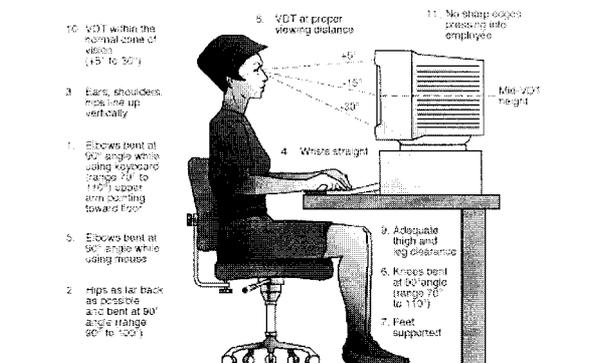
Questions:
Are my feet flat on the floor?
Are my wrists straight?
Are my elbows at a 90 degree angle?
Are my knees at a 90 degree angle?
Is my back support by the chair?
Is the monitor about arm's length away?
Is the mid-monitor height about four inches below eye height?
Do I feel comfortable?

If you answer No to any of these questions contact the ergonomic staff.

Figure 1. Self-assessment card.

of the card. The employees turned in these cards to a designated employee or us and received an award.

In conjunction with the self-assessment card, ergonomic demonstration offices were setup in three different locations on the INEEL. If the employees had questions concerning a piece of equipment (i.e. an "ergonomic" keyboard), s/he could go to the demonstration office and try one out before buying it to ensure it worked. The concept of having employees perform a self-assessment is very consistent with the principles of the Voluntary Protection Program (VPP) that the Department of Energy was beginning to promote. In mid-1996 we did a written survey and found that up to 30% of the employees were experiencing some symptoms of a CTD. Also, the demand for ergonomic assessments was so great we needed more help than just the self-assessment.

3. BIRTH OF THE ERGONOMIC BUDDY SYSTEM

Cheryl was the one who conceived of the EBS. In essence, the system involves four elements: 1, hierarchy of individuals who perform ergonomic assessments (self-assessments, ergo-buddies who are trained to do simple assessments industrial hygienists who perform more complex assessments, and ergonomists who perform the most complex assessments); 2, ergonomic committee that approves ergonomic equipment and directs the ergonomic effort; 3, ergonomic demo-offices where employees can try equipment; and 4, ergonomists on staff who serve as consultants to the organization. These are discussed below.

3.1. Ergonomic Assessments

In the EBS there are four groups who do ergonomic assessments: 1, employees who do self-assessments; 2, the ergo-buddies; 3, industrial hygienists; and 4, trained ergonomists. Figure 2 shows this hierarchy. The unique aspects of the EBS are the self-assessment process and the ergo-buddies, so they will be the items discussed here.

3.1.1. Self-assessments

The self-assessment process began as the self-assessment card being given out to employees, but this was found to not be as effective as we would like. To further improve the self-assessment process and the EBS we developed the concept for Buddy Ergonomic Software. Mountain Bluebird Products (Idaho Falls, ID Falls, USA) market it. It is a tool designed to:
- Guide employees through a self-assessment of their offices.
- Remind employees to take fatigue breaks.
- Show employees how to do stretching exercises.
- Show solutions to workplace design problems.
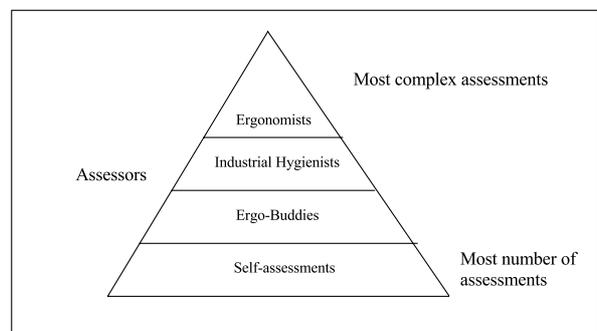
Figure 2. Diagram of ergo-buddy concept.

The primary benefit of Buddy is that it guides employees through a self-assessment of their office. It provides a simple three-step self-assessment process that is designed to be simple and effective. The software is interactive by design and provides the necessary helps to successfully guide the employee through the assessment. One of the unique features of Buddy is that it provides not only information as to what is wrong with the workstation, but also how to fix problems found. In addition, it provides a range of options so you have a choice as to how to fix any problems. For instance, sore neck from telephone usage, then the software says to try a headset phone.

One of the attributes of the Buddy Ergonomic Software is it generates a Move Card. A Move Card (Figure 3) contains the measurements necessary to set someone up in a workstation in the event they get transferred to a new location. The Move Card helps eliminate the need for a new ergonomic assessment every time an employee changes work locations. Their data goes with them and, so, movers can set them up right the first time, rather than setting them up using a general set of measures that do not fit anyone.

3.1.2. Ergo-buddies

Because we could not be everywhere, all the time we developed the concept for an ergo-buddy. Ergo-buddies at the INEEL are volunteers. The purpose of the ergo-buddy is to do ergonomic assessments for individuals who need some extra help or who feel uncomfortable doing their own assessments. The ergo-buddies receive 8 h of training on ergonomics. The training consists of:

- basics of ergonomics;
- the nature of CTD;
- proper workplace design; and
- environmental factors.

The ergo-buddies were selected based on their desire to help and so that there is one in the majority of the sites at the INEEL.

3.1.3. Ergonomic committee

The Ergonomic Committee at the INEEL is made up of employees with an interest in ergonomics, procurement personnel who are responsible for buying ergonomic equipment, ergo-buddies, industrial hygienists, ergonomists and management. The committee meets once per month to discuss issues relating to ergonomics on the site and, on a periodic basis, assess equipment for inclusion as "approved ergonomic items." We have found the ergonomic committee to be an integral part of our program and some of the best ideas have been formulated through the committee efforts.

3.1.4. Ergonomic demo-offices

The ergonomics demo-offices have been in place for almost 2 years. The demo-offices contain all the approved ergonomic furniture and equipment. A need for a mobile demo-office was seen because of the size of the INEEL (840 square miles). In response to this, the ergonomics demo-trailer was born. The ergo-trailer has evolved within the past few months. A trailer was purchased and modified for the purpose of holding office furniture and ergonomic equipment in a stable manner. The distances are great at the INEEL. We could drive 60 miles from our home office to the farthest reaches of the site. The ergo-trailer makes it easier to demonstrate equipment to those located on the desert, at the various facilities. Local vendors of equipment were happy to loan items for the trailer.

The trailer currently visits facilities on a monthly basis. The concept is that a series of ergonomic assessments are planned for a facility, the ergonomist and trailer are at the facility at the same time. The employee and manager then look at equipment in the trailer immediately after the assessment so that the correct items are purchased. This enables the employee to try chairs and other equipment and get what is right for him/her. This avoids the "looks good" syndrome.

Ergonomic Self-Assessment

5/25/99

Item	Initial Measurements		Calc. Measurements	Final Measurements
A	48	*	45	46
B	27	*	18-24 in.	23
C	43		42	43
D	20	*	15	17
E	8	*	5	8
F	15		15	15
G	26	*	23	24
H	30	*	16	30
I	28	*	24	26

* initial measurement was over or under the calculated value by more than 5%

Name: Cheryl Wilhelmsen

Height: 65

Figure 3. Move card diagram.

3.1.5. Trained ergonomists

No ergonomic program can manage without trained individuals. Though a PhD ergonomist is not always needed, someone who has the level of knowledge needed to become a certified ergonomist is needed to ensure the program goes in the right direction. Currently, we have one certified individual and two more individuals are becoming certified in the near future (there are four more certified ergonomists/human factors professionals, but they do not work INEEL ergonomic issues).

4. IMPLEMENTING THE ERGONOMIC BUDDY SYSTEM

The steps for implementing the EBS were:
- Work out the logistics of the program.
- Seek volunteer employees to become ergo-buddies.
- Train the ergo-buddies.
- Develop the database.
- Perform awareness training for all employees.

Table 1. Implementation Steps for the Ergonomic Buddy System

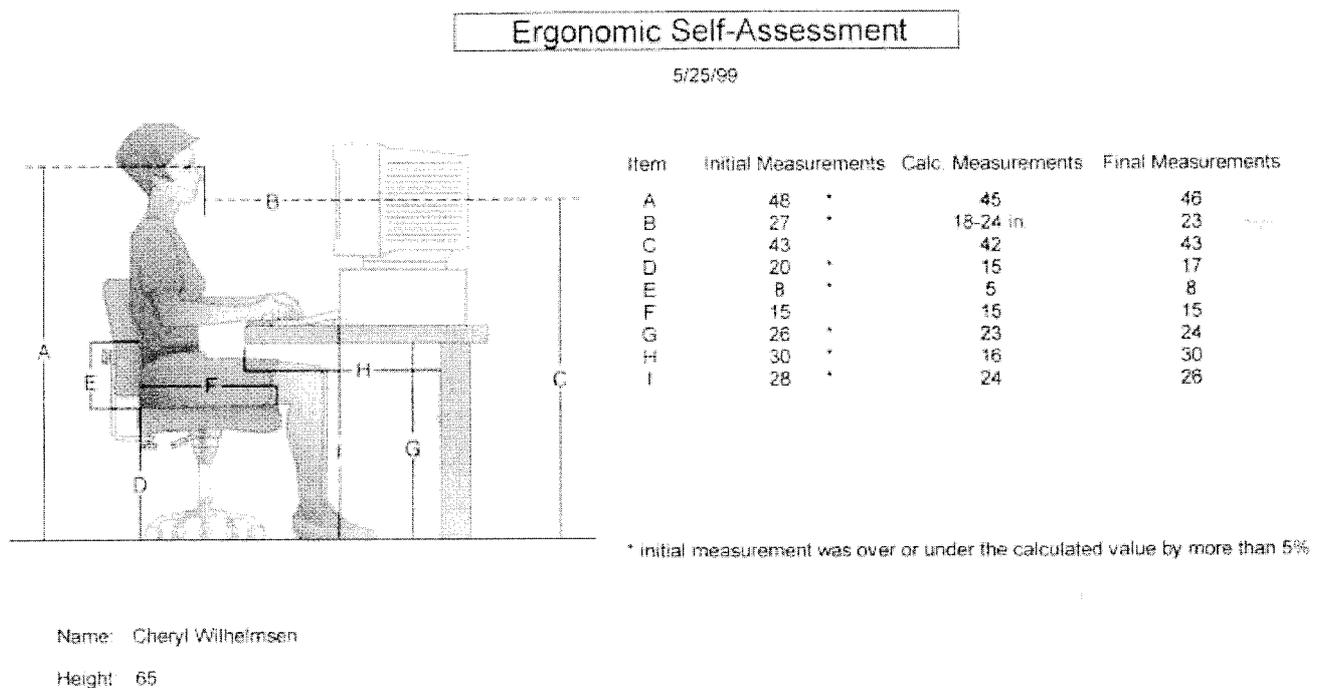| Element	Description
Workout the Logistics of the Program	The logistics that need to be worked out are: 1) How will and employee notify a buddy that he/she needs an assessment, 2) How will the buddy schedule the assessment, 3) Once performed, how will the data be collected in a central location, 4) Who will maintain the database,
Seek Employee Volunteers	Employees are asked to volunteer to participate in the program. The employees should have expressed interest in the process and are willing to do what it takes to implement it. These employees will be called Ergo-Buddies.
Training	Employees are given training on the basics of ergonomics and how to perform an ergonomic assessment. The training program used at the INEEL comprises two days and includes the priniciples of ergonomics, anthropometry, what are cumulative trauma disorders, workplace design and how to perform a practical ergonomic assessment.
Develop the Database	A database is developed that will contain: 1) The results from assessments, 2) Keep track of who has/hasn't been assessed, 3) Keep track of those who still need solutions, and 4) Keep track of workstation measurements and generate a "move-card." A "move-card" shows a diagram of an employee at a workstation with the measurements for the specific employee. Figure 1 shows the diagram for the move card.
Perform Awareness Training	Awareness training is performed for all employees. This training consists of: 1) What are CTDs, 2) Elements of workstation design, 3) How to report a problem, and 4) How to request an assessment

These steps are discussed in Table 1. The EBS was not implemented all at once in all areas of the INEEL. We began in the in-town facilities and at some of the site facilities. Currently, the program has been implemented in almost all facilities to one degree or another.

5. EFFECTIVENESS OF THE PROGRAM

We feel the EBS is successful. To date, there have been around ~3500 ergonomic assessments performed on site. All employees have received information on ergonomics, if they chose to look at it. Most of the awareness training was performed during 1996 and 1997. Ergonomic assessments are on going, however, a large number were performed during 1997 and 1998. During these assessments a large number of individuals who were experiencing symptoms of a CTD were identified. In fact, ~30% of those assessed were experiencing one or more symptoms. This caused a spike in the accident statistics, which management did not like (Figure 4). This told them the program was not working. However, all it was doing was uncovering employees who had pain and needed help. As with any new program the awareness factor tends to elevate the reporting of problems. In Reason's nomenclature this is called a reporting culture and is an integral part of a VPP (Reason 1990). This, in other words, is a good thing. Our goal is to have employees report their symptoms to medical, the ergo-buddies, or us as soon as they begin to feel them. The earlier one reports an injury, the more ergonomic and medical interventions can be implemented to prevent it from getting worse and to begin the healing process. Our data shows a spike of recordable injuries during the awareness training and initial assessment phase of the EBS. Then it shows a drastic downward trend during the early quarters of 1999. However, we begin to see an increase in first aid cases (Figure 5), which is what we want to see. We feel this is demonstrating that we are developing the reporting culture we want.

6. SUMMARY AND CONCLUSIONS

The EBS appears to have many benefits over the current methods of performing ergonomic assessments. Small battles have bee won in the implementation of an effective ergonomics program, but the war is still being fought. We were lucky to find a champion in management to help fight the battles and get management to understand the importance of providing funding now to invest in healthy employees which in turn elevates productivity. This is a key point to the whole program. Once that battle is won you are on your way to tuning the program to the needs of the employees under any working environment.

REFERENCES

GRANDJEAN, E., 1980, *Fitting the Task to the Man: An Ergonomic Approach* (London: Taylor & Francis).

OSTROM, L., 1994, *Creating the Ergonomically Sound Workplace* (San Francisco: Jossey-Bass).

OSTROM, L.T. and WILHELMSEN, C.A., 1991, *1991 INEL Safety Norm Survey*. Internal Technical Report, EGG-HFRU-9635, April (Idaho: EG&G).

PUTZ-ANDERSON, V., 1990, *Cumulative Trauma Disorders* (London: Taylor & Francis).

REASON, J., 1990, *Human Error* (Cambridge: Cambridge University Press).
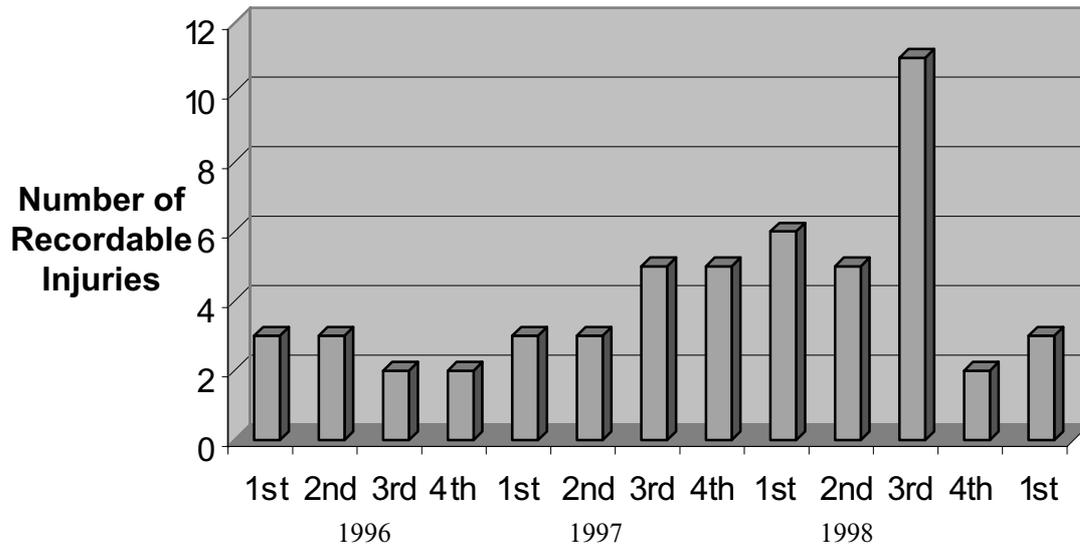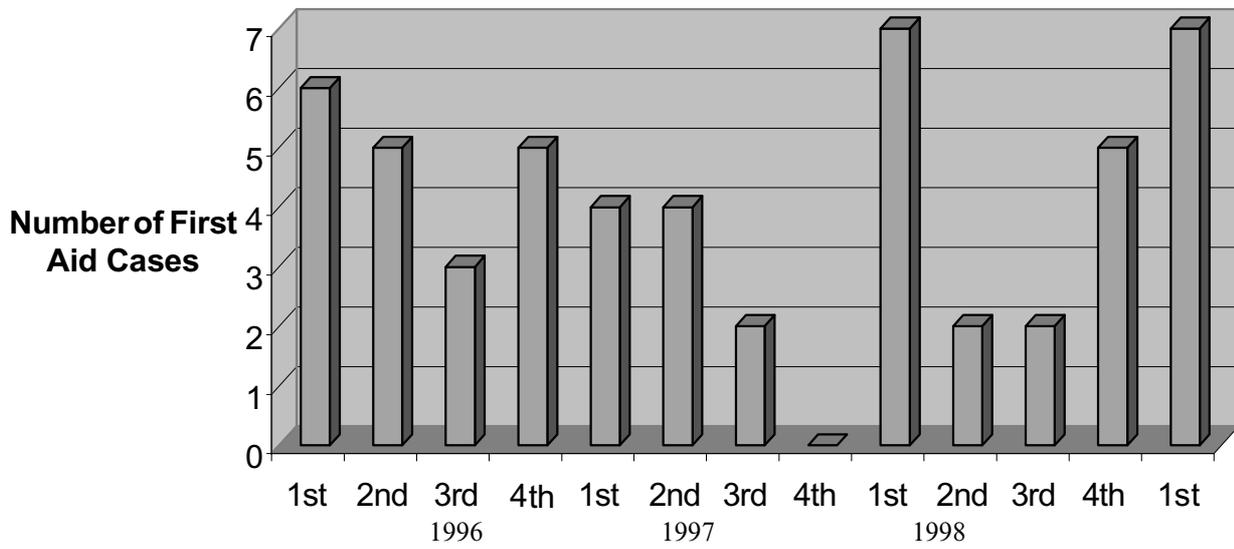
Figure 4. Recordable injuries.

Figure 5. First aid injuries.

Ergonomic Process in Small Industry

C. Stuart-Buttle

Stuart-Buttle Ergonomics, PO Box 18928, Philadelphia, PA 19119, USA

1. INTRODUCTION

An ergonomics process for a small business is different from that of a large corporation. Although the main components of an ergonomics process are similar, the characteristics and limitations related to the size of a business or industry need to be taken into consideration. Several of the issues raised may pertain to a small facility that is part of a large corporation. The main difference between a small company and a small plant of a larger entity is the latter usually has a safety and health process defined or mandated at the corporate level. The small facility may turn to the corporation for guidance and information and perhaps assistance and expertise in ergonomics. This chapter focuses on the issues faced by small businesses.

2. SMALL BUSINESS DEFINITION

In the USA small businesses account for 99.7% of the workplaces and employ 54% of the workforce. About 6.5% of all businesses (20% of the workforce) are in manufacturing, 98.5% of which have £ 500 employees (7.4% of the workforce) (Armstrong 1995). Typically the US Small Business Administration (SBA) defines a small business as having £ 500 employees.

Small businesses make an important contribution to the overall economy. In the USA since 1989 ~100% of new jobs have been created by businesses of less than or equal to five employees (Armstrong 1995). Also, in other industrialized nations, small-to-medium-sized companies are viewed as vitally important to the economic health.

3. SMALL INDUSTRY CHARACTERISTICS

Implementing and evolving an ergonomics process are difficult for any company. Small industries have unique characteristics that carry associated constraints. These constraints impact the decisions that are required to develop the best ergonomics program model to fit the culture and business needs of the company. Small industries often have the following characteristics (Stuart-Buttle 1999):

- Informal.
- Responsible for several positions.
- Responsive.
- Less specific knowledge.
- Greater management involvement.
- Less data-oriented approach.

4. REASONS FOR AN ERGONOMICS PROCESS

Injury and illness costs can be a primary drain on a company's productivity. However, ergonomics offers more to industry than reduced injuries and illnesses. It is a science that applies knowledge about people to the design of the workplace. When the workplace is designed so that people can perform at their best and machines work at their most effective, then productivity is optimized.

4.1. Competitiveness

The increase in the bottom line through production and quality is the greatest return from ergonomics. This may be partly realized by the reduction in injuries and illnesses and related costs or by the increase in work by those no longer in discomfort. However, making the job easier for the worker by improving the work layout or reducing re-handling also improves production.

Quality is part of the productivity equation. Quality is influenced by human performance, and that performance is directly impacted by design of the job and workplace. Insufficient time to perform the task, unsuitable lighting and environment and ineffective training are examples of aspects that can affect quality. Excessive reaching, awkward posture and fatigue increase the time it takes to perform a task, and quality standards may not be met.

4.2. Injury and Illness Costs

Days away from work are expensive for any company, but the burden may be greater on a small industry because there are fewer employees. The hidden or indirect costs, such as training a temporary replacement, that go hand in hand with a workplace injury should also be remembered in the cost–benefit equation.

4.3. Compliance

Compliance to government regulations should not be the sole reason for implementing an ergonomics program. If the workplace is designed well for most people to perform their jobs effectively on a long-term basis, then the workplace will be safe and healthy. If an ergonomics program is implemented with compliance as the only goal then redesigns are likely to fall short of the best solutions to the problems and limit improvements in efficiency. At present in the USA, apart from California, there is no law for implementing an ergonomics program. Management guidelines issued by the Occupational Safety and Health Administration (OSHA) have been adapted for small business in the OSHA Handbook for Small Businesses (OSHA 1992). There are many other sources of guidance for establishing an ergonomics program (NIOSH 1997, Stuart-Buttle 1999).

If a job is improved so that more of the population, including women, can perform it, then costs associated with finding the right people are reduced. Specific accommodation for the disabled also is easier because the workplace is more flexible when designed for a general population. Therefore, there are many benefits to designing well for a wide sector of the population, so compliance in itself need not be a burden to industry.

5. ELEMENTS OF AN ERGONOMICS PROCESS

An ergonomics process sustains the program using methods such as systematic evaluation and revision based on effectiveness. In the long-term, a program without a process is not successful. For success, it is important for a company to find an approach that is the most suitable and effective for the organization's needs and culture. Whatever the model, the ergonomics program goal remains the same: design the workplace so that it is healthy, safe and optimally productive.

5.1. CHOOSING THE APPROACH TO ERGONOMICS

The success of ergonomics in a company depends on how well

the process is structured and carried out. The following factors influence the development of a program and should be taken into consideration.

- Company size.
- Company culture.
- Resources.
- Type of industry.
- Types of ergonomics controls implemented.
- Compatibility with other programs and processes.

The above factors and particularly the characteristics of small industry, determine the appropriate extent of integration of an ergonomics program into existing processes. An ergonomics process is dynamic and regular evaluation and revision are necessary to maintain or improve effectiveness.

5.2. Process Elements

The primary elements of an ergonomics process are:

- Management commitment and employee involvement.
- Medical management.
- Education and training.
- Surveillance.
- Job analyses.
- Controls and improvements.

First, both management commitment and employee involvement are essential for any degree of success with an ergonomics program. If there are many work-related medical cases, quickly establish or enhance medical management. Next, acquire some knowledge about ergonomics and initiate a plan for training. Then having learnt more about what to monitor in the workplace, establish a surveillance system. Conduct job analyses as prioritized by the surveillance system and proactively assess new designs. Finally, implement and measure effectiveness of improvements for a safe, healthy and productive business.

5.2.1. Management commitment and employee involvement

5.2.1.1. Management commitment.

An ergonomics program is easily undermined by lack of management commitment. Management commitment is essential and must be communicated through example. Management should allow for employee involvement, setting clear responsibilities and provide the resources in both time and money for the process.

5.2.1.2. Employee involvement.

Employee involvement should be supported and facilitated by management. Employees are an excellent source of ideas for improvements and are the ones required to work with any implemented changes.

5.2.1.3. Written program.

A written program is often emphasized as essential for a successful program. Although a small business may react to this as "more paper work," there are many advantages to putting down in writing the basics, as it helps to:

- get the program started more efficiently;
- organize thoughts and the best plan of action;
- clearly communicate the process;
- make it easier to introduce the process to a new comer; and

- establish the goals and achievements by which the program can be assessed for success and improvement.

5.2.2. Medical management

A primary responsibility of a company is to respond to work-related medical problems reported by employees. Medical management per se is not in the realm of ergonomics, however, the quality of medical management has a direct bearing on the recorded injuries and illnesses which are often used as a measure in ergonomics. Company personnel have some control over the quality of the services rendered. The following suggestions may help to establish a good working relationship with the medical community.

- Establish close communication.
- Communicate the expertise expected of the medical group.
- Introduce the medical group to the plant's culture and processes.
- Consider the medical community as a training resource.

5.2.3. Education and training

First, some general knowledge in ergonomics is needed by someone in the company so as to plan an ergonomics program. In-depth training is recommended for a specified team, and general awareness training is recommended for managers, supervisors and line workers. The extent of investment in training may depend upon company financial resources and the turnover rate.

5.2.3.1. In-depth training.

In-depth training is recommended for the person or team primarily responsible and involved with ergonomics. The training objectives are to:

- understand the overall program objectives, goals and process;
- understand the injury and illness system for treatment, return to work, and job modifications;
- be able correctly to record and interpret medical records and OSHA logs for surveillance purposes;
- know how to conduct basic problem-solving job analysis for ergonomics issues;
- recognize risk factors for injury and illness in workplace design;
- be able to develop, implement and affirm effectiveness of solutions to basic problems;
- understand basic ergonomics principles to apply to solutions and new designs; and
- be familiar with outside resources and methods for finding resources.

5.2.3.2. Awareness training.

An awareness level of ergonomics training prepares employees to participate in the ergonomics process. The objectives are for the employees to:

- understand generally the ergonomics program;
- appreciate their role and responsibilities in the program;
- recognize the early indicators of physical problems;
- understand the company medical management system;
- understand basic risk factors for injuries and illnesses;
- know basic ergonomics principles; and
- understand their participation in job analyses.

New employees should be trained to maintain the knowledge base in the workforce. It is not unusual to give separate programs for supervisors, management and production workers as there may be differences in educational level and perspectives of each group.

5.2.3.3. Refresher training.
Refresher sessions maintain interest in ergonomics. As with the ergonomics program or process itself, the refresher sessions can be part of other processes in the company, such as continuous improvement or safety and health.

5.2.4. Surveillance
Identifying the areas with problems or the potential for improvement is the first step to improve the workplace. This entails looking at data that have already been collected by the company. Such data include information on injuries and accidents, production and quality measures and personnel records.

5.2.4.1. Data
- Medical information.
- Discomfort surveys.
- Absenteeism and turnover.
- Production and quality data.
- Accident investigation.
- Audits.
- Interview and employee reports.

5.2.4.2. Prioritizing.
Problem areas need to be prioritized to develop an effective action plan. One approach is to develop a spreadsheet with types of indicators or data by departments, areas or jobs. The amount and extent of the indicators can be used to qualitatively rank the problem areas in conference with other team members or employees. Consider factors such as the anticipated scope and difficulty of the project or whether the area or job is about to be changed for production reasons, because these factors also influence the priority.

5.2.5. Job analyses
Job analyses can be conducted from a reactive or proactive stance. A reactive analysis is evaluating a job known to have problems. A proactive analysis is looking at a new design, recent installation or redesign to anticipate problems and ensure that the workstation incorporates ergonomic principles. A proactive approach should be implemented as early as possible in an ergonomics program not only to prevent new problems arising, but also to avoid remaining in a reactive position which is more expensive in the long run, compared with designing well initially.

5.2.5.1. Responsibility for conducting analyses.
A team approach can be used in which the group collects the data, looks at the problem, brainstorms for the root causes and generates potential solutions. It is not uncommon, particularly in a small company, for a process to develop in which one person collects data, analyses it and generates the solutions with informal input from others. Caution should be used when adopting an isolationist approach, as over time, any joint responsibility for ergonomics might lessen and the program might be perceived as one person's responsibility.

Some companies choose not to invest in in-depth education of an employee but rather work with an external expert who gets to know the company process and culture. The approach chosen also depends on the amount and complexity of the issues to be addressed. There may be benefit in having more in-house knowledge if the production process changes frequently. The company needs sufficient understanding of ergonomics to know when assistance is needed, how to find it and where to get further information to address the issues.

5.2.5.2. Problem-solving.
The root causes of poor design are identified through careful analyses. A good problem list helps generate the best possible solutions. The traditionally informal structure of small industry does not preclude good problem-solving, nor does careful, detailed problem-solving always require extensive quantification.

5.2.5.3. Quantification.
Quantification assures that what is perceived as a problem is in fact a problem, and to what extent. Quantification also helps to assess improvements to determine the best alternative, and it provides a measure of effectiveness and cost–benefit. When the situation is more complex, selective quantification assures appropriate design decisions are made and helps prevent the generation of new problems.

5.2.5.4. Analysis.
Discussion of the many analysis methods and tools is outside the scope of this chapter. An overview of the basic steps in an applied analysis is provided as a guide.
- Workstation analysis.
- Clearly define the job function and the tasks, so that they are understood in context with the overall system.
- Collect pertinent job information such as performance rates and quality expectations.
- Interview employees.
- Describe the component actions of each task (possibly videotape them if the task is complex or fast).
- Identify the risk factors and job components that place excessive demand on the worker or that make the job awkward or inefficient.
- Assess the risk factors and job demands quantitatively and qualitatively.
- Determine the root causes of the risk factors, job demands and awkward methods.
- Develop a primary problem list with possible causes.
- Brainstorm for several short and long-term solutions to the problems, especially ones that are inexpensive.
- Assess the cost–benefits of alternative solutions.
- Develop an implementation plan including a trial stage if necessary.
- Reassess the solution after implementation to determine its effectiveness.
- Record the project.
- New design.

Incorporate ergonomics into the process of purchasing new equipment, designing new layouts or redesigning an existing area or workplace in the plant. A checklist may help to consider all the ergonomic aspects, but the interaction of the operators with

the equipment or layout should be especially anticipated and critiqued. Company engineers should have at least basic knowledge of ergonomics and work closely with other personnel to ensure that all safety and quality standards are met.

5.2.6. Controls and improvements

Effective problem-solving usually generates more than one solution, typically long-term and short-term ones of various expense. Engineering solutions are perceived as more permanent and less dependent upon human behavior, but it is common to need an administrative measure to accompany an engineering change.

Follow-through is important. A small company has the advantage of being project focused. Roles must be defined clearly for accountability, particularly if there is a team approach. In a small industry the cost–benefit of an improvement cannot be based on injury and illness alone because usually there are fewer and more scattered incidents throughout the facility. Therefore, gains in productivity and quality become important in the cost–benefit equation.

REFERENCES

ARMSTRONG, J., 1995, Personal communication (US Small Business Administration).

NATIONAL INSTITUTE FOR OCCUPATIONAL SAFETY AND HEALTH (NIOSH), 1997, *Elements of Ergonomics Programs: A Primer Based on Workplace Evaluations of Musculoskeletal Disorders*. DHHS (NIOSH) Publication no. 97-117 (Cincinnati: NIOSH).

OCCUPATIONAL SAFETY AND HEALTH ADMINISTRATION (OSHA), 1992, *OSHA Handbook for Small Businesses*. No. 2209 (Washington, DC: Department of Labor/OSHA).

STUART-BUTTLE, C., 1999, How to set-up ergonomic processes: a small industry perspective. In W. Karwowski and W.S. Marras (eds), *Handbook of Occupational Ergonomics* (Boca Raton: CRC Press).

Ergonomics and Production Philosophies

J. Eklund†‡ and C. Berggren†§

†Centre for Studies of Humans, Technology and Organization, ‡Division of Industrial Ergonomics, §Division of Industrial Management, Linköping University, Linköping, Sweden

1. INTRODUCTION

Industrial production may be organized according to a set of principles, or philosophies. Major determinants of the philosophy chosen are the character of the production and the context in which the company operates. New philosophies develop over time. A number of them have gained recognition and spread across the world, but eventually many become fads. The choice of production philosophy has a great impact on the technology, the organization and the personnel policy chosen, and it will have a profound influence on the performance of the company as well as on the working conditions of the employees. Few industrial companies are strictly organized according to a generic philosophy: normally, they constitute a mixture of several philosophies. The purpose of this chapter is to highlight and make a clearer distinction between different production philosophies. The characteristics of four generic production philosophies are summarized and presented in order to offer guidance in the choice and evaluation of production philosophy with regard to the resulting ergonomic consequences.

1.1. Definition of Concepts Used

A production philosophy refers to a generic way of designing technology and the organization in order to achieve specific objectives. A given production philosophy affects a number of components in the manufacturing system: production layout, machine design, work organization, skill requirements and means of managerial control. Not included in the concept as it is used below are the specific elements of human resource and personnel management, such as labor contracts, wage systems, training and career patterns, etc.

Although there is no universally agreed definition of ergonomics, most ergonomists agree that the modern concept of ergonomics includes aspects of humans at work on an individual as well as group and organizational levels. Interactions with technology, the environment and the organization are the focus. The main aims include elimination of accidents and the creation of healthy and comfortable working conditions, from a physical, psychological and social point of view. Recently, emphasis on learning and personal development has been added. In addition, ergonomics aims at improving the performance of the human–machine system, including productivity and quality.

Many attempts have been made to define the term "rewarding work." Motivational theories, socio-technical theory and ergonomics research have proposed a number of criteria for this, some of which are summarized as:

- Meaningful work content.
- Continuous learning at work.
- Decision making possibilities.
- Social interaction and support.
- Work perceived as useful.
- Work leading to a desirable future.
- Obtainable and realistic challenges.
- Recognition.
- Communication and feedback.
- Physical and mental task variety.
- Substantial cycle time.
- Authority and responsibility.
- Freedom to control one's own work.
- A job demanding skills.
- Security of employment.
- A work environment free from the risk of ill-health and accidents.

These criteria may be used as one basis for evaluating the character of a particular production philosophy.

1.2. A Brief Historical View

Both ergonomics and production philosophies are concepts of the 20th century. Before the industrial revolution, physical goods were typically manufactured in two ways:

- In small-scale craft shops, run by independent, highly skilled workers supplying individual customers or middlemen.
- In systems of large-scale production, where large numbers of unskilled workers turned out simple items for military customers or major construction projects. In these systems, high output was achieved through the multiplication of effort rather than the implementation of different manufacturing methods.

The industrial revolution of the late 18th century brought the mechanization of simple operations, and a detailed division of labor. During the following century, elaborated production philosophies emerged on the basis of these advances in the USA. A basic principle in the American System of Manufacture, pioneered by the arms manufacturers, was interchangeability of parts. Another important element was flow production, first introduced in the food industry. Third, Frederic Taylor and his associates fine-tuned systematic time-and-motion studies in the steel industry. A fourth important component was the development of sturdy, easily operated machine tools for high-volume production. Henry Ford brought all these elements together when designing the world's first manufacturing complex for large-scale production of automobiles in 1913. As a result of a long process of trial and error, a new production philosophy, Fordism, emerged. The principal aspects of Fordism were:

- Standardized products composed of standardized parts, eliminating the need for manual fitting.
- Special-purpose machines, operated by unskilled workers, designed according to the principle "one feed, one speed." New safety devices minimized breakdowns in high-volume production.
- Mechanically paced flow systems, feeding assembly workers with materials and minimizing handling time.
- Minute division of labor in the production process, rigid separation between unskilled operatives and skilled tool and die makers, quality inspectors and engineers.
- Fragmented jobs with short cycle times, £ 1 min.

Fordism realized enormous efficiency gains, as well as productive employment for millions of immigrant workers. From an ergonomics point of view, Fordism was a mixed blessing: on

the one hand, there were improvements in physical safety and pay, on the other hand, a proliferation of monotonous and machine-paced jobs resulting in work devoid of any intellectual challenge, with high strains, repetitive tasks, high levels of fatigue and occupational disorders (Hounshell 1984, Ford 1991).

2. SOCIO–TECHNICAL ALTERNATIVE

During the first half of the 20th century, Fordist production principles quickly diffused into every industry producing complex goods in high volumes, from consumer durables to furniture. In the 1960s, the long postwar boom and a tight labor market brought the deficiencies of Fordist jobs from a human point of view to the fore. New socio-technical principles for job design were developed in Britain and rapidly spread in Scandinavia in the 1970s, and later also to other parts of Europe. The socio-technical alternative started from an analysis of psychological job demands in order to arrive at a joint optimization of technical and human requirements. The socio-technical production philosophy consisted of four elements.

- Decoupling of workers from rigid machine pacing through increased use of buffers and new forms of production layout.
- Expansion of "horizontal" work content and variety by lengthening work cycles, job rotation, etc.
- Enriching of "vertical job content" by combining direct and indirect tasks and eliminating indirect and supervisory job roles.
- Delegation of decision-making power by substituting autonomous group work for tightly controlled individual jobs.

Socio-technically inspired production philosophies played an important role in major manufacturing sectors in Northern Europe in the 1970s and 1980s, especially in the automotive industry. New techniques and systems for materials handling were developed to support individualized and parallel work stations in high-volume production, and innovative ergonomics aids were introduced to solve problems of difficult posture and static loads. To assist workers in learning longer and more complex work cycles, new principles of occupational training were developed around principles of functional understanding and combination of mental maps and manual skills. Socio-technical systems significantly ameliorated the problems of Fordist working conditions, but sometimes led to difficulties in achieving sustained levels of higher production performance. Also, this approach did not focus how to integrate customer focus in operations and work organization (Pasmore 1988, Berggren 1992).

3. TOYODISM AND LEAN PRODUCTION

When the Japanese auto industry began to expand under heavy state protection after the war, the production system of the American Fordist car industry was taken as the reference point. Both the product and the labor market conditions in Japan were very different, however. Volumes were radically lower: in 1950, for instance, only 30 000 vehicles were produced in Japan, equal to 1.5 day's production in the USA. Furthermore, the Japanese companies had a wide product spread, indeed the auto firms had begun as light truck manufacturers. It was therefore necessary to adapt US methods to fit the efficient manufacture of lower volumes and, despite limited resources, to expand. The decisive defeat of independent Japanese labor unions in the early 1950s

gave the auto companies the opportunity to develop precisely these forms of low-cost rationalization. First, companies got a free hand in matters of shop floor organization and labor deployment. Second, the defeat resulted in a development of enterprise unions and corporate welfare systems, which linked workers' interests closely to the future of their company. The restricted product market and the new managerial opportunities formed the basis for the Toyota Revolution of the 1950s and 1960s. As at Ford, the Toyota system developed through a process of trial and error. Small-lot manufacturing replaced the Fordist philosophy of maximal batch sizes. One consequence was that setup times in press shops and welding lines had to be reduced dramatically. Buffers and inventory were squeezed out of the system, which made the process very sensitive to disruptions. Thus, manufacturing quality, "right the first time," acquired prime importance. Furthermore, small-batch manufacture led to the need for a highly flexible workforce. Toyota emphasized minute standardization and visual control of every operation, at the same time as workers were trained in several different tasks in order to make it easy to shift between various standardized jobs. Routine inspection was integrated with the production line as the responsibility of lead hands and operators. After work, they were mobilized in various small group activities such as quality control circles. In this respect, the problem-solving capacity of workers was effectively utilized. However, cycle-time rationalization, multi-machine tending and the integration of inspection work also contributed to considerable increases in work intensity. In summary, four specific features distinguished Toyodism from the Fordist production philosophy (Monden 1994):

- *Flexible mass production.* Instead of one standard product and dedicated lines, several products were integrated in a continuous flow, making use of quick die changes, Kanban systems for materials control and in-line inspection of quality.
- *Modification of narrowly defined job roles.* Rotation of workers through several standardized routine jobs and transfer between different sections to accommodate demand changes.
- *Continuous improvement* activities directed at reduction of all types of waste (defects, idle time, inventories, etc.).
- *Employee involvement* through quality movements and emphasis on customer satisfaction.

4. COMPARISON WITH SURVIVING FORMS OF MODERN CRAFT PRODUCTION

The three production philosophies reviewed are all basically concerned with high-volume production, Fordism especially so. However, in various industry niches producing sophisticated products for exclusive markets, or customized capital equipment for industrial customers, modern forms of craft production survive. This production philosophy is typically characterized by long work cycles at stationary objects (in contrast to the moving conveyor belt of Fordism and Toyodism), considerable worker skills and a high degree of autonomy on the shop-floor. In contrast to traditional, pre-industrial forms of craft work, the modern forms make use of high-precision tools, standardized parts and elaborate quality control and documentation systems. Furthermore, modern craft production borrows the Toyodist practice of involving workers in continuous improvement activities — as compared with the more static nature of traditional craft forms. Thus, the imprint of the individual on the end

product, so typical of older forms, is reduced in modern craft production.

Below, the four different production philosophies are compared, in terms of technological and organizational characteristics and ergonomics consequences in a broad sense.

When comparing the criteria of rewarding work with the production philosophies discussed here, it can be seen that there are agreements and disagreements within all philosophies. Since hardly any company is strictly organized according to one generic production philosophy, and most companies develop a mixture between several philosophies, there is room for changes and improvements. The above classification may be used as a tool when evaluating existing and planned production systems. Also, it allows a more conscious choice and enables consideration of ergonomics aspects as well as economics and production aspects. In this way, ergonomics may contribute to improving the production systems of the futu

REFERENCES

BERGGREN, C., 1992, *Alternatives to Lean Production* (Ithaca: Cornell University Press).

FORD, H., 1991, *Ford on Management: Harnessing the American Spirit* (Oxford: Blackwell; repr. from *My Life and Work*, 1922, and *My Philosophy of Industry*, 1929).

HOUNSHELL, D., 1984, *From the American System to Mass Production 1800–1932* (Baltimore: Johns Hopkins University Press).

MONDEN, Y., 1994, *Toyota Production System: An Integrated Approach to Just-in-Time* (2nd ed.) (London: Chapman & Hall).

PASMORE, W., 1988, *Designing Effective Organizations: The Socio-technical Systems Perspective* (New York: Wiley).

Table 1. A comparison of various characteristics for four production philosophies, referring to production work.

	Fordism	Sociotechnical	Toyodism	Modern Craft
Volume	High	Medium-high	High	Low
Production control	Volume Cost	Volume, Quality Inventory, Cost	Volume, Quality Customization Continuous improvement	Quality Quality
Type of production equipment	Dedicated & rigid Mechanical flow systems	Modified human interface Operator control of production pace Parallel flows and buffers	Flexible machines Tightly controlled flow system, Zero buffers	Versatile machines Stationary production
Technical autonomy	production	Very low- Machine paced and machinery	Medium Decoupling of workers production	Very low High Machine paced
Administrative control	Tight supervision	Several responsibilities delegated to groups	A few responsibilities delegated by external goals	Self regulation
Cycle time	Short, one minute or less	Enlarge, one hour or more	Short, one minute or less	Long, days or weeks
Job content	One repetitve task	Several manufacturing and administrative tasks	Multi-tasking	Highly varied
Individual/group	Individual	(Semi-) autonomous group	Production team	Individual craftsmen or team
Skill requirements	Low	Medium	Low to medium	High and manifold
Feedback	Little	Medium	Medium	High
Responsibility/authority	Low	Medium	Medium	High
Physical work load	High repetitive strains Fatigue	Medium repetitive strains	Repetitive strain Time pressure Fatigue	High workload Accidents Long hours
Participative problem solving	Very low	Medium-high	Medium-high	High
Labor relations	Bad	Aims at improving	Aims at improving	Varied

Ergonomics/Human Factors Audits

C. G. Drury

Department of Industrial Engineering, University at Buffalo, 342 Bell Hall, Buffalo, NY 14260, USA

1. WHAT IS A HUMAN FACTORS/ERGONOMICS AUDIT?

Auditing is "testing and checking the records of an enterprise to be certain that acceptable policies and practices have been consistently followed" (Carson and Carlson 1977). The "acceptable policies and practices" can be interpreted as to ergonomics/human factors just as well as the more usual financial interpretation. Quantitative measurement is implied by "testing and checking," while emphasis on "consistently followed" implies that the measurement is relative to some standard of perfection or close to it. In ergonomics much of the auditing takes the form of measuring the characteristics of a sample of jobs or workplaces to determine what fraction meet ergonomic criteria. In the end, an audit must make an informed statement about how closely a company's jobs or workplaces adhere to good ergonomic practice.

To meet this goal of auditing, many questions must be answered about when to use audits, what audit tools to use and how to analyze and interpret the audit data to make this informed statement.

2. WHEN TO USE AN AUDIT

Although much auditing activity focuses on measuring workplaces, an audit should not be used when the issue is "how well does an individual workplace meet ergonomic criteria." For this case ergonomists have many measurement tools and methodologies which can examine in detail human/workplace fit. Rather, an audit is designed to answer questions that go beyond the individual workplaces, and draw conclusions about a *set* of workplaces.

The need for ergonomics audit programs has arisen as ergonomics has moved from a workplace level activity (e.g. design a better job or workplace for *this* task) to a systems-level activity (e.g. design an ergonomics program to improve safety/quality/productivity/operator well-being throughout the organization). Thus, the audit must measure many units (jobs or workplaces) to draw valid conclusions at the systems level. In a typical application an ergonomics audit is used to measure the on-the-ground achievements of an on-going ergonomics initiative, for the purpose of evaluating that initiative at a specified point in time.

An audit typically supplements rather than replaces outcome measures. Many outcome measures for ergonomics/human factors programs are long-term, for example industrial injuries or airline accidents. As with quality programs (see Quality entries), the aim of audits is to measure the precursors of major problems rather than the problems themselves. We can measure outcomes, such as reduced errors in aviation maintenance following a human factors training program (Taylor and Christensen 1998), but

should not have to wait for such rare events before evaluating our program. In another example, Drury *et al.* (1998) audited a multi-plant ergonomics program early in program implementation to make recommendations for program changes. Only later were data available to show that the injury rates that the program was designed to address were reduced by ~50%.

3. HOW TO SELECT AUDIT TOOLS

The tools used in human factors/ergonomics audits vary widely depending upon the audit requirements. Some are one-page checklists, others are checklists the length of a booklet, while yet others are structured interviews. An audit tool can be classified by breadth, depth and application time. Breadth indicates the number of aspects of ergonomics covered, ranging from single topics such as thermal audits to wide-ranging job audits. Depth is a function of how fully each aspect is audited. A topic may be covered by a single question (e.g. "Is lighting adequate for the task?") or at the other extreme by a detailed set of measurements (e.g. a complete analysis of task visual aspects from color temperatures of light sources to calculation of the glare index). Clearly, application time is a function of both breadth and depth, growing longer as either or both increase. In a real sense, you get what you are prepared to pay for. Some applications need a rapid evaluation of many ergonomics topics, while others need more depth and focus.

An audit tool can be applied to a workplace, to a job or to a specific personnel within an organization. Sampling workplaces is appropriate where people spend most of their time at one workplace, for example a data entry clerk at a computer workplace or a garment worker at a sewing machine. Measuring aspects of the workplace automatically gathers data to show how well the job fits the incumbent. Where one job involves many workplaces, such as a maintenance task or an order picker in a distribution center, we must sample the job, usually at a number of times throughout the shift, to find similar information. Finally, looking only at worker/workplace fit will not help us much with system issues, such as how an ergonomics program is utilized by management. Here, the appropriate tool would be structured interviews with a sample of management personnel.

Examples of these different audit tools, with various levels of breadth and depth, are available in the literature. A comprehensive listing, with many examples, is given in Drury (1997). Two examples of the structure of such audits are given. Table 1 shows the structure and typical points from Ergonomic Checkpoints, a checklist developed for the IEA by Kogi (1994). The whole list covers 128 points (called "checkpoints"), in areas from Materials Handling to Work Organization, with each checkpoint having a structure which allows the analyst to see possible improvements. While Ergonomic Checkpoints audits workplaces, ERNAP (Table 2) uses the job as the unit. ERNAP was developed for aviation inspection and maintenance functions where technicians perform many different tasks (Koli *et al.* 1998) and is in the form of a computer program. (Note: the sponsors of ERNAP, the Federal Aviation Authority, have made this tool available at http://www.hfskyway.com). ERNAP was formed by combining two earlier checklists (ERGO and EEAM), both of which had their reliability measured, and were validated against the findings of ergonomics/human actors professionals viewing videotapes of audited workplaces.

4. CONDUCTING THE AUDIT

At this point we have established the purpose of the audit, what units we are measuring (workplace, job, personnel) and what tool to use. In carrying out the audit we must now determine how to sample the available units, i.e. which workplace? which job? which manager? Choice will depend upon the objective. Do we need to make pronouncements about the whole organization, a specific factory, all inspection jobs, the new production line, or what? The answer determines our sampling frame, i.e. the set of all units we *could* choose.

From this sampling frame we must then sample the individual units in a logical manner. This could be random sampling, where all units in the sampling frame are equally likely to be chosen. It could also be stratified sampling where we sample randomly from particular subsets (production, shipping, office, etc.) and later combine the results into a single audit sample. Or we could use cluster sampling where we choose clusters (new production line, old production line, etc.) and sample randomly within clusters. All are logical sampling plans, with particular uses, and are usually covered in detail under survey research methods.

An audit as we have defined it measures on-going activities, but we can supplement it with archival data to obtain a more longitudinal view of the organization's ergonomic effectiveness. Thus, audits of workplaces can be combined with analysis of records such as scrap, injuries, errors or employee turnover. Often the two sources can be mutually supportive, for example in an aviation maintenance environment both audits and error analysis can lead the auditor to the same problems with job instruction design.

Finally, in conducting an audit it should be standard operating procedure to use good ergonomic practice. Ensure that job incumbents are informed of the purpose of the audit and the anticipated uses of the data. Ensure too that where job incumbents must give responses, the auditor does not bias or lead the respondent.

5. DATA ANALYSIS AND PRESENTATION

Ergonomics is design for the user, so that a well-designed audit outcome should be designed to fit the user's needs. Of course, there may be multiple users with varying needs. The Chief Operating Officer may only need overall evaluation of the results (is the ergonomics program working adequately?) while the plant manager may need much more specific information (is the ergonomics program concentrating on reduction of errors in shipping? Is the new warehouse ergonomically better than the old one? Do we need to install task lighting throughout?). The most specific questions are the easiest to answer. We can tell the plant manager about the visual characteristics of the workplaces with statistical summaries of light levels, luminance values, glare indices and so on directly from the audit data. We can even break these distributions down by production lines or by job.

At the higher level there can be potential problems. Higher levels of management need to make programmatic decisions involving resource allocation between competing initiatives. We can answer questions such as "Is the ergonomics program working well in the men's apparel divisions?" or "should next year's budget cuts come from ergonomics or quality programs?," but the answers require that we use our ergonomics expertise in a managerial rather than strictly technical context. The issues

involve combining data from different aspects of human factors into higher-level indices of performance. Instead of showing the distribution of illuminance levels, or WBGT indices, we need to establish standards or criteria so that we can combine data in a meaningful way. The only way we can combine, for example, visual and thermal data is to classify each data point as either meeting or not meeting ergonomic criteria. We can then count the fractions of workplaces where specific criteria are met (e.g. 47% meet visual standards), and overall fractions of workplaces meeting all ergonomic criteria (e.g. overall, 17% of workplaces were free from ergonomic problems). To do this we have to assume that all failures to meet ergonomic criteria are equally bad, or provide some weighting scheme. Any choice of weights would be difficult to justify, so the assumption of equality is usually the most practical one.

A useful technique is to provide an overall index of ergonomic effectiveness with some qualifier about its constituents. For example, "38% of the workplaces in this division were ergonomically acceptable. The most frequent cause of non-acceptable workplace was poor posture caused by ill-designed conveyors."

This gives the decision-maker an assessment of the gravity of the situation, and an idea of what might be required (e.g. redesigned conveyors) to make improvements.

Percentage of ergonomically acceptable workplaces (or jobs) is a useful measure of overall ergonomic effectiveness, but it does presuppose ergonomics standards. We can collect data on jobs and workplaces, but eventually we must compare what is measured with what is appropriate for human use. Audit tools must do this either in the data collection or the data analysis: At times, the standard is built into the tool, for example by asking whether illuminance is greater than 500 lux for moderately difficult visual tasks. But measurements can be difficult to make reliably, especially when a standard of acceptability is implied. At many workplaces, careful placement of the light meter can be used to miss or achieve the 500 lux standard. It is usually preferable to write specific instructions for data collection, e.g. "record illuminance in horizontal plane at four points within the central 10∞ field of view." Data can later be compared with standards during analysis, particularly computer analysis.

Having decided what data to present, the report (or other presentation) should use human factors good practice to format graphs and figures for user comprehension. Suitable guidelines are given in Gillan *et al.* (1998). The most effective case is made when the report is written to address the user's issues rather than the writer's issues. An example is the ERNAP output (Koli *et al.* 1998) which is in the form of a memo to a Quality Assurance Manager listing overall findings for each aspect measured, and specific requirements for those places where ergonomic criteria were not met.

6. BEYOND THE AUDIT

Use of an audit tool and sampling plan can address the specific objectives used audit design. But the same data are also available for later use. Data at one location and one time can be extended across locations to make plant comparisons possible. They can also be extended over time to provide a time history or trajectory of ergonomic effectiveness. Specific measures can be analyzed

statistically to determine when plants or times differ significantly. By counting the reasons for workplaces not meeting human factors criteria, we can assess whether the same problems occur at all plants or whether different plants have different unmet ergonomic needs. Thus, Drury *et al.* (1998) found that most measures were consistent across plants and jobs, although postural and lighting problems were significantly job-specific.

Finally, as noted earlier, audit information can be combined with other data, often archival, to obtain a more complete understanding of the human factors issues. For example, Simpson (1994) and Fox (1992) used an error-based audit to improve many aspects of safety in coalmines. The error analysis prompted measurements (audits) of specific aspects of mine activity, such as locomotive design.

REFERENCES

CARSON, A.B. and CARLSON, A.E., 1977, *Secretarial Accounting*, 10th edn (Cincinnati: South Western).

DRURY, C.G., 1997, Human factors audits. In G. Salvendy (ed.), *Handbook of Human Factors and Ergonomics* (New York: Wiley), 1503–1616.

DRURY, C.G., BRODERICK, R.L., WEIDMAN, C.H. and REYNOLDS

MOZRALL, J.L., 1998, A corporate-wide ergonomics program: implementation and evaluation. *Ergonomics* (in press).

FOX, J.G., 1992, The ergonomics audit as an everyday factor in safe and efficient working. *Progress in Coal, Steel and Related Social Research*, 10–14.

GILLAN, D.J., WICKENS, C.D., HOLLANDS, J.G. and CARSWELL, C.M., 1998, Guidelines for presenting quantitative data in HFES publications. *Human Factors Bulletin*, 28–41.

KOGI, K., 1994, Introduction to WISE (work improvement in small enterprises) methodology and workplace improvements achieved by the methodology in Asis. In *Proceedings of the 12th Triennial Congress of the International Ergonomics Association*, vol. 5 (Toronto: Human Factors Association of Canada), 141–3.

KOLI, S., CHERVAK, S. and DRURY, C. G., 1998, Human factors audit programs for nonrepetitive tasks. *Human Factors and Ergonomics in Manufacturing*, 8, 215–31.

SIMPSON, G.C., 1994, Ergonomic aspects in improvement of safe and efficient work in shafts. In *Ergonomics Action in Mining*. Eur 14831 (Luxembourg: Commission of the European Communities), 245–56.

TAYLOR, J.C. and CHRISTENSEN, T.D., 1998, *Airline Maintenance Resource Management* (Warrendale: Society of Automotive Engineers).

Ergonomics in a Design Engineering Environment

D. Imbeau†, M. Bellemare†, J. Courville‡, S. Bergeron§ and L. Desjardins§

†Institut de recherche en santé et sécurité du travail du Québec, Canada
‡Commission de la santé et de la sécurité au travail du Québec, Canada
§EFOR Groupe Conseil, Canada

The subject of ergonomists in a design engineering environment is discussed. This chapter presents a short overview of the contributions and difficulties faced by ergonomists in an engineering design environment. The following chapter illustrates and further describes some of the concepts covered here. Both chapters are concerned primarily with the design of manufacturing or industrial production systems (MIPS).

1. APPLICATIONS OF ERGONOMICS IN ENGINEERING DESIGN

1.1. Value of Ergonomics

MIPS are becoming more complex as technology evolves, but the presence of human operators and workers remains a constant in every system. Humans implement, operate, reprogram and maintain the systems. Hence, MIPS should be designed to accommodate human characteristics. The system designer is expected to optimize system performance by giving proper consideration to human and technical components. The discipline of ergonomics encompasses the knowledge, tools and methods used to fulfill the human aspect of the system optimization task.

The reported benefits of considering ergonomics in the system design process include increased product quality, system reliability and productivity. Such increases can be achieved through reductions in human error, production time, time lost, product reworks and rejects, workforce turnover and absenteeism, the duration and complexity of worker training, start-up time and cost, and other factors. It should be remembered, however, that in most MIPS it is difficult to translate many of these advantages into financial terms, mainly because of a lack of accurate and reliable data. Incomplete or absent a priori financial justification weakens the arguments that might convince decision makers and key players in the system design process to accept the extra expense generated by the addition of an ergonomist to the design team. On the other hand, health, safety and human performance concerns are taken into account more systematically in industries where the costs or consequences of design errors are high or obvious (e.g. the atomic industry, continuous processes, aviation). In such industries, failure to consider the human aspects during system engineering design can have a significant impact on future insurance claims, for instance. Thus, the benefits of ergonomics are real, but are often difficult to express in financial terms.

1.2. Use of Ergonomics Knowledge and Tools In Engineering Design

For a number of reasons, ergonomics knowledge and tools are not used as extensively by system designers as might be expected (Campbell 1996), a community largely represented by engineers. In fact, Daniellou (1988) reported the existence of a significant gap between the availability of ergonomics knowledge and the actual application of this knowledge in engineering design.

First, most engineering training programs do not include formal and comprehensive training in ergonomics and safety aspects. Comparatively few engineers are aware of the existence of this field of knowledge, and those who are aware of it often do not know it well. To them, ergonomics boils down to table heights and selecting suitable chairs in the office environment, or to eye-catching automobile instrument panels. At the other end of the scale, most ergonomists have non-engineering or non-technical background training. As a result, they generally have a poor understanding of the technical requirements associated with engineering design.

Second, ergonomics analysis methods and tools are reported to produce output results that are difficult to incorporate in traditional engineering design methods. They are often too time-consuming for the fast pace and short deadlines that are typical of MIPS design projects. Also, designers with little training in ergonomics often find it difficult to implement the recommendations made by ergonomists and to apply the abundant information on human characteristics made available by the ergonomics community. Many guidelines and recommendations are said to be too vague to be usable, or in need of conversion to system-specific design specifications. This would be the case in particular for guidelines concerning the specifications of products or equipment in order to match or compensate for human cognitive or task performance abilities and limitations.

Third, ergonomics tools and methods are often perceived as more suited to the evaluation of a prototype or candidate system rather than to the initial design process. Moreover, there is no systematic approach to help designers incorporate ergonomics and safety considerations at the various steps of the engineering design process. Many methods and tools exist, including methods for risk analysis, task analysis, severity of consequences analysis and reliability analysis. However, they have never been integrated to form a unified design engineering framework. Designers are left with a plethora of methods and tools but there is nothing to help them select and apply the most effective and efficient methods for their particular design needs.

Despite these problems, experience indicates that considerable benefits can be achieved by applying ergonomics knowledge in MIPS design. Ergonomists who have been involved in MIPS design can confirm that transferring ergonomics and safety knowledge to designers through training and participatory approaches is an effective way of communicating ergonomics knowledge and fostering its use in the design process. Given the current state of tools, methods and knowledge in ergonomics and engineering, the successful design of a MIPS requires the contribution of an ergonomist. Meanwhile, initiatives aimed at increasing the utility and actual use of ergonomics information and methods by designers must continue.

2. DESIGN PROCESS AND ITS DIFFICULTIES FOR THE ERGONOMIST

2.1. From Problem to Solution

In industrial production systems, engineering design is generally concerned either with re-design (e.g. integration of new technology) or with improvement (e.g. correction) of existing and operational systems. The designer's task is therefore to define the problem space and develop solution alternatives. Definition of the problem space entails documenting the existing problem dimensions and the new opportunities presented by the design project. Problem space definition will usually begin with the problem definition provided by either the design project sponsor or client (or both), and the end user of the system (e.g. operators). Problem space definition is generally achieved gradually by developing solutions, and thus spans several stages of the design process. The problem is seldom defined in its entirety before the search for solutions begins, since the data required for complete problem definition is often lacking. Each new element of solution developed solves part of the problem, and at the same time introduces new aspects of the problem.

If different designers begin with the same problem statement but use different sets of criteria and work at different levels of detail, they will arrive at different feasible solutions. Indeed, the development of solutions will depend largely on the criteria used by the designers to define and tackle the problem, as well as on the importance given to these criteria. If the designers have a predominantly technical background, then the design process will emphasize technical criteria, and this will have an impact on both the data collected and the solutions developed. Level of design detail is also important. In MIPS design, the effort will necessarily focus on the highest level, represented by production means (e.g. building, machines, robots, conveyors), and to a lesser extent on the intermediate level, represented by workstation design (e.g. bench, hand tools, jigs). In many situations, the design effort will focus little if at all on the lower level, represented by work activities (i.e. how end users will actually interact with the workstations and production means to achieve expected production goals). Studies report that different designers within the same design team generally have partial and different views about the work activities being performed in an existing system and in a future system. This is even more the case when collective work is involved. For a design to be successful, the different views must be reconciled. Where work activities are concerned, this usually proves to be both time consuming and resource consuming.

The engineering design process is often described as opportunistic and intuitive. Iterations to earlier stages (i.e. redefinition of portions of the problem space) may be frequent, and changing goals are commonplace. In fact, the MIPS engineering design process is seldom purely sequential, well organized, preplanned, and totally controlled in nature. It is subject to many constraints, since it involves different people with different backgrounds and objectives who interact to achieve a common goal.

2.2. Contribution of an Ergonomist

The role of the ergonomist in an engineering design environment is usually threefold. First, s/he is an expert in human characteristics. The ergonomist can tap into a considerable amount of literature pertaining to human behavior and characteristics in various contexts, and then extract pertinent information from which to formulate recommendations for the systems designers to use in their work. Second, the ergonomist is an expert in the activities of end users (e.g. operators and maintenance workers). Access to methodological tools designed specifically to analyze work activities and interact with end users, often coupled with extensive experience in particular work settings, means that the ergonomist becomes very knowledgeable about the work performed in a variety of production system implementations. S/he is therefore a valuable resource when different design solutions must be considered. In these two roles, the ergonomist is somewhat passive, providing input only when asked.

The third role of the economist is as an active member of the design team, whose participation can influence both the design process and the design itself. Part of this role will certainly consist in ensuring that other designers listen to the opinions of end users in the design process.

2.3. Difficulties for the Ergonomist

One of the main difficulties faced by ergonomists is that their contribution is generally solicited too late in the design process, i.e. when technological choices have already been made (Oland 1991). Since these choices probably did not take the human component into consideration, new constraints are added to the ergonomist's own problem space, making the task even more difficult and less effective in the end. This situation may be due to the fact that designers are often skeptical about the usefulness of the ergonomist's contribution to the design process, or do not understand the importance of including the ergonomist at a very early stage of the process. The fact that ergonomists and designers have different formal background training may exacerbate an already difficult situation, often plagued with communications problems, preventing the ergonomics specialist from making an active and effective contribution to the design process. Case studies report that where the contribution of the ergonomist is clearly understood and accepted through adequate communication, the design process usually proceeds well.

In manufacturing systems, an error in the design of a system may result in the system being inoperable because of technical problems. The result of such an error sends immediate feedback to the designer, so that s/he can correct it. However, if the design error relates to an ergonomics aspect, then the system will probably still be operable because of the unique and formidable ability of humans to compensate. However, this type of error usually leads to dysfunctional states of the system, degrading the working conditions of the employees. Degraded working conditions have been linked to productivity losses, including occupational accidents and diseases.

An illustration of this situation, and of the sometimes conflicting objectives of designers, is the use of the Pareto distribution, which virtually all engineers are taught. The Pareto concept is used by engineers to help focus on important aspects of a problem. The bulk of a problem can be solved with a comparatively small investment of effort (e.g. a 20% effort can solve 80% of a problem) (Konz 1990). Working on the most important problems or aspects of a problem leads to a more efficient use of engineering resources. Engineers apply this basic

concept to a variety of problems. Thus, an engineer designer will usually be content with an operational design that is stable (or works well or is within specified parameters). S/he has little incentive to spend valuable resources in order to achieve a near perfect system, since at a certain level the effort required to further improve performance greatly outweighs the gains that can be expected. Besides, the engineer knows that the end user will compensate in any case for the few remaining system inefficiencies. In contrast, the ergonomist is interested in the sources of variability in the system, because they cause dysfunctions that render the end users' work activities and lives more difficult. It is precisely the variability, which is not considered by the designer because it occurs in a comparatively small portion of cases, that forces human interventions. The ergonomist is aware that these dysfunctions will probably generate costs in the long-term. Unfortunately, such costs are very difficult to estimate at the design stage. It follows that, in the engineer's mind, the ergonomist is spending, not to say wasting, time and valuable resources on aspects or conditions of the system that are not worth the effort.

The ergonomist is often included in a design team in the hope that his or her contribution will help reduce occupational accidents and diseases. In the long-term, this type of contribution is hard to sell, since its main pitfall is that resources will be invested for "non-events", i.e. events that are expected not to materialize (e.g. occupational accidents and diseases). This is counter-intuitive, in the sense that humans usually expect an investment in resources to produce some form of result, not the absence of something. Hence, the contribution of the ergonomist will not be observed or measured unless a similar and comparable existing system can be used to benchmark the new system. In theory, both the old and the new systems should be operated together long enough to allow for accidents and diseases to materialize in the new system - a situation that is rarely possible in reality. Hence, the ergonomist's contribution should not be tied solely to health and safety questions, but also to productivity, quality and reliability issues.

Design problems are often complex and are divided into subproblems, each of which is assigned to a different design team. Since it is common practice to have one ergonomist for the whole project, this person has to cope with the burden of interfacing and dealing with several teams at once, because ergonomics issues will sometimes cut across several subproblems. This may help make the ergonomist's task more difficult.

3. CONDITIONS TO FACILITATE THE WORK OF AN ERGONOMIST

The literature suggests that the following five conditions should be present if an ergonomist is to be successful in an engineering design environment. First, the ergonomist must convince the other technically oriented designers of the importance of the field of ergonomics. This ongoing effort is important, since the ergonomist's participation in the design project is bound to generate new problems for the other members of the team - problems that would otherwise have gone unnoticed. Top management in particular must be convinced of the value of ergonomics from the outset. Second, the ergonomist must play an active role in the very early stages of the design process, so that the human side of the system to be designed is addressed adequately. Third, the ergonomist must constantly describe the human aspects to the other members of the design team, and link them to the other technical components of the system. This is necessary, since the other members of the design team, owing to their predominantly technical background, will often unconsciously evacuate the human aspects to focus on technology. Fourth, since the choices made throughout the design process predetermine the work activities of the future system's end users, the ergonomist must constantly help the other designers to anticipate those activities, so that they can orient their design choices accordingly. Finally, ergonomists must adapt. On the one hand, they must adapt to the constant changes imposed by new technologies, which more than ever before are placing emphasis on the cognitive rather than the physical aspects of work activities. On the other hand, they must adapt to the constant changes that occur in the design process environment (e.g. changes of objectives, changes of design team members).

REFERENCES

CAMPBELL, J.L., 1996, The development of human factors design guidelines. *International Journal of Industrial Ergonomics*, 19, 363–71.

DANIELLOU, F., 1988, Ergonomie et démarche de conception dans les industries de processus continu: quelques étapes clés. *Travail humain*, 51, 185–94.

KONZ, S., 1990, *Work Design: Industrial Ergonomics* (3rd ed.) (Scottsdale: Horizons).

OLAND, J.S., 1991, Prevention of ergonomic problems in the stages of planning and projecting. In *Proceedings of the 11th Congress of the International Ergonomics Association*, 2, 1353–5.

Exposure Assessment of Upper Limb Repetitive Movements: Work Reintegration Criteria

D. Colombini, E. Occhipinti, R. Bergamasco*, G. De Vito, and G. Molteni

Research Unit on Ergonomics of Posture and Movement, Via Riva Villasanta, 11–20145 Milan, Italy and *Embraco-Aspera SRL, Via Buttigliera, 6 Riva presso Chieri, Turin, Italy

1. INTRODUCTION

The return to the workplace of individuals affected by work-related musculoskeletal disorders (WMSDs) of the upper limbs, represents a critical problem in work settings which feature a multitude of tasks liable to biomechanically strain the upper limbs. Workers suffering from disorders are largely considered to be temporarily or permanently unfit for the jobs which are likely to have caused or aggravated the disorder. Some relevant criteria, procedures, and solutions for returning workers with limited fitness to tasks and jobs with lower exposure will be described. For successful reintegration we suggest an intensely participatory approach involving technical staff, medical staff, and affected workers. This should allow the affected workers to remain productive, at least to some extent, while safeguarding their health.

2. PROCEDURES AND CRITERIA

2.1. Involvement of Technical Staff in Redesigning Jobs

The technical staff should be trained to submit the relevant jobs to the specific risk assessment procedure, so as to more effectively redesign tasks for both "healthy" and diseased workers. The training should cover aspects such as definition of repetition injuries of the upper limbs and their relationship to work stresses; introduction to the pathogenesis of the principal upper limb disorders related to repetitive movements; methods and criteria for rating occupational risk; techniques for measuring action frequency by analyzing slow-motion films, methods of applying an upper limb strain index using the Borg scale, criteria for detecting awkward postures and evaluating recovery times, identification and measurement of additional risk factors.

The technical staff should be asked to inspect their departments and single out jobs and duties that would be immediately suitable for WMSD-affected workers, or could be adapted with minimal changes. When jobs are not entirely suitable for affected workers, job redesign recommendations should be drafted. Any jobs or tasks that would be too expensive or too slow to modify should be rejected. An ergonomic supervisor should carry out a final check on the task analysis and evaluation performed by the technical staff and a final check on the jobs they have selected.

2.2. Criteria for Locating Suitable Tasks and Jobs for Affected Workers

The recommendations for redesigning jobs for affected workers could be based on the following criteria: frequency should be less than 20 actions per minute; minimal upper limb exertion (less than 5% of the maximal voluntary contraction, or a score of 0.5 on the Borg scale); posture and movements not requiring "intense involvement" of the main joints; adequate recovery times during each shift. If tasks already include adequate recovery times, no further allowances are added.

If posture is not optimal, the recommendation is to slow down the frequency of the actions, with special attention devoted to the clinical findings of the worker involved. If the job involves occasional manual material handling, the following weights should be considered, assuming the handling is carried out with the correct posture:

- For 4–5 kg loads, lifted alone, lift no more than 2–3 times every 30 minutes.
- For the occasional 7–10 kg load, lifted by 2 workers, lift no more than once every 30 minutes.
- For the occasional 14–15 kg load, lifted by 2 workers, lift no more than once every 60 minutes.
- Avoid lifting loads of more than 15 kg.

2.3. Matching Tasks and Jobs with Disorders

To facilitate communications between the technical staff and the medical staff within the factory, the results of the analysis of the various jobs and tasks should be classified. It is also useful to classify upper extremity disorders in relation to degree and severity. This makes it easier to match the affected workers with the most appropriate jobs.

The following ratings are suggested for the task: excellent with no limitations (no modifications needed, suitable for all affected workers); excellent with some limitations (no modifications needed, but not suited to all affected workers); excellent with modifications (suitable for all affected workers, even the most severe, provided that the recommended modifications are adopted); very good with modifications (suitable for workers with a moderately severe condition, provided the recommended modifications are adopted); unsuitable (job difficult to adapt, therefore not suitable for affected workers).

The following ratings are suggested for disease degree: severe (only one part of the upper limb affected severely, or several parts affected moderately); moderate (only one part of the upper limb affected moderately, or several parts affected mildly); anamnestic (only suspect symptoms detected, with no clinical or instrumental findings). It is also important to check the sequence of the upper limb involved — scapulohumeral joint (shoulder), elbow, wrist, hand/finger.

2.4. Timetable for Returning WMSD-affected Workers to the Workforce

The principal aim of reassigning workers with upper limbs disorders to new or redesigned jobs or tasks is to alleviate the clinical signs and symptoms associated with the disorder while maintaining the worker's productivity. To ensure the results of the various decisions could be monitored continuously, it was essential to have close cooperation between the technical staff,

the plant medical staff, and above all the workers themselves. Here is a recommended schedule.

2.4.1 Medical history

Workers known to have work-related disorders should be interviewed by medical staff regarding their symptoms upon starting the redesigned job. The interview should be carried out using a clinical questionnaire in order to standardize the data collection.

2.4.2 Enhancing worker awareness

Meetings should be arranged between the medical staff and the workers involved in the study. The meetings should be designed to make the workers more aware of how best to cooperate with the medical and technical staff in verifying how successfully the jobs are redesigned and allocated. During the meetings (for groups of up to 10 workers), the following points should be discussed:

- Type and pathogenesis of the most common biomechanical strain disorders affecting the upper limbs; reasons for changing and shifting workers to new or redesigned tasks and jobs.
- Correct use of the modified job, particularly the correct use of recovery times; importance of maintaining a regular pace without accelerating the frequency of the actions; importance of avoiding the needless lifting of harmfully heavy loads.
- Need to pay careful attention to any disorders affecting the upper limbs, and in particular to any flare-up of existing symptoms (if symptoms worsen, the worker is invited to notify the plant physician, even between regularly scheduled follow-up appointments).
- Need to pay careful attention to the new method adopted for performing the job, and inform the supervisor whenever any unforeseen problem arises (use of excessive force, inability to keep up with required speeds, etc.).

2.4.3 Supply and use of orthopedic orthesis

During the meeting to discuss the worker's medical history, the medical staff should recommend the use of specific orthopedic ortheses whenever they are appropriate. The ortheses are recommended primarily for workers with carpal tunnel syndrome; initially they should be worn only at night.

3. HEALTH CARE MONITORING

Close monitoring is required every time jobs and tasks are redesigned; this should be carried out almost continuously, so that any technical and organizational modifications can be made promptly. A special health care program is used to validate the decisions made about the various diseases that are present; this aims to monitor (at close intervals) the clinical condition of the workers in their newly designed jobs, as well as their degree of acceptance of the jobs.

A simple questionnaire should be used to obtain information regarding changes in relevant symptoms (classified as symptoms completely disappeared, symptoms improved, situation unchanged, symptoms worse, appearance of new symptoms or disorders); use of the orthopedic appliance, in terms of compliance and degree of tolerance; assessment of acceptability of new job or task, with details concerning any inadequacies

deriving from awkward joint segment positions, excessive muscle force, etc.

Moreover, the questionnaire should provide a diagnostic link, perhaps for requesting further clinical or instrumental tests; and a useful operational link for the production engineers. Based on the results of the study, the physicians are able to provide the engineering staff with practical information, which may be summarized as follows: the worker need no longer be classed as "affected" (remission of disease); the worker may continue to perform his/her duties in the current job, even if not redesigned for affected workers (anamnestic case); the worker must remain in his/her current workstation redesigned for affected workers (clinical case); the worker must be transferred to a redesigned workstation (new case or worsening of anamnestic case); the workstation to which the worker has been assigned must be redesigned for diseases other than those related to repetitive movements. Checkups should be scheduled as follows:

- 2 weeks after changing jobs
- 3 months after changing jobs
- 6 months after changing jobs
- 12 months after changing jobs
- every 12 months thereafter

The checkup at 2 weeks is to give a rough assessment of whether or not the situation is satisfactory, and to reinforce the instructions given to the worker.

4. RESULTS OF THE HEALTH CARE MONITORING PROGRAM

Cross-tabulation is used to analyze the preliminary results of the monitoring program on workers affected by upper limb disorders who were transferred to a redesigned job or task one year earlier. The tables show disease, degree of severity, sex, location, bilateral location, multiple joints, relative complaints (improvement or worsening), appearance of new disorders, department and job redesign. To elucidate the most plausible phenomenon, it is usually advisable to evaluate the worst percentiles and the better percentiles.

Table 1 shows the trend for disorders affecting wrists analyzed in two different departments: the job in one department (assembly line) had been redesigned earlier than in the other department (electrical engines) due to technical problems. A large proportion of the workers in the second department were still performing jobs that were only partially redesigned. The table clearly illustrates the difference between the wrist symptoms in the two departments: the results are highly positive in the first area. Even without a thorough statistical analysis, there is no question the redesign has had positive effects. Special attention should be paid to carpal tunnel syndrome and the use of orthopedic ortheses; workers who wear an orthesis frequently report an improvement.

5. CONCLUSIONS

- It seems essential to reduce the amount of exposure to repetitive tasks of the upper limbs among workers with WMSDs.
- Reducing exposure to risk factors associated with repetitive tasks of the upper limbs would seem to be a sufficiently

Table 1. Condition of workers with specific wrist disorders 6 months and 1 year after starting re-designed job, in two different workstations.

WRIST SYMPTOMS	RIGHT WRIST					LEFT WRIST			
	ASSEMBLY BAY WORKSHOP		ELECTRICAL ENGINES			ASSEMBLY BAY WORKSHOP		ELECTRICAL ENGINES	
	No.	%	No.	%		No.	%	No.	%
Disappeared	10	33.3	2	11.1		5	20	3	18.7
Improved	8	26.6	4	22.2		8	32	4	25
Unchanged	10	33.3	7	38.9		9	36	5	31.2
Worsened	0	0	5	27.5		1	4	3	18.7
Onset of new symptoms	2	6.6	0	0		2	8	1	6.2
TOTAL POSITIVE CASES	30	100	18	100		25	100	16	100

adequate measure. In fact, it produces a distinct improvement in symptoms.

- The decision to allow workers to perform tasks with a frequency of less than 20 actions per minute, in the absence of any other risk factors, proved to be a good starting hypothesis which deserved to be further verified and, eventually, validated. This approach moreover preserves the WMSD-affected worker's residual productivity. Expressed as an OCRA exposure index (Occhipinti 1998), this decision would imply that workers with upper limb disorders are on the whole better suited to jobs and/or workstations with a strain index score less than or equal to 0.7.

- For the program to be successful, it is vital to ensure active participation and excellent communications between all those involved (technical staff, plant medical staff, workers).

REFERENCES

BATTEVI, N., MENONI, O., and VIMERCATI, C., 1998, The occurrence of musculoskeletal alterations in worker populations not exposed to repetitive tasks of the upper limbs. *Ergonomics*, 41(9), 1340–46.

OCCHIPINTI, E., 1998, OCRA: a concise index for the assessment of exposure to repetitive movements of the upper limbs. *Ergonomics* 41(9), 1290–1311.

Healthy Work Organization

D.M. DeJoy, M.G. Wilson, and C.S. Griffin-Blake

Workplace Health Group, Department of Health Promotion and Behavior, University of Georgia, Athens, GA 30602, USA

1. INTRODUCTION

Work organization refers to the way work processes are structured and managed, including scheduling, job design, interpersonal aspects, management style, organizational characteristics, and related topics. Healthy work organization is a logical outgrowth of work organization and represents the idea that (a) it should be possible to identify a set of job and organizational dimensions or factors that characterize the healthy organization, and (b) such workplaces should have safer, healthier, more productive workers, and superior profitability and market success. In most respects, the basic concept of healthy work organization falls within the scope of macroergonomics. Macroergonomics emphasizes a top-down sociotechnical systems approach to the overall design of organizations, work systems, jobs, and related human-environment interfaces (Hendrick 1991). Microergonomics, in contrast, focuses more on the immediate human-machine interface and the design of specific tasks, jobs, and workstations.

Part of the attractiveness of healthy work organization is that it represents the integration of research and practice in several areas, including human resources and organizational development, occupational stress, occupational safety and health management, and worksite health promotion. Each is discussed briefly below.

1.1 Human Resources and Organizational Development Perspective

Case studies of successful organizations suggest that healthy people and healthy relationships are at the very core of business, and that many business problems and failures to remain competitive can be traced to traditional, hierarchical forms of management and employee relations (Rosen and Berger 1991). From this perspective, creating healthy work organizations involves a redefinition of the relationships, expectations, obligations, and interactions between employees and organizations.

1.2 Occupational Stress Perspective

A number of prominent stress researchers have attempted to identify the job and organizational attributes or factors which characterize healthy or low stress work environments. Almost without exception, these authors emphasize the importance of organizational or contextual factors in diagnosing and remedying the causes of stress within organizations. For example, Cooper and Cartwright, in their "front-end" approach to addressing organizational stress (Cooper and Cartwright 1994) suggest a variety of organizational change interventions, in addition to the more traditional individual-oriented strategies typically used to manage job stress. Karasek and Theorell (1990) argue that a fundamental shift in the relative distribution of decision latitude or control is essential to creating healthy work organizations.

1.3 Occupational Safety and Health Management Perspective

Three areas of occupational safety and health research, in particular, underscore the breadth of interest in organizational or contextual factors. First, research on safety and health climate or culture points to the importance of the broader organizational context in shaping employee behaviors. Second, recent research on human error, most notably Reason's work on active and latent failures (Reason 1995), calls attention to the importance of organizational action or inaction in creating the substrate or backdrop for human error by frontline workers. And third, research on workplace self-protective behavior is consistent in highlighting organization-level factors as facilitating or hindering safe behavior in the workplace.

1.4 Health Promotion Perspective

Work in this area argues that worksite health promotion programs should be expanded to include interventions designed to improve organizational and environmental conditions. DeJoy and Southern (1993) have proposed an integrative model of workplace health promotion that features three interactive systems: job demands and worker characteristics, work environment, and extraorganizational influences. The basic architecture of this model suggests that attempts to preserve and enhance worker health should develop from a total work situation analysis.

1.5 Cross-cutting Themes

Viewed together, the four perspectives reveal three cross-cutting themes:
- The increased salience of organizational or contextual factors in the work-health relationship.
- The importance of organization-level action in producing positive change.
- The need for modification of the traditional employer-employee relationship in terms of increased opportunities for employee involvement and input.

2. DIMENSIONS OF HEALTHY WORK ORGANIZATION

Reviews of the literature cited earlier suggest that a set of dimensions of healthy work organization can be compiled and conveniently organized into three categories or domains: (1) job-related factors, (2) organizational structure or climate, and (3) career development (Figure 1). Although it is impossible to argue that the dimensions listed in Figure 1 represent a perfect or all-inclusive list, there is sufficient consistency across the literature to make a persuasive case for the importance of each dimension.

2.1 Job-related Factors

The rather extensive literature on job stress provides much of the foundation for a general agreement on the job factors that are most important in preserving or enhancing worker health and well-being. These factors include workload, autonomy and control, role clarity, job content, work scheduling, and environmental conditions.

Workload consists of the daily demands of the work situation or job design. Workload is associated with job satisfaction and an assortment of psychological, physiological, and behavioral strain symptoms. However, workload per se may not necessarily

Figure 1. The three domains of healthy work organization

be the defining factor in producing worker strain. Research on machine-paced and other externally controlled tasks indicates that strain may most likely occur when control is not commensurate with job demands.

Autonomy and control refers to the degree to which the job provides substantial freedom, independence, and discretion to the individual in scheduling the work and determining the procedures to be used in carrying it out. Employees who perceive relatively high levels of control at work have been associated with higher levels of motivation, performance, job satisfaction, involvement, and commitment, as well as lower levels of absenteeism, turnover, emotional distress, and physical symptoms and conditions, including exhaustion, headaches, and cardiovascular disease.

Role clarity involves the extent to which an employee's work goals and responsibilities are clearly communicated, and the degree to which the individual understands the actions and processes required to achieve these goals. Role ambiguity, the opposite of role clarity, exists when there is a high level of uncertainty about job expectations. It can lead to burnout, increased stress, work dissatisfaction, and decreased productivity and occupational commitment.

Job content involves the meaning, value, and worth that employees attach to their jobs. Jobs which are fragmented, highly repetitious, and monotonous, or that otherwise provide little stimulation or skill utilization have been linked to job dissatisfaction, worker strain, and poor mental health. Work scheduling encompasses a variety of options: full-time or part-time, fixed schedules or rotating schedules, and day work versus shift work. Employees who work their preferred schedule have increased job satisfaction, work commitment, and positive work-related attitudes. Rotating shifts and permanent night work appear to be most problematic in terms of psychological, social, and physical well-being.

Environmental conditions cover a variety of physical conditions and work requirements. They can affect the employee's well-being through exposure to factors such as loud noise, inadequate ventilation, bad lighting, variable temperatures and/

or smoke, toxic chemicals, and infectious diseases. Musculoskeletal symptoms and injuries have also been found in conjunction with various ergonomic hazards, such as highly repetitive tasks and frequent bending, twisting, and heavy lifting.

2.2 Organizational Structure and Climate

The dimensions within this domain emphasize the way in which the workplace is organized, particularly in terms of how it impacts interpersonal relations and the general social environment. They include organizational support, participation and worker involvement, and feedback and communication. Healthy work organizations should provide opportunities for meaningful interpersonal interaction and communication; this is because they give emotional support and they also give instrumental or tangible support in fulfilling job tasks and other assigned responsibilities.

Organizational support can be viewed as any action taken by the organization or its representatives that indicates a concern for the well-being of its employees. Employees' commitment to their organization and their job satisfaction, productivity, and morale are influenced by their perceptions of the organization's support and commitment to them. In addition, social support from supervisors and coworkers have significant effects on job stress and burnout.

Participation and worker involvement refer to situations in which employees have some meaningful input into job-related decision making. Considerable research indicates that increased levels of participation and worker involvement can increase job satisfaction, motivation, productivity, and mental health, and they can reduce job stress and role ambiguity.

Feedback and communication includes the degree to which employees receive direct, clear, and timely information about the effectiveness of their performance and the relative existence of ongoing and two-way communication within the organization. Having an effective two-way communication system may be one of the best methods to involve and empower employees, and it has shown positive effects on role ambiguity, job performance, job stress, worker satisfaction, safety and health program effectiveness, safety climate, and a variety of safety-related behaviors and outcomes.

2.3 Career Development

The third category shifts attention from the attributes of specific jobs and the characteristics of the organizational social environment to job security and career development considerations. In healthy work organizations, employees should be clearly informed about opportunities for improving their job skills and career opportunities, as well as about the organizational and economic developments that may alter their employment situation. The dimensions in this category include job security, advancement and learning opportunities, equitable pay and benefits, and flexible work arrangements.

Job security has been defined as an employee's perception of potential threat to continuity in his or her current job. Perceiving that one's job is secure has been consistently related to job satisfaction, work commitment, quality of life, and mental and physical health.

Advancement and continuous learning opportunities have assumed increasing importance in today's era of mergers, corporate downsizing, and rapid technological change. This

dimension involves opportunities for employees to broaden their job skills and knowledge, and to increase their resources and ability to cope with organizational change.

Equitable pay and benefits is a fundamental issue for employees and the organization. There are two important issues: the actual distribution of rewards and the perceived fairness of the procedures used to decide this distribution. Decreased job satisfaction, leadership, and working relations have been associated with inadequacies in the pay system.

Flexible work arrangements involve the extent to which job requirements allow employees the opportunity to accomplish tasks such as arranging childcare, shopping, scheduling appointments, and meeting other personal obligations and responsibilities. Positive relationships have been found between flexible work arrangements and depression, anxiety, job satisfaction, and work-family conflict for both men and women.

3. MODEL OF HEALTHY WORK ORGANIZATION

Figure 2 presents a conceptual model of healthy work organization that builds upon the dimensions presented earlier. A key aspect is that organization-level action is critical to creating and maintaining the healthy organization. The dimensions of healthy work organization may genuinely be considered as effects more than as causes; that is, well-designed jobs, supportive organizational climates, and positive career development options exist largely as a result of the policies established and actions taken by the leaders of the organization. So, in developing a workable conceptual model, there needs to be some entity, some construct, or some set of constructs that enable these effects to occur. The research literature which support the model provides ample evidence that organization-level support and action is fundamental to organizational effectiveness in terms of employee well-being and financial performance.

The *core organizational attributes* in the model include organizational values, beliefs, and behaviors. These three variables delve into the leadership resources of the organization at three different levels. Organizational values are defined as relatively enduring beliefs about what kinds of behaviors or end-states are preferable to others. Recent studies of successful organizations indicate that success typically comes from paying attention to all stakeholders. In terms of organizational values, healthy organizations and successful organizations should both show good balance between production and employee orientations. Organizational beliefs involve how the organization views its commitment to and responsibility for employee health and well-being.

Of central importance is the idea that employees represent more than units of cost in the business equation. Organizational

behaviors are managerial actions related to employee health and well-being, which essentially reflect how the organization's values and beliefs are translated into policies and programs. The inclusion of core organizational attributes in the model emphasizes the importance of organizational-level action as a leverage point for producing change in the workplace. The inclusion of the three levels (values, beliefs, and behaviors) is intended primarily to establish points of attachment for interventions designed to enable organizations to become healthier.

The job design, organizational climate, and job future components include many of the dimensions discussed earlier. The job design component refers to the dimensions related to job and task demands which consist of workload, control and autonomy, role clarity, job content, work scheduling, and environmental hazards.

The organizational climate component includes the dimensions of organizational support, coworker support, involvement, communication, and physical and psychological danger. Organizational support and coworker support are treated as separate variables in the model to reflect the importance of formal (organizational) and informal (coworker) social support in the workplace.

Feedback and communication are blended into one dimension, and as suggested by recent research on safety climate and culture, physical and psychological danger is included to represent the importance of personal safety and security on the job. The job future component consists of job security, equity, advancement and learning opportunities, and flexible work arrangements.

Figure 2 also features a component labeled *worker perceptions and expectations*. This component is defined as the beliefs, expectations, attitudes, and goals that employees have concerning the organization in which they work. It recognizes that the way in which workers perceive and evaluate events is important for understanding the impacts of various job and organizational factors. Four variables define this component of the model: job satisfaction, organizational commitment, empowerment and stress symptoms. They may also be viewed as intermediate outcomes that are ultimately linked to organizational effectiveness. These particular variables are included because of their demonstrated relevancy to the various dimensions of healthy work organization, and because satisfied, committed, empowered, and low stress employees contribute to organizational effectiveness.

Finally, the *organizational effectiveness* component includes financial and health-related outcome measures. The key to healthy work organization is that it impacts the financial performance of the organization as well as employee health and well-being. Financial measures might include outcomes such as return on equity, sales growth, absenteeism, and turnover. Health and well-being measures might include health care costs, health risk behaviors, loss prevention results, and work-related injuries and illnesses.

4. SUMMARY AND IMPLICATIONS

The idea of healthy work organization has come about largely as a logical extension of the interest in work organization within the field of occupational safety and health. Work organization sometimes means different things to different people, but the
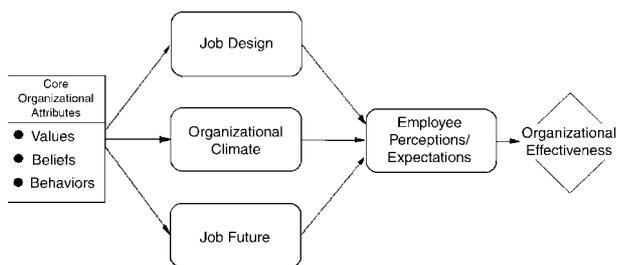
Figure 2. Conceptual model of healthy work organization

basic theme is that in trying to make workplaces safer and healthier, it is important to examine contextual and organizational factors, particularly how work processes are structured and managed. Although its roots may be in occupational safety and health, the concept of healthy work organization also reflects a convergence of research and thinking in human resources and management, occupational stress, and worksite health promotion.

Review and synthesis of this literature suggest that a profile of healthy work organization can be derived in terms of a defined set of job and organizational characteristics or dimensions. The dimensions have each been linked to various health or behavioral outcomes in the workplace, and some of these outcomes, such as job satisfaction and absenteeism, have been linked to traditional indices of organizational effectiveness. Moreover, these dimensions can be organized into a conceptual framework that provides a workable blueprint for auditing workplaces and organizations and for designing healthier ones. This model makes a strong case for the criticality of organizational change in creating and maintaining healthy organizations.

However, in spite of progress in describing healthy work organization, the actual predictive or explanatory capacity of the healthy work organization model has not been comprehensively evaluated. The model is tentative in the sense that it remains to be determined whether the various elements are both necessary and sufficient to explain either health or financial outcomes in real-world organizations. The dual-benefits aspect of healthy work organization is also largely untested. Very few studies have examined worker well-being and organizational financial performance simultaneously, even with respect to single variables or dimensions.

Using autonomy and control, one of the job dimensions in the model, as a case in point, a number of studies show that jobs characterized by low levels of worker control are associated with a variety of adverse health and behavioral consequences. Interventions designed to increase worker control also have demonstrated beneficial effects in terms of worker well-being. But shifting to the financial side, traditional research on organizational effectiveness has paid relatively little attention to control and autonomy as a variable of interest in examining outcomes related to productivity, efficiency, and financial performance. Even fewer attempts have been made to examine both worker well-being and financial outcomes in the same study.

Many researchers and workplace advocates would argue that positive behavioral and health outcomes alone provide sufficient justification for organizational redesign efforts, but the importance of testing the parallel-benefits hypothesis should not be diminished. Positive findings would provide significant and much needed impetus for redesign efforts aimed at maximizing human capital and creating safer, healthier, and more fulfilling workplaces and organizations.

REFERENCES

COOPER, C.L. and CARTWRIGHT, S., 1994, Healthy mind, healthy organizationæa proactive approach to occupational stress. *Human Relations*, 47(4), 455-71.

DEJOY, D.M. and SOUTHERN, D.J., 1993, An integrative perspective on worksite health promotion. *Journal of Occupational Medicine*, 35, 1221-1230.

HENDRICK, H., 1991, Human factors in organizational design and management. *Ergonomics*, 34, 743-756.

KARASEK, R. and THEORELL, T., 1990, *Healthy Work: Stress, Productivity, and the Reconstruction of Working Life* (New York: Basic Books).

REASON, J., 1995, A systems approach to organizational error. *Ergonomics*, 38(5), 1708-21.

ROSEN, R.H. and BERGER, L., 1991, *The Healthy Company: Eight Strategies to Develop People, Productivity, and Profits* (New York: G.P. Putnam).

Historical Development of Macroergonomics: the Development of Human–Organization Interface Technology and its Application to Work System Design

H. W. Hendrick

Hendrick and Associates, Englewood, CO 80111, USA,

1. INTRODUCTION

Ergonomics, or human factors as it alternately is known in North America, is considered generally to have had its formal inception as an identifiable discipline in the late 1940s. From the beginning, ergonomics has been concerned with the design of sociotechnical systems to optimize human–system interfaces. For the first three decades of its formal existence, the focus of this design concern was on optimizing the interfaces between individual operators and their immediate work environment, or what today often is referred to as *microergonomics*. Initially this focus was labeled as *man–machine* interface design. In the 1970s, as industrialized cultures became more sensitive to gender issues, it became known as *human–machine* interface design. In general, the objective was to apply scientific knowledge about human capabilities, limitations, and other characteristics to the design of operator controls, displays, tools, workspace arrangements, and physical environments to enhance health, safety, comfort, and productivity, and to minimize human error via design.

Beginning with the development of the silicon chip, a subsequent rapid development of computers and automation occurred, bringing a whole new set of human–system interface problems to the ergonomics discipline. As a result, a new sub-discipline of ergonomics emerged which centered around software design and became known as *cognitive ergonomics*. The primary focus remained on the individual operator, but with an emphasis on how humans think and process information, and how to design software to dialog with humans in the same manner.

In the late 1970s, ergonomists began to realize that our traditional predominate focus on individual human–system interfaces, or what herein is referred to as micro-ergonomics, was not achieving the kinds of improvements in overall total system functioning that both managers and ergonomists felt should be possible. More specifically, ergonomists began to realize that they could effectively design human–machine and human–software interfaces, and still have a poorly designed work *system*. Out of this realization developed the sub-discipline of *macroergonomics*.

2. THE BEGINNING: SELECT COMMITTEE ON HUMAN FACTORS FUTURES, 1980–2000

Although one can identify a number of precursors, the most direct link leading to the formal development of macroergonomics as a distinct sub-discipline can be traced back to the US Human Factors Society's Select Committee on Human Factors Futures, 1980–2000.

In 1978, Arnold Small, Professor Emeritus of the University of Southern California (USC) and former President of The Human Factors Society, noted the many dramatic changes occurring in all aspects of industrialized societies and their built environments. He believed that traditional human factors/ergonomics would not be adequate to respond effectively to these trends. At Professor Small's urging, the Human Factors Society (now, the Human Factors and Ergonomics Society) formed a "Select Committee on Human Factors Futures, 1980-2000" to study these trends and determine their implications for the human factors discipline. Arnold Small was appointed chair of this select committee. One of those appointed to the committee, Professor Hal Hendrick of USC, was specifically charged to research trends related to the management and organization of work systems.

After several years of intensive study and analysis, the committee members reported on their findings in October 1980 at the HFS Annual Meeting in Los Angeles, California. Among other things, Hendrick noted the following six major trends as part of his report:

1. *Technology*. Recent breakthroughs in the development of new materials, micro-miniaturization of components, and the rapid development of new technology in the computer and telecommunications industries would fundamentally alter the nature of work in offices and factories during the 1980–2000 time-frame. In general, we were entering a true information age of automation that would profoundly affect work organization and related human–machine interfaces.

2. *Demographic shifts*. The average age of the work populations in the industrialized countries of the world will increase by approximately six months for each passing year during the 1980s and most of the 1990s. Two major factors account for this "graying" of the workforce. First, the aging of the post-World War II "baby boom" demographic bulge that now has entered the workforce. Second, the lengthening of the average productive lifespan of workers because of better nutrition and healthcare. In short, during the next two decades, the workforce will become progressively more mature, experienced, and professionalized.

   As the organizational literature has shown (e.g. see Robbins 1983) as the level of professionalism (i.e. education, training, and experience) increases, it becomes important for work systems to become less formalized (i.e. less controlled by standardized procedures, rules, and detailed job descriptions), tactical decision-making to become decentralized (i.e. delegated to the lower-level supervisors and workers), and management systems to similarly accommodate. These requirements represent profound changes to traditional bureaucratic work systems and related human–system interfaces.

3. *Value Changes*. Beginning in the mid-1960s and progressing into the 1970s, a fundamental shift occurred in the value systems of workforces in the United States and Western

Europe. These value-system changes and their implications for work systems design were noted by a number of prominent organizational behavior researchers, and were summarized by Argyris (1971). In particular, Argyris noted that workers now both valued and expected to have greater control over the planning and pacing of their work, greater decision-making responsibility, and more broadly defined jobs that enable a greater sense of both responsibility and accomplishment. Argyis further noted that, to the extent organizations and work system designs do not accommodate to these values, organizational efficiency and quality of performance will deteriorate.

These value changes were further validated in the 1970s by Yankelovich (1979), based on extensive longitudinal studies of workforce attitudes and values in the United States. Yankelovich found these changes to be particularly dramatic and strong among those workers born after the World War II. Of particular note from his findings was the insistence that jobs become less depersonalized and more meaningful.

4. *World competition*. Progressively, US industry is being forced to compete with high quality products from Europe and Japan; and other countries, such as Taiwan and Korea, will soon follow. Put simply, the post-World War II dominance by US industry is gone. In the light of this increasingly competitive world market, the future survival of most companies will depend on their efficiency of operation and production of state-of-the-art products of high quality. In the final analysis, the primary difference between successful and unsuccessful competitors will be the quality of the *ergonomic* design of their products and of their total work organization, and the two are likely to be interrelated.

5. *Ergonomics-based litigation*. In the US, litigation based on the lack of ergonomic safety design of both consumer products and the workplace is increasing, and awards of juries have often been high. The message from this litigation is clear: managers *are* responsible for ensuring that adequate attention is given to the ergonomic design of both their products and their employees' work environments to ensure safety.

One impact of this message, as well as from the competition issue, noted above, is that ergonomists are likely to find themselves functioning as true management consultants. A related implication of equal importance is that ergonomics education programs will need to provide academic courses in organizational theory, behavior, and management to prepare their students for this consultant role.

6. *Failure of traditional (micro-) ergonomics*. Early attempts to incorporate ergonomics into the design of computer workstations and software have resulted in improvement, but have been disappointing in terms of (a) reducing the work system productivity costs of white collar jobs (b) improving intrinsic job satisfaction, and (c) reducing symptoms of high job stress.

As Hendrick noted several years later, we had begun to realize that it was entirely possible to do an outstanding job of ergonomically designing a system's components, modules, and subsystems, yet fail to reach relevant systems effectiveness goals because of inattention to the *macro*ergonomic design of the overall work *system*. Investigations by Meshkati (1986) and Meshkati and Robertson (1986) of failed technology transfer projects, and

by Meshkati (1990) of major system disasters (e.g. Three Mile Island and Chernobyl nuclear power plants, and the Bhopal chemical plant), have all led to similar conclusions.

Based on the above observations, Hendrick concluded in his 1980 report that, for the human factors/ergonomics profession to truly be effective and responsive to the foreseeable requirements of the next two decades and beyond, there is a strong need to integrate organizational design and management (ODAM) factors into our research and practice. It is interesting to note that all of these predictions from 1980 have come to pass, and are continuing. These needs would appear to account for the rapid growth and development of macroergonomics that since has occurred.

3. INTEGRATING ODAM WITH ERGONOMICS

As a direct response to Hendrick's report, an ODAM technical group was formed within the Human Factors Society in 1984, and similar groups were formed that year in both the Japan Ergonomics Research Society and the Hungarian society. Less formal interest groups were formed in other ergonomics societies internationally. In 1985, the International Ergonomics Association formed a Science and Technology Committee comprised of eight technical groups. The first eight groups formed were based on the input from the various federated societies as to what areas could most benefit from an international level technical group. One of those first eight was an ODAM Technical Group (TG). This TG has consistently been one of the IEA's most active groups. For example, the IEA ODAM TG has helped to organize six highly successful IEA international symposia on Human Factors in ODAM, with the proceedings of each being commercially published by North-Holland (Elsevier), and it is planned to continue this activity indefinitely on a biennial basis.

In 1988, in recognition of its importance to the discipline, ODAM was made one of the five major themes of the 10th IEA Triennial Congress in Sidney, Australia. In recognition of both its importance and rapid growth, it was one of twelve themes for the 11th Triennial Congress in Paris, France in 1990. At the 12th Triennial Congress in Toronto, Canada in 1994, and again at the 13th Congress in Tempare, Finland in 1997, a major multi-session symposium on Human Factors in ODAM was organized. For both of these congresses, more papers were received on macroergonomics and ODAM than on any other topic. A similar symposium is being planned for the 14th Triennial Congress in the year 2000.

4. THE CONCEPT OF MACROERGONOMICS

By 1986, sufficient conceptualization of the ergonomics of work systems had been developed to identify it as a separate sub-discipline, which became formally identified as *macroergonomics* (Hendrick 1986). In 1998, in response to the considerable methodology, research findings, and practice experience that had developed during the 1980s and 1990s, the Human Factors and Ergonomics Society's ODAM TG changed its name to the "Macroergonomics Technical Group" (ME TG).

Conceptually, Macroergonomics can be defined as a top-down, sociotechnical systems approach to work system design, and the carry through of that design to the design of jobs and related human–machine and human–software interfaces. Although top-down conceptually, in practice it is top-down,

middle-out, and bottom-up. The Macroergonomics design process tends to be non-linear and iterative, and usually involves extensive employee participation.

For an understanding of the sociotechnical systems approach, see *Sociotechnical systems theory* in this encyclopedia; for a better understanding of Macroergonomics and work system design, see *Sociotechnical analysis of work system structure* in this encyclopedia.

REFERENCES

ARGYRIS, C., 1971, *Management and Organizational Development* (New York: McGraw-Hill).

HENDRICK, H. W., 1986, Macroergonomics: a concept whose time has come. *Human Factors Society Bulletin,* 30, pp. 1–3.

MESHKATI, N., 1986, Human factors considerations in technology transfer to industrially developing countries: an analysis and proposed model. In O. Brown, Jr. and H. W. Hendrick (eds), *Human Factors in Organizational Design and Management, Vol. II* (Amsterdam: North-Holland), pp. 351–68.

MESHKATI, N., 1990, Human factors in large-scale technological system's accidents: Three Mile Island, Bhopal, and Chernobyl, *Industrial Crisis Quarterly*.

MESHKATI, N. and ROBERTSON, M. M., 1986, The effects of human factors on the success of technology transfer projects to industrially developing countries: a review of representative case studies. In O. Brown, Jr. and H. W. Hendrick (eds), *Human Factors in Organizational Design and Management, Vol. II* (Amsterdam: North-Holland), pp. 343–50.

ROBBINS, S. R., 1983, *Organizational Theory: The Structure and Design of Organizations* (Englewood Cliffs, NJ: Prentice-Hall).

YANKELOVICH, D., 1979, *Work Values and the New Breed* (New York: Van Norstrand Reinhold).

Human Factors and Total Quality Management

C. G. Drury

Department of Industrial Engineering, University at Buffalo, Buffalo, NY 14260, USA

1. INTRODUCTION

To understand and use the relationship between quality and ergonomics/human factors, the quality movement must first be reviewed, then where ergonomics and quality have worked together must be indicated and finally some prescriptions to enhance both disciplines must be provided. A fuller account of these relationships is found in Drury (1998).

2. THE QUALITY MOVEMENT

The quality movement began in the 1920s and 1930s with the application of statistical techniques to industrial operations. These included control charts for determining when a process had left its controlled state, sampling schemes for determining whether to accept a batch of product and industrial experiments to improve products and processes. Acceptance sampling is no longer used, but control charts have evolved into statistical process control (SPC) and industrial experiments into the Taguchi methods of robust design.

A major advance in the quality movement began when Japan began to develop quality techniques in the 1950s and 1960s to meet the challenges of producing quality goods with few natural or economic resources. Two themes characterize this movement: variance reduction and management. The first sees variation from all sources as a strong cost driver and therefore seeks to find, reduce and eliminate sources of variation in products. This has led to the redesign of products and processes using the philosophy of continuous improvement, i.e. continuous (often small) changes to reduce variance. Management's role in quality is now seen as a key to success. Quality has moved from being a technical discipline delegated to specialists to a managerial discipline of concern to the whole organization. The drive for continuous improvement is usually led by top management commitment, but implemented at much lower levels. Teams of employees, comprising operators, engineers and supervisors, are the preferred change agents. Underlying this management philosophy are the ideas that driving down variability will reduce costs as well as increase quality, and an emphasis on implementation rather than just statistical analysis (e.g. Evans and Lindsay 1995).

When these ideas were applied, quality moved from quality control, a largely on-line activity, through quality assurance to quality management, a largely off-line activity. Thus, the responsibility for quality moved both up and down within the organization, and (through robust design) backwards in the design/production/sales cycle. Quality management became an important tool throughout the world in the 1980s and 1990s with numerous successes claimed and many parallel activities, such as quality awards and international quality standards (e.g. the ISO-9000 series). "Successes" need to be treated with some caution, however, as isolating a single factor as causal in a complex, open system such as a manufacturing organization is logically quite difficult.

3. QUALITY AND ERGONOMICS

There are many sources of information on the quality movement, with an excellent overview and analysis provided by Hackman and Wageman (1995). Table 1 is adapted from their work to summarize the fundamental beliefs of the quality movement, particularly total quality management (TQM). The rest of this encyclopaedia provides equivalent summaries of the fundamental beliefs of the ergonomics movement, which has evolved over a similar time span into an industrially useful discipline. A summary of ergonomics/human factors tenets is given in Table 2. The parallelism between ergonomics and quality invites comparisons, which turn out to be useful in showing how these disciplines can work together more effectively.

Tenets of TQM parallel those of ergonomics/human factors in many ways. Both disciplines believe in the essential integrity of the human operator as a person who desires quality and avoids errors. Equally ergonomics and quality both rest on a foundation of measurement, with a belief in quantification rather than intuition. Finally, quality and ergonomics use teams and active operator participation as a way of finding problems and implementing change. However, ergonomics differs from TQM in some fundamental ways. It has often focused on the minutae of work design (microergonomics) whereas quality management has operated at a higher level. But microergonomics is not fundamental: ergonomics has at least as much to say about socio-technical systems as about biomechanics, and macroergonomics is now advocated as a higher-level approach.

The tools used by the quality movement, such as quality

Table 1. Tenets of TQM (adapted from Hackman and Wageman 1995)

Assumptions	1. Good quality is less costly to an organization than is poor workmanship
	2. Employees naturally care about quality and will take initiatives to improve it
	3. Organizations are systems of highly interdependent parts: problems cross functional lines
	4. Quality is viewed as ultimately the responsibility of top management
Change principles	1. Focus on the work processes
	2. Uncontrolled variability is the primary cause of quality problems: it must be analyzed and controlled
	3. Management by fact: use systematically collected data throughout the problem-solving cycle
	4. The long-term health of the organization depends upon learning and continuous improvement
Interventions	1. Explicit identification and measurement of customer requirements
	2. Creation of supplier partnerships
	3. Use of cross-functional teams to identify and solve quality problems
	4. Use scientific methods to monitor performance and identify points for process improvement
	5. Use process-management heuristics to enhance team effectiveness

Table 2. Tenets of ergonomics/human factors

Assumptions	1. Errors and stress arise when task demands are mismatched
	2. In any complex system, start with human needs and system needs, and allocate functions to meet these needs
	3. Honor thy user: use measurements and models to provide the detailed technical understanding of how people interact with systems
	4. Changing the system to fit the operator is usually preferable to changing the operator to fit the system. At least develop personnel criteria and training systems in parallel with equipment, environment and interface
	5. Design for a range of operators rather than an average; accommodate those beyond the design range by custom modifications to equipment
	6. Operators are typically trying to do a good job within the limitations of their equipment, environment, instructions and interfaces When errors occur, look beyond the operator for root causes
Change principles	1. Begin design with an analysis of system and human needs using function and task analysis
	2. Use the task analyses to discover potential as well as existing human/system mismatches
	3. Operators have an essential role in designing their own jobs and equipment, and are capable of contributing to the design process on equal terms with professional designers
	4. Optimize the job via equipment, environment and procedures design before optimizing the operator through selection, placement, motivation and training
	5. Use valid ergonomic techniques to measure human performance and well-being before and after the job change process
Interventions	1. Prepare well for any technical change, especially at the organizational level
	2. Involve operators throughout the change process even those in identical jobs and on other shifts
	3. Use teams comprising operators, managers and ergonomists (at least) to implement the change process

function deployment, design of experiments, statistical process control and cause-and-effect charts, focus on the technical processes. In contrast, the tools of ergonomics treat the human operator explicitly and centrally within a process context. Thus, function and task analysis goes beyond process description to discover potential human/system mismatches, which can lead to errors. To do this, ergonomics calls on explicit models of human functioning.

A final point of departure between ergonomics and quality is the consideration of the effect of the system and process on the operator. Ergonomics has operator well-being as a central concern, whereas this is only an implicit consideration in quality. The quality movement avoids errors, which can be logically extended to include injuries, stress and disease states as design errors. In practice, such extensions are rare. In contrast, ergonomics deals directly with questions of human injury, operator stress, job satisfaction and even industrial hygiene.

It is interesting that neither the quality movement nor ergonomics automation are the central issue, although automation is perhaps the most pervasive change agent in society. In quality, automation is treated, if at all, as an extension of mechanization, and thus as a way to reduce variability from human sources. In ergonomics, the design of automation plays an important, but hardly central, role, with function allocation seen as a way to make automation serve human needs. The effects of automation on healthy work design, stress and error rates are a specific area of ergonomics concern.

4. JOINT APPLICATIONS OF ERGONOMICS AND QUALITY

A number of case studies have been reported where ergonomics and quality were used jointly (for details, see Drury 1998). Two examples will be given to show that such joint activities can indeed be successful.

Eklund (1995) reported a major on-going study in automobile manufacturing where ergonomic and quality aspects were considered together. He showed a close correspondence between jobs with ergonomic problems and those with quality problems. This Quality, Working Environment and Productivity

(QPEP) project provided much evidence of the interrelationship between ergonomics and quality, and also questioned whether the quality movement alone could design healthy, safe and satisfying jobs.

In one of the USA's large mail-order companies, Rooney *et al.* (1993) started an ergonomics program under the auspices of an on-going TQM initiative. The focus was first on workplace design, but it later extended to include work organization and payment systems. Great reductions in workplace injuries and other measures of economic and ergonomic performance were reported.

There is clear evidence that ergonomics and quality can be implemented together with mutually beneficial results.

5. NINE PRESCRIPTIONS FOR COOPERATION

The following list of factors needed for joint ergonomics/quality efforts is modified from Drury (1998). It comes from a comparison of the tenets of quality and ergonomics (e.g. Tables 1 and 2) and a consideration of the literature in both disciplines. At first, similarities (points 1 and 2) where the movements are quite closely aligned are highlighted; then the focus is on areas where ergonomics can learn from quality (points 3–5), then ending with places where quality can learn from ergonomics (points 6–9).

5.1. Study and Measure the Process

Start from a systems focus rather than the current process, an approach also advocated in Business Process Reengineering. Use this as the basis for a detailed quantitative understanding of the process. Standard quality techniques should be used to measure process parameters, and models of human performance and well-being are used to measure and understand the role of the operator in the system. Use these measurements as the basis for directing and quantifying continuous improvement.

5.2. Honor Thy User

Respect the operator in the system as a person trying to do their best, and having an inherent stake in performing well. Do not necessarily blame the operator alone for poor quality/productivity/safety. Tap the potential for operator-empowered improvement by giving real power to small teams that include operators. The

rewards will be improvements in performance, safety and job satisfaction.

5.3. Consider the Strategic Level

Understand the forces beyond the process within the factory, such as requirements of the ultimate customer, and active management of the supply chain. Ensure that ergonomic interventions are truly customer-driven by explicitly measuring customer needs.

5.4. Understand Leadership

Any change activity needs responsibility of managers, up to the highest level. Do not take the mechanistic view of an organization that defines each manager by function. Understand the principles of leadership, recognize leaders and practice leadership. All change projects need a powerful champion.

5.5. Use Well-developed Team Skills

TQM, and many other change disciplines have standard methods of starting, organizing and running successful teams. Use these methods where appropriate. At least understand the methods so that the teamwork-training legacy within the organization from TQM programs can be built on.

5.6. Use Allocation of Function Techniques

A basic building block of human factors is the concept of function allocation, i.e. permanent or flexible assignment of logical functions between human and machine. This has been used by ergonomists at levels ranging from the whole complex system to a single human–machine system. Without an explicit treatment of function allocation, technology can easily fail.

5.7. Error-free Manufacturing/Service

While TQM is calling for drastic reductions in error rates, human factors is coming to grips with the causes of human error. Many military, industrial and service systems have classified errors and derived logical interventions, moving from a consideration only of the accident-precipitating event to a study of root causes and latent pathogens.

5.8. Interface Design

From the workplace layout to reducing injuries to the interface between software and user, human factors engineers have been designing less error-prone interfaces between people and systems. This set of techniques is largely ignored in the TQM literature, despite the latter's emphasis on error reduction, parts per million and six-sigma processes. Ensure that quality programs have access to, and actively use, human factors techniques.

5.9. Maximize Operator Well-being

Respect for the operator should lead to explicit use of operator well-being as a design criterion. System performance (quality, productivity) must not be achieved at the expense of human injury, disease, stress or dissatisfaction. Although designing for long-term performance should eventually ensure that these goals are met, they need to be a visible part of the criterion set if the short-term tradeoffs between the good of the organization and the good of the operator are to be avoided.

As can be seen, ergonomics and quality have enough points of similarity to ensure that they can work together successfully. Additionally, the strengths of each tend to compensate for specific weaknesses of the other. A joint approach should be, and has been measured to be, beneficial to the organization, its workforce and its customers.

REFERENCES

DRURY, C.G., 1998, Human factors and TQM. In Marras, W. and Karwowski, W. (Eds), *Handbook of Industrial Ergonomics*, chapter 78.

EKLUND, J.A.E., 1995, Relationships between ergonomics and quality in assembly work. *Applied Ergonomics*, 26, 15–20.

EVANS, J. R. and LINDSAY, W. M., 1995, *The Management and Control of Quality* (3rd ed.) (Minneapolis/St Paul: West), p. 143.

HACKMAN, J. R. and WAGEMAN, R., 1995, Total quality management: empirical, conceptual, and practical issues. *Administrative Science Quarterly*, 40, 308–42.

ROONEY, E. F., MORENCY, R. R. and HERRICK, D. R., 1993, Macroergonomics and total quality management at L. L. Bean: a case study. In Nielsen and Jorgensen (Eds), *Advances in Industrial Ergonomics and Safety V* (New York: Taylor & Francis), pp. 493–8.

Inspection

C. G. Drury, PhD, CPE
Department of Industrial Engineering, University at Buffalo, NY 14260, USA

1. THE ROLE OF INSPECTION

In manufacturing, inspection has traditionally been an activity associated with customer protection: the inspector protected the customer from defects in the product. This role has gradually changed as we move towards techniques of in-line quality assurance (QA), and, more recently, off-line QA. The aim now is to monitor the process, and its output, to determine whether it is in a state of control, i.e. will not produce defects (Vardeman and Jobe 1998). Inspection is now seen as a detection process within this QA system, so that the inspector is largely freed from ever seeing a defect. Indeed, if a production stream needs to be checked for defects, this is often performed automatically. Finally, any inspection activity has usually been reassigned to the production operator for the purpose of rapid control of the production process. Hence, there is unlikely to be a person with the title "inspector". The natural question is whether there is indeed any role for a human inspector?

Human inspectors have a natural role where the inspection task is complex — for example, finding one of a large number of possible visual defects in a casting, or when uniquely human sensing is required as in wine-making. They also have a role in hybrid inspection, where a display system presents manipulated images of objects so that human pattern-recognition and decision skills can be utilized. In addition, with the inspection functions now assigned to production operators more, not less, people are likely to be engaged in inspection activities. But manufacturing is not the only home of inspection. It is a feature of medical diagnostic techniques (X-rays, CAT scans), of continuing airworthiness of aircraft (crack detection, systems testing) and of many security systems, such as screening airline passenger baggage for weapons or explosives. If the inspector at the end of a production line has passed, the inspection activity in manufacturing is still a valuable asset, and the inspector in service functions is still very much a growth job.

Inspection as an activity should above all be accurate: a baggage screener who misses weapons is ineffective. Inspection should also be efficient in that it should take the minimum of resources, usually inspector time, to achieve the desired level of effectiveness. Finally, the inspection should be flexible, i.e. able to detect a variety of threats/defects/targets each with a high level of accuracy and efficiency. It should be noted that none of these attributes specifically imply human inspection.

2. THE FUNCTION OF INSPECTION

While we can specify the desired attributes of the inspection activity, we must examine the functions comprising inspection more closely if we are to understand and improve an inspection system. There have been a number of listings of the functions comprising inspection, but a generic list from Drury (1992) will be used here as it can apply to many different inspection systems. Table 1 shows the generic function list, with particular examples for industrial inspection, security screening and civil aircraft in-service inspection.

From Table 1, it is clear that inspection comprises some functions which are mainly mechanical (Access, Respond), some which have considerable cognitive components (Search, Description) and some which can include significant amounts of both (Initiate). The functions listed in Table 1 can be used to predict and classify the errors which can arise during inspection. Each function can fail in different ways, and for different underlying causes. While a complete analysis is not possible here (see e.g. Drury and Prabhu 1993 for more detail), typical errors and typical causes are presented in Table 2 for each function. Because the main objective of inspection is accuracy, the possible outcome errors at each function are also listed in Table 2.

At the most aggregated level, inspection errors can be classified as either a Miss or a False Alarm:

Miss — Error of failing to respond correctly to a true defect, i.e. concluding "No Defect" instead of "Defect".

False Alarm — Error of responding to an item without a defect as if it had a defect, i.e. concluding "Defect" instead of "No Defect".

Table 1. Generic inspection functions with examples

Generic function	Industrial inspection	Security screening	Civil aircraft inspection
1. INITIATE	Calibrate inspection equipment. Set up for specific product.	Calibrate X-ray screener. Run self-test of computer system.	Read work interruption card for particular area to be inspected. Obtain lighting equipment, etc.
2. ACCESS	Select item from production system. Bring item to inspection position.	Next bag passes through X-ray on conveyor.	Locate area to be inspected on aircraft. Position self and equipment so area is visible.
3. SEARCH	Move eyes across selected item. Move item if needed to ensure visual coverage. Stop if an indication is found.	Visually scan display of bag image. Stop if an indication is found.	Move eyes, flashlight, and mirror so as to cover area to be inspected. Stop if an indication is found.
4. DECISION	Compare indication to standards for rejection of item. Reject if indication beyond standard.	Manipulate image if needed to compare indication to memorized images of threat items. Decide 'threat' if indication exceeds standards.	Manipulate lighting, mirror and, if needed, indication itself to determine if it exceeds memorized or written standards.
5. RESPOND	Mark item as defective, and re-route from process. If item not defective, return to search or next item.	If threat item, invoke defined security procedures, e.g. hand search of bag or call for law enforcement personnel.	If defect found, mark defect location and write up repair order. If not a defect, return to searching area.

Table 2. Errors associated with each generic function for the example of industrial inspection

Generic function	Typical errors	Typical causes	Outcome errors	
			Miss	False alarm
1. INTITATE	Equipment not correctly calibrated	Poor procedures. Setup for wrong product.	Miss	False alarm
2. ACCESS	Wrong item selected for inspection.	Poor procedures. Correct item not immediately available.	Miss	False alarm
3. SEARCH	3.1 Fails to locate indication	Inadequate coverage indication not recognized	Miss	-----
	3.2 Locates false indication	Poor briefing on indications. Visual similarity to true indication	---	False alarm (if decision also incorrect)
4. DECISION	4.1 Accepts true defect	Poor briefing on standards. Visual difficulty of judgement.	Miss	---
	4.2 Rejects false indication	Poor briefing on standards. Visual difficulty of judgement.		False alarm
5. RESPOND	Makes response which does not correspond to decision	Skill-based slip. Interrupted task. Difficult to make one of the responses.	Miss	False alarm

These outcome defects usually have quite different consequences. For example in aircraft inspection a Miss can result in an in-flight failure while a False Alarm can result in unnecessary delay and cost, e.g. to replace an avionics unit. Clearly, misses are the more serious error, but false alarms are not without costs. It has been found in avionics inspection that about half of all units removed for repair have "No Fault Found" on later detailed testing and analysis. These were presumably false alarms which incurred unnecessary additional testing and replacement costs.

While logically false alarms and misses can arise at any of the listed functions, the evidence of many studies (Drury 1992) shows that the most error-prone functions are Search and Decision, i.e. the more cognitive functions. The Initiate function is largely procedural, while the Access and Respond are typically physically limited functions. Brief descriptions of the roles of search and decision in inspection are given here, while the entries on visual search and signal detection contain more detail.

Search in inspection is often visual, although examples of tactile search and procedural search can be found. Visual search proceeds as a sequence of eye movements between fixation points, with the aim of searching the whole required area. Around each fixation point is an area within which the target or indication can be seen: the visual lobe. This area depends upon the characteristics of the target and background, being small when they are similar and large where they are very different. Some target/background combinations may result in visual lobes so large that they effectively cover the whole search area. In these cases, there is no serial search process and the target appears to "pop-out" from the background. An example would be a target of a color substantially different from the background. Provided this effect does not occur, then the inspector must choose at each fixation whether an indication has been found or not. If so, the Decision function is invoked: if not, the inspector typically continues searching. To continue searching the inspector must have some strategy for choosing the next fixation location. This can range from highly systematic, where the inspector never returns to the same fixation point again to quite random, where the inspector may well-refixate an area. There has been much modeling activity of these strategies (e.g. Morawski, Drury, and Karwan 1980) but the effect of strategy on outcome may not be that large unless the task is very simple. Most skilled inspectors tend to have a repeated search pattern (e.g. Megaw and Richardson 1979), which is modified by what is found during the search.

The models of the search process noted above can be used to predict how the probability of detection (POD) changes with the time spent searching the area. Obviously, longer times will give more complete coverage, with a consequent increase in POD. Typical curves for a single task are shown in Figure 1. Because search performance depends on the time allocated, search is a Resource Limited task. A frequent cause of inspection failure is search failure (Sinclair and Drury 1979), often due to insufficient coverage of the area. As the inspector searches the area for longer and longer, there will come a time when he or she decides to stop searching and move on to the next area or item. Choice of stopping time in search can again be modeled (Morawski, Drury, and Karwan 1980), with the models appearing to predict search performance of human inspectors quite well (Chi and Drury 1995). If the inspector does stop searching without finding an indication, the Decision function is trivial: the item should be classified as "No Defect." If an indication *was* found, then decision needs more careful consideration.

The Decision function is a choice between alternatives: either the indication is judged to have exceeded the standard for a defect or it has not. Such decisions are often modeled using Signal Detection Theory, which considers an evidence variable to be generated proportional to the perceived severity of the indication. This evidence variable is typically subject to noise, so that whether a true signal was present or not cannot be uniquely determined. The best an inspector can do is to choose a fixed level of the evidence variable as a criterion, and always report "signal" (i.e. "Defect") if the evidence variable exceeds the criterion and "No Signal" (i.e. "No Defect") if it does not. Decision performance is a function of the signal-to-noise ratio (or detectability) of the defect and the choice of criterion position, which is based on the costs and probabilities of the outcomes in the Signal Detection Theory model (McNichol 1972). Typically inspectors choose a criterion related to these costs and probabilities, but not the same criterion as an economic maximizer would (Chi and Drury 1998).

As the inspector's criterion is varied, so the balance between the two errors (miss and false alarm) changes. A low criterion will give few misses and many false alarms, while a high criterion will reduce false alarms at the expense of increased miss rate. Note that no signal detection issues can be invoked unless the search function succeeds in finding an indication to judge. Thus, it is important in understanding an inspection task to analyze
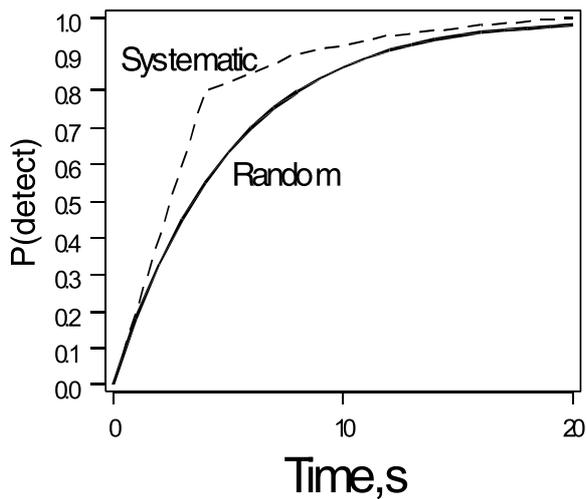
Figure 1. Probability of detection vs searching time.

only the Decision function in terms of theories such as Signal Detection Theory. If we use Signal Detection Theory on a task with a non-trivial Search function, factors affecting search will be wrongly interpreted as affecting decision. Because the overall wrong outcomes of inspection are "Misses" and "False Alarms", many have been tempted to apply decision theories such as Signal Detection Theory to the overall results of inspection performance measurements, with consequent misinterpretation of the findings.

The advantage of considering such models of search and decision in inspection is that they allow detailed quantitative modeling of the inspection task, and hence prediction of the variables affecting performance and the magnitude of the effects of these variables. For example in visual search, visual lobe size is proportional to the conspicuity of a target in a background, and mean search time is inversely proportional to visual lobe size. Hence the effect of enhanced conspicuity on inspection performance can be predicted, either as a known decrease in mean search time or a known increase in accuracy (POD) for any given search time. Similarly for decision, the Signal Detection Theory model shows that low defect probabilities tend to increase the chosen criterion, meaning that as product quality improves, the inspector can be expected to generate more misses and fewer false alarms.

3. ALLOCATION OF FUNCTION IN INSPECTION

With a listing of inspection's functions such as Table 1, it is possible, indeed inviting, to consider how these functions should be allocated between human and machine in an overall inspection activity. In this way, we can have a rational approach to inspection automation, replacing the currently practiced ad hoc approach. The primary determinant of function allocation in inspection is the outcome of the sensing element of search. If this sensing gives a continuous measurement, we have what is known in quality assurance as Variables Inspection. Examples would be measures of length, electrical resistance or color values. In contrast, the outcome can be sensing of discrete defects, such as scratches, rust, missing components or reversed assembly. This is known as Attributes Inspection.

Unless there are compelling reasons for including humans,

all functions in variables inspection are probably best allocated to machines. Computers can process continuous data easily, and produce reliable outcomes based on even complex rules and calculations. These outcomes can easily be coupled to the process itself to produce feedback control or even predictive control without human intervention.

Attributes inspection is where human inspectors prove their value. They can detect, perhaps with the aid of sensors, a wide variety of discrete defects, even those not considered by the inspection system's designers. Thus, security screeners can recognize as a "gun" a whole variety of firearms, most of which they have never seen before. A machine allocated the same function would perhaps test for the total projected area of some degree of X-ray opacity typical of gun material. Any area beyond some threshold would be flagged, missing very small guns or those with high plastic content, while alarming on metallic objects unrelated to weapons. Humans and machines will both generate misses and false alarms, but under particular circumstances humans can outperform machines.

Within attributes inspection, the relatively poor performance of people at visual search tasks suggests that the Search function be given primary consideration for machine allocation. The Decision function, in contrast, can often benefit from human pattern recognition abilities and human capability for responding to unprogrammed situations. In general, if only one or two well-characterized defects are possible, then machines can inspect well. Machines are, of course, constantly improving as computing power increases. There is a tendency among automation engineers to see their objective as ridding inspection of error-prone human inspectors. The result has often been highly automated devices which either do not perform even as well as the people they try to replace, or which are later found to need a human after all as the final decision-maker.

A better way is to choose between humans and machines for each separate function, and later combine the functions into an integrated, hybrid automation system. Often this means having an automated search function presenting displays of indications to a human decision-maker who is equipped with well-designed displays and job aids. Hou, Lin, and Drury (1993) show how such hybrid systems can out-perform either fully manual or fully automated systems.

4. DESIGNING INSPECTION JOBS

A system design must include specifications of those factors which control performance variability. For inspection, as for most systems, these can be classified under Task, Operator, Machine and Environment. In this entry they can only be outlined, but fuller treatments available, e.g. Drury (1992), a reference which also contains a systematic design procedure for inspection tasks based on the function list of Table 1 and a list of the attribute defects which must be detected. Here, an example of a listing of the factor affecting inspection performance is given in Table 3 for the example of baggage screening inspection. For each generic function factors are listed under the Task/Operator/Machine/ Environment headings.

From this list come some straightforward conclusions. First, there are large individual differences in the abilities necessary for inspection performance. These are unfortunately not constant across inspection tasks so that the attributes of a good circuit board inspector may not be those of a good X-ray technician.

Table 3: Factors affecting screener and system performance

Function	Task	Operator	Machine	Enviroment
Set-up	1. Calibration procedures 2. Maintenance procedures	1. Ability to follow procedures accurately 2. Training (quality and recency) for calibration and maintenance	1. Operator interface for calibration 2. Operator interface for maintenance	1. Management support to calibration and maintenance 2. Location of equipment 3. Availability of procedures, job aids
Present	1. Baggage selection procedure	1. Training (?)	1. Physical layout of equipment	1. Management control over baggage selected 2. Perceived and actual passenger pressure
Search*	1. Defined wand search pattern	1. Training (quality and recency) for search pattern 2. Operator dexterity	1. Human interface, e.g. wand, alarm indicator	1. Management support to ensure procedures are followed 2. Perceived and actual passenger pressure
Decision	1. p (true threat) 2. p (false alarm) 3. Time available	1. Knowledge of threats 2. Knowledge of potential false alarm items 3. Experience, overall and recent 4. Decision ability 5. Resistance to passenger pressure	1. Operator interface to display design 2. Operator interface to acquire additional information	1. Management support for following correct decision procedures 2. Perceived and actual passenger pressure
Response	1. Alarm resolution procedures 2. Management provision of adequate feedback	1. Knowledge of alarm resolution procedures 2. Resistance to passenger pressure	1. Interface to other systems (LEO, management, security)	1. Management support for alarm resolution procedures 2. Perceived and actual passenger pressure

Some individual difference measures do predict part of the performance variability across a number of tasks — for example, Field Dependency or Peripheral Visual Acuity. Despite considerable research, we still do not have universally effective selection tests for inspectors.

Training, however, is much better understood, and effective training procedures are available. These usually separate search and decision training (Gramopadhye, Drury, and Prabhu 1997) and go far beyond the simple lectures-plus-job experience found in most industries.

Task design must consider both the time allowed for inspection, which has a large effect on search, and the probabilities and costs of outcomes, which predict decision performance. Beyond these, factors such as time-on-task can play a role. Laboratory evidence shows that vigilance is difficult to sustain on repetitive tasks with low signal rates, similar to inspection tasks. There is not much evidence for such vigilance decrements in actual inspection tasks, but in fairness it should be noted that vigilance effects have not been sought in most inspection measurements in the field. At present, reasonable advice is that it would be prudent to allow for possible vigilance effects by using inspectors only for relatively short periods, perhaps up to 30 minutes, before they perform other duties.

Machine or equipment design can play a major role in inspection effectiveness. Workplace ergonomics do improve inspection performance. Also the design of the display of indications to the inspector can have a major effect. Now that we can manipulate images on a computer screen, we need to examine carefully how we manipulate them for maximum performance. This may not always be obvious to designers, as evidenced by the bizarre color mappings seen in inspection imaging equipment. Just because we *can* use hundreds of bright colors does not mean that this will enhance measured performance.

Finally, there is environmental design. In visual inspection the most obvious environmental influence is the provision of adequate lighting. However, not all lighting increases or "improvements" actually have an effect on measured performance. An example is the widespread use of green lighting in sheet steel inspection, despite direct evaluations showing that it gives performance no different from white light of the same intensity. Recent work, however, shows that sheet metal inspection can be improved by careful lighting design (Lloyd 1998).

5. CONCLUSIONS

Human factors has a long history of application to the inspection process and certain solid findings emerge. All inspection tasks appear to have the same generic list of functions, a finding which can be used to bring detailed ergonomic knowledge (e.g. of visual search) to bear on inspection systems design. In the design process, such knowledge can be used both to point the way towards intelligent use of automation, and to design the human-allocated tasks to fit the inspector's capabilities.

REFERENCES

CHI, C.-F. and DRURY, C.G., 1995, A test of economic models of stopping policy in visual search. *IIE Transactions (1995)*, 27, 392–3.

CHI, C.-F. and DRURY, C.G.,1998, Do people choose an optimal response criterion in an inspection task? *IIE Transactions* (1998), 30, 257–66.

DRURY, C., 1992, Inspection performance. In Salvendy, G. (ed.), *Handbook of Industrial Engineering* (2nd ed.) (New York: Wiley) 2282–314.

DRURY, C.G. and PRABHU, P.V., 1993, Human factors in test and inspection. In Salvendy, G. and Karwowski, W. (eds), *Design of Work and Development of Personnel in Advanced Manufacturing* (New York: Wiley), 355–402.

HOU, T.-S., LIN, L., and DRURY, C.G., 1993, An empirical study of hybrid inspection systems and allocation of inspection function. *International Journal of Human Factors in Manufacturing*, 3, 351–67.

GRAMOPADHYE, A.K., DRURY, C. G. and PRABHU, P., 1997, Training for aircraft visual inspection. *Human Factors and Ergonomics in Manufacturing*, 3, 171–96.

KARWAN, M., MOROWSKI, T.B., and DRURY, C.G., 1995, Optimum

speed of visual inspection using a systematic search strategy. *IIE Transactions,* 27, 291–9.

LLOYD, C.J., 1998, *Towards the Development of More Effective Surface Inspection Methods* (Morrisville, NC: Visual Performance).

McNICHOL, D., 1972, *A Primer of Signal Detection Theory* (Allen & Unwin).

MEGAW, E.D. and RICHARDSON, J., 1979, Eye movements and industrial inspection. *Applied Ergonomics,* 10, 145–54.

MORAWSKI, T., DRURY, C.G., and KARWAN, M.H., l980, Predicting search performance for multiple targets. *Human Factors,* 22(6), 707–l8.

SINCLAIR, M.A., and DRURY, C.G., l979, On mathematical modeling in ergonomics. *Applied Ergonomics*, l0(4), 225–34.

VARDEMAN, S. and JOBE, J.M., 1998, *Statistical Quality Assurance Methods for Engineers* (New York: Wiley).

Kansei Engineering and Kansei Evaluation

M. Nagamachi

Kure National Institute of Technology, 2-2-11 Agaminami Kure Hiroshima 737-8506 Japan

1. INTRODUCTION

Kansei Engineering was founded at Hiroshima University ~30 years ago and it aims at an introduction of consumers' thoughts on product design as well as function. When purchasing a product the consumer has a psychological image and feeling about the product, and if the product does not fit their expectations, then they do not want to buy it. The psychological image and feeling with regard to a product are called "*kansei*" in Japanese, and nowadays *kansei* is one of the most important design points for the development of a new product. Today consumers are more technology smart and they have a variety of goods at home. However, they will buy new goods if the new products have the characteristics coincident with the their feeling and demand.

Kansei Engineering is an ergonomic technology aimed at consumer-oriented product development which has the following roles: (1) identifying the consumer's *kansei* for a coming new product using ergonomic and marketing survey tools, (2) transferring the consumer's *kansei* to design field and (3) constructing an ergonomic system using computer science technology for a product development by designers or for selection of products by consumers' feeling. Figure 1 shows a diagram of the flow of Kansei Engineering in which the consumers' *kansei* is translated into the design domain through Kansei Engineering

2. KANSEI EVALUATION

2.1. Methods of Kansei Evaluation

The first point about Kansei Engineering is how one can evaluate the consumers' *kansei* accurately. The *kansei*, namely the consumers' feeling, will be expressed on their face and behavior as well as in their opinion. Therefore, the consumer's face and behavior is observed in a variety of situations. To develop a new type of home electric refrigerator, we videotaped a housewife's behavior when using it to prepare a meal, and we found a scene representing her workload when she bent her back. A Japanese home appliance maker, Sharp, developed a new type of home electric refrigerator that has a freezer at the bottom and vegetable room on top of the freezer. This change of position enabled the customers to pick up vegetables more easily than the old style freezer.

The estimation of physiological phenomena is also useful
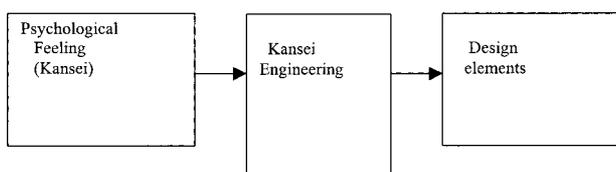
Figure 1. Diagram of Kansei Engineering.

index for product development concerning Kansei Engineering. We measured electroencephalograms (EEG) to estimate "Comfort," while heart rate and an electromyogram (EMG) improved movement and operation. All measurement tools used for ergonomic research are applicable to *kansei* with regard to Kansei Engineering. Face feature, face temperature, eye-open width, eye fixation, etc. are all used to estimate the consumer's emotion. Nose temperature was an accurate estimate of body temperature in room air-conditioning control.

Concerning the psychological survey, we often used the semantic differential method (SD). We collected a variety of word of nouns or adjectives which were regarded to *kansei* product features and constructed a five-point scale with those *kansei* words. The *kansei* words of "graceful–not graceful," "beautiful–not beautiful," etc. were used in the SD scale. The subjects of designers or customers were asked to represent their feeling on the SD scale with *kansei* words for a number of the products, and the obtained data on the scale are analyzed by statistical methods. When a sports type vehicle is a target as a new product, many sports cars are collected from the different makers and the subjects are asked to express their feeling about each car on the SD scale for each *kansei* word.

2.2. Statistical Methods of Relations between the Kansei and Design Elements

2.2.1. Factor analysis

First, we used factorial analysis to find the meaning structure of *kansei* to a specific design domain. For sports car exterior design, we evaluated a number of sports cars collected from the different makers on the *kansei* SD scales and these evaluated data are calculated by factor analysis that provides reduced factors with factor loads. This shows the meaning structure of sports car exterior domain and the obtained factors illustrate the design components essential to sports car exterior designing.

2.2.2. Quantification Theory

Second, we applied the statistical method by C. Hayashi "Quantification Theory Type I" to find the relation between *kansei* words and design elements. This method is a kind of multiple regression analysis applicable to qualitative data like the *kansei* and is a very powerful tool to find the design characteristics fit to the specific *kansei*. If our strategy on the new design might be "high class and compact," we followed the design side data from these *kansei* words and obtained several design elements which will contribute to our strategic mission.

2.2.3. Other methods

Neural Network Model was utilized to find the design element from *kansei* words in multiple layer NNM and the Genetic Algorithm (GA) can derive the decision trees connected to the design elements. Fuzzy Integral and Measure Theory is a useful model to make a relational structure between the *kansei* and design domain.

3. TYPE OF KANSEI ENGINEERING

We have developed so far six technical styles of Kansei Engineering methods: Types I to VI.

3.1. Type I: Category Classification

The category classification is a method by which a *kansei* category at the starting point of a planned product is broken down in a tree structure to determine the physical design details. At the beginning phase of a new product development, the project team in general attempts to decide the product concept. In the case of Kansei Engineering, the team decides "the zero-stage *kansei* concept" using whole surveyed data and this *kansei* concept is broken down to subconcepts at the first stage and to more meaningful details at the second concepts, finally to reach the physical specifications. This is like a brainstorming procedure, but in Kansei Engineering the top concept is the zero-stage concept and several concept layers construct a tree structure consisting of group subconcepts. It is said that this procedure is similar to QFD (Quality Function Development).

The physical specifications follow the ergonomic experiments to decide the design specifications and they are integrated to design a whole sports car at the final stage. The final whole design is in general reflected the zero-stage *kansei* concept in Kansei Engineering. The typical example used this technique for the new product development is "Miata" made by Mazda.

3.2. Type II: Kansei Engineering Computer System

Type II is a computer-aided system for Kansei Engineering that has the *kanei* database, image database and an inference engine and that supports the transfer of the consumers' feeling to design domain using an expert system.

The Kansei Database consists of statistically analyzed data of factor analysis and cluster analysis about *kansei* evaluation. The image database is constructed by the relational data between the *kansei* words and the design elements. The inference engine can estimate the design features including a product function through the rules of the expert system.

3.3. Hybrid Kansei Engineering System

Type II of Kansei Engineering means the flow to derive the design elements from the *kansei* words, which is called the "Forward Kansei Engineering System." This has the supportive roles for a designer to get a new idea of product and for a customer to select a good product fit to his/her *kansei*. While the designers want to create a new product based on the candidate that the computer system proposes, in some cases they want to check their new idea in reference to the *kansei* engineering system if it is best fit to the *kansei* database. For this purpose, the backward flow system was combined with Forward Kansei Engineering System, which is named "Backward Kansei Engineering System. Hybrid Kansei

Figure 2. Diagram of Hybrid Kansei Engineering.

Engineering System is an integrated computer system of Forward and Backward Kansei Engineering System (Figure 2).

In Hybrid Kansei Engineering System, a designer inputs his/her *kansei* in the Forward Kansei Engineering System to get the candidates proposed by the system, and the designer inputs the sketch into the system which he/she drew based on the candidate. The computer presents the result of the Backward Kansei Engineering System, which means a diagnosis about the inputted sketch. This technique was used to make the most recently developed steering wheel at Nissan.

3.4. Kansei Engineering Modeling

Kansei Engineering Type III implies a mathematical modeling constructed in a *kansei* system using a mathematical model. In this system, the mathematical model works as if it is a kind of logic like the rule-base. Sanyo Electric Co. attempted to implement Kansei Fuzzy Logic as a machine intelligence in a new color printer and it was successful in a sophisticated color printer which can produce more beautiful color print than the original picture. Nagamachi has developed a computerized naming system for a new product based on Kansei Fuzzy Integral and Measure Logic.

3.5. Virtual Kansei Engineering System

Virtual Reality Technology is now very popular technology in which people can walk through virtually a graphic design created by the computer and displayed by means of HMD (head-mounted display) and data gloves. We tried to combine Kansei Engineering System with Virtual Reality Technology regarding kitchen design in which a candidate of kitchen is determined by the Kansei Engineering System with a specific *kansei* and the customer can have a virtual experience by walking through displayed kitchen. Nagamachi extended this system to a whole house design that is called "HousMall."

3.6. Collaborative Kansei Designing System

We have many *kansei* databases. Then, we decided to implement these databases in the Website which has an intelligent 3D computer graphic system. The designers in separate rooms or in separate area collaborate with each other to create a new design of a product using the intelligent Internet with a variety of *kansei* databases. They talk to each other and correct a partner's proposed design through the Website.

4. REMARKS

There are many applied cases of Kansei Engineering. For instance, automotive, home electric appliance, construction machine, costume and house industries have introduced Kansei Engineering in producing the new products. The names of "Miata" made by Mazda, "Good-up bra" produced by Wacaol and "Ellesse" produced by Golwin are very well known products with applied Kansei Engineering. The introduction of Kansei Engineering has been spread out in the world and the researchers of Kansei Engineering are increasing. Kansei Engineering is an ergonomic technology of consumer-oriented product development and it is very powerful technology to be successful in producing very popular products.
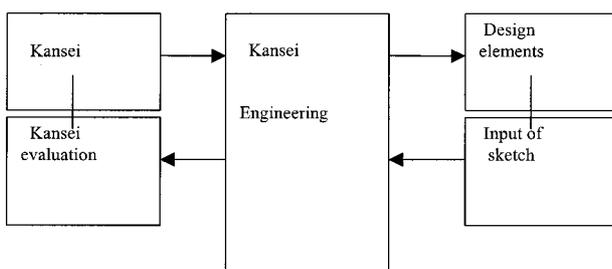
REFERENCES

NAGAMACHI, M., 1989, *Kansei Engineering* (Tokyo: Kaibundo).

NAGAMACHI, M., 1991, *Comfort Science* (Tokyo: Kaibundo).

NAGAMACHI, M., 1995, *Introduction of Kansei Engineering* (Tokyo: Japan Standard Association).

NAGAMACHI, M., 1998, Kansei Engineering: a new consumer-oriented technology for product development. In W. Karwowski and W.S. Marras (eds), *The Occupational Ergonomics Handbook* (New York: CRC Press), 1835–48.

Management of Work-related Musculoskeletal Disorders: Clinical perspective

M. J. Sanders† and R. Strickoff‡

†Department of Occupational Therapy, Quinnipiac College, Hamden, CT 06518, USA

‡Occupational Solutions, LLC, 40 Northwood Road, Monroe, CT 06468, USA

1. INTRODUCTION

Cumulative trauma disorders (CTD) have increasing challenged clinicians in the past two decades as escalating numbers of clients present in medical clinics with neuromuscular problems related to prolonged ergonomic exposures in the workplace. While some CTD conditions require surgical intervention, the majority of CTD requires conservative treatment that addresses multi-level problems. Such problems include not only diagnosis-specific tissue damage, but also proximal muscle imbalances that develop from prolonged awkward postures and potential behavioral problems associated with chronic pain.

Failure fully to address these issues may prevent clients from ultimately re-establishing their capacity for functional tasks.

This chapter will discuss the clinical treatment of CTD utilizing a comprehensive approach that emphasizes a thorough physical assessment, client education, ergonomic intervention, communication with the employer and client responsibility for treatment. It will discuss the goals of treatment at acute and chronic stages along with treatment modalities commonly used to treat CTD.

2. DEFINITION OF CUMULATIVE TRAUMA DISORDERS AND APPROACH TO TREATMENT

CTD, as termed in the USA, refers to a group of injuries affecting tendons, nerves, joints and blood vessels that develop insidiously as a result of longstanding ergonomic exposures. Although protocols for treatment of specific CTD diagnoses exist, the treatment of CTD is rarely protocol-based. Intervention must be considered in the full context of medical, workplace, insurance and family systems. Members of each system must actively communicate and share a perspective that appreciates the complexity of the problem. The following assumptions underlie a comprehensive treatment approach.

2.1. Multi-level Treatment

"Multi-level" connotes that CTD conditions must be examined within the context of multiple systems within the body and multiple levels within each system. These levels include multiple anatomic levels, multiple physiologic systems and multiple psychosocial contexts within the client's realm of functioning.

2.1.1. Multiple manatomic levels

The clinician must address not only local, diagnosis specific pathology but also anatomy both *proximal and distal* (above and below) to the identified injury. As prolonged use of the body in awkward or static positions alters the biomechanics of the entire limb, treatment of the discreet, identified area of pathology may provide only short-term relief.

2.1.2. Multiple physiologic systems

More than one body system may be involved in a CTD diagnosis. For example, while the pathology of lateral epicondylitis is generally regarded as inflammation and degeneration of the extensor tendons as they insert into the lateral epicondyle, the health and mobility of the radial nerve may also be compromised by the inflammatory process (Butler 1991). This example relates to a diagnosis of one system (tendon) which also includes multi-system involvement (nerve and tendon).

2.1.3. Multiple psychosocial contexts

Finally, both work organizational (work-related risk factors for job stress) and psychosocial (attributes of both the job and the individual that contribute to job stress) factors are believed to contribute to the development of CTD. Psychosocial factors may include job autonomy, job uncertainty, work pace and complexity, relationships with authority and peer groups, personal workstyle, job satisfaction and negative affectivity. Researchers believe that *psychosocial demands* may influence the perception and development of CTD in the following manner: (1) these demands increase muscle tension and thus exacerbate the biomechanical strains inherent in task performance; (2) psychosocial demands may impact clients' perceptions of the cause, awareness and reporting of musculoskeletal symptoms especially in VDT work; or (3) that in monotonous jobs, stress-related arousal of the central nervous system may sharpen sensitivity to normally subthreshold musculoskeletal stimuli (Moon and Sauter 1996).

2.2. Holistic Approach

Because CTD are multi-level, treatment must look at the whole person in the context of industrial, social, legal and personal situations, also termed a biopsychosocial approach. In the evolution of a CTD, more acute conditions may involve fewer systems. As CTD become chronic, the interplay between systems becomes more complex. Job and/or family stresses interact with workplace stressors to impact attitude, coping mechanisms and successful adaptation to work following an injury.

2.3. Client Empowerment

In medical models of treatment physicians traditionally take responsibility for "patient cure." These problem-oriented models efficiently treat tissue pathology but rarely prepare clients to prevent and manage future problems. The word "empowerment" builds upon the Latin root *passe*, from which one derives the words power and freedom. In empowerment models of treating, clients are responsible for influencing their health through making informed choices that affect the quality of their lives. Empowered clients influence change in their own personal behavior and social situations. They gain skills at empowerment through information exchange and gradually increasing competency at controlling their lives. Clients and therapists participate in the treatment process as equals and co-learners.

Table 1. Comparison of medical model and empowerment models of client-practitioner interaction. Adapted from Berger, *et al.* (1997).

Medical model	Empowerment model
practitioner-centered	client-centered
authoritarian relationship	collaborative relationship
practitioner holds responsibility to "cure"	client takes equal responsibility with practitioner for health
client compliance in treatment	client commitment to treatment
information giving	information exchange
practitioner must motivate the client	client must motivate self toward goals
client resistance to treatment is harmful	client resistance to treatment yields information

Table 1 outlines a comparison between the medical models and client-empowerment models of treating.

3. CLIENT EVALUATION

Clients are referred for therapy both through the worker's compensations and personal health insurance systems. As managed care expands, clients' access to specialists is limited with primary care physicians orchestrating much of the treatment for CTD. Depending on the physician's experience CTD diagnoses may be adequate but often incomplete. Thus, therapists must perform skilled physical examinations for the upper quadrant (neck, shoulder, upper back and upper extremities) truly to define the levels of injury.

3.1. Client Interview/History

Therapists must take a thorough *medical history*, which includes the presence of systemic conditions affecting soft tissue function, a history of respiratory, vascular or cardiac disorders relative to the patient's aerobic potential and previous or current orthopedic or neurological conditions. The consumption of caffeine and nicotine are addressed as both are vasoconstrictors and may affect target tissue nutrition.

The client *interview* frames clients' perceptions of their condition in terms of the etiology, mechanism and severity. Clinicians request information regarding the onset date, tasks associated with the onset, as well as descriptions of the location and intensity of pain. Pain scales are often used to qualify and quantify client's pain levels. Clients attempt to attribute the causal factors to tasks performed at work, home or both. In the more acute cases of CTD, clients commonly relate specific activity to localized pain. As conditions become more chronic, clients describe more diffuse discomfort throughout the upper quarter that is superimposed on a more tender area. The discomfort in such chronic, multi-level CTD, may be due to altered biomechanics and a combination of chronic and acute tissue damage that often coexist (Higgs and MacKinnon 1995).

3.2. Physical Examination

Clinicians are often misled by anatomically specific diagnoses and fail to inspect the entire extremity, scapula and trunk for multiple tissues involved in the kinetic-chain. This kinetic chain refers to a series of joints interconnected by soft tissues, in which altered structure in one region influences the structure and function of distant areas in the chain (Magee 1997).

A physical examination begins with observations of clients as they enter a room focussing on upper extremity use, guarding of an extremity, posture and facial grimaces. Posture assessments provide valuable insight as to the dynamics of the entire kinematic chain. For example, a forward head, elevated and rounded shouldered posture with pronated arms may indicate tightness of the sternocleidomastoid, scalene muscles and internal rotators. This information will key therapists into further analyzing muscle imbalances later in the assessment. Other components of the physical assessment include tenderness, range of motion, strength, sensation, neural dynamics, reflexes and edema in addition to provocative tests that stress specific tissue and reproduce symptoms.

3.3. Functional Assessment

The myriad roles that a patient assumes (homemaker, spouse, parent) must be examined in equal importance with role of a worker. A functional assessment identifies a client's occupational performance in self-care, leisure and homemaking chores. Activities that the client has identified as relevant and meaningful (such as grooming, gardening or cooking) are used as the benchmark for improved functional independence. Generally, clients with acute conditions complain less vehemently about difficulties performing self-care activities; those with more chronic conditions state that majority of self-care activities are unwieldy. While *pain does not necessarily relate to function*, therapists need to understand the extent to which self-care activities are limited due to the overall context of the illness.

3.4. Job Analysis

Therapists assess the vocational status of the injured worker for length of employment, work restrictions, and worker's compensation history. The job analysis identifies the essential job functions and specific ergonomic exposures related to the job such as awkward and static postures, forceful exertion, degree of repetition, mechanical stresses, vibration and work environment. Importantly, the productivity standards and organizational aspect of the job give insight as to the workplace dynamics. Ergonomic tools such as the NIOSH lifting index and strain index are used to quantify findings and prioritize need for intervention. While valuable information may be gathered through structured interviews, whenever possible the therapist should conduct an onsite worksite evaluation in order to verify information and ensure communication regarding treatment goals and work modification needs with the employer.

3.5. Psychosocial Status

The extent to which an injury impacts the client is closely related to pre-morbid psychosocial coping mechanisms and one's ability to adapt to stress. Clinicians interview clients regarding their home and work responsibilities, stress in their lives, and their means for identifying and coping with these stressors. Generally, those with balanced lifestyles and good support systems manage the added stress of an injury better than those without external supports. Certainly any client that appears depressed or highly anxious is referred for psychological counseling.

Table 2. General goals and treatment modalities for acute treatment of a CTD. Adapted from Duff (1997) with experience of the authors.

Goal for acute stage	Treatment modality
Reduce pain and inflammation	Eliminate activities that provoke signs and symptoms Ice Iontophoresis with dexamethosone Pulsed ultrasound Static splinting
Protect and rest affected regions	Eliminate stress-inducing activities Static splints Adapt workstation Encourage frequent breaks
Decrease proximal guarding of affected area	Promote circulation and good motor control with proximal exercises Breathing exercises Relaxation techniques
Educate client	Discuss risk factors Identify with client activities that *do* and *do not* produce symptoms Problem-solve means to modify activities collaborate on treatment goals
Participate in work-related and self-care activities as able	Provide adapted and ergonomic equipment. Modify home and workstation. Provide activity prescription

4. TREATMENT INTERVENTION FOR ACUTE INJURIES

Once the evaluation is complete the client and therapist set goals for treatment intervention. Goals are based on realistic expectations for tissue recovery and therapists/clients' expectations for functional and vocational task resumption. While the goals and corresponding treatment interventions in Table 2 reflect accepted standards for conservative care, interventions may vary according to the client and stage of recovery.

4.1. Goals and Progression Concepts in Acute Injuries

At the *acute stage* of a CTD, the primary consideration in therapy is to reduce the pain and inflammation while protecting the involved area from further injury. This task is accomplished through eliminating the stress-inducing activities at work and at home and by applying appropriate physical agent modalities in therapy. Numerous methods exist for treating CTD at this acute stage, usually driven by the experience of the therapist or physician. For example, in treatment of carpal tunnel syndrome, iontophoresis, ultrasound, splinting, nerve mobilization and tendon gliding exercises may be used in various combinations according to the individual response.

For therapy to be effective, therapists must communicate the need for job modifications to the employer up-front. Workers often attend therapy during scheduled work hours and return to the originating ergonomics stressors after treatment. If the job is not modified, the therapy process becomes a self-defeating cycle that is scrutinized by employers. Finally, clinicians need to reinforce client's participation in life roles with or without adaptation. Continual participation in customary life habits and work lessens the chances of a client transitioning from the worker role toward the patient or sick role.

4.2. Treatment Interventions in Treating Acute CTD

The following four concepts are essential in treating CTD at *both the acute and chronic stages*. These components address the client's role in treatment, the need to modify activity, balance musculature and address the importance of everyday physical activity.

4.2.1. Client education

Clients' level of knowledge, personal experiences, belief system and culture together shape clients' understanding of their conditions. Clients must accurately conceptualize the pathophysiology involved with a CTD and the rationale for the treatment before they can participate in the recovery process. For example, if a client's believes that passive inactivity will cure a condition, compliance with a home exercise program will be minimal at best. Clients must also understand the relationship between pain perception and anxiety, overly negative thinking and maladaptive coping strategies (Moon and Sauter 1996). Time taken to explain the mechanism of injury often improves clients' commitment to the treatment process.

4.2.2. Activity modification

Activity modification is essential during all phases of recovery and work reintroduction. Clients are taught techniques to unload vulnerable structures by using their upper extremities in a neutral position or by substituting stronger muscles groups for smaller groups. Clients may be taught energy conservation techniques at home and at work to simplify tasks and eliminate unnecessary actions. Activity modification also includes workstation adaptation, work schedule changes and the use of ergonomics tools.

Important to activity modification is monitoring the worker for *proper technique*. Some worker's technique probably contributed to the tissue injury; other workers may have altered their work methods to compensate for pain. In functional retraining clients practice proper technique until these movements are integral to their motor behaviors.

4.2.3. Muscle imbalance

Muscle imbalances must be corrected for longstanding soft tissue health prior to functional and postural retraining. Imbalances can be corrected by stretching tight, hypertonic muscles (typically the scalene, upper trapezius and pectoral minor muscles) followed by strengthening weaker groups (middle trapezius, rhomboids and serratus anterior) (Keller *et al.* 1998).

4.2.4. Physical activity

While the benefits of physical activity are well documented for healthy populations, physical activity potentially benefits those with existing CTD or at-risk for CTD. Physical activity is hypothesized to prevent cInjuries and assist recovery due to the improved oxygen delivery to peripheral tissues, increased capillary density, and increased strength of tendons at the myotendinous unit (junction of the tendon and bone)that occur as a result of

exercise (Magee 1997). Those in good physical condition use a smaller percentage of their total strength to overcome work-related forces as strength improves. Many clinicians recommend a walking program in conjunction with therapy to restore circulation and balance biomechanics.

The following interventions are additionally used at the acute phase of an injury to decrease pain and enhance tissue recovery.

4.2.5. Physical agent modalities

Physical agent modalities are effective short term agents that hasten the recovery process by reducing pain (elevating pain threshold), inflammation, and muscle spasm so that a client may return to functional activities in the most expedient manner. A specific modality will not be effective for all clients; therefore, the use of modalities should be monitored judiciously and evaluated for efficacy. *Cryotherapy* (cold packs, ice cubes, ice massage) is traditionally employed for pain relief and edema reduction in acute injuries. The most common use of cold therapy, cold packs, reduces inflammation by decreasing the metabolism and vasoactive agents at the injury site (Michlovitz 1996). While applications of 15–20 min are considered adequate to effect desired changes, prolonged application of ice may decrease muscle strength and impair nerve function over peripheral nerves. Cold packs may be self-applied during the day or used during treatment.

Iontophoresis utilizes electrical current to deliver ionized drugs such as anti-inflammatory medication transdermally to a specific area. This electrical transfer of medication is highly effective in the relief of acute pain for site specific tissue damage such as in lateral epicondylitis, deQuervain's disease and shoulder impingement syndromes. *Ultrasound* is a form of acoustic energy that stimulates protein synthesis for healing, increases lymph drainage and circulation, reduces edema and pain. Pulsed ultrasound (a mode that utilizes 50% power) produces non-thermal effects primarily used at the acute stage.

4.2.6. Splinting

Splinting can be an effective treatment technique during the acute phase of a CTD to protect soft tissue structures from further injury via compressive forces, mechanical stresses or repetitive strains. For example, a wrist stabilization ("cock-up") splint in a neutral position minimizes compressive forces on the median nerve in carpal tunnel syndrome; a radial gutter or thumb spica splint helps to decrease inflammation of tendons in the first dorsal compartment for deQuervain's disease. Custom splints (those fabricated by therapists) may also enhance function by immobilizing a painful joint yet allowing full mobility throughout the rest of the extremity.

Therapists must be cautious about recommending the use of splints during work. Splints may alter the normal joint biomechanics, alter functional prehension patterns and translate forces to another body part. Splinting should be combined with daily active range of motion to prevent joint stiffness. In short, the goals for splinting should be focussed and the implementation monitored conscientiously as prolonged use can be both physically and psychologically damaging.

4.2.7. Taping

Taping the upper extremity provides the therapist with a temporary means to inhibit over firing of certain muscle groups. For example, a client who subconsciously elevates her shoulder while typing may benefit from inhibitory taping to the upper trapezius. Taping has been used to correct joint instabilities, increase kinesthetic awareness of posture, unload weak painful muscle and assist with motor planning for activity modification.

4.2.8. Nerve mobilization

The circulation and excursion of nerves through tissue becomes compromised with limitations in daily movement. Clients with CTD benefit from careful instruction in techniques designed to promote mobility and circulation of neural tissues. Therapists instruct clients to sense the tension point for a specific nerve and then mobilize the nerve to restore its elasticity and circulation.

4.2.9. Breathing exercises

Treatment of the chronic CTD often begins with breathing exercises to promote circulation throughout the upper quarter. When patients are in chronic pain, they breathe shallowly and overuse the muscles in their neck. Nerves and vessels pass through these muscles and become constricted as patients continue this paradoxical breathing pattern. Diaphragmatic breathing is often the first tool that a client learns to relax and decrease pain.

5. TREATMENT INTERVENTION FOR CHRONIC INJURIES

5.1. Goals for Treating Chronic Injuries

At the *chronic stage* of a condition, symptoms may be provoked with activity but are generally not symptomatic at rest. Clients who have developed a chronic condition need intervention to remediate tissue length, circulation and strength as well as to manage chronic pain that potentially resurfaces with a resurgence of biomechanical or psychosocial stressors. Numerous theories exist about the diffuse, non-localized pain some clinicians associate with chronic, multi-level CTD. Biomechanical theories suggest that adaptive shortening of musculature due to prolonged awkward positioning causes pain as tight muscles are stretched and neurovascular structures are compressed (Higgs and Mackinnon 1995). Neural dynamics theories suggest that in CTD, the absence of normal movement causes venous stasis in the extremity and increased pressure within the neural structures. This increased pressure limits the normal mechanical strain (movement) that is necessary for healing; the central nervous system becomes bombarded with abnormal impulses from these tissues. Overtime, the nervous system becomes sensitized to pain that has spread beyond the original site (Butler 1991). In either situation, a pain cycle may begin that effects functioning of the entire kinematic chain.

5.2. Treatment Intervention for Chronic Injuries

Treatment for chronic injuries must include *patient education, activity modification, muscle balance restoration, and physical activity* as discussed above. Treatment may continue certain therapies from the acute stages but will integrate techniques to remodel and strengthen tissues for function. A sports medicine approach that assumes "pain is gain" is vastly different than the approach

Table 3. General goals and treatment modalities for the chronic phase of a CTD. Adapted from Duff (1997) with experience of the authors.

Goals for chronic phase	Treatment modalities
Restore length and mobility of tissue	Superficial heat Continuous ultrasound Soft tissue mobilization Gentle AROM with stretching Manual techniques Nerve mobilization Massage therapy
Restore strength to premorbid level	Graded isotonic strengthening Proximal followed by distal strengthening
Educate as to disease process and body awareness	Identify activities that *do* and *do not* produce symptoms with client Discuss risk factors Problem-solve means to modify activities Journal keeping
Demonstrate proper use of involved muscles	Provide body mechanics education Practice proper technique at worksite Recognize and avoid compensation patterns
Return to work	Perform worksite evaluation Identify physical limitations Modify job as necessary
Return to independent functioning in self-care, home-related and avocational activity	Modify self-care and home maintenance activities as needed Discuss choices in hobbies
Establish sense of control over life	Cognitive behavioral approaches Learn energy conservation Support groups

to progressing a client with a CTD. Gains are measured by slow stable increments so as not to irritate the originating problem or cause any biomechanical impedance in other areas. Therapist should observe the following:

- Indicate to clients that treatment should not elicit pain except normal muscle soreness due to exercise, called delayed onset muscle soreness (DOMS).
- Note changes in vasogenic and neural activity that are provoked with movement.
- Restore soft tissue length (including muscles and joints) before strengthening.
- Strengthen throughout the entire range after soft tissue length is restored.
- Empower clients to listen to their bodies and not exercise when ill or fatigued.

Table 3 provides common treatment interventions utilized as this stage.

5.2.1. Physical agent modalities
Moist heat is classically applied to more chronic injuries to increase tissue extensibility prior to gentle stretching, decrease pain, and to enhance the healing process by increasing tissue metabolism. *Continuous ultrasound* is another form of heat that can be effective for both superficial and deeper tissues (Michlovitz 1996).

5.2.1. Stretching
Stretching is consistently advocated in subacute and chronic CTD to restore the length of injured tissues. Great care must be undertaken when developing a stretching program; stretches must be safe, easily graded and simple to reproduce at home. Meticulous attention must be paid to proper techniques to avoid a painful reaction. While conventional wisdom supports that stretching may prevent injuries to non-damaged tissue by increasing tissues metabolism and facilitating a glide of tissue, studies have yet to indicate that stretching in isolation of other components to an ergonomics program decreases injuries

5.2.2. Strengthening
Strengthening applies a controlled strain to tissue that promotes tissue regeneration. Strengthening may be implemented when pain levels are under control and the client performs stretches and breathing exercises with good technique. Therapists will follow a progression from active range of motion (AROM) to isometrics, to progressive resistive exercises (PRE) at midrange and throughout the entire range (if indicated for that client). The last step, functional retraining, incorporates broad movements through full range of motion with good posture in all planes of movement. Treatment programs focus on building endurance before strength and allowing adequate recovery time between sessions (24–48 h).

While proximal strengthening is integral to CTD strengthening programs, distal strengthening is cautiously reintroduced, since injured tissues may be more vulnerable to stresses of lower intensity. Therapists educate clients in signs of overuse and many rely on gains through functional tasks to indicate increasing tolerance for work-related activities.

5.2.3. Postural retraining
Proximal stabilization of the trunk, shoulder girdle and scapulothoracic muscles is critical to healthy movement patterns and proper technique (Keller *et al.* 1998). A strong hereditary/behavioral component is inherent to posture. However, most clients benefit from learning to correctly position their bodies and maintain good alignment of joints. This is achieved through proprioceptive exercises, mobilization of the scapulothoracic joint, taping techniques, rib mobilization and graded spinal movements with breathing.

5.2.4. Manual therapies
Manual therapy techniques are "one to one" techniques that help restore integrity to soft tissue systems in the body. Manual therapies include trigger point therapy, soft-tissue mobilization, muscle energy, massage therapy, craniosacral therapy, neuromuscular education and more. Each technique requires execution by an experienced therapist who can discern the level of appropriateness for the client.

5.2.5. Cognitive behavioral therapies
Cognitive Behavioral Therapies emerged in the 1960s with heightened interest in behavioral therapies. The central concept is that a person's affect and behavior are largely determined by the way in which the individual construes the world (Turk *et al.* 1996). For example, some clients that are required to learn new skills at work may respond with anxiety, muscle tension and a

heightened awareness of pain; others view new skill acquisition as a welcome challenge, with no accompanying cascade of physical behaviors. Therapists assist clients in recognizing the connections among cognition, affect, and behavior. Clients learn to become aware of the role that negative thoughts and images play in the maintenance of maladaptive behavior. Change is noted when a client commits to replacing old ineffective behaviors with new positive approaches.

6. INDUSTRIAL REHABILITATION

While all therapies facilitate return to functional and vocational tasks, industrial rehabilitation provides treatment specifically focussed on acquiring or re-establishing work-related skills.

6.1. Functional Capacity Evaluations
Functional capacity evaluations (FCE) are standardized assessments that determine a client's safe physical capacities for weighted activities (carrying lifting), tolerances for body motions (standing, reaching, handling), and manual dexterity for the upper extremity. FCE are used to compare clients' capacities to job demands and thus determine the need for further therapeutic intervention or job modification (Isernhagen 1995).

6.2. Work Hardening Programs
Work hardening programs involve workers in functional activities that attempt to simulate the actual job movements and tasks. Work hardening programs include work conditioning, job simulation and body mechanics educational components to condition injured tissue and practice proper workstyle. Work hardening programs are particularly valuable in transitioning clients into work tasks and work roles who have been absent from the workforce for an extended duration.

6.3. Return to Work Programs
The transition to full duty work should occur over a period of weeks or longer. An activity prescription, usually provided by the physician, outlines the guidelines for modified duty. Activity prescriptions should include (1) specified duration of time in tasks, (2) restrictions for weighted activities, (3) the use of short breaks and stretches during the day and (4) modifications to the workstations as needed. The company determines the actual modified duty jobs in accordance with these guidelines.

7. CONCLUSION

Treatment of CTD requires communication and coordination of medical, industrial, insurance and psychosocial providers. While some clients will necessitate treatment intervention to remediate site specific tissues damage and correct muscle imbalances, other clients will require more complex interventions focussed on gradually gaining strength, self-care independence, improved posture and managing chronic pain. All clients will benefit from a thorough educational program that reviews the injury process and examines causal factors in light of both home and work contexts. This collaborative and client empowering approach will facilitate a smooth transition to work and strive to prevent future injuries from reoccurring.

REFERENCES

BERGER, B., HARPER, H. and KAVOOKJIAN, J., 1997, Patient empowerment. *US Pharmacist*, 12, 3–14.

BUTLER, D.S., 1991, *Mobilisation of the Nervous System* (Melbourne: Churchill Livingstone).

HIGGS, P.E. and MACKINNON, S.E., 1995, Repetitive motion injuries. *Annual Review of Medicine*, 46, 1–16.

ISERNHAGEN, S.J., 1995, *The Comprehensive Guide to Work Injury Management* (Gaithersburg: Aspen)

KELLER, K., CORBETT, J. and NICHOLS, D., 1998, Repetitive strain injury in computer keyboard users: pathomechanics and treatment principles in individuals and group intervention. *Journal of Hand Therapy*, 11, 9–26.

MAGEE, D.J., 1997, *Orthopedics Physical Assessment* (3rd ed.) (Philadelphia: W.B. Saunders).

MICHLOVITZ, S.L., 1996, *Thermal Agents in Rehabilitation* (Philadelphia: F.A. Davis).

MOON, S.D. and SAUTER, S.L., 1996, *Beyond Biomechanics: Psychosocial Aspects of Musculoskeletal Disorder in Office Work* (Bristol, PA: Taylor & Francis).

SANDERS, M.J., 1997, *Management of Cumulative Trauma Disorders* (Boston: Butterworth-Heinemann).

TURK, D.C., MEICHENBAUM, D. and GENEST, M., 1996, *Pain and Behavioral Medicine* (New York: Guilford).

Organizational Change and Supporting Tools

I. Hamburg[†], H. Paul[†] and J. Bahlow[‡]

†Institute for Work and Technology, Gelsenkirchen, Germany

‡GITTA mbH, Berlin, Germany

1. INTRODUCTION

In many European countries organizational, technical and personal development changes take place due to aggressive market conditions and the development of new technologies. The efforts made to implement cooperative work are examples for such changes. Cooperative work structures require not only organizational measures, but also an adequate technical support and a well-trained staff with the corresponding qualifications like social competence, the ability to deal with decentralized responsibility and a workplace comprehensive vision.

An increasing number of production companies with complex business processes have understood that the implementation and use of tools that help to deal more efficiently with workflows and business processes are requirements which ensure the success of an organizational change. Therefore, workflow systems (WFS), respectively workflow management systems (WFMS), have been developed to realize these requirements.

The development of a gas turbine at ELSTER Produktion GmbH is first described as an example for efficient and effective cooperative work, while afterwards a workflow procedure based on Petri-nets (used for an ecologically-friendly modeling of gear manufacturing within a TEMPUS cooperation) is described as an example for new concepts in the design and use of WFMS. Then some reflections on the special needs of WFMS in relation to successful, object-oriented companies – with a special look on the limits of current WFMS – are presented.
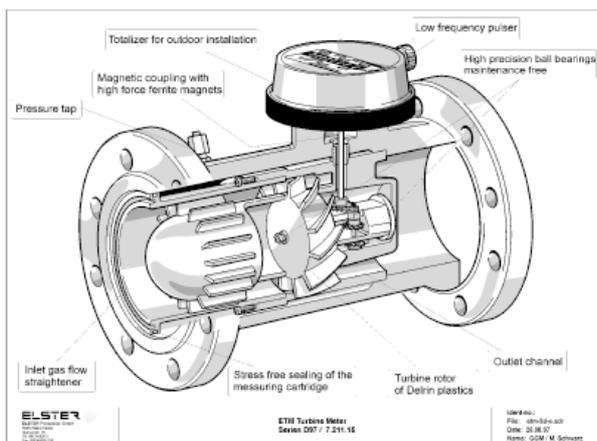
Figure 1. Turbine gasmeter developed by ELSTER Produktion GmbH during the RAMONA project.

2. COOPERATIVE PRODUCT DEVELOPMENT UNDER PRESSURE

In March 1996 aggressive market conditions and related factors led to a project on "gas transportation" at ELSTER Produktion GmbH, Mainz-Kastel, Germany. (ELSTER Produktion GmbH and GITTA mbH are partners in the project association RAMONA, "Frame conditions and modelling of new work structures." RAMONA is supported by the Federal Ministry for Education, Science, Research and Technology as a part of "Produktion 2000" and is supervised by the project management agency for manufacturing technologies and quality management at Karlsruhe Research Center. TEMPUS is part of the EU programme for the economic and social restructuring of the countries of Central and Eastern Europe for ELSTER.) It was necessary to develop – within the shortest possible time – a turbine-gasmeter for selected international markets. This especially meant the very fast development of a completely new product under high competitive pressure.

The new turbine-gasmeter was required (Figure 1):
- to enable reduction of costs by one-third compared with the existing products;
- to fulfil the high standards of all ELSTER products on accuracy and reliability; and
- to be available for market introduction within 6–9 months.

It was quickly realized that these highly set goals could only be reached through a team organization that would allow the early participation of all concerned company domains. It was necessary to find and establish an effective project management and a project team organization, being a process supported by the research and consulting company GITTA mbH in Berlin. Thus, a joint concept for organizing the teamwork and for distributing it into subteams was developed:
- Early inclusion of staff from marketing, distribution, buying, controlling as well as production and assembly in the project structure.
- Participation of team members in several teams.
- Restructuring of the teams for production start-up.

In monthly meetings of the whole team, the different backgrounds of experience and varying views on project aims and results were discussed and integrated. These meetings had the function of a bridge between the different subcultures of the company and their corresponding information categories (Schein 1992). The main goals were:
- the fast and uncomplicated cooperation among the team members; and
- the continuous exchange of their individual representations of the project progress.

The FMEA method (Failure Mode and Effects Analysis; Verband der Automobilindustrie 1986) identified possible difficulties and errors in the product concept as soon as possible. FMEA also enabled design engineers at early stages of the project to communicate and elucidate design ideas downstream. The discussion about concrete specifics of the not yet existing product proved highly motivating both for the production and the assembly staff.

Finally, interviews were held with the team members to evaluate the effects of team organization on the success of the project. Not only were the original goals of cost reduction, quality assurance and timely market introduction reached, but also:

- a frictionless production start was achieved;
- possibilities were recognized immediately to improve the customary products; and
- a basis was established for the continued collaboration of the concerned company domains.

The team members had learned a lot about the different problem views in the different company domains. Whenever problems or difficulties concerning product development arise, the team members now know the right persons to address.

The way the product was developed caused an increase in the abilities, qualifications and potential of the company and its staff – a further development that pays more than the success of the product.

3. WORKFLOW MANAGEMENT SYSTEMS BASED ON PETRI–NETS APPROACHES

WFMS are more complex than other information systems (IS) because their modeling objects – workflow processes – are much more complex than the modeling objects of other IS, e.g. data, functions. One of the aspects to be considered in WFMS is which information is used by the processes, from where and when does it come, and to whom and when does it go, as well as its influence on the workflow.

What is required is the capture of major aspects of the processes in a precise and concise manner. Furthermore, it is useful to have a direct relationship between the graphical and analytical representations. This is particularly advantageous in the analysis, simplification and verification of large and complex systems. Finally, the representation should describe abstractions and refinements in the process.

WFMS use optimized organizational structures to automatize workflow. They also influence the ways of individual work, e.g. through the assignment of corresponding work steps (Lehman and Ortner 1998).

Some objectives to aim for in a production company through the implementation and use of WFS (Oberweis 1996) are:

- To increase the productivity, e.g. by taking away some routine tasks from the staff and by decreasing the number of inefficient manual document searching which is needed for the corresponding task.
- To minimize costs, e.g. by minimizing the process running time.
- To improve the quality of products, e.g. by decreasing the error rate of the process' order.
- To increase client satisfaction, e.g. by the correct determination of the state of the current processing task, of the right person dealing with the task and of the task's length as well as by fast and flexible reaction to clients wishes.
- To guarantee openness for organizational changes in the company.

One of the requirements of WFMS in connection with modern cooperative, distributive forms of work is to coordinate the workflow in different company locations. So WFMS have to be developed as distributed systems which run in Internet and support distributed workflow.

From their theoretical basis, WFMS are based on different approaches like Petri-nets (PN), flowcharts or decision trees. Here, extended PN approaches are referred to, to use them for the modeling of business processes taking in consideration different aspects, e.g. also ecological ones.

The basis elements of a classic PN – which is a graph – are two node types called transitions, t, and input and output places, p (containing dots called tokens). The transitions and places are connected via directed arcs.

The generalized PN transition (GPNT) is characterized by a precondition function, functions of selection and functions of token color (for more on GPNT, see Condorovici and Hamburg 1998). To represent the objects (documents, resources, humans) more clearly and readably, some information is attached to the tokens in a PN with GPNT. This information is used, for example, in the analysis of the PN, for producing graphical or alphanumerical reports about the PN's evaluation.

An analyzer of a PN with GPNT has been developed that also facilitates the evaluation of costs and of some ecological aspect of the manufacturing process. In the application it was referred to the electrical energy consumption but, with small modifications, the PN can be used also for other ecological factors.

A workflow procedure was built that is based on the GPNT-concept, using a world-wide web (WWW) browser as a platform-independent user interface which is connected to a HTTP server (Condorovici and Hamburg 1998). The analyzer was developed as a stand-alone application – written in GNU C++ under a Unix operating system – and is automatically run by the server with parameters collected by client application.

The workflow procedure was used for the modeling of gears manufacturing in a Romanian machine tool company, taking in consideration the consumption of electrical energy. These developments are the result of a European cooperation between the Institute for Work and Technology in Gelsenkirchen (Germany) and the University of Iasi (Romania).

4. INSTRUMENTS OF ORGANIZATIONAL CHANGE

Here some findings about organizational change, organizational flexibility – important characteristics of successful companies (Fröhlich and Pekruhl 1996) – and the deriving special needs for supporting computer applications are reflected (Paul 1998).

Organizational change and organizational development should neither be seen as extreme exceptions, nor as nothing but *re*-action to external events. Successful companies can act in an anticipatory and prospective manner. Nevertheless, organizational change can be a risky process – even for an experienced company. Any support and every removed obstacle means an important aid for a changing organization. A workflow management system could be helpful as a supporting instrument of change.

The decisive questions are:

- Who is deciding on the organizational change?
- Who does initiate the organizational change?
- Who is bearing the processes? and
- Who is in control of the processes or who thinks that he is in control of the processes?

Hierarchically organized companies with a centralized way of decision-making differentiate among strategic, tactical and functional tiers: WFMS are tools for defining the functional tier. They are regarded as useful in the support of well-structured tasks with well-defined decision situations. On the superordinate tiers, where decisions about changes in the organizational structure and of its processes are made, WFMS cannot be found. WFMS

are used to bring those decisions to reality – they are not used to support the process of coming to decisions.

The underlying acceptance of the rule that decisions are made "upstairs" and decisions are executed "downstairs" makes this point of view a problematic one. This way of thinking proves counterproductive in organizations with object-oriented units and explicit decentralization of structures of decision-making.

What will happen in one of these autonomous object-oriented units when the unit comes to the conclusion that it has to restructure itself? What will this unit do when it finds out that it has to reshape its *internal* organizational structure and its *internal* processes? How will it handle the possible need to reconfigure the "interface" to its neighboring object-oriented units? How will it deal with tasks that will become unnecessary or that will be swapped out to other units or swapped in into the unit? The most popular way to handle these problems would be simply to do nothing but continue with "business as usual" and to ignore the findings – because one will never succeed with ideas in "higher" places.

There is a clear need for modular, flexible, scalable, tailorable and object-oriented WFMS that are able to deal with frictions. The main characteristic of this "next generation" of WFMS is that it can help the unit to re-structure itself and also to elaborate adequate interactive computer systems to the new structure.

A unit modifying itself needs answers to questions such as: why does the unit has to perform a certain task for who?; what is its function in the general structure of the organization? A WFMS has to provide the answers to these questions, but, therefore, it has to contain more than just data interfaces to applications. For an object-oriented unit, decentralized structures of decision-making and extended scopes of action also mean the treatment of tactical and strategical questions in a decentralized way and also one has to decide about innovation and not just everyday problems on the spot. And it also means change to the local organizational structure and its processes locally when it is needed. WFMS that are real tools of change will support the unit in such a situation – and will not hinder it like most of today's tools do.

It is a special quality of a WFMS as a tool of change that it is a useful aid even before structures and processes are well established and 100% fixed. It is their special task to support the development of the new structures and processes: they will help the organization to re-organize itself.

Organizations need an efficient and effective set of tools to be able to adapt themselves flexibly to changing conditions and the new possibilities. In fact, they need a set of tools that allows them to deal faster and more efficiently with their workflows and business processes, to reach goals more easily and to make more profit out of them. They have to be real WFMS: managing the workflow, and not simply making them flow somehow. They must help the organization to change and to reshape itself, e.g. to set up, run and modify new business branches, to move tasks within the branches, to modify structures of decision-making or to establish new forms of cooperation.

Although important findings were made, especially in the second half of the 1990s, the knowledge about this domain is not complete, e.g. we are still learning about group dynamics and what is triggered off by decentralized structures of decision-making. Little is known about the initialization of organizational change and the possibilities to "control" such processes. There are many more theories and assumptions about hindrances and stimulating factors of organizational change processes than proven facts. Instruments of organizational change are not just around the corner.

5. CONCLUSIONS

Successful organizational change and the use of supporting tools have helped manufacturing companies to survive and prosper in a turbulent market with continuous changes and changing clients' requirements. By being developed from national and international programs, in cooperation with research and industry, new approaches to organizational change have proven that these new concepts are more than just theoretical frameworks. They have increased the competitiveness of SME = Small and Medium (sized) Enterprises and their effectiveness and efficiency. They have helped handicapped companies to find new solutions to their current problems such as controlling, monitoring, optimizing and supporting business processes by using WFMS.

But it has also become clear that the finishing line has not yet been passed. There are many unsolved problems particularly referring to the special needs of object-oriented, decentralized organizations. There are still more questions than answers when it comes to helping companies to change and re-shape themselves.

The answers can be found neither in isolated "think tanks" nor by the companies themselves. Further intensive national and international cooperation among research and industry is required to develop and transfer findings to industrial practice, so that they are verified under realistic conditions.

REFERENCES

BAHLOW, J., SCHWARZ, M. and SIEFKES, M., 1997, Schnell und kostengünstig zum Markt. Praxisbericht aus einem bereichsübergreifenden Entwicklungsteam. In Brödner, P. Hamburg, I. and Schmidtke, T. (eds) *Informationstechnik für die integrierte, verteilte Produktentwicklung im 21. Jahrhundert* (Gelsenkirchen: Institute for Work and Technology), pp. 13–26.

BRÖDNER, P., HAMBURG, I. and SCHMIDTKE, T., 1997, *Informationstechnik für die integrierte, verteilte Produktentwicklung im 21. Jahrhundert* (Gelsenkirchen: Institute for Work and Technology).

CONDOROVICI, S. and HAMBURG, I., 1998, Implementation of a Petri-nets-based Workflow Procedure. *Else Software*, 1, 17–27.

FRÖHLICH, D. and PEKRUHL, U., 1996, *Direct Participation and Organisational Change – Fashionable but Misunderstood? An Analysis of Recent Research in Europe, Japan and the USA* (Dublin: European Foundation for the Improvement of Living and Working Conditions).

LEHMAN, F. and ORTNER, E., 1998, Die umfassende Bedeutung der Workflow-Management-Technologie. In Paul, H. and Maucher, I. (eds) *Integration von Mensch, Organisation und Technik. Beiträge zum EMISA-Fachgruppentreffen*. GR 1998-04 (Gelsenkirchen: Graue Reihe des IAT), pp. 61–73.

OBERWEIS, A., 1996, *Modellierung und Ausführung von Workflows mit Petri-Netzen* (Stuttgart: B.G. Teubner).

PAUL, H., 1998, Workflow-management systems and organizational development – tools of change? In Karwowski, W. and Goonetilleke, R. (eds) *Manufacturing Agility and Hybrid Automation – II* (Santa Monica: IEA Press), pp. 97–100.

PAUL, H. and PEKRUHL, U., 1997, *Aspects of Integrated Company Renewal* (Gelsenkirchen: Institute for Work and Technology).

SCHEIN, E.H., 1992, *Organizational Culture and Leadership* (San Francisco: Jossey Bass).

VERBAND DER AUTOMOBILINDUSTRIE E.V., 1986, *Sicherung der Qualität vor Serieneinsatz* (2nd ed.), vol. 4: *Qualitätskontrolle in der Automobilindustrie* (Frankfurt a.M.: VDA).

Organizational Culture and Development

E. Järvenpää and E. Eloranta

Helsinki University of Technology, Department of Industrial Engineering and Management, PO Box 9500, FIN-02015 HUT, Finland

1. INTRODUCTION

Today companies look for competitive advantages in several ways. One of the ways is to develop organizational processes and practices to be as effective and efficient as possible. Development of business processes, organizational structures, demand–supply chains and networking with other companies are approaches companies often use to improve their performance. The aim of development activities is to increase productivity, shorten time-to-market, simplify processes, facilitate information and knowledge sharing and at the same time, increase employee well-being.

In organizational development activities, characteristics of the organization to be developed are not always sufficiently taken into account. Development processes are implemented without a deep understanding about the culture of the organization. This means that the values, norms and behavioral patterns the organization members share are neglected. However, these may have influence on how organization members react, accept and participate in the development activities of their organizations.

People who are responsible for organizational development activities should ask before and during the implementation of organizational development activities the following questions. What kind of organization are we developing? What is its culture like? To what kind of culture are employees socialized? The most important question is how organizational culture and its characteristics support or hinder organizational development. In this chapter, the focus is on the characteristics and effects of organizational culture that should be taken into account when designing and implementing organizational development activities.

2. ORGANIZATIONAL CULTURE

The concept organizational culture is used and studied in several disciplines, such as sociology, management and organizational sciences, work and organizational psychology, and anthropology. Each discipline has its own definitions for the concept and at the same time, different interests in organizational culture. In sociology and cultural anthropology, research deals with characteristics of organizational culture, and how culture is expressed and shared. Work and organizational psychology deals with characteristics of organizational culture and recently organizational learning. Management and organization sciences are interested in connections between organizational culture and other characteristics of organization. One goal is to find out if organizational culture affects business performance and, especially, what these effects are.

The concept organizational culture indicates that organizations have their own culture, i.e. that organizations have their own special features that differentiate them from other organizations. A variety of definitions for culture is proposed by different schools of cultural anthropologists and sociologists. In this chapter, the definition of culture by Hofstede (1997) is adopted. According to Hofstede (1997), culture means the collective programming of the mind that distinguishes one group or category of people from another. Culture can be found on different levels, such as national culture, occupational culture or organizational culture.

Organizational culture indicates how people are used to think, act, make decisions and participate in an organization. In everyday language organizational culture means "the way things are done in an organization." The ways the members of the organization are used to think or act as well as how the values and norms of the organization may affect their acceptance of organizational changes and the willingness of the personnel to participate in planning and implementing changes.

Schein (1985) has defined organizational culture as a set of core values, behavioral norms, artifacts and behavioral patterns, which govern the way people in an organization interact with each other and invest energy in their jobs and in the organization at large. For organizational culture, two levels are usually differentiated (Schein 1985). Culture refers to rather permanent values, norms and basic assumptions about the organization. Climate means procedures, practices and behavioral models according to which things are done in an organization. Organizational climate is also indicated by external signs, such as buildings, decoration, logos or how people dress in the organization. Since organizational culture consists of rather permanent values, norms and basic assumptions about the organization, changes take place only very slowly (De Witte and Van Muijen 1994). Therefore, to change organizational culture, one must first change organizational climate.

Argyris and Schön (1978) differentiate espoused and enacted content themes in studying organizational culture. Exposed content themes mean expressed opinions, i.e. what people think, believe or do themselves. Enacted content themes are abstractions that capture aspects of how people actually behave, rather than how they say they behave. This distinction may be useful in understanding the connections between organizational development and organizational culture.

A great variety of models and dimensions of organizational culture and organizational climate are proposed. Quinn (1989) has developed a model of organizational culture, and DeCock *et al*. (1986) a model of organizational climate. Both models describe culture and climate as follows. First, the organization may be focused on internal or external factors. Internal factors refer to personnel and processes, and the organization itself. External factors are defined as contact to environment, e.g. to customers. Second, organizational culture and climate can be linked to control or to flexibility. Dimensions of organizational culture and organizational climate consist of the combinations of these four elements. They are supportive, rule-oriented, goal-oriented and innovative organizational culture. Harrison (1993) has defined four dimensions of organizational culture: role culture, achievement culture, support culture and power culture.

Ekvall (1996) has differentiated organizational culture from organizational climate, and defines organizational climate as the core concept. It is an attribute of the organization, i.e. a combi-

nation of attitudes, feelings and behaviors that characterize life in the organization. Organizational climate exists independently from the perceptions and understandings of the members of the organization. It means the organizational reality in an "objectivistic" sense. This concept corresponds to the definition by Schein (1985), in which climate is regarded as a manifestation of organizational culture. According to Ekvall (1996), dimensions of the organizational climate that stimulate creativity and innovation are challenge, freedom, idea support, trust/openness, dynamism/liveliness, playfulness/humor, debates, conflicts, risk-taking and idea time. These characteristics of organizational culture can also be expected to support organizational development activities.

The culture within an organization can be considered in terms of consistency or diversity. Martin (1992) differentiates three perspectives in research on organizational culture. They are an integration perspective, a differentiation perspective and a fragmentation perspective. In integration perspective, all cultural manifestations are interpreted as consistently reinforcing the same themes, and all members of an organization are viewed as sharing an organization-wide consensus. An organization is seen to share a common culture. A differentiation perspective describes cultural manifestations as sometimes inconsistent. Consensus occurs only within the boundaries of subcultures; yet, these subcultures may be in conflict with each other. Accordingly, an organization may have several different cultures, e.g. different departments may have different cultures. A fragmentation perspective on ambiguity as the essence of organizational culture. Consensus and dis-sensus are seen as issue-specific and constantly fluctuating. An organization has no organization-wide common culture or subcultures. When designing and implementing organizational development activities for the entire organization, the possible subcultures within the organization should be taken into account. This understanding may help in focusing the development efforts to better fit the employees' "mind-set" and views of their organization, job roles and expectations.

3. ORGANIZATIONAL DEVELOPMENT

In this chapter, organizational development means active, goal-directed development and re-design of work processes and organizational structures regarding, e.g. management processes, business processes and demand–supply chains. The goal of organizational development is to improve different kinds of processes or structures of activities of the organization. The scope of the development varies from development of one work process to the change of the entire organizational structure.

Cultural factors and their effects on organizational development have so far not been extensively studied. Hames (1991) has shown the importance of cultural factors in implementing quality programs. Successful transformation to quality was found to need a fundamental remodeling of organizational and management culture. Siehl and Martin (1990) have discussed the complicated relationships between organizational culture and financial performance of organizations.

Contemporary organizational development seems to be holistic and highly business driven compared with the traditional human resources-oriented efforts or organizational development (OD) approach (French and Bell 1973). In this vein business process re-engineering (BPR) has been among the most popular hypes of the decade. This hype started in the consultancy business and progressed first as a major industrial movement and later also as an area of research interest. One of the first definitions of BPR by Hammer and Champy (1993) reflect this pragmatic drive: "BPR is about fundamental rethinking and radical redesign of business processes to achieve dramatic improvements in critical, contemporary measures of performance, such as cost, service and speed."

The academic interests have been diversified primarily towards hermeneutic research work on one hand – what is this phenomenon about (Davenport 1995) and is it worth the efforts? – and towards constructive research work on the other – how can the practitioners be facilitated by appropriate methods and tools for the successful implementation of re-engineered processes (Laakso 1997, Rajala 1997)? The latter builds on the former. The hermeneutic studies have provided us with evidence that BPR projects tend to be risky and the failure rates of the projects are high. The causality, or at least the related factors behind successes/failures, have been elaborated (Grover *et al*. 1995, Childe *et al*. 1996).

What remains still unanswered is the connection between business process development and organizational development. Is organizational development an enabler for business process development or visa versa? Is one of the underlying enablers for organizational and/or business process development appropriate organizational climate building on appropriate organizational culture? These questions are complex and employ a multitude of complex variables. In spite of this, we have tried to shed some vague light on these relevant issues.

Organizational change processes are usually difficult, time-consuming and expensive processes for companies and their personnel (Hammer and Champy 1993). To manage and facilitate change processes in organizations, different approaches, methods and tools have been developed. They include, for example, participatory methods for design and implementation of organizational changes, and simulation games for teaching and simulating new work processes. Moreover, management involvement and support have been found to be very important in successful organizational changes.

By understanding different dimensions of organizational culture, and their level in a particular organization, the organizational development activities can be focused on those issues in organization that are possible and most important for change. Organizational culture may affect practically any aspect in organizational development.

The approach used in developing organizations should be adapted to the organizational culture. In traditional, rule-oriented cultures, the participatory, bottom-up approach with employee participation in development activities, may not be applicable. More top-down approach with rules and guidelines may be more appropriate. At the same time, the employees should be activated step by step to participate in development efforts.

The time frame of the development activities, i.e. how fast changes and improvement can be expected may depend on organizational culture. In traditional, rule-oriented cultures the time frame may be longer than in dynamic, goal-oriented and innovative organizations.

In multinational organizations organizational culture and climate is highly intertwined with different cultural dimensions. Global business organizations do employ virtual teams that

represent multiple nationalities and social cultures, where, for example, Asians, Americans and Europeans are striving towards customer satisfaction by joint efforts, even without the help of co-location. We do have some inductive evidence in our research material that in such conditions it is extremely hard to isolate the impact of organizational culture from the impact of the culture of the surrounding society and nationality. However, at least inductively it looks apparent that the deeper cultural context (i.e. society and nationality) is extremely hard to be bypassed by any organizational culture. Team members representing a rule-oriented society tend to look after detailed guidelines also in a company with organizational culture emphasizing high employee participation and empowerment. Such behavior is amplified in a business process re-engineering situation where the guidelines are in a state of change by definition and where an individual cannot lean on familiar well-practiced rules.

In designing and implementing organizational development activities in global companies it is important to understand the existence of national culture and organizational culture. Our case study showed that organizational culture at five production sites of a global company differed across countries and production sites (Järvenpää and Eloranta 1996). This means that organizational development activities in each production site should be designed locally, and a centralized, headquarter-directed development program may not be applicable. Moreover, organizational culture within each production site differed according to departments, functions and occupational groups. This indicates the need for more tailored development efforts within organizations.

The results of one of the production sites showed that the organizational culture of the company was in general rather balanced, with the main emphasis on rules orientation. The balanced culture means that none of the organizational dimensions measured was dominating. The culture was rather stable, slow responding and rather slow in decision-making. The goals and focus of the company may be unclear, and the employees may not know if they should concentrate on internal aspects, such as following the rules or external issues, such as customer satisfaction. Rules orientation as the dominating value in the organization may have implications for development activities. The personnel may expect that the development process is a top-down process with rules and guidelines. In general, decision-making in a balanced, somewhat rules-oriented organization may be slow and management-driven. The personnel may expect similar decision-making procedures for development processes. Therefore, the personnel should be activated and encouraged to actively participate in development activities.

In a balanced organizational culture, the resistance to changes may be expected. The implementation of and adjustment to new ways of doing things may need much time. Therefore, the approach and tools used in development activities should motivate and activate the personnel. At the same time, the personnel have to get information on the forthcoming changes and how they affect their job. Active management and leadership are needed, together with personnel participation. The personnel need to know the goals of the development process as well as the importance of their participation in the development project.

At the case company, organizational climate was more innovative than culture. This means that the personnel had new ideas even if the innovativeness was not a prevailing value in the organization. Innovativeness is a good basis for development activities. The personnel may have new ideas and motivation to change their jobs and the organization. However, due to socialization to the rules-oriented organizational culture, they may need empowerment, and even training about how to participate and develop their work and organization.

4. CONCLUSIONS

In organizational development activities, the understanding of the current state of the organization is of crucial importance. This means not only the understanding of the goals, mission and visions of the company, but also how people in the organization behave, and the reasons and rationality behind their organizational behavior. The organizational culture and climate indicate how the employees in organizations act in their daily work. For global, leading-edge companies and other organizations, competent personnel is the most important asset. To enable and guarantee optimal working conditions, and development possibilities to their personnel, organizations should understand the meaning of cultural contexts. This is especially true when organizations are in change, due to organizational developmental activities. National culture, organizational culture and occupational culture with their different characteristics affect how people react, become motivated and are willing to change their behavior and their organizational environment.

REFERENCES

ARGYRIS, C. and SCHÖN, D.A., 1978, *Organizational Learning: A Theory of Action Perspective* (Reading, MA: Addition-Wesley).

CHILDE, S., MAULL, R. and MILLS, B., 1996, *UK Experiences in Business Process Reengineering: A Scoping Study for the Innovative Manufacturing Initiative*. Research Report for Grant No. GR/K67328 (Plymouth: School of Computing, University of Plymouth).

DAVENPORT, T., 1995, Business process reengineering: where it's been, where it's going. In V. Grover and W. J. Kettinger (eds) *Business Process Change: Re-engineering Concepts, Methods and Technologies* (Harrisburg: PA Idea Group Publ.).

DE WITTE, K. and VAN MUIJEN, J., 1994, Organizational climate and culture in Europe. A theoretical and practical introduction to the Focus questionnaire. K.U. Leuven and V.U. Amsterdam (unpublished working paper).

EKVALL, G., 1996, Organizational climate for creativity and innovation. *European Journal of Work and Organizational Psychology*, 5, 105–23.

FRENCH, W.L. and BELL, C.H., 1973, *Organizational Development* (Englewood Cliffs: Prentice-Hall).

GROVER, V., LEONG, J., KETTINGER, W. and TENG, J., 1995, The implementation of business process re-engineering. *Journal of Management Information Systems*, 12, 109–44.

HAMES, R.D., 1991, Managing the process of cultural change. *International Journal of Quality and Reliability Management*, 8 14–23.

HARRISON, R., 1993, *Diagnosing Organizational Culture, Trainer's Manual* (Amsterdam: Pfeffer & Co.).

HAMMER, M. and CHAMPY, J., 1993. *Reengineering the Corporation – A Manifesto for Business Revolution* (New York: HarperBusiness).

HOFSTEDE, G., 1997, *Cultures and Organizations – Software of the Mind* (New York: McGraw Hill).

JÄRVENPÄÄ, E. and ELORANTA, E., 1997, Organizational culture and organizational development: a case study in process industry. In Seppälä, P., Luopajärvi, T., Nygård, C.-H. and Mattila, M. (eds) *From Experience*

to Innovation. IEA '97, vol. 1: *Organizational Design and Management* (Helsinki: Finnish Institute of Occupational Health), pp. 76–8.

LAAKSO, T., 1997, A performance evaluation and process innovations – a method for business process development. *Acta Polytechnica Scandinavica, Mathematics, Computing and Management in Engineering Series*, No. 86.

MARTIN, J., 1992, *Cultures in Organizations: Three Perspectives* (New York: Oxford University Press).

QUINN, R.E., 1989, *Beyond Rational Management: Mastering the Paradoxes and Competing Demands of High Performance* (San Francisco: Jossey-Bass).

RAJALA, M, 1997, A framework for customer oriented business process modelling. *Acta Polytechnica Scandinavica, Mathematics, Computing and Management in Engineering Series* No. 89.

SCHEIN, E.G., 1985, *Organizational Culture and Leadership* (San Francisco: Jossey-Bass).

SIEHL, C. and MARTIN, J., 1990, Organizational culture: a key to financial performance? In B. Schneider (ed.) *Organizational Climate and Culture*. (San Francisco: Jossey-Bass), pp. 241–81.

Participation

A.M. de Jong

TNO Work and Employment, PO Box 718, 2130 AS Hoofddorp, The Netherlands

1. INTRODUCTION

In this article theory and practice of participatory ergonomics will be discussed. Participatory ergonomics is a way of improving work in an organization by involving employees and other important parties. The knowledge that technological changes are seldom without a change in organizational aspects implies a more active attitude from management towards improving work in organizations to meet the consequences of introducing changes in an organization (Brown and Hendrick 1992). Four forces can be distinguished that change management practices to a more participative character. First, societal forces come from legislation that protects the rights of employees. Employees feel the need and the right to speak for themselves and to demand that the organization is run in a reasonable and fair way. The second force concerns business environment. Increasing competition and the ever-apparent danger for loss of quality or productivity makes a constant awareness necessary. Third, the type of product has changed to service orientation and jobs that require specialized knowledge. Traditional management styles are aimed at industrial production models and are no longer relevant. Instead, human resources and organizational efficiency are major themes. Finally, the workforce generally has a higher educational level and has higher expectations from work and future careers.

These forces argue for alternative management styles. Participatory ergonomics is one of those styles that can be used in certain situations to manage change processes.

2. THEORY OF PARTICIPATORY ERGONOMICS

2.1. Definition

The results of a round table session concerning participatory ergonomics (Vink *et al.* 1992) showed a lack in theory and categorization of the different varieties of the method. However, all parties agreed on the importance of investigating the theory and use of it, since the need for participation constantly occurs in history and modern life.

Some definitions of the methods have been determined in literature. First a definition of the method will be proposed. Noro and Imada (1991: 3) proposed a first definition of the method: "A new technology for disseminating ergonomic information." However, this definition does not specify the amount of influence that an employee has during the process. Also, the definition emphasizes the use of ergonomics, though this is merely a means to achieve a goal, which itself is much more important.

Wilson (1995: 331) recently defined participation within an ergonomics management program: "The involvement of people in planning and controlling a significant amount of their own work activities, with sufficient knowledge and power to influence both processes and outcomes in order to achieve desirable goals." In his definition, Wilson emphasizes the importance of power of the employee. Participation in development can only be successful if the participative parties have a share both in the decision-

making process as in the selection process of the outcome of the project. Furthermore, the amount of participation is mentioned in the word "significant," which sets the boundaries for the participation. Participation is a means, not a goal. The goal of participative ergonomic projects is to improve work activities through joint action and decision-making.

2.2. Approaches

Here some basic approaches of participative ergonomics will be distinguished. Every project initiator will have to decide what type and amount of participation is best suited for a specific project or situation. Employees can be directly involved in a project, or indirectly through representatives. These are two extremes; there is also another alternative, in which employees are directly involved, but not in all stages of the process and decision-making is done by management. The approaches are (Noro and Imada 1991):

- Consultative participation: employees are merely consulted to collect necessary information but they are not present at meetings or decision-making.
- Substantive participation: employees are organized in work teams that delegates a representative for participative problem solving and decision-making.
- Representative participation: employees are part of the management team and are present at meetings and decision-making.

This classification of approaches shows three standard theoretical approaches with varying type and amount of participation. The approaches consist of properties that may vary on absolute or continuous scale. To determine the exact composition of the standard approaches, the following three properties of participation have been defined (Dachler and Wilpert 1978):

- Formal–informal: "a system of rules … imposed on or granted to the organization" as opposed to "nonstatutory consensus emerging among interacting members".
- Direct–indirect: "immediate personal involvement of organization members" as opposed to "some form of employee representation".
- Levels of decision-making; 6 levels were defined ranging from "no information given to employees about a decision" to "decision is completely in the hands of the employees."

If one defines the properties for the standard approaches the following definitions appear:

- Consultative: informal–indirect no decision-making.
- Substantive: formal–indirect decision-making to some extent.
- Representative: formal–direct–full decision-making.

However, participative projects do not have to use a standard participative approach, but may vary the properties to suit the project and situation.

2.3. Benefits and Disadvantages

The method of participatory ergonomics can be used for many reasons. For instance, the involvement of people may increase the acceptance of the results. But participation also implies more efforts of the project organization to execute the project. Therefore, benefits and disadvantages of the project must be carefully weighed before starting a participative project.

Gjessing *et al.* (1994) defined the benefits of participation as follows:

- Enhanced worker motivation/job satisfaction: not only extra salary will improve one's satisfaction over work, but also even more the opportunities to influence your own work.
- Added problem-solving capabilities: the person doing the job has the best knowledge of work methods, resulting in higher quality of the result.
- Greater acceptance of change: a better understanding of the needs for change and opportunities to help structure the change may improve commitment to successful implementation.
- Greater knowledge of work organization: not only knowledge of one's own work, but also the relation to the overall company operation can be improved resulting in improved communication and coordination among members of the organization.

Some case studies have been performed to assess the potential benefits and disadvantages of the method. Although most cases report on good results, other authors start emphasizing the risk of failures (Wilson 1991). These failures can appear at three levels: (1) the results were not achieved for production, (2) working conditions were not actually improved and (3) participants had no benefit from participation.

In spite of these potential failures, the method may still be successfully applied if an approach is chosen which is adjusted to the project characteristics.

2.4. Effectiveness

In literature a broad discussion on the effects of the method of participatory ergonomics can be found. Since innovative projects are depended of many variables, it is difficult to point the results of the projects to participation alone. There are many reasons to explain the positive effect of participatory ergonomics on work. For instance, attention which is paid to employees by management can have positive effects on performance (e.g. Hawthorne experiments). However, other reasons focus on a more cooperative and self-consciousness image of employees. They stress the improvement of quality of result or even more improvement of the implementation process as a result of the commitment of employees.

Furthermore, the effects of the method are difficult to assess since the projects often last a couple of years and changes can only be measured after the first start-up problems.

However, this section will give a resume of the discussion that exists in literature on the effectiveness of the method. Most authors agree on the fact that the way a change process is organized is by far the most important cause of the effectiveness of the project (e.g. Vink *et al.* 1992, Haims and Carayon 1996). Research on the effects of different approaches, as discussed above, of organizing the process showed that indeed some approaches resulted in better effects than others.

Relationships were shown between participation and satisfaction of the user of the result. However, other research proved that these relationships showed only small significant relationships between participation and satisfaction. These small relationships are based though on change processes in general and not exclusively on ergonomic changes. Because of the applied nature of ergonomic changes positive effects on satisfaction are more likely to occur for this type of processes.

More significant relationships might exist between participation and the quality of the results. These relationships should therefore be determined for future participative projects to assess the effectiveness of the method.

3. PARTICIPATORY ERGONOMICS IN PRACTICE

This section gives an inventory of a number of participatory ergonomic methods that are used in practice. The methods in the following overview differ in their set-up and whether the change is of a continuous kind or not. They will be mentioned hereafter alphabetically without further classification. Only publications that define the method used to sufficient extent will be taken into account in this overview. For the specifics of each method the reader is pointed to the references given in each section.

3.1. Ergonomics Coordinator Program (Haims and Carayon 1996)

This is a permanent, internal participatory program that involves increased levels of control, acquisition of knowledge and skill, and developmental growth for individuals involved on a continuos and cumulative basis. This makes outside experts gradually abundant in time as the involved individuals are capable of continuing to program themselves. This transfer of control can be related to higher levels of instrumental, conceptual and organizational job control.

3.2. Health Trak (Moir et al. 1997)

Health Trak combines rapid appraisal (popular education and participative techniques) and participatory ergonomics to perform interventions in the construction industry. This method is developed around 8-week cycles, which utilize three major components. The first components is divided in four steps: (1) problem identification and definition (weeks 1–2), (2) analysis of hazards and strategies for intervention (weeks 3–5), (3) implementation of proposed interventions (weeks 6–7) and (4) evaluation by participants (week 8).

3.3. Participatieve Ergonomie (Vink et al. 1997)

The essentials of the approach are a strong commitment of the management and a direct worker participation as well as continuous feedback to participants and summary feedback after each step. The method has been used in a number of projects to reduce the physical strain of workers in a variety of trades. All projects use six main steps: (1) preparation: initiation and strategy of the project, (2) analysis of work and health: definition of user requirements, (3) selection of improvements: development and selection of solutions, (4) pilot study: testing and adjusting solutions, (5) implementation: applying solutions to work situations and (6) evaluation: evaluating results and process.

3.4. Participatory Ergonomics (Noro and Imada 1991)

This method is the first as such defined for participatory ergonomics, which is a technology for disseminating ergonomic information. The method is designed to be applied by non-experts, using approaches and methods which differ from ergonomics intended for experts. A number of steps have been

determined to be used by small groups: (1) select theme, (2) set goal, (3) grasp present situation and analyze factors, (4) identify problem, (5) develop and improve measures to solve problem and (6) confirm effect of measure taken.

3.5. Problem Solving Group (Wilson 1995)

This program is based upon the Design Decision Group (reported by Wilson) and has seven stages that should finally result in continual improvement after change. The seven stages are: (1) familiarization with tasks, jobs, workplace and team, (2) field visits to similar sites, (3) Design Decision Group (A); drawing, discussion and idea generation, (4) Design Decision Group (B); layout building, discussion and idea testing; (5) lighting and visibility simulation at the site, (6) sourcing and costing of solutions by participants and (7) continual improvement after the change.

3.6. System Group Development Technique (Andersson 1990)

A systems approach, based on the product development process, adjusted to the marketing system and the "initial stages" of the product development process. Participants are drawn from the entire marketing system representing all major functions, such as manufacturing and distribution. The stages are: (1) need session; from ergonomic problem to defined need, (2) idea session; from defined need to alternative solutions, (3) screening session; from alternative solutions to selected concept, (4) concept session; from selected concept to defined concept, (5) design session; from defined concept to design options.

3.7. Tuttava (Laitinen et al. 1997)

This method was designed to support the creation of a participatory ergonomics process intended for continuous use in an organization. First an ergonomic knowledge package was presented to participants including a method manual with among others the description of a dialogue model. This model can be used to structure the dialogue between operators, manufacturing engineers and suppliers, which should result in a proposal for change.

Literature reports on several methods of participatory ergonomics on the basis of case studies. Most methods are basically the same and distinguish themselves only in the number of phases or the goal to be achieved. It seems that in a national context, every country has developed its own way of applying participatory ergonomics and no international standard exists.

In conclusion, can be said that there are a number of aspects in which the methods show some similarity. These are:

- A systematic (stepwise) approach.
- Some form of participation.
- Ergonomic analytic elements.

The selection of a method for a project will be based on the availability on information of the specifics of a method and the suitability of a project for participation. Furthermore, positive experiences of the project initiator will increase the chances of the application of a participative method.

4. CONCLUSIONS

In short, theory on participatory ergonomics has been reviewed to show the experiences with the method from literature. Participatory ergonomics is one of many methods to successfully implement ergonomic changes in organizations to improve work by involving employees. The method is merely a means to achieve a goal and not a goal in itself.

Major advantages of the method are better quality of the innovation and more satisfied employees and major disadvantages are chance of failures and more efforts of management and project organizers needed. The advantages have to be greater than the disadvantages the justify the use of the method. The effectiveness of the method has not yet been proven in literature, but experiences show that especially in the ergonomic context the method will show good results. A number of participatory ergonomics methods are available internationally, which have a number of similar aspects. Namely, a systematic (stepwise) approach, some form of participation and the use of ergonomic analytic elements. The selection of a method depends on the project characteristics and the availability of information on the specifics of the method.

Nowadays, organizations have to innovate to keep up with expectations of their environment. Participatory ergonomics is a way to optimize these innovation process in order to improve the chances of success of the innovation projects.

REFERENCES

ANDERSSON, E.R., 1990, A systems approach to product design and development. *International Journal of Industrial Ergonomics*, 6, 1–8.

BROWN, O. and HENDRICK, H.W., 1992.

DACHLER, H.P. and WILPERT, B., 1978, Conceptual dimensions and boundaries of participation in organizations: a critical evaluation. *Administrative Science Quarterly*, 23, 1–39.

GJESSING *et al.*, 1994.

HAIMS, M.C. and CARAYON, P., 1996, Implementation of an "in-house" participatory ergonomics program: a case study in a public service organization. In O. Brown (ed.) *Human Factors in Organizational Design and Management – V*, 175–80.

LAITINEN, L. *et al.*, 1997, Initiating an innovative change process for improved working conditions and ergonomics with participation and performance feedback: a case study in an engineering workshop. *International Journal of Industrial Ergonomics*, 19, 299–305.

MOIR, S. *et al.*, 1997. Health Trak: a participatory model for intervention in construction. In Proceedings of the 13th IEA Conference, Finland, 151–3.

NORO, K. and IMADA, A., 1991, *Participatory Ergonomics* (London: Taylor & Francis).

VINK, P. *et al.*, 1992, Experiences in participatory ergonomics: results of a roundtable session during the 11th IEA Congress, Paris, July 1991. *Ergonomics*, 35, 123–7.

VINK, P. *et al.*, 1997, A participatory ergonomics approach to redesign work of scaffolders. *Safety Science*, 26, 75–85.

WILSON, J.R., 1995, Solution ownership in participative work redesign: the case of a crane control room. *International Journal of Industrial Ergonomics*, 15, 329–44.

Participation and Collaboration in Workplace Design

M. Launis

Finnish Institute of Occupational Health, Unit of Ergonomics, Topeliuksenk. 41 a A, FIN-00250 Helsinki, Finland

1. INTRODUCTION

1.1. Reasons for User Participation in Designing

Involvement of the operators and users in design and development activities has become increasingly important in organizations, in order to improve the quality of production and services, to increase the flexibility of their functions, and to prevent disturbances in systems performance. The enhancement of participation is not only a means for achieving direct gains in the efficiency of major activities, but also a means to improve production technology, the work environment, work content and the employees' well-being.

The idea of the users participating in the designing is not new, but only during the past two decades have the overall preconditions in organizations evolved sufficiently to apply it on a wider scale. Participation in designing is not only a question of design methodology, but also it is always an organizational and psychological issue, and strongly affected by the work culture. The present trends in organizations, such as a flattening of hierarchies, the establishment of work teams and the expansion of work tasks and responsibilities, support greater participation, as does the general rise in the educational level of the employees. In addition, the production technology and processes have become more complex, requiring the utilization of all available knowledge and experience within the organization.

1.2. Participative Design Approaches

At present there are several well-established design approaches in which participation is an integrated element. Some of them are aimed to develop production activities, such as quality circles and continuous improvement, some to advance products or systems, e.g. user-centered design and usability testing, and, more emphatically to improve jobs and working conditions, the participatory ergonomics approach. In addition to these, the concept of participative design can be mentioned as a comprehensive term for the basic idea of users' participation in designing. In principle, all these approaches are parallel in aiming at the improvement of activities, equipment and conditions of use by involving operators and users in the development processes.

For more information on participatory ergonomics, see Noro and Imada (1991), and Wilson and Haines (1997).

1.3. Towards a Broader Collaboration in Designing

Development tasks, as a part of the operators' daily activities, such as improving the work tasks in a production team, or as an operator of a complex machine, seem natural expansions of these activities. Workplace design is, however, to a great extent part of an isolated and fragmented process of designing production technology, led by technically oriented designers, and participation in this hasty process does not seems obvious. In the technical design process, the work tasks and working conditions are largely determined, but likely to stem from different starting points and values than those of the people acting in the production. The designers are responsible, first of all, for the technical functioning of the systems at the lowest costs. Thus, in the rigor of the design process the consideration of work activities remains secondary.

The integration of ergonomic impact throughout this design process requires, in addition to user participation, the contribution of many other personnel groups and interest groups in the organization, e.g. that of supervision, of maintenance department, of safety and health specialists, of the human resource management department, etc. Seen from a common viewpoint, it is rather a question of collaboration in the design process instead of participation, which, as a concept, refers to unequal roles of the different actors in the process.

1.4. Aim

The aim of this chapter is to give an overview of the arguments for and against participation in workplace design, of its application, of criteria for the methods and tools, and of preconditions for carrying it out.

2. ARGUMENTS FOR AND AGAINST GREATER USER PARTICIPATION IN WORKPLACE DESIGN

2.1. General

From the ergonomic and human factors point of view there are several arguments in favor of the users' participation in workplace design. Some of them are associated with the results of designing, some with the expected gains in the users' qualification and motivation, and some with the functioning of the entire work organization. When interacting with each other, these effects, if successful, likely cumulate the total advantages of participation.

However, participation is not always beneficial, neither for the outcome nor for the participants themselves. Nor can it be applied in all circumstances. An overall negative aspect of participation can be an obscuring, or even rejection of the role of expert knowledge in the process. The participants' views may be limited to the visible and immediate everyday experience, and seeing the future demands, or the entirety of the design (e.g. in designing entirely new objects or in the development of large systems), may additionally require other forms of expertise.

2.2. Utilization of the Users' Knowledge and Experience

Utilization of the users' experience and knowledge in the activities at the workplace is obviously the first argument for participation. Presently, in complex work processes, it is ever more likely that the user knows best the problems in task execution or the disturbances in the performance of the system. In particular, the user knows his/her own work activity in detail, i.e. the operations and functioning of technical solutions which are potential objects of ergonomic improvements. Furthermore, the load, and difficulty or ease of the work performance experienced by the participants constitute the human criteria for these improvements.

2.3. Accommodation of the Design to the Individual

In the participative design process, the work activity and the design can be accommodated to fit the characteristics of individuals or of a specific user group. In tests and evaluations, the demands of the actual tasks are proportioned to the actual users' capabilities, the result of which is beneficial for individuals who are in some respect physically less capable compared with the average, e.g. for women and aging users. Similarly, the users' task-related expert competence is taken into account.

On the other hand, designing for individuals or small groups can lead to generally inappropriate solutions as regards essential variations within the user population, e.g. in respect to the variation in qualifications, anthropometric dimensions or muscle strength. Also over-motivated participants may tend to demonstrate their capabilities in the design sessions, and this may lead to solutions which are not suitable for the others or inappropriate in the long run.

2.4. Expansive Learning of all Concerned

In participative and collaborative design sessions a lot of essential information is naturally being transmitted between the users and designers, and other groups concerned. The users bring their experience and feedback regarding usage to the designers, and the designers provide understanding of the system's way of operation. This ever-growing general understanding helps the designers to design more operable systems, and the operators to operate the systems more efficiently and reliably.

2.5. Fulfillment of Individual Needs

Well-trained people want to utilize their skills and capabilities and to experience self-realization in their work; participation offers a channel to achieve these. People also generally want to have more control over their work tasks, and to take part in molding their work environment. Apart from the functional aspects, there may also be aesthetic or self-determining needs concerning the environment, and they become evident, for example, when making the work site home-like or to reflect the personality of the user.

2.6. Achieving Commitment and Acceptance

One of the greatest benefits of users' participation is the feeling of ownership regarding the solutions, and the resulting increase in motivation and commitment to the usage. This basically positive consequence involves ethical questions of participation, too, especially if the outcome of the design is not expected to be beneficial for the user, or if it does not entail any positive change compared with the earlier design. If the only purpose of participation is the manipulation of users, it is evidently questionable, and forms an obstacle for later initiatives to involve people in these activities.

2.7. Effects on Organizational Behavior

Increasing the skills of collaboration, added to improved comprehension of the goals and development processes very likely contributes to the functioning of the entire organization. On the other hand, the participation of selected persons or groups may cause envy among those who cannot participate. The participative process may also be unsuccessful, with no useful results, or results

that may be unfavorable to some individuals. All such reasons can potentially cause frustration among the employees.

2.8. Integration of Different Aims

Generally participation serves many purposes at the same time. All in all, the participants assess their workplaces as functional entities, and it may be even difficult to separate e.g. productional and health-related aspects from each other. As a strategic notion, it may be easier to motivate the management to support participation with all-inclusive gains as compared with purely health-oriented or humanistic ergonomic interventions. This situation may lead to the misuse of participation, however, if the designers or management utilize the multiple outcome of participation at the expense of the users' well-being.

3. LEVELS AND MODES OF PARTICIPATION

3.1. General

When participation is applied in designing, it should be made clear to all the parties involved what is the object of design, the structure of the design process and design organization, and the goals and limits of participation. Participative actions can be organized by different groups, e.g. designers, the management, occupational health and safety personnel, or the participants themselves. Participation can take place as a separate project, a part of an ergonomics program, incorporated in continuous improvement activities, or, as part of design projects.

Important issues of participation in these respects are the questions of who should participate, and what are the participants' areas of influence, decision-making power, and responsibility concerning the results. It is also important to find useful and feasible ways of action for different design settings and purposes. These questions need to be solved locally, by mutual consultations of the parties in question, and preferably preceded by the assessment of the overall situation in the organization (e.g. workplace atmosphere, cultural environment, relations between the management and unions, threats of downsizing, etc.).

3.2. Representative versus Direct Participation

In indirect, representative participation only authorized representatives (e.g. shop steward, safety delegate of the union, team leader) of the users participate, in direct participation the actual user of the workplace being improved participates. In some countries it is customary for the union representatives to participate in the meetings of larger workplace design projects. They are not necessarily acquainted with the object of design, however, and their concern is, to a great extent, limited to the general and legally determined aspects of design. As regards the basic principles of participation described earlier, only the direct mode could be considered "genuine" participation. However, this mode is more vulnerable to misuse as there is less control from the union representatives.

3.3. Influence, Power and Responsibilities of the Participants

How determining or influential the participants' role in designing is, varies between two extremes in the different modes of participation. The participants may act merely as sources of information or test subjects, having less power to make decisions concerning the outcome. They may be members of the design

team, sharing the decision-making on some level or, form a separate participative team making analyses and generating solutions or suggestions for the designers, normally within prescribed limits, however. At the other end, the users are designers themselves, which is often true in the designing of personal office environments in ready-made components.

Regardless of their decision-making power, the participants have an opportunity to influence the outcome greatly, whenever they can freely express themselves. When they are given power and responsibility, the question of compensation will arise, which is still a largely unsolved issue of participative design.

3.4. Participation Varies According to Object of Design

The appropriate ways to participate or collaborate depend greatly on what is being designed. The object of design may be a large entity or only a small part of it, or, an entirely new design or merely a correction to an existing design. Also the complexity of the work place (geometric, structural, functional) and its relation to the entire work system dictates the design approach and the associated design methodology, and sets requirements or limitations for involving users in the process. For example, the design objects of work places vary from simple furniture and work equipment, allowing easy or even improvised participation, to complex systems requiring the application of special methods with careful guidance.

3.5. Modes of Participation in Different Stages of Design

The contribution of the users and other groups concerned is needed in all those stages of the design process in which the characteristics of the work tasks and workplaces are determined. Such stages are e.g. gathering and transmitting knowledge and information, setting objectives, making analyses and evaluations, preliminary sketching of the design, assessing drawings, building models and prototypes, assessing them, trying out and testing the designs or completed workplaces, or gathering feedback from ready workplaces and transmitting it to the designers.

There are several ways to participate and collaborate in these stages of the design projects, e.g. in design sessions, formal meetings, reviews and audits. Participation in the project requires open flow of information regarding the schedules of the projects, and the object of design at each time, so that the participants (or collaborators) know when and how to contribute. Also it must be made clear to the designers at the beginning of the projects, who can potentially contribute usefully to each design task. The designers can also develop their practices and methods so as to facilitate and encourage greater participation and collaboration (Launis *et al.* 1996).

In the design projects, new work systems are usually constructed, but still large parts of them will be mere modifications of earlier designs. The human activity will not change entirely, either, and therefore it is possible to convey earlier experiences to new designs through participation and collaboration. In addition to careful analyses of the earlier designs and associated work activities, also "reference work sites" similar in some important aspects, can be visited and analyzed fully to utilize the existing experience in the new designs (Daniellou and Garrigou. 1992).

4. THE NATURE OF USER KNOWLEDGE

The conception of the users' knowledge is the essence of participative design. This knowledge is based mainly on experience, and often can not be expressed as concepts, not even verbally. This tacit knowledge can not always be recalled, either, and therefore, e.g., actual walk-throughs of the work processes at the workplace are needed for processing and voicing it. Also the users' mental models of the functioning of the technical systems arise from experience, and therefore they differ considerably from those of the designers.

5. REQUIREMENTS FOR THE METHODS OF PARTICIPATION

5.1. General

The diversity of the above-mentioned conditions of participation makes it difficult to establish common procedures or even methodology for participative designing. As a matter of fact, in participative design, many well-known design methods are applied in a way that takes into account the nature of the participants and the requirements for the participative design situations. From this point of view, some criteria on the procedures and methods of participation will be presented in the following sections.

5.2. Clarification of the Basic Concepts

The methods used must help to clarify the concepts concerning the work process and functions of the work system. This is a prerequisite for ensuring understanding and fruitful discussion between the designers and the participants. The designing of more complex systems requires also common conceptual models of all functions and their relationships, linked to the associated human activity. This can be done in joint discussions by describing verbally the work processes, the purpose of each task and technical element, and the activities performed by the operators. The use of the wallboard technique can be appropriate for generating such a representation of a larger work system.

5.3. Making Invisible Visible

Work systems and work processes have to be illustrated clearly, especially if they are not obvious or visible, e.g. chemical processes or information systems. Schemes and diagrams may well serve this purpose. Walk-throughs and task simulations are effective in persuading the users to bring out their hidden knowledge. Sketches of the workplace and simulation of layout with cardboard models are well known illustrative means. Computer-aided design with human models, as well as virtual reality technology, can offer means of illustrating designs of the physical environment for purposes of participation, but this requires additional effort from the designers, which is not always feasible in single-occasion workplace design.

5.4. Focus on Interaction between Work Activity and Technology

The emphasis in participative design is on work activity, which is the expertise area of the participants, but the focus must be shifted also to the assessment of the associated technical solutions in their use. Methods of ergonomics for analyzing and restructuring work tasks are useful, but they need to be complemented with user trials and tests of the models, mock-

ups and prototypes. In respect to this main consideration, purely technical aspects, as well as such human aspects for which clear criteria and data have been established, e.g. environmental hygiene, anthropometric variability, or maximum recommended weight limits, may be beyond the scope of participative designing.

5.5. Aim at Consensus Decisions

The users' views and values vary from those of the technical personnel, and in order to achieve appropriate compromises, methods of dealing with different suggestions are needed. Again, the wallboard technique can be applied for gathering ideas, for discussing them, for combining ideas, and for rating alternatives. The resulting group decisions are beneficial for overall acceptance of the design, but they involve the risk of neglecting important factors (e.g. concerning safety) which are not known or only poorly understood in the group. A common understanding of the design goals and the importance of various factors is one prerequisite for a successful group decision process.

5.6. Ease and Efficiency

The limited time provided for participation and inexperience of the participants in designing requires that the methods must be exceptionally easy and the threshold of starting the work must be low. Methods requiring much documentation and verbal formulation are less suitable for participative designing. Specific methods intended for participative inquiries, such as mind-mapping and round robin questionnaire, are examples of cost-effective means. Also ready-made tools supporting designing, such as agendas for meetings, forms for gathering information, checklists and other clear tools can be efficient in participative designing.

6. PRECONDITIONS OF PARTICIPATION

6.1. General

In addition to favorable overall preconditions of participative designing (support from parallel activities, higher level of education of the workforce, democratization of work life, etc.), some specific requirements are apparent. User participation is a voluntary extension of people's daily activities, and thus it is sensitive to both positive (support, motivation) and negative influences (resistance, territorial behavior) prevailing in the organization.

6.2. Commitment of those Concerned

Participation cannot be successful unless all the groups involved in the events accept it. Understandably, however, for several reasons, it is not possible to motivate everybody in these activities. Extra effort, unsuitable timing, or threat of losing power are potential reasons for a negative attitude towards participation. The management and the unions have a decisive role in supporting the basic idea of applying participation in workplace design.

6.3. Consideration of Cultural Differences

In organizations, different groups have adopted different values and principles guiding their action; these may severely hamper the development of workplace design. Designers aim at neat technology with the least use of human work, which makes it difficult for them to consider work activity throughout the design process. The top management is involved with financial survival and marketing processes, and they may experience the development of production activities as remote (Schein 1996). Workplace design, and collaborative activities in particular, must therefore prove their beneficially to the decision makers. A practical strategy that has proven useful is to start from a limited area or project and to spread out to other areas after trying out the methods and obtaining good results.

6.4. Provision of Resources

Participation cannot be put into practice without allocating sufficient time for it. This is important especially when involving persons from production duties in the participation. In most cases participation means costs; this stresses the importance of efficient methods. Also extra effort may be needed on the part of the designers and supervisors to take care of additional actions.

6.5. Guidance

Participative design actions need careful planning, organizing, guidance and many other forms of support. The haste inherent in design activities obviously prevents the designers from concentrating on these time-consuming activities. Participation requires the purposeful contribution of a designated person, a "facilitator" or "internal consultant." The qualifications of such a person are crucially important for the success of participative designing.

7. CONCLUDING REMARKS

Participation in designing has been considered an "easy way" of applying ergonomics in work organizations, because it seems, at least to some extent, to replace the need to employ ergonomics or human factors expertise in the designing. Experience shows, however, that participation can be a fairly laborious way to proceed, but it nevertheless has many positive effects that go beyond the design output itself. So far this way of promoting conventional technologically oriented designing may have been afflicted by start-up problems, but growing demands and the integration of similar trends are creating pressure to make also designing a more collaborative attempt. The evident goal is a design practice and culture, which incorporates participation and collaboration in continuous activities naturally and unnoticeably.

REFERENCES

DANIELLOU, F. and GARRIGOU, A., 1992, Human factors in design: sociotechnics or ergonomics? In M. Helander and M. Nagamachi (eds) *Design for Manufacturability — A Systems Approach to Concurrent Engineering and Ergonomics* (London: Taylor & Francis), 55–63.

LAUNIS, M., VUORI, M. and LEHTELÄ, J., 1996, Who is the workplace designer? Towards a collaborative mode of action. *International Journal of Industrial Ergonomics*, 17, 331–41.

NORO, K. and IMADA, A.S. (eds), 1991, *Participatory Ergonomics* (London: Taylor & Francis).

SCHEIN, E.H., 1996, Three cultures of management: the key to organizational learning in the 21st century, Working Paper 10.011, MIT Organizational Learning Center.

WILSON, J.R. and HAINES, H.M., 1997, Participatory ergonomics. In G. Salvendy (ed.) *Handbook of Human Factors and Ergonomics* (New York: Wiley), 490–513.

Participation of Users in Architectural Design

J. Å. Granath

Department of Architectural Design, Chalmers University of Technology,
SE-412 96 Gothenburg, Sweden

1. INTRODUCTION

1.1. The Scandinavian Experience

Participatory design in architecture has been practiced in the Scandinavian countries for three decades. Such design has different stages for user participation, and each stage is treated here separately. In real life one does not find clear distinctions between these stages of participation since they often overlap. Nevertheless, one can discern a sort of evolution of the notion of participation in Scandinavia from what the present author calls "power-oriented" to more "knowledge-oriented" processes. In terms of the outcome, one can discern a shift from an "object-oriented" to a more "process-oriented" view. Simultaneously, one can discern a global movement from "producer orientation" towards "customer orientation" that has put participatory design on the agenda today.

1.2. Users

The term *users* denotes those who actually use the building in their everyday activities. In this sense all people working in a building including staff, management and service personnel are users. A kind of user not included in the *user category* here is those who are some sort of visitors or use the building as a part of a service provided in the building. Such groups are students, patients and visitors. Groups also excluded here, if they are not users in the above sense, are owners, politicians, union representatives and public officials. As representatives for important groups their participation is important and their roles are, therefore, discussed in relation to user participation.

1.3. Inherent Tension in the Architect's Profession

The attitude to user participation is ambiguous among architects. Architecture and the profession of architecture embody both an artistic dimension and a social dimension. The artistic dimension can sometimes inhibit users from involvement in the design process of architecture. This is a result of the conception that art is a private and not a collective activity. On the other hand, the social dimension of architecture and the social visions of the profession of architecture encourage architects constantly to try new methods to involve users in the design activity so that the resultant architectural artifacts might attain a more appropriate and effective design.

2. PARTICIPATION IN DESIGN AND PARTICIPATION THROUGH DESIGN

Participation of users in the design process can be interpreted in two different ways:

- To design architecture in such a way that it supports partici-

pation in the use of architecture – participation through design.
- Participation of users in the actual design process – participation in design.

2.1. Participation through Design

Architecture in itself can, however, prevent or support participation from users. This has to do with accessibility, understanding, appropriation and ownership. It is important to understand that the built environment is designed both by expert designers and by us. We use it and live in it and it is by doing so that we design the environment.

2.2. Participation in Design

Participation of users in the design process could be discussed from different points of departure. In architectural practices and research in Scandinavia there have been three stages of development in the participation process motivated by democracy (Figure 1), quality (Figure 2) of the product and improvement (Figure 3) of the client groups through learning.

Participation of users in the design process has developed over the past three decades. In the late 1960s Professor Johannes Olivegren was one of the pioneers in the field of housing design. His involvement in user participation had mainly democratic motives but he was also searching for more appropriate housing design. The process of user participation has not been linear but there is, at least in Scandinavia, a major line from participation as driven by democratic reasons, through quality-driven, towards participation for organizational improvement. If one looks at single design projects, however, one can find early, single instances of all three kinds, and at the moment one sees participation of users as a way of achieving quality in terms of all three aspects.

2.2.1. Benefits of participation

Democratic involvement is in itself an important factor in society. The main reason for setting up legislative procedures in the
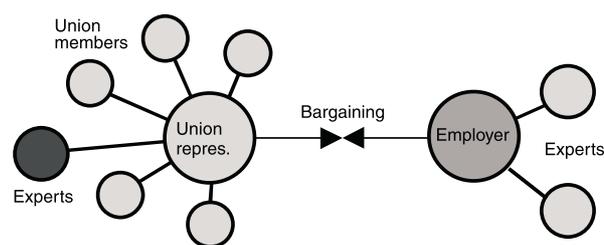

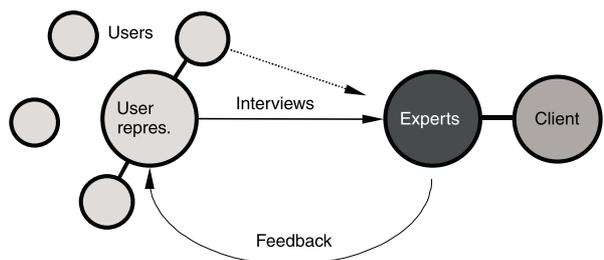**Figure 1. Bargaining model of communication.**

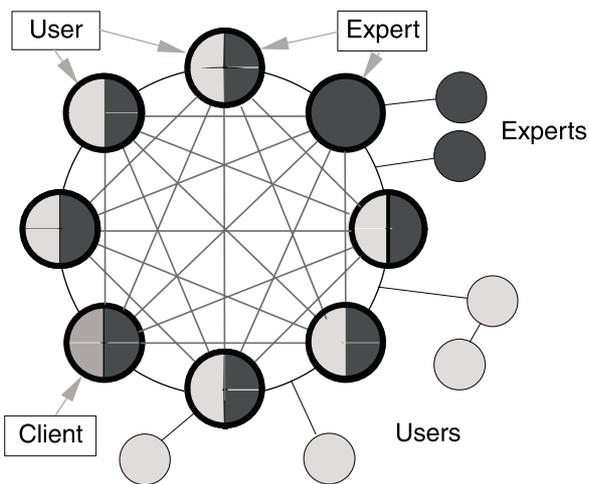**Figure 2. Data collection model of communication.**

Figure 3. Collective learning model of communication.

1970s was to ensure that the basic work environment quality was met in the design of work places. This was an improvement in terms of original goals as early 1970s participation praxis had attempted to turn participation into a matter of power instead of an activity that added value to the outcome of the design. The last step in the development of the notion of participation is to improve the performance of the user groups. One can find explanations for endeavor in the fact that information society has today come to see the employee as a company's most valuable resource; this in contrast to the view of the employees in the 1960s and 1970s. In the information society context, the participatory design perceives the employees as a more dedicated, more knowledgeable group. Employees are seen as capable of designing, re-designing and managing production recourses such as the built space they occupy. Participatory design has become a step on the way to a learning organization. A participatory process has developed where the user through *participation in design* can achieve a design solution that supports *participation through design*; on the way they construct knowledge that makes them able to take active part in the re-design and management of the designed environment as demands change.

3. PARTICIPATION AND DEMOCRACY

3.1. Importance of the Context

Two contexts that somehow overlap will now be studied: direct participation and participation through representatives. Typical for the first context is that the stakeholders are well defined. In the second context, a number of stakeholders is involved, with different possibilities to act. The first context is represented by the workplace design in relatively small or local private companies. The design of municipal buildings and workplaces of large companies, where the decision-making process is centralized, represents the other. The reason for distinguishing between these two contexts is the democratic systems that work within them. In private workplace design, one is mainly concerned with two parties: the employer and the employee.

In municipal building design projects it is often more complicated. Here there are politicians who represent the people in society but they might also be the owners and financiers. There

is also the manager of the activity that takes place in the building and the employee. These two last groups are the actual users of the building. In most cases, such as for hospitals, schools and other public buildings, there is a third group, namely the patients, students or visitors, who is also users in a different meaning. In many cases the actual users are not yet employed when the design takes place and those who will visit the building when it is finished is not a well-defined group. Therefore, the participation mostly becomes a matter of formalized democracy involving politicians and municipal officials in public building or representatives for managers and employee from the corporate headquarter in private companies.

In practice it has proven difficult to deal with participation by the actual users in this second context. There are, however, cases both in municipal building design projects and private industrial workplace design where the actual users have taken active part in briefing and designing the buildings. In many such cases there exists an ambiguity for the designing architects concerning the input from the actual users and that from the representatives of different groups.

3.1.1. Shortcoming of a formal participation process

In formal participatory workplace design, participation takes place through union representatives, the counterpart being the employer. (*Formal* here means a participation process that fulfils the procedures in the legislation.) The crucial point in this setting is the overall relationship between the two parties. In many cases where there is a fruitful the relation, the design process has also been rewarding to both parties. In other cases the participation in the design process has been a matter of power (Figure 1).

In municipal building design projects a formal participation process might be sufficient from a political point of view. Direct user participation in this context is more difficult than it is in private companies if one is talking about participation as a way to real influence on the outcome. To accommodate direct user participation, an adaptive design process has been developed. This means that the architects base design proposals on information from elected representatives. These proposals are then presented to the users, if there are any, or to people in the community on specific occasions throughout the design process so that they may have the opportunity to comment on the presented proposal. The architects can then adapt the design to accommodate these comments – if these do not have a major impact on cost and time or other policy decisions made.

One of the shortcomings of this participatory process is that the users seldom get involved in the project early enough to have a chance to influence the conceptual design phase. They may only suggest detail changes for a more or less fixed design. This entails that the more important conceptual input from the users is seen as negative from the point of view of architects and elected representatives since it affects the costs, delays the project and may even be perceived as ignorant or at least out of place. From the architects' point of view, the users' comments often affect the architectural quality of the project as it forces them to make changes to their conceptual design. These changes could easily have been accommodated within the design if they had been part of the brief or taken into account during the conceptual phase.

4. PARTICIPATION AND QUALITY OF THE OUTCOME

Another argument in favor of user participation is that it improves the quality of the outcome. The democratic ambitions to involve users in the design process resulted in the production of design handbooks, pedagogical tools and education programs. These enabled users to take part in design processes. Many practicing architects realized what a tremendous source of information and knowledge the users represented. New methods of retrieving this information from the users into the design projects were therefore developed (Figure 2).

Most commonly architects had extensive interviews with users to understand the essence of the organization they were designing for. They soon realized that the group of involved users was not restricted to union representatives, and the focus of the participation process was shifted from the democracy motivation and the rights to influence the design effort to achieve an appropriate high quality building.

When one uses participation as a tool of achieving better quality, the architect's ambition as an artist might also be an obstacle that needs to be overcome. Artistic values might positively influence the architect's interpretation of the information communicated from the users, but it may also create conflicts between the architect's artistic values and the user's perception of the design.

The strength of focusing participation of the user on quality of the outcome is that both employees and employer can get something out of it. The employers can communicate with the architect directly and, therefore, may have an impact on the outcome. To influence the quality of the building is in the interest of employers wherever they own or rent the building. The problem is, however, that the focus of participation is restricted to the quality of the building. It will be discussed how participation of users can enhance the quality of the outcome of the design process in more ways than getting a better building.

5. PARTICIPATION AND LEARNING

5.1. Participation – An Adaptive or a Generative Process

The experience of user participation in architecture design has been mixed. Many writers on the subject argue that the outcome of participatory processes have not always been received better by users than outcomes of a more "artistic" design process where the architect has played the most dominant role. To understand this we have to elaborate two issues:

- What qualities represent a positive outcome of a participatory design activity?
- What are the significant qualities of interaction between different actors in participatory design processes.

5.1.1. Value of the Outcome

First one has to realize that the above-described stages of participation in architectural design all have their own standards for value of the outcome. From a democratic point of view the procedure of contacts between the parties and a result in terms of a having an impact on the result might be enough. A satisfying result in terms of quality of the product might be better working conditions, e.g. lower noise levels, ergonomical equipment. From the point of view of the employer it might be important that these improvements are advantageous for the production and for the overall social atmosphere of the company. If one, however, looks at the outcome of participation from a learning point of view, both parties will value individual growth, creativity and better overall performance. The focus has shifted from the building to the design process and the development of the organization that uses the building (Figures 1, 2 and 3).

5.1.2. Interaction between actors

User participation based on formal democratic procedure often lacks direct contact between users and architects in the early conceptual phases of the process. Politicians, local officials or union officers may represent the users in the dialogue with the architect. In this case the discrepancy between the users' needs and what they get is mainly a matter of how well the representatives represent the organization they are meant to represent.

Often, however, the participation activity is set up mainly to gain knowledge from the users. In most cases the architects collect information though interviews with personnel and through studies of the previous situations, if there are any. They then *interpret* this information into design solutions that goes back to the users for their comments. Even in these cases there may be problems.

Users are often badly prepared for participation in changing processes. The point of departure for their conceptions of the future is often limited to the existing situation with its restrictions and possibilities. Being asked about what they want, they may have problems *conceptualizing* their wishes, *articulating* them even to themselves and, even more, *communicating* them to colleagues. To communicate something in an interview with architects is more difficult. Architects, on the other hand, often have very early preconceptions of the design solution or they will formulate them after a few interviews. This is not only one of architects' strongest professional abilities, but also one of the obstacles in participation. They will have these preconceptions in their minds and they will, together with architects' professional and cultural values, function as *interpreters* to what the users try to communicate.

Architects' tacit understanding of the situation will be imbedded in the designs that they feed back to the users. These designs are mostly beautifully packaged as rendered perspective drawings, elaborated layout drawings produced in CAD or by hand. Maybe even a three-dimensional real model or a virtual computer model accompany the architect's interpretation. Mostly it is very hard for the users, given their problems of conceptualizing, articulating and communicating their needs, to determine whether the suggested design will really facilitate their future activities

From research and practice in Scandinavia one knows that the users often do not understand the suggested designs until they have moved in and started to use them. Disappointments at this stage not only discourages the users from further participation in architectural design, but also discourages the architect, who may have had strong intentions to listen to the users in order to design a well-functioning building for them.

5.2. The User as Designer

Another kind of user participation is to recognize the user as a co-designer. This process is sometimes called *the collective design process*. The idea is to form cross-disciplinary groups within a company. This group, which includes architects and other external experts as resource persons, drives the design process. In this design setting the focus is taken away from the building. The focus becomes the performance of the organization and how to design a building that supports better performance. The integration between the design of built space, technical systems and organization of work becomes natural in such a setting (Figure 3).

However, such a settings calls for the development of new behavior and new methods. The creation of a mutual language within the design group and mutual understanding of the situation might be the single most crucial factor. The roles of the participants in a collective design process are dual. All participants are experts in their own professional field. Whether they be nurses, teachers, assembly workers or maintenance workers, they can contribute hard facts and experience to the design process. The in-house participants are also users in the traditional sense and as such they have demands on the future design. Experience has shown that the group views the participants *as users* and allows opinions on almost anything concerning the design, based on the fact that their situation will be directly affected by the outcome of the design. If the group participants, on the other hand, see each other only as experts they tend to be more polite and respectful of each other's expertise. They

are, however, both experts and laymen. This makes it easy to everybody's good ideas into account, even if it is not based on expert knowledge. The usual interdisciplinary tacit understanding that often stands in the way of creative solutions can be overcome by the intervention of the kind of "disrespect" only laymen can show towards state-of-the-art solutions. In their role of field experts, can the participants directly influence details in the outcome without the interpretation by architects or other external experts?

There are, however, problems involved in this kind of participatory design. One of these problems is related to the fact that the design group becomes very skilled to the disadvantage of those not participating. They develop group knowledge with its own repertoire of experiences and sometimes its own internal language. The pedagogical problem is then to ensure that everybody knows what is going on and has the opportunity to contribute to the process. It must not become a closed process. The design activities should preferably take place on the premises of the user organization or close to the user organization. It is especially important that management, unions or other groups that are policy makers or have power to manipulate the process or the end result can participate in the manner they feel is appropriate. These groups might otherwise be alienated from the design process as well as the design solutions with all the negative effects that might have.

Another problem is to set up the actual design process. Professional language barriers, professional roles and the preconception of knowledge and design behavior are obstacles to these kind of participatory processes.

Participatory Ergonomics

J. R. Wilson and H. M. Haines

Institute for Occupational Ergonomics, University of Nottingham, Nottingham, NG7 2RD, UK

1. INTRODUCTION

Organizations increasingly appreciate the benefits associated with applying ergonomics to the design of workplaces and jobs, but there is a recognition that ergonomics experts alone cannot achieve widespread and successful implementation of ergonomics at work. Practically, there are simply not enough ergonomics experts to meet the needs of industry, and their employment may anyway not be cost effective for many smaller organizations. Outside experts may also be too isolated to be the core of an ergonomics and safety culture within a company. The alternative is an internally managed participatory ergonomics program.

There has been a significant growth of interest in participatory ergonomics during the past 15 years. Many of the rapidly expanding number of "participatory" initiatives reported worldwide attempt to use this kind of approach with the aim of improving the health and safety of their employees, or else to enhance the chances of success for any implementation of change. Consequently, there are many different types of participatory ergonomics initiative (see Haines and Wilson 1998 for a framework of these; much of the content of this entry is drawn from this source).

2. DEFINITIONS OF PARTICIPATORY ERGONOMICS

There is no general agreement about what the term participatory ergonomics actually means. It can be seen as variously a philosophy, an approach or strategy, a program or even a set of tools and techniques. (Neither is this lack of agreement limited to participatory ergonomics; similar views have been expressed in the participative management field.) A range of views of participatory ergonomics as a mixture of content and process can be seen in its definitions. In fact, participatory ergonomics should not be thought of as a unitary concept; Batt and Appelbaum (1995), for example, make a distinction between "off-line" and "on-line" participation where off-line refers to "off the job" problem-solving processes; "on-line" refers to where decision-making about work is part of the job (e.g. through work teams). Similarly, Zink (1996) draws a distinction between "selective participation" and the attempt to achieve continuous improvement.

Participatory ergonomics is a complex concept involving a number of different dimensions. At its most basic it consists of "stakeholders" contributing to an ergonomics initiative or sharing ergonomics knowledge and methods. Reflecting the broad range of potential ergonomics initiatives across workplaces, jobs and work organizations, a working definition is: "The involvement of people in planning and controlling a significant amount of their own work activities, with sufficient knowledge and power to influence both processes and outcomes in order to achieve desirable goals" (Haines and Wilson 1998). It is clear that employee participation in general, and participation in ergonomics initiatives in particular, can vary widely. Which kind of participation is appropriate in any given context? Cohen (1996) suggests a form of contingency approach, and that a number of so-called "shaping factors" should be considered in making such decisions namely:

- the nature of the issues requiring consideration;
- whether the matters are broad-based or specific to a local operation or group;
- whether the needs for response or action are time-limited or necessitate continuing efforts;
- the abilities of the group most affected; and
- the organization's prevailing practices for joint labor-management or participative approaches in resolving workplace issues.

3. THE CONTEXT FOR PARTICIPATORY ERGONOMICS

Why has there been such a significant rise in interest in participatory ergonomics in recent years? In general, the developed world has become more participative over the past 40–50 years; the public and the workforce in many countries will not accept government or management practices, which were widespread half a century ago. Disappointment with technical investments and recognition that motivation and performance are complex issues has bred human-centered manufacturing initiatives with emphasis upon an educated, involved and responsible workforce; there is a greater emphasis on attributes of quality, flexibility and customer service, rather than merely quantity of output, which in turn supports the need for greater workforce participation. High-performing companies are said to have employee empowerment as one of the enablers to be successful, and this has often been in tandem with teamwork; "people involvement" programs have grown out of quality circles and action groups as a part of Total Quality Management (TQM), although TQM and its derivatives can be viewed as subtle forms of management control, and empowerment and participation can be interpreted as insidious means by which worker power is eroded in the context of industrial disputes and attempts to downsize. Further drive for participatory ergonomics comes from health and safety and workplace regulations across the world.

4. THE PROS AND CONS OF PARTICIPATORY ERGONOMICS

Of all the advantages of participatory ergonomics, there are two direct benefits that are commonly referred to. First, employees have unique knowledge and experience of work and their involvement should, therefore, provide a clearer understanding of both the types of problems being encountered and the solutions that will be appropriate. Second, involving people in analysis, development and implementation of a change should generate greater feelings of solution ownership and may generate a greater commitment to the changes being implemented.

Participation may often be a learning experience for those involved. For workers, knowledge of their own work and organization may be increased, and involvement in a development or implementation process can mean faster and deeper learning of a new system and, hence, decreased training costs and improved performance. Participatory ergonomics may also be a

learning experience for designers or other technical specialists, and a departure from conventional thinking can improve design effectiveness.

At a wider level, by spreading understanding and expertise in ergonomics among those involved, and by attracting interest among work colleagues, solutions can be generalized and ergonomics transferred throughout an organization. If people are involved in making ergonomics changes at work, they are more likely able to adapt when circumstances change or to use the ideas elsewhere, increasing the in-house knowledge base of ergonomics.

The process of participation itself may provide benefits, with participants developing more self-confidence, competence, independence, personal development, social contact, feedback, influence, challenge and variability, the very characteristics that have been identified as contributing towards "good" work and reducing stress at work (Karasek and Theorell 1990).

The effects of worker participation on satisfaction and productivity have been the subject of debate, with some agreement about gains for both and that the greatest benefits come from the more extensive forms of participation. Also, basic workplace ergonomics and health and safety interventions can be carried out very efficiently and effectively through work teams participating together to identify problems and implement improvements. Involvement in such initiatives can help give the work teams a specific focus and strengthen the team basis and functioning. The socio-technical principle of compatibility dictates that it is desirable for the groups that develop ideas for new workplaces, work tasks and work structures to subsequently form into work teams themselves and vice versa.

Despite the considerable potential benefits of participatory ergonomics, it may also bring with it some difficulties. It is not always easy to instigate or support; management and workers might both be reluctant or resistant, because of suspicion, fear or the feeling they lack knowledge, motivation, time and energy. There are potential problems associated with the *process* of participation. Group decision-making can often be fairly unreliable or limited. Planning and developing new systems participatively may be slower, more complex and require greater effort. The process may encourage employees to develop unrealistic expectations about the timescale for management approval, implementation and benefits from their recommendations, or else the extent to which jobs or equipment will be changed. Sometimes the process or content of participation impacts on other parts of the organization in unhelpful ways: increasing the roles of one group may reduce job content for others; a participative process that improves the workplace for some may make those in other departments envious or dissatisfied; also, other groups may wish to be similarly involved and this may be neither possible nor even desirable.

Some projects referred to as participatory ergonomics may be so in name only, being little more than exercises in information provision or even manipulative programs with a covert agenda (Reuter 1987). Cynical use of participative methods may yield some limited dividends in the short-term but will not provide any of the potential greater long-term benefits. This sort of use may partially explain participation's ambiguous tradition within industrial relations, and particularly the suspicions of trade unions.

Finally, there may be problems in providing proof of results; it may be hard to show that participation has truly brought about a better system or system change, or that more autocratic methods would have been less effective. This lack of evidence may contribute to a lack of face validity in the eyes of management.

Clearly a participatory approach to ergonomics problems at work is not a panacea. Some organizations are simply not ready for participative practices. However, while participatory ergonomics will certainly not be universally appropriate, if careful consideration is given to the time, place and the individuals involved, as well as to the structure and methods being used, then its potential benefits may be realized.

5. PARTICIPATORY TOOLS AND METHODS

A variety of different tools and methods have been reported as useful within participatory ergonomics initiatives. Some have been borrowed, adapted or developed with participatory ergonomics specifically in mind, whereas others have been taken from more "traditional" ergonomics initiatives and then applied within a participative exercise. The variety of ways in which participatory ergonomics may be applied means that certain techniques will be much better suited to certain situations. For example, if we are looking to instigate and support participatory ergonomics as a macro-ergonomic strategy, we are likely to be interested in tools to sell a participative approach to stakeholders or to facilitate group-working and improve interpersonal skills. Tools and methods will be employed in various stages of a participatory ergonomics exercise, for instance organization, support, problem analysis, idea generation and concept evaluation.

A major issue in the selection and use of appropriate tools and techniques concerns participants' "expertise." Depending on the structure chosen for the participatory initiative, experts may play a number of different roles. They may restrict their intervention to largely supporting or guiding the participatory process, they may be involved with a variety of other workers as members of a multifunction group, or they may undertake much of the analysis themselves. Generally it is better to avoid using overly complex or technical analytical tools; if one of the aims of participatory ergonomics is to free institutions from reliance on the outside expert then the use of highly complex tools may prove self-defeating. This may restrict the use of some computer-based modeling and simulation tools for the time being (the sophistication here lying not so much in their use as in the interpretation of the results they produce).

Many techniques such as word maps, round-robin questionnaires, layout modeling and mock-ups form part of Design Decision Groups (DDG) (Wilson 1991), themselves derived from the work of O'Brien (1981), who adapted theories and techniques from market research and the literature on creativity and innovation. DDG have been used in a large number of applications in manufacturing, transport, service and process industries, and as part of much wider systematic approaches to engineering and design throughout an organization. An extension of a DDG approach includes reference visits by participants to other, similar, sites and their sourcing and costing of solutions within a budget set by management. This initiative can be extended into a process of continual improvement (Wilson 1995). Role-playing techniques and simulation games are of potential value and there is increasing interest in the place of simulation

games within organizational development, particularly from the perspective of developing work processes or group interaction and teamwork.

6. EVALUATION OF PARTICIPATORY PROGRAMS

Although there are difficulties involved in proving the effectiveness of participatory projects, there are a number of ways in which a participatory program may be evaluated. The most obvious is to measure the extent to which the original, anticipated outcomes have been achieved; is the incidence of musculoskeletal disorders falling, are employees more satisfied with their work, has productivity increased? There are, however, potential difficulties with using such outcome measures: it may take some time for any effects to show, or else effects might be significant but small, or else effects may be confounded by changes in company environment, market, structure, processes or people during the life of the implementation and evaluation.

Alternative approaches to evaluation may use "process measures," such as the number of changes implemented, participants' satisfaction with their involvement or the spread of the program throughout an organization.

For many managers the success of a participatory ergonomics program will be measured in terms of cost savings. However, cost–benefit analyses of ergonomics projects are not always easy to calculate. Accurate economic estimates for all factors are often very difficult to produce and some positive outcomes may not have direct economic consequences. Few cases in the participatory ergonomics literature evaluate their success in financial terms, although a special issue of the *American Industrial Hygiene Association Journal* (vol. 58, 1997) features company-wide ergonomics programs incorporating some degree of worker participation where the primary aim is to reduce musculoskeletal disorders, and attempts have been made to calculate the resulting benefits. One of the cases looks at some of the costs associated with setting up an ergonomics program suggesting that, over 2 years, almost $108 000 was spent on training and consultation and significantly more ($510 000) was spent on ergonomics interventions; during the same period over three times that amount was spent on worker compensation costs (Mansfield and Armstrong 1997). Another case comes from a collection of examples reported by Hendrick (1996). Structured almost like a controlled experiment, it featured seven companies that received training on how to set up a participatory program, and where all but one went on to implement them. Eighteen months later the six companies could report a total saving of over $1.25 million due to reductions in "strain-type injuries" whereas the seventh company reported an increase in injury levels.

7. A FRAMEWORK FOR PARTICIPATORY ERGONOMICS

Participatory ergonomics is not a unitary concept, and participatory ergonomics initiatives can take a wide variety of forms, from a single re-design exercise to development of a full culture of employee involvement. Table 1 identifies some of the dimensions across which participatory ergonomics initiatives might vary.

The first dimension of participatory ergonomics is its *extent or level*, whether participatory ergonomics is applied across an organization, a work system or a single workstation or product. The second dimension concerns *purpose*. Is it being used to implement a particular change or to be *the* method of work organization? Participatory techniques within a design exercise are often part of implementing a change, but may also be used as part of a product design process. A further dimension is provided by the *continuity* of use of participation. Has the process got a continuous or discrete timeline: is participation to be used as an everyday part of an organization's activities or is it applied from time to time as a one-off exercise?

The next dimension of participatory ergonomics — *involvement* — concerns who will actually take part in the process. Full direct participation is when all stakeholders directly affected become participants (however, restrictions on resources may dictate representation by a subgroup of those affected — partial direct participation). Representative participation generally occurs in two situations, where worker representatives are involved in product or equipment or job design, but may not necessarily be the eventual users, or where trade union personnel represent the interests of their members at work. The fifth dimension acknowledges that the *formality* of worker participation in ergonomics may vary, from mechanisms such as teams or committees with attendant procedures, to informal, almost casual use. The sixth dimension is the *requirement* for participation — is it voluntary or compulsory? Voluntary participation is the most usual form, in the sense that participation works best where the workforce volunteers its contributions and is involved in setting up the process. Compulsory participation is seen, for instance, in companies with compulsory quality circles or production groups, where involvement in troubleshooting and continuous improvement is an obligatory part of job specification and roles; it is arguable whether such compulsory participation is truly participative.

The seventh dimension considers *decision-making* structures. At one end of the scale, workers are consulted, although decisions rest with management. At the other end of the scale, workers make decisions. A third category concerns the situation where

Table 1. Dimensions of participatory ergonomics (adapted from Haines and Wilson, 1998).

Extent/level	Organization ——— Worksystem —————— Workplace ——— Product
Purpose	Work organization ————————Design———————— Implementation
Continuity	Continuous ———————————————————— Discrete
Involvement	Direct (full/partial) ———————————————— Representative
Formality	Formal ———————————————————————— Informal
Requirement	Voluntary ———————————————————— Compulsory
Decision-making	Workers decide ——————— Consensus ——————— Consultation
Coupling	Direct ———————————————————————— Remote

input is obtained from a range of those affected and decisions are made by consensus. Finally, how directly participative methods are applied (a concept referred to as *coupling*) may vary. Participative methods may be directly coupled (work groups re-designing their own workplace, for example). Remote coupling involves some filtering of participants' views, for example through the use of company-wide questionnaires.

Figure 1 shows a framework to illustrate the bases on which an organization might initiate and structure participatory ergonomics initiatives. It starts at the point where an organization makes the decision to employ participatory ergonomics and identifies some of the main factors which may motivate this decision. Some form of initiative will then be implemented, the structure of which may be defined across the eight dimensions of participatory ergonomics. Criteria influencing both the structure of the initiative and the selection of participatory methods have been identified. An evaluation of the initiative will contribute to any further motivation to employ participatory ergonomics. The whole process will take place within an organizational and environmental context, which will influence all its elements.

The participatory ergonomics literature shows some agreement about the requirements for a successful ergonomics

initiative. Because participatory ergonomics can operate at many levels, from a single design exercise up to the basis of a shop-floor culture, some requirements will be more relevant than others depending on the type of initiative being undertaken. The requirements can be summarized as:

- Establishing a climate and support for participation.
- Structuring a participatory initiative.
- Enabling participatory processes and methods.
- Evaluating participatory ergonomics initiatives.

8. CONCLUSIONS

Enthusiasm for participatory ergonomics belies the fact that the concept is less than straight forward and people call initiatives "participatory" that, in reality, are not. Participatory ergonomics is complex and diverse, an umbrella used to cover a fairly broad range of ideas and practices. This means that there is a range of models and ways of doing participatory ergonomics, and a multiplicity of tools and methods employed within participatory ergonomics initiatives.

Yet despite these variations, it would appear that most commentators see participatory ergonomics as offering a common set of advantages (Figure 2). First, it is seen as exploiting the detailed knowledge and experience of those who inhabit the very

Factors which may motivate an organisation to implement participatory ergonomics

- availability of expert advice
- legislation
- external recommendation
- workforce / union negotiation
- management philosophy
- awareness of importance of (workplace) ergonomics problems
- worker complaints or compensation claims
- product / market advantage

Decision to implement participatory ergonomics

Structure of initiative

- extent /level
- purpose
- continuity
- involvement
- formality
- requirement
- decision-making
- coupling

Criteria influencing structure of initiative

- size of organisation
- organisational culture
- nature of (workplace) problems
- timeframe available
- stakeholders (who is affected)
- resources available (expertise, time, finance)
- workforce education and training

Implementation process

Criteria influencing selection of methods

- type of problem
- resources available (expertise, time, finance)
- knowledge or experience of methods
- number of participants

Evaluation

Wider social, economic, legal and technical factors

Figure 1. Framework for structure of participatory ergonomics initiatives (adapted from Haines and Wilson, 1998).
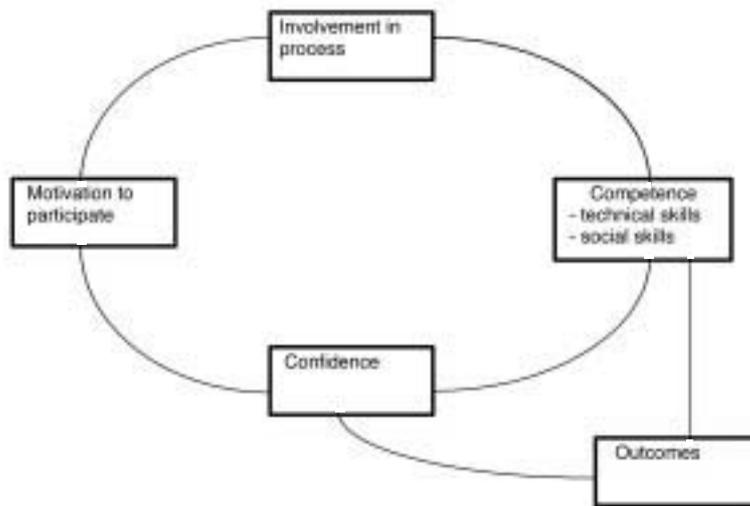
Figure 2. The participatory ergonomics cycle (Haines and Wilson, 1998).

workplace under investigation — getting most from those who, in a sense, should know best. Second, it is felt to encourage a sense of ownership among the participants, such that it helps to secure at least some degree of commitment both to the process itself and to any changes that may result. A third and related advantage of participatory ergonomics is thought to center upon psychosocial factors. Irrespective of the specific details of how these effects operate, it is not surprising that a workforce whose views are sought and taken seriously might feel more positively about both themselves and their workplace. Against such benefits, there may be some concerns about the time, cost and ease of implementing participatory programs, and about inappropriate or poor quality implementations. The potential benefits, though, can far outweigh any possible costs and problems.

REFERENCES

BATT, R. and APPELBAUM, E., 1995, Worker participation in diverse settings: does the form affect the outcome, and if so, who benefits? *British Journal of Industrial Relations*, 33.

COHEN, A.L., 1996, Worker participation. In A. Bhattacharya and J.D. McGlothlin (eds) *Occupational Ergonomics Theory and Applications* (New York: Marcel Dekker), 235–58.

HAINES, H.M. and WILSON, J.R., 1998, *Development of a Framework for Participatory Ergonomics*. Health and Safety Executive Report 174/1998 (London: HMSO for HSE Books).

HENDRICK, H.W., 1996, Good ergonomics is good economics. In *Proceedings of the Human Factors and Ergonomics Society 40th Annual Meeting*. Human Factors and Ergonomics Society, Santa Monica, California.

KARASEK, R. and THEORELL, T., 1990, *Healthy Work: Stress, Productivity and the Reconstruction of Working Life* (New York: Basic).

MANSFIELD, J.A. and ARMSTRONG, T.J., 1997, Library of Congress Workplace Ergonomics Program. *American Industrial Hygiene Association Journal*, 58, 138–44.

O'BRIEN, D.D., 1981, Designing systems for new users. *Design Studies*, 2, 139–50.

REUTER, W., 1987, Procedures for participation in planning, developing and operating information systems. In P. Doherty *et al.* (eds) *System Design for Human Development and Productivity: Participation and Beyond* (Amsterdam: North-Holland), 271–6.

WILSON, J.R., 1991, Design decision groups – a participative process for developing workplaces. In K. Noro and A. Imada (eds) *Participatory Ergonomics* (London: Taylor & Francis).

WILSON, J.R., 1995, Solution ownership in participative work design: the case of a crane control room. *International Journal of Industrial Ergonomics*, 15, 329–44.

ZINK, K.J., 1996, Continuous improvement through employee participation. Some experiences from a long-term study in Germany. In V.O. Brown, Jr and H.W. Hendrick (eds) *Human Factors in Organizational Design and Management* (Amsterdam: Elsevier), 155–60.

Participatory Ergonomics – A Scandinavian Approach

P. L. Jensen

Department of Technology and Social Sciences, Technical University of Denmark, Denmark

1. INTRODUCTION

Participatory ergonomics is a new concept within ergonomics that refers to a practice well known in Scandinavian countries. For years the general, espoused theory on improving working conditions has emphasized participatory approaches. Many studies on change processes based on workers' participation has been conducted resulting in recommendations for tools and procedures. But studies have shown that in practice participation in change processes is not guaranteed and that many difficulties have to be overcome while implementing these approaches. This article will present the status of participatory ergonomics in Scandinavia emphasizing central issues to be developed and researched.

2. DEFINING CENTRAL CONCEPTS

In Anglo-American literature the two concepts "ergonomics" and "human factors" are used to denote activities concerned with "fitting the task to the man." Scandinavians tend to associate ergonomics with applied physiology of work whereas the concept of "working environment" covers analytical, descriptive approaches to working conditions as well as approaches to establish normative recommendations for fitting tasks, workplace and working conditions to man. In this text all three concepts are seen as synonyms

The concept of "participatory ergonomics" was launched in the 1980s (Noro and Imada 1991). Within ergonomics it can be seen as a dissociation from the traditional, expert-based approach to analyses and change of workplaces. Based on an examination of more definitions Haines and Wilson (1998) define participatory ergonomics as "the involvement of people in planning and controlling a significant amount of their own work activities, with sufficient knowledge and power to influence both processes and outcomes in order to achieve desirable goals."

While the promotion of participatory approaches might be a new issue within ergonomics, it has been a recommended practice in the Scandinavian countries at least since World War II. This approach is based on a combination of direct and indirect participation. Indirect participation is based on participation of elected or appointed representatives of workers while direct participation is based on the participation of the workers, having to live with the changes.

There are different degrees of participation. The following typology makes it possible to differentiate between them:

- Information from management to workers on plans for action.
- Gathering of information and experience from workers on present working conditions.
- Consultation allowing workers to make suggestions and present points of view on new work sites.
- Negotiations to define problems and find solutions.
- Joint decision-making in agreement between involved parties on problems and solutions.

On the basis of these concepts the development of the Scandinavian approach can be described.

3. PARTICIPATION IN REBUILDING SOCIETY AFTER WORLD WAR II

After World War II the main issue for most European countries was to rebuild an effective production sector. Several strategies were followed of which one was based on local cooperation between workers and first-line managers to improve production. The existence of a generally highly qualified work force was an important precondition for this strategy. Under the heading "work smarter not harder" simple tools were presented. The aim was to use workers' and first-line managers' detailed experience of every-day activities to improve the production process and the work activities. Simultaneously changes were accepted. This direct participation approach was embedded in a representative, participatory system-based on national, collective agreements. The employees were represented by their shop stewards, which were protected against dismissal through a general agreement. This design was made to guarantee the general legitimacy of the process and the results of the direct participation procedures. In the representative system agreements were made on conditions of participatory schemes and for the allocations of the economic results between workers and employers.

In the 1950s and 1960s the main issue of these arrangements was to achieve increased productivity and higher wages. But simultaneously disputes on working conditions in these rationalized productions appeared. First, the central labor market organizations recommended a more efficient use of the representative bodies before and during organizational and technological change processes. They stressed the importance of joint decision-making or at least efficient consultancy of workers and not only information of their representatives. But in a response to a general discussion of working conditions particularly in industry resulting in wild strikes during the economic boom at the end of the 1960s and the beginning of the 1970s, a new act on working environment was passed in all Scandinavian countries.

4. THE 1970s WORKING ENVIRONMENT ACTS

The acts on working environment legalized employees' participation in issues concerning working environment. The general principle in the acts in accordance with the general prerogatives of management is to make the employers and top-management responsible for establishing safe and sound working conditions. With reference to the Danish act this implies that in planing and implementing chemical substances, technical aids, the design of the workplace, the excursion of work and the scheduling of working time management should prevent risk of accidents, work-related diseases, the observance of ergonomic principles and reduction of psychological stress.

Additionally the act states that employer's responsibility should be fulfilled in cooperation with the employees. To support this cooperation and to ensure a continuous monitoring of the working environment the act orders enterprises of more than 10 employees to form safety groups comprising the local super-

visor and a safety representative elected among the employees in the department. In enterprises of > 20 employees (and more safety groups) a safety committee has to be established including two supervisors elected by the supervisors of the safety groups, two safety representatives elected by the safety representatives of the safety groups and one person appointed from top management. The tasks of management, supervisors, employees, safety groups and safety committees are specified in general terms. The departmental orders under the act presents the aims to follow and a series of duties for all parties is enumerated.

From the beginning of the 1990s additional demands on the management process of working environment has been requested. In Norway and Sweden an "internal control system," i.e. a formalized system for managing working environment is demanded, whereas in Denmark a procedure for workplace assessment implying action plans is requested. In both cases the importance of a participatory approach is emphasized.

The intention of the legislators can be described as, first, to insist on the claim on employers to ensure safe and sound working conditions. Second, to legitimate that a room for a discourse on working environment between top-management, first-line managers and workers represented by safety representatives is set up in enterprises. Judged by the text of pamphlets the recommended approach was based on joint decision-making, but from a legal point of view consultations were demanded.

The majority of the legislators expected this to entail preventive actions in firms.

5. PRESENT STATUS

Many tools for participatory approaches to work place design have been developed and tested in the Scandinavian countries. There is no single textbook neither in one of the Scandinavian languages nor in English, but Haines and Wilson (1998) gives an overview of many of the tools also used in Scandinavia. Recommendations on tools and processes have been widely disseminated by public authorities, labor marked organizations, occupational health services and other agencies. Many stories and studies of success and how difficulties have been overcome have been spread for inspiration. Nevertheless professional in working environment generally find that the achievement of the enterprises do not fulfill the intentions of the act.

There are different interpretations of the problems of establishing effective participatory schemes. It is not perceived as a problem of having adequate tools but rather of the problem of integrating participatory processes into the existing decision-making structure of the enterprise and of qualifying the employees to participate in the process. The safety organization as a platform for a working environment discourse has generally been accepted. It is, however, often segregated form the central actors responsible of the operation and development of the enterprise and from the organizational units involved in the design of changes in productions processes (Jensen 1997).

Second, many cases prove that most of the important decisions are made before the plans are presented to workers and their representatives. The participatory arrangements primarily offers management a possibility to inform the employees' representatives of plans for action and maybe as a final possibility to get information on potential risks for malfunctioning. Seen from the employees' point of view they or their representatives mainly

have the role as a rubber stamp. If they are involved in change processes the representatives often regard themselves as taken hostage by the technical staff and management. When employees are involved in the early phases of a change project they often find it difficult to present their interests in accordance with the present level of conceptualization within the adequate conceptual frame.

6. ORGANIZATIONS AS POLITICAL SYSTEMS

It necessitates a more profound analysis to be developed to remedy this situation. Here the paradigm of organizations as political systems is considered a fruitful approach. This paradigm understands organizations as political systems, i.e. systems of actors and groups of actors, each of them more or less explicit trying to take care of their own more or less explicitly formulated interests. This implies that organizations are regarded as coalitions of stakeholders in different positions of power.

In such a system participation is not power equalization. First, because the different stakeholders have different possibilities to influence the design of and phases in the decision-making processes and the structure in the organization. Second, because the stakeholders have different mental model of the life in the organization. These mental models define the frame in which problems, causes, possible actions and legitimate strategies are identified.

7. IMPLICATIONS FOR ACTION

Persons involved in activities to improve the working environment by participatory approaches often suggest structural elements (like committees and task forces), processes and tools. But the analytical approach presented points to some central elements of the basic understanding of their activities.

First, it is a problem for workers to be formally involved in organizational change processes. It is important that they co-operate with strong stakeholders (among others top management and actors responsible of change programs and processes) in formulating participatory approaches not only as an espoused theory but as a theory in action. Second, it is difficult to keep up with management's initiatives to adapt to the new market conditions, rationalization initiatives and other initiatives entailing organizational and technological changes. Therefore, in implementing structural elements, procedures and tools for participation it is crucial to pay attention to the mental models of the stakeholders involved. This applies, of course, to the employees who should have a basic understanding of the general context of the change processes, the character of the planing processes, and how to make working environment an issue in the different phases of a planning process and how legitimately to present attitudes and make suggestions. But it concerns also the technical staff involved in conceptualizing and planning the changes. They often find it difficult to comprehend the relation between their projects and working conditions. Second, they should be capable of establishing a dialogue on equal terms with other stakeholders.

Owing to legal demands on mandatory courses for elected safety representatives and first-line managers the development of suitable mental models for employees has for years been an issue. Whereas strategies to increase the understanding of the technical design staff have in recent years been subject to research

and development processes with no final recommendations reached yet (Broberg 1997).

Third, it may be difficult for the appointed representatives to prove that they are representing the workers' interest. A combination of representative and direct participation might be a solution to this problem. Direct participation allows ideas for and views on problems, on their position in a causal web and on possible solutions to be presented. It offers a possibility for the representatives to insist on an indebt analysis of ideas and point of views and their implementation.

Fourth, studies of change processes based on participation show that firms may lose motivation if change processes implying a high degree of participation is entailed by normal operation implying a low degree of workers' participation. So participatory ergonomics should be comprehended as an issue both in the design of change processes and as well as in work design for normal operation, which the definition given by Haines and Wilson also reflects.

8. STATUS OF THE SCANDINAVIAN SITUATIONS

The Scandinavian countries have provided platforms for representative participation on a legal basis to legitimate participation and give some protection for the representatives. But the possibilities within the legal frame for this platform remain to be fully utilized.

Legal demands for internal control (i.e. demands for formal procedures for working environment) are presently under implementation in Norway and Sweden, as well as a demand for formalized workplace assessment procedures in Denmark. This may promote the development of participatory ergonomics. The first Danish studies give rise to some optimism (Jensen 1998), but more experience has to be gathered.

Also, the central organization of unions has as a central policy to promote what has been labeled "developmental work" that emphasizes the importance of qualifications and control in work in some of the new concepts for rationalization. Participatory ergonomics is incorporated in this approach.

Present experience show that to benefit from available possibilities an analytical approach stressing the political elements of the process and the importance of reaching compromises between interests seem to be a promising path and in this connection the development of the mental models play an important role not only for the weak parties (the workers) but for all parties involved.

In the Scandinavian countries (and according to Walters 1995 in the European Communities too) participatory ergonomics is a central element of the "espoused theory" on workplace design and change. But it is still not a warranted element of "theories in action" on workplace design and changes. Recommendations for tools and procedures are available in accessible forms, and agencies to promote participation such as occupational health services and the labor inspectorate do exist. It is, therefore, essential to make programs for participatory ergonomics and to analyze both successes and failures to improve our understanding of the conditions of this approach. So a continuation of the cooperation between firms, research agencies and working environment professionals including ergonomists is crucial.

REFERENCES

BROBERG, O., 1997, Integrating ergonomics into the product development process. *International Journal of Industrial Ergonomics*, 19, 317–27.

HAINES, H.M. and WILSON, J.R., 1998, *Development of a Framework for Participatory Ergonomics*. Contract research report 174/1998 (Norwich: HSE Books).

JENSEN, P.L., 1997 Can participatory ergonomics become "the way we do things in this firm" – the Scandinavian approach to participatory ergonomic. *Ergonomics*, 40, 1078–87.

JENSEN, P.L., 1998, Workplace assessment – a tool for developing ergonomics in firms? In Vink, P., Koningsveld, E.A.P. and Dhont, S. (eds) *Human Factors in Design and Management – VI* (Amsterdam: Elsevier).

NORO, K. and IMADA, A.S. (eds), 1991, *Participatory Ergonomics* (London: Taylor & Francis).

WALTERS, D., 1995, Employee representation and occupational health and safety: the significance of Europe. *Journal of Loss Prevention in the Process Industries*, 8, 313–18.

Participatory Ergonomics at the Shop Floor Level

I. Kuorinka

Lieu dit Sénil, F-46160 Gréalou, France

1. PARTICIPATORY PROJECT VERSUS AN ORDINARY PROJECT

Participatory ergonomics at the shop floor level usually takes the form of a project, an ergonomic intervention, which is limited both in time and by its scope. It is, therefore, interesting to have a closer look at how participatory ergonomics projects differ from other projects in the workplace.

The specificity of a participatory project would be easy to outline if there were a typical industrial project, which is not the case. Handbooks of organizational practice usually describe a project as a chain of events that converge towards an endpoint. The links in the chain usually include elements such as entry, diagnosis, action planning, implementation and termination. Although many theorists and probably all consultants admit that in reality any project is full of unexpected events, the process is nevertheless theoretically supposed to move linearly from phase to phase in a coordinated manner.

A participatory project has specific characteristics (either planned or spontaneous) that are iterative and based on experience and learning. It is not specialist-centered but relies as much as possible on available local competence (both formal and experience-based). Even the goal of a participatory project may be loosely formulated at the start, and may change during the project. In fact, the main emphasis is on the process itself: a successful participatory project creates, to a large extent, both goals and methods as result of the process itself. The nature of a participatory project may look odd to someone trained to follow project management step-by-step. It must be emphasized, however, that the aim of a participatory project is the same as that of a traditional project: creating coherence in the design process, reducing uncertainty, and producing a worthwhile end result. It just does it differently. A participatory project is not a social gathering for fun.

A participatory project has subsidiary benefits that go beyond the goal of the project itself: it is largely a learning experience, as will be explained later. It is also an effective means of dealing with and controlling uncertainty in a rapidly changing production environment. This issue has been discussed (e.g. Shipley 1990).

It should be noted that participatory ergonomics projects are usually limited in scope. They usually deal with problems close to the everyday production tasks, involving individuals and small groups. If a participatory project becomes large, it will be closer to participatory management types of issues, which are beyond the scope here.

2. PARTICIPATION, PARTICIPANTS, AND HOW IT WORKS

Participation is not a well-defined issue; it even seems to defy any definition. We may, however, add some qualifications that help us to grasp the concept better. Participation in this context means direct participation, where those concerned try to solve the problems they encounter. Participation combined with ergonomics deals with practical everyday problems in the workplace, not major societal issues. It is down-to-earth problem-solving where learning is an important element.

Many other terms to describe participation in workplaces have been coined: empowerment has been used as a term to describe the employee's right to have more say about his/her own working conditions. Promoting creativity through group work has been explored, especially in white-collar environments.

In this context, participatory ergonomics (project) is defined as an ergonomic intervention that relies on the participation of those who have firsthand experience with the problem to be solved. This contrasts with formal, representative participation, as will be discussed later.

In most cases, participatory activity involves a group. Thus, the methods available and often employed are not necessarily much different from those used in any group activity. Groups whose activity involves some notion of participation have been called, for example, a "design decision group" (Wilson 1991), "quality circles" (e.g. Nagamachi 1991), "focus groups" (Caplan 1990), "participatory design groups," "problem-solving groups." Group techniques have been used in the context of "problem based learning," "value analysis," "quality management," "value analysis and engineering", etc. The list is endless. Depending on personal preferences and the nature of the participatory group, an ergonomist may choose whatever seems to be appropriate in the case in question. Participatory ergonomics has some characteristics, however, that require special attention. This is discussed in the chapter on "ethics."

Members of a participatory ergonomics group do not form a natural group, although they usually come from one production unit, department, etc. They must be selected by taking into account the needs of the intended project. Participant selection is not without problems: the aim is to recruit those and only those who have actual knowledge about the problem to be solved in the project. Formal representatives of management, labor and other social groups may not have the necessary firsthand experience about the problem at hand, which is the requirement of a participatory ergonomics project. It has in fact been shown that participation through representation is less efficient than direct representation (Cotton *et al.* 1988). However, the social actors have a legitimate right to be integrated in some way into the project. Bringing together experienced group members requires diplomatic negotiation. In some cases, separating the "decisional/representative" level and the "operational" level has been a solution, with the latter level reporting to the former.

A participatory ergonomics project may be run in very different ways. It is helpful to think of a project as a learning experience. The ergonomist-facilitator should understand at least the core issues in andragogy (adult learning, e.g. Knowles 1987). Observing modern methods of training at work which rely on (cognitive) psychological principles (e.g. Vartiainen 1987) is not only useful but necessary in designing a participatory ergonomics project.

Learning involves both the participatory process itself and the subject-matter related to the problem at hand. In fact, formal training at the beginning and during the project is desirable. If the group members have no experience with a participatory

process, significant time must be devoted in preparing the group to work together. The principles and general rules to be followed during sessions must be explained. Concepts like consensus, right of veto, dealing with divergent views, responsibilities, etc., must be made clear.

Issues related to the problem's subject-matter must also be taught. At the beginning, it seems useful to present a general picture of the problem. A knowledgeable person, preferably from the company itself, should explain the context of the problem. A description of the technological, organizational and social issues as well as experiences from other companies will help group members understand the problem in a larger context.

The participatory process also aims to elicit knowledge in order to uncover hidden, tacit knowledge. The participatory group is the core actor in seeking such information but should have help from external specialists and the ergonomist himself. Methods such as auto-confrontation are available for discovering important tacit information.

How should a participatory project be run? There is no commonly accepted standard procedure; there is a lot of variation. An approach that is successful in a small company may be inappropriate in a large company with a formal organization. Different professional groups require different approaches. Sen (1987) has presented approaches to be used in a technical professional environment. Wilson (1991) presents practical approaches for designing workplaces.

Some elements are specific to a participatory project, however. It is down-to-earth and deals with everyday problems. The process aims at stimulating participants and their reference groups as well as other organizational groups directly linked to the project. Before the project can even start, there must be a commitment from management and labor as well as from those directly involved. Information and communication are an important part of the project. The aim is to widen the basis of participation and spread knowledge among those not directly concerned by the project.

3. PARTICIPATORY ERGONOMICS, CONCEPT AND DEFINITIONS

The concept of participatory ergonomics is fairly recent. Noro and Kogi (1985) proposed the notion, which in fact seems to have been anchored in the Japanese experience in small group activities and "suggestion" schemes. Several ILO projects under the direction of Kazutaka Kogi have developed the notion further. The 1985 IEA congress in Bournemouth, UK, brought together other persons interested in the theme. The same ideas were undoubtedly discussed and tested well before 1985 but the concept and name of "participatory ergonomics" seem to have been established by that time.

Participation has been promoted as an important element in industrial democracy. We can trace some elements of workplace participation back to the debate on "industrial democracy" (e.g. ILO 1981) and other sociologically colored issues which were largely debated in the 1960s and 1970s. Questions like participation systems, labor unions and their representative role, participation in decision-making, participation in the company's capital, etc., were some of the issues.

Some institutions seem to have had more influence on the debate on participation than others (e.g. the Tavistock Institute in UK).

Also, trends in industrial organization have advanced some forms of participation. Quality management programs and concurrent engineering (Zhang and Zhang 1995) are among these trends (Daniellou and Garrigou 1992).

As explained earlier, progressive trends in society have created interest in participatory approaches. The field has, however, proven to be complex and contradictory. In many cases, participation has been ostensibly promoted, but has in practice met tough resistance from those who fear loss of power and control over individuals.

Participation as a basic principle for organizing work is not unknown but is probably not too extensively practiced, either. In some cultures, collective planning and problem-solving have been natural ways of organizing work. In others, social conflict has been seen as prime mover for solving problems. Some industrial cultures underline a leader's role. Participatory ergonomics proposes participation as a natural way of organizing work while focusing on common goals.

Participatory ergonomics has not been defined authoritatively and may defy any attempt to do so in the future. Cultural differences between countries and organizations are such that even a modest level of employee participation, which under other conditions may look trivial, may be seen as counterproductive and unacceptable.

Martin and Baradat (1998) have analyzed in interesting way the paradoxes of a participatory project.

4. THE ROLE OF AN ERGONOMIST IN A PARTICIPATORY PROJECT

In any project, an ergonomist may serve as a project leader, a specialized resource person or a consultant. If the project is organized in a participatory framework, an additional role is necessary: the ergonomist must become a facilitator. A facilitator's task is to help control the participatory process itself.

Any person, regardless of his profession, may serve as a facilitator if he has the necessary skills to work with a group and has sufficient knowledge about the problem's subject-matter. A well-trained ergonomist has the advantage of being able to deal with work-related matters across professional boundaries. It is thus possible to take into account and balance the organizational, physiological and psychological, working-condition-related, subjective and objective issues in a participatory project.

The requirements of an ergonomist/facilitator are not trivial, as can be deduced from the list above. The task is demanding and requires both theoretical knowledge (about participation itself and the problem's subject-matter) and practical experience dealing with workplace problems.

5. ETHICAL ISSUES OF PARTICIPATION

Participation in the workplace is a complex process where each participant is supposed to contribute constructively to solving the problem in question. It contrasts with the process where participants come together only to defend their interests. In practice, these two modes often exist simultaneously, although constructive contribution dominates the participatory process as defined above. It means that a participatory process is safe from

manipulation and organizational intrigues. It would be helpful to understand participation as an intellectual exchange: the participants bring something valuable and original to the negotiating table but assume that they will receive something equivalent in exchange.

A participatory process can in some cases be a profound subjective experience. Thus all attempts to manipulate, exploit or otherwise take advantage of the situation may provoke a violent reaction among the participants. As a result, the participatory process stops and may be difficult to revive. In one case known to the author, a serious labor dispute was the consequence of such a situation.

Respect for ethical principles is a precondition for a successful participatory process. Some of the conditions are:

- The goals of a participatory project must be transparent and honest. Differences in opinion must be dealt with openly.
- Rules on how to deal with disagreements must be established and agreed to.
- The formal position of the participatory project in relation to other organizational functions, collective agreement, etc., must be defined and agreed to.
- The "ownership" of the results must be defined at the start.
- Rules for dealing with possible unexpected results must be defined.

REFERENCES

CAPLAN, S., 1990, Using focus group methodology for ergonomics design. *Ergonomics*, 33, 527–33.

COTTON, J.L., VOLLRATH, D.A., FROGGATT, K.L., LENGNICK-HALL, M.L. and JENNINGS, K.R., 1988, Employee participation: diverse forms and different outcomes. *Academy of Management Review*, 13, 8–22.

DANIELLOU, F. and GARRIGOU, A., 1992, Human Factors in design: sociotechnics or ergonomics? In M. Helander and M. Nagamachi (eds) *Design for Manufacturability: A Systems Approach to Concurrent Engineering and Ergonomics* (London: Taylor & Francis), 55–63.

KNOWLES, M.S., 1987, *Adult Learning in Training and Development Handbook: A Guide to Human Resource Development* (New York: McGraw-Hill), 108–79.

MARTIN, C. and BARADAT, D., 1998, Les paradoxes de la participation dans les projets. *Actes des Journées de Bordeaux sur la Pratique de l'Ergonomie,* Université Victor Segalen Bordeaux, 2, 114–20.

NAGAMACHI, M., 1991, Application of participatory ergonomics through quality-circle activities. In K. Noro and A. Imada (eds) *Participatory Ergonomics* (London: Taylor & Francis), 139–64.

NORO, K. and IMADA, A., 1991, *Participatory Ergonomics* (London: Taylor & Francis).

SEN, T.K., 1987, Participative group techniques. In G. Salvendy (ed.) *Handbook of Human Factors* (New York: Wiley), 454–69.

SHIPLEY, P., 1990. Participation ideology and methodology in ergonomics practice. In E.N. Corlett and J.R. Wilson (eds) *Evaluation of Human Work* (London: Taylor & Francis), 819–34.

VARTIAINEN, M., 1987, *The Hierarchical Development of Mental Regulation and Training Methods*. Helsinki University of Technology, Industrial Economics and Industrial Psychology, Report 100. Otaniemi, TKK Offset.

Workers Participation in Decision Within Undertakings (Geneva: ILO, 1981).

ZHANG, H.C. and ZHANG, D., 1995, Concurrent engineering: an overview from manufacturing engineering perspectives. *Concurrent Engineering: Research and Applications*, 3, 221–36.

Prevention and Compensation of Shiftwork Effects

G. Costa

Department of Public Health, University of Verona, I-37134 Verona, Italy

1. INTRODUCTION

Shiftwork, in particular that including nightwork, is a well-known risk factor for health and well-being, as it interferes with biological rhythms and social relations, with consequent negative influence on work performance, family and social life, and health (see the chapter on "Shiftwork Health Consequences").

To prevent such negative consequences it is important to adopt some preventive and protective measures aimed, on one hand, at creating more favorable working conditions and, on the other hand, at preserving workers' health.

Many kinds of interventions aimed at compensating for shift and nightwork inconveniences have been introduced in recent years, in most cases very empirically according to different work conditions and specific problems arising in different companies, work sectors and countries.

According to Thierry's model, such interventions can be defined in two different ways in relation to the domain they deal with:

1. The "counter-weights," that is, the interventions adopted to compensate only the inconveniences. Monetary compensation is the main counter-weight worldwide adopted as basic reward for shift- and nightwork. It is a simple translation of the multidimensional aspects of the problem into money, and its value may vary according to different factors, some of which may not be strictly connected to the severity of the inconveniences (e.g. Unions' power, economic conditions, productive needs). Other counter-weights may be represented by measures aimed at improving other aspects of the work organization, such as work environment and/or job enlargement/enrichment, and fringe benefits.

2. The "counter-values," that is the interventions adopted to reduce or eliminate the inconveniences by dealing with the themes in which the complaints occur; they can be directed:

   - towards the causes of inconveniences, such as: reduction of the amount of night- or shiftwork, planning more ergonomic shift schedules; transfer from nightwork to daywork; reduced work load at night;
   - towards the consequences of inconveniences, such as: canteen and sleep facilities, coping strategies, medical control and re-habilitation, early retirement, extra days off, longer holidays, more breaks at work, better social services (transportations, housing, children care); and
   - towards the psychological meaning of the problems, such as: education, physical and psychological training and social support.

2. ORGANIZATIONAL INTERVENTIONS

Particular attention has to be given to the organization of the shift schedules, taking into account not only economic reasons, but giving priority also to the workers' needs, in particular as concerns the human physiology, the psychological and social problems, and the possible negative health consequences.

The main guidelines for designing shift systems according to ergonomic criteria are:

- Quickly rotating shift systems are better than slowly rotating ones, since they interfere less with circadian rhythms and minimize the extent of any cumulative sleep deficit.
- Clockwise rotation (morning–afternoon–night) is preferable to the counterclockwise (afternoon–morning–night), since it parallels the endogenous circadian rhythms, which show a periodicity slightly longer than 24 h in "free-running" experiments; it also avoids quick changeovers (e.g. morning and nightshift on the same day) and allows longer rest periods for the immediate recovery from fatigue and sleep deficit.
- Early starts of the morning shift should be avoided to reduce the truncation of the sleep (REM phase in particular) and the consequent fatigue and sleepiness.
- Prolonged work shifts (9–12 h) should only be considered when the workload is suitable, there are adequate pauses, and the shift system is designed to minimize the accumulation of fatigue and the exposure to toxic substances.
- Shift systems should be as regular as possible and able to guarantee as many free weekends as possible, to allow people to plan and enjoy their leisure time more conveniently.
- Permanent nightwork can be acceptable only for particular working situations that require a complete adjustment to nightwork to guarantee the highest levels of safety.
- Rest days should come preferably after the night duty period to consent a prompt recovery from the sleep deficit and an easier return to the normal sleep–wake cycle.
- Some flexibility in working times is desirable to give the workers the possibility to combine better work duties with family and social life.
- Supplement crews should be provided to reduce the amount of nightwork of the individual worker.

However, it is worth stressing that there is no "best" shift system to be recommended in general, but each shift schedule should be planned and adapted according to the different job demands, and the specific characteristics, social habits and cultural background of the workers involved. This implies a careful strategy for the arrangement of the shift schedules, which requires the workers' participation in the analysis, design, implementation and assessing of the shift rota chosen. This is of paramount importance, not only for taking into account the suggestions of those who have direct experience of the problem, but also for promoting the right motivation for accepting the changes and, consequently, improving the psychophysical tolerance.

3. COUNSELING

It is very important for shiftworkers to be aware of the possible negative consequences and to receive useful suggestions and guidelines on how to cope best with shift and nightwork.

Therefore, counseling and training has to be done both at individual level and among groups of shiftworkers through educational programs. These should deal with improving self-care strategies for coping, in particular with sleep, diet, stress

management, physical fitness, off-job activities and exposure to bright light.

Basic information should be given on body rhythms and the necessity to maintain as much as possible their normal pattern.

Good sleep strategies and relaxation techniques should be adopted to minimize de-synchronosis and fatigue. Shiftworkers should try to keep tight sleeping schedules and avoid disturbances (e.g. by arranging silent and dark bedrooms, using earplugs, making arrangements with family members and neighbors). The timing of diurnal sleep after a night duty should be scheduled taking into consideration that sleep onset latency and length are strongly influenced by the phase of the temperature rhythm, so that sleep starting in the early morning, during the rising phase of the temperature rhythm, shows longer latency and shorter duration than that starting in the early afternoon.

In addition, a proper use of naps can be very effective in compensating sleep loss and enhance alertness. Useful naps can be taken before nightshift or extended duty periods, during night-work to alleviate fatigue and sleepiness, or after early morning and nightshifts to integrate normal sleep. A proper combination of nap and sleep times makes it possible to avoid the abuse of hypnotics, which can have a negative effect on the process of re-synchronization and cause a transient impairment on vigilance and performance.

People should also try to maintain stable meal times, which can act as co-synchronizers of body functions and social activities.

A proper timing and composition of meals can help the adaptation. During nightwork in particular, assuming that meals with high carbohydrates content facilitate sleep by stimulating serotonin synthesis, whereas meals at high proteins content favor wakefulness and work activity by stimulating catecholamine secretion, it would be preferable that shiftworkers have a pro-tein meal in the early part of the nightshift and then take light snacks with carbohydrates and soft drinks. They should also try to minimize the intake of stimulating substances (coffee, tea, cola drinks) for fighting drowsiness, as they may disrupt the subsequent sleep pattern and have negative effects on the digestive system. To favor a good digestion it is convenient to take a sufficiently long break and have a hot meal by arranging proper canteen facilities.

Keeping physically fit has been proved to be helpful in improving shiftwork tolerance by positive effects on synchroni-zation of circadian rhythms, performance efficiency and restora-tive properties of sleep, thus reducing fatigue and health problems. This can be achieved both by favoring individual regular exercise and sport activity, and by promoting positive fitness programs at work, both as prevention and rehabilitation. Also relaxation tech-niques (e.g. massage, yoga, autogenous training) are becoming more and more popular among people who feel under stress and who need help to control restlessness, anxiety, muscular tension, insomnia and other symptoms of stress.

On the other hand, people should avoid ineffective ways of coping (i.e. smoking, alcohol, drug consumption), which can have an apparent short-term positive effect, but in the long run can add further problems for health and well-being.

The consumption of drugs (e.g. stimulants, tranquilizers, anti-depressants), in particular, can have a positive effect only if they are used (under medical supervision and paying attention to their negative effects on vigilance and performance!) as auxiliary and temporary support for the organization of effective strategies aimed at removing the causes of stress.

A proper exposure to bright light, both through outdoor physical activities and increasing indoor artificial lighting (> 1000 Lux), has certainly a beneficial effect on psychophysical condi-tions. Bright light suppresses the secretion of melatonin by the pineal gland, which plays an important role in the entrainment of the circadian system. Therefore, when necessary, a proper tim-ing of light exposure can influence the direction and the magni-tude of the entrainment of circadian rhythms (e.g. light exposure in the morning causes a phase advance, whereas light exposure in the evening causes a phase delay). Moreover, bright light has a direct stimulating effect on mental activity and has been proved to reduce the symptoms of seasonal affective disorders: and some of the negative effects of nightwork can be likened to a mild form of endogenous depression.

It is also desirable to take into consideration and try to com-bine some psychophysiological characteristics and working con-ditions that may be expected to favor adaptation and tolerance to shiftwork. For example, "morning" types and aged people cope better with morning shifts and worse with nightshifts; on the contrary, "evening" types have fewer problems on nightshift, but more difficulties in the early morning shifts.

4. MEDICAL SURVEILLANCE

Workers' fitness for shift- and nightwork should be evaluated by occupational health physicians before their assignment, at regular intervals, and in case of health problems.

It appears reasonable to consider exemption from nightwork for people suffering from severe disorders that can be either directly related to shift- and nightwork, or can be worsened by irregular working hours because of their interference with sleep, diet or drug treatment, Such conditions are: clinical sleep disor-ders (i.e. insomnia, sleep apnoea, narcolepsy); severe gastroin-testinal diseases (i.e. chronic gastritis, peptic ulcer, chronic active hepatitis, cirrhosis, chronic pancreatitis); ischemic heart disease, hyperkinetic syndromes and severe hypertension; insulin-dependent diabetes; severe thyroid and suprarenal pathologies; severe nervous disorders (i.e. brain injuries with sequelae, epilepsy requiring medication, chronic anxiety and/or depression); chronic renal impairment; malignant tumors; women in pregnancy.

Moreover, particular attention should also be paid to particular workers for which shift and nightwork may represent a possible risk factor under certain circumstances, in particular those exposed to toxic substances; aged people; women with menstrual disorders or small children; people suffering from mild digestive disorders, asthma and chronic obstructive bronchitis, or afflicted by marked hemeralopia or visual impairment; persons taking drugs affecting the central nervous system (e.g. benzo-diazepines, high intake of alcoholics); people with unsatisfactory housing conditions and with long commuting times; persons with high levels of neuroticism or prone to internal de-synchronization of circadian rhythms.

Many of these conditions may represent, in principle, abso-lute or relative contra-indications to night- and shiftwork, thus requiring a temporary or permanent assignment or transfer to daywork. However, it can not be directly assumed that the sole removal of such risk factor can be certainly or totally beneficial to the worker's health. In fact, it has to be considered that most

of the complaints of shiftworkers pertain to the psychosomatic domain and recognize a multifactorial origin, related to family heritage, life styles, social conditions, other occupational risks and intervening illnesses. Consequently, maladaptation and intolerance to shift and nightwork are the result of complex interactions among the above cited risk factors, and can manifest differently in each shiftworker during his/her working life both in terms of severity and time of manifestation.

Moreover, advances in clinical diagnosis, pharmacology and rehabilitation now offer better possibilities for treatment of some diseases (e.g. peptic ulcer, hypertension, myocardial infarction) and, therefore, may consent the permanence on shiftwork of some workers, for whom the transfer to daywork can be problematic, due to either other risk factors or personal resistance to change job.

The complexity of the problem requires a careful attention to the occupational health physicians to avoid a risky attitude towards an a-critical selection of many workers, thus reducing their possibility to find a job (that could be more dangerous for their health and well-being!). It is also to be taken into account that sometimes shiftworkers do not report completely their health troubles, as considered "part of the job," or even mask them because they are more afraid of losing the economic benefits connected to shift- and nightwork.

Periodical checks should be scheduled for detecting early signs of difficulty in adjustment or intolerance to shift and nightwork (e.g. sleep troubles, digestive problems, psychosomatic complaints, accidents, perturbed menstrual cycle, drug consumption), which may require prompt interventions both at the organizational and individual level. Standardized questionnaires, checklists, diaries of the daily activities and recordings of some physiological parameters (e.g. body temperature, cortisol, melatonin, performance) can be helpfully used in evaluating the level of adaptation of the persons. In case of exposure to toxic substances, the biological monitoring should also take into account both timing of exposure and time-qualified reference values.

The periodicity of the health checks should not to be set *a priori*, but in relation to several factors concerning both working conditions (e.g. shifts rotas, combined risk factors) and individual characteristics (e.g. age, health). As a general guideline it appears advisable to plan a second health check not later than 1 year after starting nightwork (as the first year is crucial for adaptation and coping), and successive health checks every 3–5 years for those < 45 years of age, and every 2–3 years for those > 45 years of age.

REFERENCES

COSTA, G., 1998, Guidelines for the medical surveillance of shiftworkers. *Scandinavian Journal of Work Environmental Health*, 24, 151–5.

INTERNATIONAL LABOUR OFFICE, 1990, *Night Work Convention No. 171* (Geneva: ILO).

KNAUTH, P., 1996, Designing better shift systems. *Applied Ergonomics*, 27, 39–44.

KOGI, K., 1996, Improving shiftworkers' health and tolerance to shiftwork: recent advances. *Applied Ergonomics,* 27, 5–8.

MONK, T. and FOLKARD, S., 1992, *Making Shiftwork Tolerable* (London: Taylor & Francis).

TEPAS, D.I., 1993, Educational programs for shiftworkers, their families, and prospective shiftworkers. *Ergonomics*, 36, 199–209.

WEDDERBURN, A. (ed), 1991, *Guidelines for Shiftworkers.* Bulletin of European Shiftwork Topics no. 3 (Dublin: European Foundation for the Improvement of Living and Working Conditions).

WEDDERBURN, A. (ed), 1992, *Compensation for Shiftwork.* Bulletin of European Studies on Time no. 4 (Dublin: European Foundation for the Improvement of Living and Working Conditions).

Principles and Strategies for Team Training

E. Salas, J. A. Cannon-Bowers and K. A. Smith-Jentsch

Department of the Navy, Naval Air Warfare Center Training Systems Division, 12350 Research Parkway, Orlando, FL 32826-3275, USA

1. INTRODUCTION

Teams have become a way of life in organizations. In fact, a recent survey conducted on US industries reported that 82% of businesses with 100 or more employees used teams in one way or another to accomplish their objectives. Teams are more popular than ever and organizations are investing considerable resources in strengthening the communication and coordination mechanisms of their employees. Furthermore, many organizations are now designed around teams so that their very survival depends on effective teamwork.

This interest in teams by organizations has led to an explosion of team training research and practice. The aim of this research has been to address the question: what turns a team of experts into an expert team? That is, what developmental activities can enhance effective team functioning among highly interdependent members that have distributed expertise? To report the state-of-the-art in team training, what has been learned in answering this question will be outlined. Specifically, team training principles and strategies that have been generated in recent years will be focussed on.

To begin the discussion, a team is defined as "a distinguishable set of two or more people who interact, dynamically, interdependently, and adaptively toward a common and valued goal/objective/mission, who have been assigned specific roles or functions to perform, and who have a limited life-span of membership" (Salas *et al*. 1992: 4). There are key concepts in this definition that bound the generalizability of this discussion. The focus is particularly on teams that have distributed expertise, have high task interdependency and are organized hierarchically (i.e. there is a leader). Examples of this type of team include military command-and-control teams, medical emergency teams, some police and fire-fighting teams, cockpit crews, some sports teams and some management teams.

2. WHAT IS TEAM TRAINING?

Team training is a set of tools, methods and content that together create an instructional strategy that focuses on enhancing teamwork. Figure 1 outlines the basic elements of team training. As seen, team training is not a program or a place, as it has been popularly characterized. Team training interventions are theoretically based, built around sound instructional principles, based on carefully crafted and designed instructional strategies and focused on promoting the ease of coordination, communication and synchronization among team members. Team training is more than just team building (i.e. where team members learn about each other) and individual training (i.e. where team members learn about their task). Team training is about designing and developing instructional strategies — through tools, methods, team objectives, and competencies — for influencing team processes.
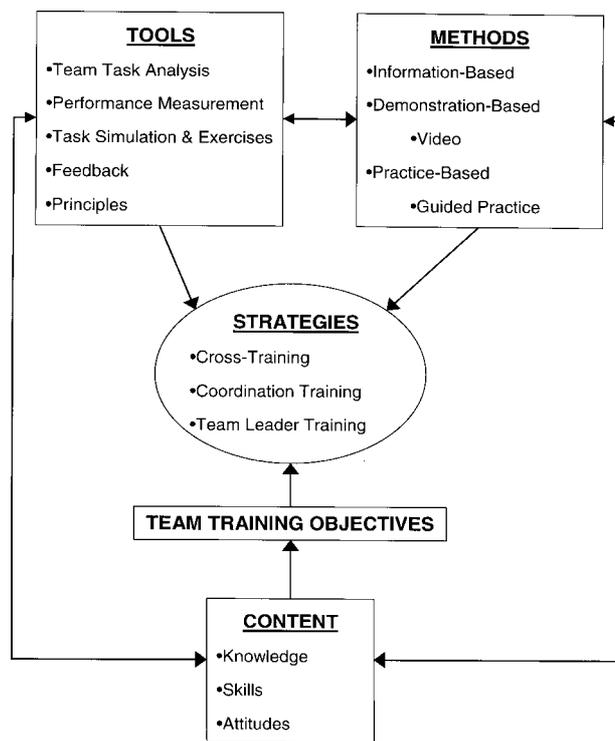
Figure 1. Structure of team training (from Salas and Cannon-Bowers 1997).

2.1. Tools

There useful tools available for designing effective team training strategies. For example, research has shown that an analysis of the coordination demands required for successful task performance can serve to focus the nature of practice and feedback. This coordination analysis requires that subject matter experts (SME) specify the criticality, frequency and difficulty of performing certain team-related tasks. This is usually done via surveys or questionnaires. Team task analysis is also an available procedure. Although much work remains here to demonstrate its utility, this tool allows the uncovering of the knowledge, skills and attitudes (KSA) necessary for effective teamwork given specific task requirements.

Performance measurement is the key to learning. Performance measurement tools embedded into training allow for learning to occur. Measurement tools for teams must consider a number of principles, which are readily adaptable/applicable to any team training environment. First, it is essential that any system of performance measurement have a strong theoretical foundation on which to draw. There are many frameworks now available that can help guide the development process (Brannick *et al*. 1997). Second, a clear purpose for designing such a measurement system must be established up-front. This can be accomplished by first performing team/cognitive task analyses to tease out the criterion (i.e. KSA) upon which the system should be based. Next, opportunities for measurement during the training must be specified. This approach allows for specific and controlled learning opportunities and a manageable context to measure complex team performance. Any measurement system should also be aimed at not only determining "what" transpired but also should explain "why" it happened to facilitate learning. Therefore, process measures (i.e. to capture moment-to-moment actions and

behaviors) must be designed and used. Usually, these measures take the form of observation protocols or scales.

Additionally, team performance measurement systems should be capable of supporting the diagnosis of performance deficiencies so as to provide relevant re-mediation to team members. Another principle that should underlie the development of a team performance measurement tool is that it must not be cumbersome to use. Users should be able rapidly to collect, evaluate and synthesize performance data so that immediate feedback can be provided to trainees. This can be realized by automating the system and by preparing instruction for the assessors.

A key element to the success or failure of any performance measurement system lies with the involvement of the user. Involving SME, of course, in the development process will help to ensure the end product is accepted, useful and effective. Finally, the tools and training necessary for instructors consistently to use the performance measurement system should be provided by way of instructor training packages as well as supporting materials (i.e. guide sheets, definitions, scenario events). Adherence to these principles will ensure that a strong team performance measurement tool to support team training is developed.

Other tools to support team training include feedback and learning principles, which are used to guide and focus the design process. These are available in the literature from not only the human factors perspective but also the education and cognitive disciplines. Finally, task simulation and exercises are also tools for team training. These provide the environment or context that allows team members to practice the requisite KSA.

2.2. Methods

Many methods are available to deliver instruction. These apply to both individual as well as team training. There are information-based methods such as lectures, handbooks, slide-presentations and the like. These are widely used and effective when combined with other methods. There are also demonstration-based methods. These are more dynamic and stimulating. These methods include video demonstrations of effective and ineffective teamwork behaviors as well as multimedia systems that illustrate key principles and performance requirements in teams. Finally, there are practice-based methods. These are more sophisticated, costly and interactive. These methods include simple-to-advanced exercises such as role-play exercises to high fidelity simulation environments (i.e. simulators) for teams to practice. These systems are visually rich and very interactive and engaging to the trainees. Moreover, it is known that simulators are effective training devices when appropriate instructional tools and features (e.g. performance measurement) are embedded in them (Salas *et al.* 1998).

2.2.1. Principles for enhancing the design of team training

Principles exist that can be used to guide the design and delivery of team training. Some representative examples are provided (Cannon-Bowers and Salas 1998, Salas and Cannon-Bowers 1997. For example, team training must foster the development of team skills such as team leadership, adaptability and compensatory behavior; link cue patterns to response strategies not only with regards to the task but also with other team members as well;

provide team members with guided practice under realistic conditions; provide team members with motivational guidance and outline factors that hinder their motivation; make team members aware of the importance of effective communication; develop simulations that allow team members to experience different courses of action; facilitate team members ability to learn about each others' roles and build realistic expectations about tasks requirements; diagnose team members teamwork deficiencies with a combination of process measures and outcome measures; provide the necessary information to team members about effective teamwork, demonstrate those behaviors and actions and allow for the practice with feedback on the requisite KSA; and exercise team members with a variety of novel situations to build adaptive mechanisms.

2.3. Team Competencies

Team competencies are the heart of team training. The task context provides the required KSA needed successfully to complete a team's tasks. KSA in teams are quite different from those found for individuals.

Team competencies describe what team members need to know (i.e. knowledge), how they need to behave (i.e. skills) and what task related attitudes (i.e. attitudes) they need to hold. Using the tools described earlier (i.e. team task analysis) team KSA must be derived as a precursor to developing team training. In fact, the KSA established for a team should provide the training objectives to be targeted in developing instruction.

Recently, researchers have focused attention on developing taxonomies of KSA for different types of teams. In general, this work has highlighted the fact that team cognition — that is, the knowledge component has been overlooked in past research. In addition, results indicate that some teamwork skills are consistent across tasks, while others may be more generic. Likewise some team competencies may be specific to the particular team at hand, while others may hold regardless of team membership (for more detail about KSA, see Cannon-Bowers *et al.* 1995, Cannon-Bowers and Salas 1999).

2.4. Examples of Team Strategies

As stated, the methods, tools and content (with the appropriate team training objectives) create a team training strategy. These take many forms and shapes depending on needs, resources and objectives of the training. Most team training strategies these days combine information, demonstration and practice-based methods.

2.4.1. Team coordination training

Team coordination training is probably the most widely used in organizations. This strategy has been used to train teams in a variety of settings including aviation, medical and military teams, to name just a few. This strategy is also referred to as crew resource management or aircrew coordination training. While the course content differs for each of these settings, there are certain commonalities inherent in all of these training programs. To begin with, this strategy concentrates on teaching team members the general components of teamwork in an effort to encourage it. Next, most of these programs use the general team performance or group dynamics literature as a framework. In addition, the

training is usually delivered through a combination of information, demonstration (e.g. video examples) and practice-based (e.g. role-plays) methods over 2–5 days.

2.4.2. Cross-training

Cross training is a strategy in which team members are exposed to the basic tasks, duties and responsibilities of the positions held by other members of the team. The intent of this strategy is to alleviate the decline in performance following personnel changes as well as to increase implicit coordination (that is, being able to coordinate without the need explicitly to communicate). This is accomplished by the subsequent increase in the interpositional knowledge held by each team member following training (i.e. a strengthening of the teams shared mental models). The training is composed of cross-role information sharing (team-mates, task, equipment, situation), enhanced understanding of interdependencies, roles and responsibilities, and cross-role practice and feedback. The training is usually delivered via lecture, multimedia systems and/or role-plays. Research has shown that cross training can be an effective means of enhancing team performance. Among other things, cross-trained teams are better able to anticipate the information needs of their team-mates, commit fewer errors and display better team process quality (for a review of this and other team training strategies, see Cannon-Bowers and Salas 1998).

2.4.3. Team self-correction

Team self correction, as defined by Blickensderfer *et al.* (1997), is a naturally occurring tendency for effective teams, following a performance episode, to review events, correct errors, discuss strategies and plan for the future (e.g. debrief themselves). This team training strategy attempts to capture this tendency and train other teams how to engage in this process. The end result is a team able to correct their knowledge-, skill- and attitude-based deficiencies without the aid of a formal instructor. This training is delivered through a combination of lecture, demonstration, practice and feedback. The target of training is the team leader (if there is one) and team members. Specifically, team members are taught how to observe their performance, how to categorize their effective and ineffective behavior into a structured format, and how to use this information to each other feedback (Cannon-Bowers and Salas 1998). When an instructor is present, he/she is taught how to probe the team members to come up with examples of effective and ineffective performance. This method of guided team self-correction has been demonstrated to improve team performance.

2.4.4. Event Based Approach to Training (EBAT)

EBAT is a general approach to the design of simulation-based exercises where training opportunities are systematically identified and introduced as "trigger" events in a training exercise. The purpose of doing this is to provide opportunities for students to exhibit specific team behaviors that are based on the team training objectives established through team task analysis. These behaviors are in turn captured, rated and than used later for debriefing and feedback. This strategy has been successfully used in multiservice distributed military training environments as well as in aviation settings.

2.4.5. Assertiveness Training

In complex team environments it is important that teams make use of all available information and resources. This means that the unique knowledge, skills, ideas, opinions and observations of each team member must be exploited to arrive at the optimal decision. In fact, it is often the willingness of team members to voice concerns assertively that can help to avoid potentially life-threatening decisions (e.g. in law enforcement teams or aircrews). Hence, efforts have been directed toward training team members to display effective task-related assertiveness. For example, team members may need to act assertively when providing feedback to other members, when stating opinions, when stating potential solutions or when offering or providing assistance. Training to improve assertiveness should involve practice and feedback to ensure that assertiveness skills will be demonstrated (Smith-Jentsch *et al.* 1996).

REFERENCES

BLICKENSDERFER, E., CANNON-BOWERS, J. A. and SALAS, E., 1997, Training teams to self-correct: an empirical investigation. Paper presented at the 12th annual meeting of the Society for Industrial and Organizational Psychology, St Louis, Missouri.

BRANNICK, M. T., SALAS, E. and PRINCE, C. (Eds), 1997, *Team Performance Assessment and Measurement: theory, methods, and applications* (Mahwah: Lawrence Erlbaum Associates).

CANNON-BOWERS, J. A. and SALAS, E. (Eds), 1998, *Making Decisions Under Stress: implications for individual and team training* (Washington: American Psychological Association).

CANNON-BOWERS, J. A. and SALAS, E. A., 1999, Team effectiveness (this volume).

CANNON-BOWERS, J. A., TANNENBAUM, S. I., SALAS, E. and VOLPE, C. E., 1995, Defining team competencies and establishing team training requirements. In Guzzo, R. and Salas, E. (Eds) *Team Effectiveness and Decision Making in Organizations* (San Francisco: Jossey Bass), pp. 333–80.

SALAS, E., BOWERS, C. A. and RHODENIZER, L., 1998, It is not how much you have but how you use it: toward a rational use of simulation to support aviation training. In *The International Journal of Aviation Psychology* (8th ed.) (Mahwah: Lawrence Erlbaum Associates).

SALAS, E. and CANNON-BOWERS, J. A., 1997, Methods, tools, and strategies for team training. In Quiñones, M. A. and Ehrenstein, A. (Eds) *Training for a Rapidly Changing Workplace: applications of psychological research* (Washington: American Psychological Association), pp. 249–79.

SALAS, E. and CANNON-BOWERS, J. A., in press, The anatomy of team training. In Tobias, L. and Fletcher, D. (Eds) *Handbook on Research in Training* (New York: Macmillan).

SALAS, E., DICKINSON, T. L., CONVERSE, S. A. and TANNENBAUM, S. I., 1992, Toward an understanding of team performance and training. In Swezey, R. and Salas, E. (Eds) *Teams: their training and performance* (Norwood: Ablex), pp. 3–29.

SMITH-JENTSCH, K. A., SALAS, E. and BAKER, D. P., 1996, Training team performance-related assertiveness. *Personnel Psychology*, 49, 909–36.

Psychosocial and Work Organization Risk Factors for Work-Related Musculoskeletal Disorders

N. Warren

Division of Occupational and Environmental Medicine, Ergonomic Technology Center, University of Connecticut Health Center, MC 6210, Farmington, CT 06030-6210, USA

1. INTRODUCTION

Ergonomic research has made significant progress in identification of biomechanical (physical) risk factors for work-related musculoskeletal disorders (WRMSDs) and in development of intervention strategies to reduce these exposures. This research may have contributed to the observed reduction in WRMSD rates in the second half of the 1990s. However, despite broad agreement on biomechanical risk factors and control strategies, the reduction has been disappointingly small. This suggests the presence of other causal factors not susceptible to intervention techniques focused solely on biomechanical risks. Particularly over the past decade, the ergonomics research community has thus devoted increasing attention to psychosocial risk factors associated with the organization of work. This article addresses these factors. It has four aims:

- Clarify the confusing terminology that has evolved in this field.
- Present the current state of knowledge concerning the contribution of psychosocial, work organization, and higher-level organizational risk factors to WRMSD etiology.
- Explore possible mechanisms for these contributions
- Summarize research directions necessary to expand knowledge in this field.

2. DEFINITIONS
2.1. Psychosocial Stressors

The definition of "psychosocial stressors" varies widely, depending on the researcher and the particular intellectual discipline. Several disciplines use the term to include personal characteristics of the worker — demographic (social class, culture, age), genetic (gender, intelligence), and psychological (personality, attitudes). Although these factors are important contributors to how the individual experiences and reacts to the work environment, this article views psychosocial stressors as nonphysical aspects of the work environment that have a psychological and physiological impact on the worker. Since the psychosocial work environment is usually assessed through self-report of the worker, these psychosocial stressors are thus measures of worker experience — integrated assessments of a wide and subtle range of work environment (and nonwork) characteristics, filtered through the respondent's response templates. Thus, psychosocial risk factors represent the intersection of the psychology of the individual with the sociological characteristics of the organization. It can be forcefully argued that it is this integrated percep-

tion of psychosocial stressor levels that is most closely related to stress and health outcomes.

The physical or biomechanical stressors associated with WRMSD are well defined elsewhere in this encyclopedia. Note, however, that the separation between psychosocial stressors and biomechanical stressors is to some extent artificial; often they are two sides of the same coin. For instance, rapid, machine-paced work is generally associated with constrained job autonomy (a psychosocial stressor) as well as with higher levels of biomechanical stressors (e.g., repetition and force). It is this close link between many psychosocial and biomechanical stressors that makes it difficult to estimate their separate associations with health outcomes.

2.2. Work Organization Stressors

Work organization refers to the way production and the work process is designed — characteristics of the work environment that give rise to the employee's perceptions of psychosocial stressor levels. These characteristics are the visible expression of the company characteristics listed below as well as the company's economic, cultural, and political environment.

2.3. Organizational Stressors

Organizational stressors are aspects of the organization's demographics, structure, technology, and culture that create and/or exacerbate levels of risk experienced by employees. The intersection of the organizational development literature with occupational health intervention literature has focused attention on the crucial role played by organizational culture in the creation or prevention of disease. A complete assessment of factors contributing to WRMSD etiology should also include macro factors that affect the company, its work organization, and the worker — cultural factors, economic pressures, regulatory pressures, etc. These factors are not addressed here.

2.4. Ergonomics

To fully incorporate psychosocial stressors into a full-spectrum definition of risk, it is useful to broaden the meaning of the term "ergonomics" itself. In too much of the literature, ergonomics is treated as the science of identifying and reducing biomechanical risk factors for WRMSD alone. But the Greek root of "ergonomics" is "the natural laws of work." For most practitioners, ergonomics translates into "fitting the work to the worker." Ergonomic practice addresses a complicated mixture of (1) biomechanical risk factors, (2) psychosocial/work organization risk factors, and (3) organization-level risk factors that affect levels of these stressors and the potential success or failure of interventions. Although the term "ergonomic" risk factor is often used synonymously with "biomechanical" risk factor in the literature, this article terms all three types of risk factor "ergonomic risk factors" and suggests that researchers more precisely define the type of ergonomic risk factor they are addressing

3. TYPES OF STRESSORS
3.1. Psychosocial Stressors

What are the psychosocial factors that research suggests are related to stress and health outcomes? The many models of job-related stress identify somewhat different sets of psychosocial stressors. But there seems to be rather broad agreement on their general

nature. A basic reference for the interested reader is the book *Healthy Work* (Karasek and Theorell 1990), which presents a detailed background for the widely used demand-control model (Karasek *et al.* 1985) as well as the expanded demand control–social support model for job stress (e.g., Johnson and Hall 1988). These models relate stress and health outcomes to the interaction between psychological job demands, the amount of decision latitude (conceptualized as a combination of autonomy and ability to learn and use skills) available to the worker, and the amount and quality of workplace social support. The extensive social support literature identifies many axes of this concept; much recent occupational health research has focused on the relative importance of coworker and supervisor support and has distinguished between so-called instrumental support (provision of information, help in completing work, etc.) and emotional support.

Johannes Siegrist (1996) more recently has proposed an effort-reward model, linking job stress to an imbalance between effort expended and rewards received in work. Finally, many researchers would add perceived job security to this brief overview of psychosocial stressors. The reference list demonstrates that each of the broad psychosocial stressor concepts listed here can be further broken down into many components.

3.2. Work Organization Stressors

There are many characteristics of the work organization that are associated with these psychosocial stressors. Kasl (1992), summarizing concepts from a wide range of research, presents a broad taxonomy of work environment factors that affect levels of psychosocial stressors:

- *Physical aspects of work*: environmental factors, biomechanical demands, etc.
- *Temporal aspects of job design*: hours, shift work, work-rest schedules, pace, etc.
- *Job content*: variety, repetitiveness, skill use, mental workload, participation in decision making, clarity of demands, etc.
- *Interpersonal relations*: group cohesion, support from co-workers and supervisors, availability of feedback, etc.
- *Organizational aspects*: structure of the organization, bureaucratic characteristics, etc.
- *Financial and economic aspects*: pay structure, benefits, etc.
- *Community and societal aspects*: status, prestige of job, etc.

Note that the job content and interpersonal relations aspects contain the same labels as some of the psychosocial stressors noted above. In this list they are included as externally measurable characteristics of the work organization. Organization-level risk factors will be explored below.

4. PSYCHOSOCIAL AND WORK ORGANIZATION FACTORS ASSOCIATED WITH DISEASE

The most extensive research connecting psychosocial stressors with disease has been in the area of heart disease. Prospective epidemiological studies, many with very large databases, have demonstrated that the combined effect of high psychological job demands with low decision latitude is predictive of increased blood pressure and cardiovascular morbidity and mortality. Reduced levels of social support increase these predictive asso-

ciations. This literature has motivated investigators to extend the model to WRMSD etiology.

5. PSYCHOSOCIAL AND WORK ORGANIZATION FACTORS ASSOCIATED WITH WRMSD

Although researchers have made great strides in detailing associations between risk factor exposure and the development of WRMSD, the natural history of these disorders is still imperfectly understood. There is controversy over whether the increase (until recently) in WRMSD is attributable to an actual increase in underlying risks, better recognition and reporting, changes in workers' compensation structure, and/or other variables such as cultural and social factors. It seems reasonable to hypothesize that at least some of the increase is related to risk increases, with risk being defined more broadly than biomechanical stressors alone. In its documents on national strategies for prevention of WRMSD, NIOSH has explicitly included the assessment of psychosocial risk and the multilevel examination of organizational characteristics. The same changes in production that have led to increased biomechanical stressors have also driven an increase in psychosocial and work organization stressors, among them increased competition, downsizing and mergers, increased division of labor, and technological changes in tools and processes. This article can only sample the many studies of psychosocial stressors for WRMSD that have appeared over the last decade. For more detailed treatments, refer to Kourinka and Forcier (1995), Bernard (1997), Moon and Sauter (1996). Bongers *et al.* (1993) have also prepared a review of the epidemiological literature. This looks at 29 studies which directly addressed work characteristics and it suggests that monotonous work, high perceived workload, time pressure, low control, and lack of social support are related to musculoskeletal outcomes. Several European projects have found task flexibility and peer contacts to be associated with a number of upper extremity diagnoses.

A 10-year prospective study in Finland demonstrated that social support and work content measures (monotony, autonomy, skill use, etc.) predict changes in upper extremity, lower back, and lower limb disorders. Analyses of the large Monitor database from the Netherlands have demonstrated that psychosocial and biomechanical exposure variables have individual associations and mild interactive associations with outcomes of musculoskeletal strain, pain, long-term sick leave, and partial disability. This research also suggests that the interactive demand-control model can explain a portion of WRMSD symptomatology. A cross-sectional analysis of the Nurses' Health Study cohort demonstrated a relationship between physical function and pain reports and psychological demands, decision latitude, and social support.

Office work has been the site of many suggestive studies (Moon and Sauter 1996), although the research is controversial. The much vaunted Second Industrial Revolution, with promised increases in worker control, innovation, and creativity, has not materialized in most offices. To the contrary, in many workplaces, highly routinized and externally controlled computer work has replaced industrial production as the machine-paced assembly line of the twentieth century. Three NIOSH studies of visual display (VDT) workers have found a relationship between psychosocial stressors and the presence of musculoskeletal symptoms. A number of workplace and laboratory studies have found that

musculoskeletal symptoms in VDT workers were related to a complex mix of biomechanical stressors, psychological demands, decision latitude, and employee relationships with the supervisor (and sometimes coworkers). Computer monitoring of employee performance and productivity has been associated with increased levels of perceived stress and increased musculoskeletal symptoms.

6. ORGANIZATION-LEVEL STRESSORS AND WRMSD

Building on a large body of research linking organizational demographics, structure, technology, and culture to work attitudes, research is just beginning on characterizing the importance of organization-level risk factors for WRMSD. Suggestively, the Canadian Institute of Work and Health found management safety commitment to be a strong negative predictor of accident rates. A number of researchers are examining aspects of "organizational response" that have impacts on the prevention and management of WRMSD. The Monitor analyses, from the Netherlands, have shown that employer and employee perceptions of risks in the company are poorly correlated.

I have done some further analysis of this database and found that the difference between employee and employer risk reports is associated with employee reports of WRMSD symptoms. A growing body of research suggests that employee perceptions of organizational support are important predictors of stress levels (hence WRMSD etiology). A recent randomized survey of the Connecticut workforce found that the employees' perceptions of organizational support showed more consistent negative associations with WRMSD symptom status than increased control, supervisor support, or coworker support. Finally, threat of job loss appears to be associated with a wide array of poor health indicators and stress measures, including musculoskeletal disorders.

7. POSSIBLE MECHANISMS FOR THE ASSOCIATION BETWEEN PSYCHOSOCIAL FACTORS AND WRMSD

At present, the mechanisms by which psychosocial stressors might cause musculoskeletal disorders are not well understood. The recent NIOSH review of the epidemiology of these disorders (Bernard 1997) presents a conservative list:

- Psychosocial factors may increase awareness of musculoskeletal symptoms. The effect of increased awareness is unclear, since increased awareness in the context of a well-organized ergonomic program results in earlier reporting and a lower rate of debilitating WRMSD.
- Psychosocial factors may affect reporting behavior. This is almost certainly true, but these factors, particularly in the arena of social support, present the worker with a complicated mixture of incentives and disincentives to report.
- Psychosocial stressors vary as the result of variation in biomechanical stressors. This article has noted the close connection between these types of stressors, but multivariate analysis of several large and diverse databases shows that the two stressor types also have independent associations with WRMSD outcomes.
- Initial pain symptoms due to biomechanical stressors may result in chronic nervous system disorders, both physiological and psychological, that may perpetuate chronic pain.

- Psychosocial stressors may produce chronically elevated muscle tension, thus predisposing soft tissues to the effects of biomechanical stressors.

The first four possibilities undoubtedly contribute to the association between psychosocial stressors and WRMSD outcomes; note that the first three are also tied to work organization and organizational culture. But this section concentrates on the last suggestion and other potential physiological pathways between psychosocial stressors and WRMSD outcomes. Although it is generally assumed that the psychosocial factors moderate the effect of biomechanical stressors on body tissues, rather than causing WRMSD directly, even in the absence of biomechanical stress, this is not yet clear.

The potential mechanism most studied is increased muscle tension related to levels of psychosocial stressors. The largest body of research has focused on monitoring electromyogram (EMG) signals from the trapezius muscles during laboratory and work-site activities. This research has consistently demonstrated that both biomechanical and psychosocial stressors can independently increase muscle tension and that the combined effect of both stressor types results in the highest tension levels. Individuals exposed to higher levels of psychosocial stressors at work also take longer to unwind after work, thus exacerbating the possible effects of sustained muscle tension. EMG research also demonstrates that psychosocial stress is associated with a reduced frequency of EMG gaps — brief (0.2–2.0 s) unconscious interruptions in the EMG signal that may represent automatic rest breaks for muscle fibers. Reduced frequency of EMG gaps is associated prospectively with trapezius myalgia.

Psychosocial stressors might actually result in an overall disorganization of muscular activity, ineffective muscle fiber recruitment, and prolonged activation of muscle fibers with low activation thresholds (whimsically termed Cinderella fibers). Disorganization of other body regulatory mechanisms may also play a part in the association between WRMSD and psychosocial stressors. Attention to a broader range of regulatory mechanisms affected by psychosocial stressors suggests other possible mechanisms for WRMSD genesis:

- Interference with blood flow to muscles
- Interference with energy metabolism and blood sugar regulation
- Impeded muscle coordination, increasing cocontraction and muscle force output

Chronic exposure to psychosocial stressors is associated with characteristic patterns of increased stress hormones, in particular, the catecholamines (adrenaline and noradrenaline) from the adrenal medulla, and cortisol from the adrenal cortex. These are the hormones responsible for the body's adaptive flight or fight responses to acute stressor exposure. At present, the effects on body tissues of chronic exposure to elevated levels of these hormones are unclear. But the ability of these hormones to break down proteins and other complex organic molecules (to mobilize available energy resources) and to affect immune responses might, over time, predispose soft tissues to injury. As in the EMG research, individuals with higher levels of psychosocial stressors at work appear to maintain elevated blood levels of these hormones long after the workday ends.

Recent research has paid attention to the tendency of individuals under stress to hyperventilate, thus removing CO_2 from

the blood. One theory (Schleifer *et al.* 1999) suggests that the resulting increase in pH of body tissues may cause heightened neuronal and muscle activity (producing increased muscle tension), paresthesias, and a shift in autonomic nervous system balance towards sympathetic dominance. Increased sympathetic activity increases responsiveness to catecholamines — secreted at higher levels under conditions of chronic exposure to adverse psychosocial factors. These proposed mechanisms require much more research. The state of the research at present suggests that these mechanisms would not operate in isolation; rather they represent a complicated and linked set of physiological responses that might combine in several ways to affect WRMSD etiology.

FUTURE DIRECTIONS

- Continued exploration of the physiological pathways through which psychosocial stressors either exacerbate the effects of biomechanical stressors or cause certain types of WRMSD directly.
- Research directed towards gender and age differences in the relationship between psychosocial stressors and WRMSD. Many existing studies suggest that males and females experience and react to different psychosocial stressors in different ways.
- Improved longitudinal studies. Although difficult to carry out, longitudinal studies are the best way to firmly establish causal direction.
- Multidisciplinary and multilevel studies examining the contribution to disease and interaction of a wider range of risk factors.
- Intervention studies to determine the protective effect of reducing psychosocial stressors and to provide organizations with practical work organization change strategies.

SUMMARY

The field of psychosocial epidemiology, especially in relationship to WRMSD outcomes, is relatively young. But an emerging body of research, combining laboratory, epidemiological, and intervention studies, suggests that psychosocial, work organization and organization-level risk factors are implicated in WRMSD etiology. Thus, to devise truly effective strategies for the reduction of musculoskeletal disease in the workplace, practitioners and organizations must address a broad spectrum of "shop-floor" risk factors (including both biomechanical and psychosocial stressors). Interventions must also take a multilevel approach, altering aspects of organizational culture, structure, and technology that create the work organization, which in turn creates this spectrum of risk.

The implications of this integrated prevention strategy are highly political. The experience of major companies indicates that employee involvement is a crucial basis for all safety programs, including ergonomic programs. This is not surprising, given what the research suggests about the importance of autonomy, skill use and learning, and social support in WRMSD etiology. But work organization and levels of psychosocial and biomechanical risk are largely the result of the extremely lopsided distribution of power and information within organizations. There are incentives to increase the employee share of these quantities, on medical, productivity, and innovation grounds. But the social, historical, and economic climate also provides strong disincentives to employers. The choice is becoming increasingly clear; it is unlikely that the high rates of WRMSD experienced by the workforce will be dramatically reduced unless these political questions are resolved and employees have greater power and influence within their organizations.

REFERENCES

BERNARD, B.P. (ed.), 1997, *Musculoskeletal Disorders and Workplace Factors*, US Department of Health and Human Services, Public Health Service, Centers for Disease Control, National Institute for Occupational Safety and Health.

BONGERS, P.M., DE WINTER, C.R., KOMPIER, M.A.J., and HILDEBRANDT, V.H., 1993., Psychosocial factors at work and musculoskeletal disease. *Scandinavian Journal of Environmental Health*, 19, 297–312.

JOHNSON, J.V. and HALL, E.M., 1988, Job strain, workplace social support, and cardiovascular disease: a cross-sectional study of a random sample of the Swedish working population. *American Journal of Public Health*, 78, 1336–42.

KARASEK, R.A. and THEORELL, T., 1990, *Healthy Work: Stress, Productivity, and the Reconstruction of Working Life* (New York: Basic Books).

KARASEK, R.A., PIEPER, C.F., and SCHWARTZ, J.E., 1985, Job content questionnaire and user's guide, revision 1.1. Developed at Columbia University; information from Dr Karasek, Department of Work Environment, University of Massachusetts, Lowell MA.

KASL, S.V., 1992, Surveillance of psychological disorders in the workplace panel. In Keita and Sauter (eds) *Work and Well-Being: An Agenda for the 1990s* (Washington DC: American Psychological Association), pp. 73–95.

KOURINKA, I. and FORCIER, L. (eds), 1995, *Work-Related Musculoskeletal Disorders (WMSDs): A Reference Book for Prevention* (London: Taylor & Francis).

MOON, S.D. and SAUTER, S.L. (eds), 1996, *Psychosocial Aspects of Musculoskeletal Disorders in Office Work* (London: Taylor & Francis).

SCHLEIFER, L.M., LEY, R., and SPALDING, T.W., 1999, Job stress, psychophysiological mechanisms and musculoskeletal disorders: preliminary results. Presented at Work Stress and Health '99, APA/NIOSH Conference, 11–13 March.

SIEGRIST, J., 1996, Adverse health effects of high-effort/low-reward conditions. *Journal of Occupational Health Psychology*, 1, 27–41.

Quality and Ergonomics in Concert

J.R.C. Axelsson[†‡] and J.A.E. Eklund[†‡]

†Centre for Studies of Humans, Technology and Organisation and
‡Division of Industrial Ergonomics, Linköping University, Sweden

1. INTRODUCTION

Recently there have been many research initiatives attempting to integrate quality and ergonomics issues. However, currently there are few examples of applications in industry and public sector that are based on an integrative approach. This chapter gives a broad overview of ergonomics preconditions for quality and presents a framework for an integrative approach to the development of quality and ergonomics in concert. To illustrate the possibilities and to visualize some of the methods and tools that can be used in this context, some examples will be presented briefly.

1.1. Definitions

There are many definitions of both quality and ergonomics. The Nordic Ergonomics Society uses the following definition of ergonomics: "Interdisciplinary field of science and application *considering integrated knowledge of human requirements and needs* in the interaction human–technology–environment in the design of technical components and work systems." The aims of ergonomics have furthermore been expressed in various contexts ranging from the well-being of employees, customers, clients, users and organizations to communities and society in whole. Promoting health, safety, comfort and satisfaction are seen as relevant for all parties and the crossovers are believed to reduce costs and to increase motivation, commitment, efficiency, productivity and quality. Quality, on the other hand, has been defined as: "The quality of a product or service is its *ability to satisfy the needs and expectations of the customers.*" As modern quality concepts emphasizes both internal customers (i.e. every employee) and external customers the overlap of the two definitions are quite obvious. This is further reflected in definitions of total quality management (TQM) such as: "A constantly developing system consisting of values, tools and methods for *increased customer satisfaction*. It builds on continuous improvement work in all the organization's processes, in which all the *employees are permitted and encouraged to become involved*" (Axelsson and Bergman 1999).

Even though the two disciplines developed separately there have been mutual influences. There are similarities and common features as well as differences. Recognizing these is important for the understanding of the basis for an integrated approach.

2. CORE CONCEPTS OF AN INTEGRATED DEVELOPMENT

An integrated development, incorporating quality and ergonomics may take place at different levels. In this respect, there are opportunities to address a number of core issues. Roughly these can be divided in to four main perspectives:

- organizational values;
- strategic management;
- operational methodology; and
- integrated applications.

As all perspectives are closely interconnected, a holistic standpoint will support and secure an integrated approach. The organizational values provide management control and gives direction regarding the way in which an organization achieves its objectives. A lack of clear values can lead to ethical problems for the organization and motivational problems. The establishment of organizational values and beliefs is part of establishing the culture of the organization, and is, therefore, a part of strategic planning and policy deployment. The strategic issues, when accurately cared for, can be seen as the foundation of a well-organized, properly established and potentially effective management system. An organization with clear objectives without explicit values is, however, in danger of encouraging the employees to use any methods they choose to achieve goals. Hence, to achieve an effective system, the organizational values must support the strategies and be reflected in the operational methodology used. Tools and methods applied must comply with this view, and with the needs and characteristics of the given context.

2.1. Organizational Values

Corporate philosophies, values, beliefs and socio-structural systems influence the specific culture of an organization. The general assumption is that managers can control the culture with policies, visions, tools and techniques, thereby obtaining a strong organizational culture believed to lead to better performance. Implementing an integrated system will depend upon existing structures and values, as will it influence the same (Axelsson 1999).

The quality movement as well as the ergonomics movement inherits their own philosophies, ideologies and values; however, in essence they do have many similarities. As the two disciplines arouse about the time we find a dynamic development of quality and work life issues (Axelsson and Bergman 1999). The interactive maturation of quality philosophies and work life doctrines has made the two concepts share several basic beliefs.

The basic assumptions of the quality movement and the ergonomics movement reflect the basic values these movements build on.

As can be seen in Table 1, the values of the two disciplines are partly in concert and partly complementary. TQM embraces certain humanistic approaches regarding the improvement of

Table 1. Basic assumptions and values of the quality and ergonomics movements.

The quality movement	The ergonomics movement
1. Good quality is profitable	1. Good ergonomics improves performance
2. Everything can be improved	2. Fit tasks and equipment to the human
3. Employees try to perform good quality	3. Employees try to perform a good job
4. Organisations are open systems	4. Organisations are open systems
5. Quality is top-management responsibility but everyone's concern	5. Ergonomics is interwoven with participation
6. Learning is essential	6. Developing work is essential

quality as well as the development of people and work life. It is commonly believed that to achieve high external quality the requirements and needs of the employees (internal customers) must first be meet. People in the organization are the source of continuous improvement and provide the driving force for it. Their knowledge, experience and involvement are needed to achieve long-term success. The socio-technical philosophy emphasizes the necessity of being able to influence the quality output in the job, and has also as a basic assumption that people strive to try to do a good quality job.

Achieving cultural change is central to a human-centered quality system. In this way, the integration of quality and ergonomics becomes an approach or a philosophy, a way of taking account of all customers in the long-term, internal as well as external. As a concept, quality management as well as ergonomics concerns realizing the potential of people — getting people involved, motivated and open to change. According to Kondo (1993), "a company is its people" and however excellent an organization is, it will be useless or even counterproductive if the people in it lack motivation.

2.2. Strategic Management

The rationale for and purpose of strategic planning has been discussed for many years and it represents a broad continuum of perspectives. In the past decade, many of the principles of quality management have emerged into principles of strategic management. For example, popular concepts of customer satisfaction, continuous improvement, process orientation, employee involvement and learning are today closely related to strategic concepts such as mission, objectives, motivation and corporate culture. The concept of worker empowerment has clearly emerged as central to the effective implementation of these changes, and the anticipated performance benefits are regarded as primarily a consequence of a more satisfied and committed workforce. In this context an introduction of ergonomics can play a major role, improving job satisfaction, performance and quality of working life. Macro-ergonomics has even been suggested as a TQM strategy and participatory ergonomics as an approach, highly applicable in the context of quality improvement (Noro and Imada 1991).

Yet, in reality we find that the quality of products or services is often considered an issue of strategic management whereas ergonomics is not to the same extent treated as an integrated part of business. There are, however, those who strongly argue that to achieve long-term success, the strategic goals must be balanced — taking the financial, customer and human perspective into consideration. If the system as a whole lacks the adequate focus on the integrational aspects of quality and human issues including ergonomics, the consequences can contradict the purpose. The main issue is, therefore, how to design effective integrated strategic systems that realizes the potential of synergies between the two disciplines.

To realize such a participatory system integrating ergonomic issues at all levels, a balanced strategic management concept based on Policy Deployment and Balanced Scorecard has been proposed (Axelsson 1999). This system provides a set of indicators that stress the need to include a wide range of performance indicators. Thus, the door is opened for ergonomic-based performance indicators and an integrative improvement approach is supported.

The holistic systems view recognizes employee satisfaction and capabilities as strong driving elements in the profit chain — linking ergonomic factors to excellent internal processes and satisfied customers, ultimately resulting in higher financial performance. Furthermore, to support a participatory planning process a "catchball" strategy is applied — taking strategic issues to the grassroots level, asking employees at each level of management to "add value" to the plan based on data analysis and experience of their functional areas or processes. An additional purpose is to ensure that strategies, objectives and measures are well communicated, understood, balanced, realistic and sufficient to achieve the objectives.

The proposed integrated management system represents a highly effective strategy for participation and enhanced motivation as well as fulfilling its obvious purpose of company alignment and customer control and improvement. By focusing on problems and possibilities in the working environment this connection can be made clear, but there is still a need for participatory problem-solving arenas supported by appropriate methodology.

2.3. Operational Methodology

As discussed above, designing and improving a system in accordance with ergonomic principles can be seen as a quality issue in which the internal customer (employees) requests of ergonomics are given a high priority. In addition, performing good quality at work can be seen as an important prerequisite of job satisfaction and job motivation. However, without a clear understanding, active commitment, involvement and support from all employees, quality performance cannot become a reality. In this context the development of people and their participation in improvement activities is recognized as a key feature — today a highly important element of TQM as well as a current concern of ergonomics.

The issue of participative problem-solving has many dimensions. First, there are various techniques for employee involvement, for instance suggestion schemes, quality circles, cross-functional teams, task forces, business councils and self-managing work-groups. Second, these techniques can be assigned to different levels of participation, being more or less user-controlled, active or passive, organization wide or group oriented, etc. Third, there is a variety of tools used to support improvements, each depending on the technique adopted, and the problems and possibilities encountered. From a quality point of view the Seven QC (Quality Control) tools are seen as the most important tools for use in participatory problem-solving — investigating the causes of why things are not working as expected, so that root causes may be identified and eliminated. Definition of the seven tools varies slightly with different writers, but Professor Kaoru Ishikawa as defines it:

1. Pareto chart;
2. Cause and effect diagram;
3. Stratification;
4. Check sheet;
5. Histogram;
6. Scatter diagram; and
7. Graph and control chart.

All these tools can be used in quality as well as ergonomics improvements activities, in conjunction often to the advantage of both areas (Axelsson 1999). The following example illustrates
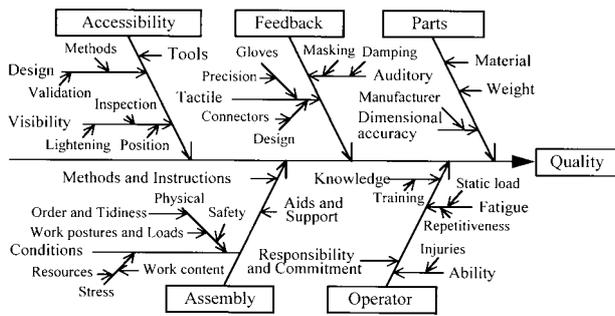
Figure 1. Cause and effect diagram illustrating problems in the subassembly of a microwave oven.

Figure 2. Scatterplot showing the relationship between physical loading and quality.

a cause-and-effect diagram used to find the root causes of quality problems in a subassembly task of a microwave oven (Figure 1).

The cause-and-effect diagram provides an excellent basis for participatory problem-solving and can in addition give a clear indication as to whether more data is required and how it is to be collected. The mapping of possible causes, in the example above, clearly indicates that quality problems in large relates to ergonomic difficulties and show that it is not enough just to rely on quality methods to carry out improvements. Many of the problems call for ergonomics methods and techniques. Noro and Imada (1991) suggest a blend of tools for participatory ergonomics initiatives in continuous improvements — containing the 7QC tools as well as ergonomics methods such as link and time analysis, layout modeling and mock-ups, check-lists, man–machine charts and five ergonomics viewpoints.

The Deming PDCA cycle (Plan–Do–Check–Act) for structured problem-solving is another example of a participative method for quality as well as ergonomics improvements. Originally, this was applied so that managers planned, blue collar workers did the job and white collar workers checked it. This was found inappropriate, since it is not until each person (on his/her level) is involved in the whole cycle that the full potential effects on learning, motivation and improvement takes place.

2.4. Integrated Applications
In Axelsson (1999), a participatory problem-solving effort performed by members of a problem-solving group is described. The improvement activity includes a variety of methods, such as SPC charts, 7QC tools, surveys, the RULA method (Rapid method for Upper Limb Assessment), video recordings, walk- and talk-throughs, and brainstorming. To illustrate the successful combination of ergonomics and quality methods the scatterplot in Figure 2 may serve as an example.

The scatterplot examines the hypothesis of a relationship between quality and physical loading on the operator, and as shown in Figure 2 the mean quality deficiency rate was found to be almost 10 times higher for the worst posture compared with the best posture.

Ergonomic tools are well suited in the quality improvement process, and participatory ergonomics can be used not only as an effective tool eliminating ergonomics problems, but also progressively to improve working conditions and performance — stimulating a positive development of quality. To identify such possibilities there is a need to take a holistic viewpoint on production. Generally speaking, most of the work environment
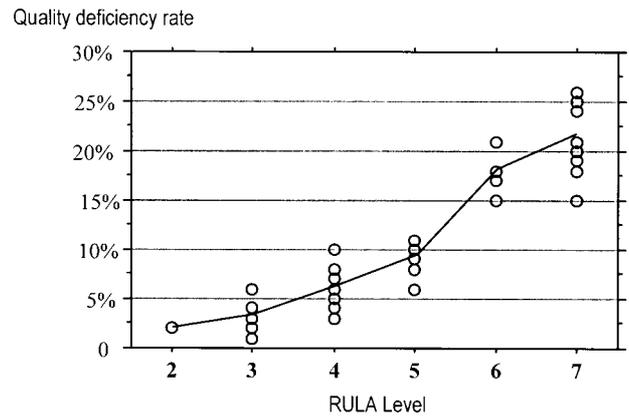
factors influencing human performance will in fact also have an effect on quality. The integrated applications are many and will of cause depend on the context. Some relations can, however, be generalized, as pointed out below.

3. QUALITY SHAPING FACTORS OF THE WORK ENVIRONMENT
There are many studies in the ergonomics literature reporting influences of work environment on performance. In particular, inspection tasks have been subjected to such investigations. Effects on product quality due to ergonomics conditions would, therefore, be expected. Also, ergonomics requirement may also be seen as a part of quality in the context of product design. Eklund (1997) and Drury (1997) have reviewed these relationships, some of which are presented below.

3.1. Noise and Vibrations
There is intentional information in many kinds of sounds, for example human speech and warning signals. Some sounds also include unintended information from work processes, such as the cooling fan of a hand tool, revealing the rotation speed of the tool. There is also pure noise, which by definition is unwanted. Being able to distinguish the sound cues that are needed to diagnose the work process is crucial to perform a good quality job. Noise will have a masking effect and can also impair verbal communication. Arousal will be increased by noise but decreased by infra-sound, a factor which is strongly related to performance. Noise has also been shown to impose stress reactions, disturb concentration, and cause distraction and lapses of attention. Of course, hearing loss and the use of noise protectors create increasing difficulties when trying to hear important information. Vibrations can impair the finger sensitivity in the short-term, and also can cause vibration white fingers and other occupational disorders in the long-term. Difficulties in both hearing and perceiving sensations with the fingers consequently cause performance decrements and increase the risk of errors. If the eyes or the viewed object vibrate, the ability to see small objects deteriorates. One such effect of vibration is that the error frequency in reading tasks may increase. In manual control tasks, vibrations can cause errors due to unwanted movement of the controls. Vibrations at frequencies > 20 Hz can interfere with the neuromuscular processes and thereby cause errors.

One example is assembly work with snap-on fasteners. Increased noise levels as well as the use of gloves increased error rate in production (see the example in Figure 1). Another example is an experienced machine operator who immediately will hear if there is a defect ball bearing, and thus be able to stop the machine for repair before a serious break down occurs. A third example is surface treatment in carpentries. The same persons that frequently use hand-held oscillating grinding machines also have the task to control the smoothness of the surface by touching it with their hands and fingers.

3.2. Climate

Several studies have identified loss of speed and precision of finger movements as a result of a relatively small lowering of finger temperature. This can occur in ordinary offices where low physical activity in combination with light office clothing may result in finger temperatures ~25°C. Outdoor work in too low temperatures might result in deteriorated coordination and difficulties also with upper limb movements. Frost bite in a finger causes deterioration of touch sensitivity in the finger years after the event, a factor the individual is not always aware of. In mental tasks, the error rate has also been shown to be higher already for small deviations from peoples preferred temperature. Further, the accuracy of learning decreases with both high and low temperatures. High temperatures cause sweating, which not only distracts the attention and affects concentration, but also might result in a slippery hand grip. In an unsuitable climate, there has also been an increased rate of accidents recorded, a factor which indicates decreased human reliability.

3.3. Chemical Health Hazards

Some chemical compounds may impair the function of the nervous system. Several other systems might also be affected, including the immune, digestive and respiratory systems, etc. Some effects might be acute and temporary, while others might be chronic and permanent. Solvents and lead are two examples of chemicals with a potential of causing a wide range of symptoms from affecting the nervous system, including nausea, dizziness, head ache, concentration difficulties, memory losses, balance disturbances, longer reaction time, impaired dexterity and even personality changes such as sudden aggressiveness. All kinds of symptoms and impaired physical well-being constitute a risk of distraction from the task and a quicker onset of fatigue. In particular, physical performance deteriorates in terms of loss of speed and precision of motor tasks.

The use of protective equipment is a secondary effect due to deficiencies in working conditions. In general, such equipment may have many adverse effects such as being heavy or uncomfortable to use, making natural movements or activities more difficult, and impairing perception, communication and information retrieval. All these aspects influence human performance and accuracy. In many cases, gloves are worn for protection against health hazards such as chemicals, low temperatures or cutting injuries. As a consequence, precision decreases and the tactile feedback, sometimes necessary to perform a good job, disappears.

3.4. Light

Many studies have identified increased rates of misjudgments and inspection errors due to insufficient light. There may be many contributing factors to this, such as too low light levels, insufficient color rendering, non-optimal luminances and disturbing reflections. Similar problems have been observed in tasks such as proof-reading of texts with poor visibility. In general, jobs including visual work that put high stress on the eye functions may lead to a lower quality of the work result. There are many documented cases from production industries where increased illumination levels have resulted in substantial reductions in rejection rates or wastage. Reductions of ~40% have been reported. In a similar way, substantial increases in the number of defects found in inspection tasks have been seen after improvements of the light. For example, surface irregularities, such as dents, wave shapes or defects in planeness, can be detected more easily with a special light source, giving light at a low angle to the surface. Judgement of the smoothness of painted surfaces improves if light is directed so that the inspector can view the reflex. Judgements of color deviations are aided by light sources with a high-color rendering index in combination with daylight properties. Scratch marks can easier be judged in a diffuse light from light sources with large light emitting areas or from indirect light.

Complaints on light conditions are rare, but an important determinator of the accuracy of quality judgements. Acute symptoms such as headache and eye strain may arise in visually demanding tasks. Further, the effects of insufficient light on melatonin levels could lead to disruption of the circadian cycle, which can cause tiredness and depression symptoms.

3.5. Physical Work Load

The precision and strength capability of body movements varies between different directions of movements and between different muscles in the body. Heavy and strenuous tasks and awkward postures often result in increasing discomfort during the working day, in addition to increasing the risks of musculoskeletal disorders. In such situations, performance of the work and extending it with additional corrections and improvements imposes even more discomfort. Thus, the willingness to improve borderline quality diminishes and quality deficiencies increase with increasing discomfort level. The quality deficiency rate increases substantially for adverse work postures compared with good postures. Several studies have identified a relationship between quality deficiencies and ergonomically problematic work tasks to the extent that ~30–50% of all quality remarks are related to or directly due to ergonomic problems (Eklund 1997, Axelsson 1999). The component design influences assembly ability and the ergonomics work situation which in turn will influence the quality outcome. Some aspects of this, which have been identified in assembly work, are given in Table 2.

Table 2. Examples of component design aspects for improved quality in assembly.

Design–features				
Visibility	Sequence	Weight	Friction	Hardness
Position	Size	Colour	Fragility	Fittings
Layout	Shape	Texture	Heat conductivity	Force

3.5. Physical Ergonomics Influences on Quality

The above results point to that there are several different ways through which ergonomics conditions influence quality. In particular, adverse environmental and physical conditions cause quality deficiencies through several mechanisms. These include the following mechanisms:

- impaired ability to perceive;
- masking of information/cues/signals;
- impaired ability to interpret and decide;
- impaired ability to perform and exert the task;
- reduced feedback; and
- distraction from the main task.

When humans feel discomfort, especially if it is caused by the task, this seems to impair performance and to increase the probability of quality errors or deficiencies. One explanation may be that discomfort may cause distraction and lapses in attention, or lead to changed activity or compensatory activities that compete with the main task.

3.6. Psychosocial and Organizational Aspects

Short work cycles in paced jobs lead to increased rates of quality deficiencies. One reason is the variability in time needed for each operation, leading to unfinished or deficient operations when slight problems occur. Repetitive and monotonous jobs give rise to symptoms of boredom or fatigue, followed by a decrease in performance in terms of longer reaction time and an increased error rate progressively during the work periods. There are also indications that increased work content in simple jobs seem to improve product quality. Status differences and tensions between categories of workers in a hierarchical work organization have been related to quality deficiencies.

Several studies have documented that production philosophy, work organization and personnel policy, as well as wage form may influence quality. It has been reported many times that motivation and job satisfaction leads to improved product quality. Among important factors for this are a continuous interest in the work, challenge, recognized achievement, ownership and communication. Further, companies known for quality are also often characterized by a concern for their personnel.

4. SUMMARY AND CONCLUSIONS

There is no short cut for how an organization is to succeed with integrating quality and ergonomics, but there are studies that describe important characteristics in the work environment as well as the development of integrated systems and processes. Integrative approaches incorporate company values, strategic management, operative methodology and applications. A top-down strategy including management support is seen as a prerequisite for successful results, but the bottom-up participative strategy is also an important prerequisite. Commitment and motivation are important factors in this development, but at the same time they make demands on how the work should be structured. The integrated approach consequently assumes the combination of a top-down and a bottom-up strategy. Thus, all employees must be given the necessary conditions for doing a good job. This also points to the necessity to integrate values and management strategies with simple tools for improving work conditions, i.e. a joint effort of quality and work development. But such a process does not run by itself. It will though be supported by a view that problematizes and develops the interactive and synergistic area that exists between concentration on quality and ergonomics, both operative and strategic.

REFERENCES

AXELSSON, J.R.C., 1999, Ergonomics and quality — towards successful integration. PhD thesis, Linköping Institute of Technology.

AXELSSON, J.R.C. and BERGMAN, B., 1999, The dynamics of quality and work life improvement. In G. Edvardsson and H. Gustavsson (eds) *The Nordic School of Quality Management* (Lund: Studentlitteratur).

DRURY, C.G., 1997, The Ergonomics Society Lecture 1996: Ergonomics and the quality movement. *Ergonomics*, 40, 249–64.

EKLUND, J., 1997, Ergonomics, quality and continuous improvement — conceptual and empirical relationships in an industrial context. *Ergonomics*, 40, 982–1001.

KONDO, Y., 1993, *Human Motivation — A Key Factor for Management* (Tokyo: 3A Corp.).

NORO, K. and IMADA, A. (eds), 1991, *Participatory Ergonomics* (London: Taylor & Francis).

Quality Inspection Task in Modern Manufacturing

J. A. PESANTE-SANTANA[+] and J. C. WOLDSTAD[+]

†Department of Industrial Engineering, Polytechnic University of Puerto Rico, San Juan, Puerto Rico 00919-2017

‡Department of Industrial Engineering, Texas Tech University, Lubbock, TX 79409-3061, USA

1. INTRODUCTION

The development of numerically controlled machines, group technology, cellular manufacturing and just-in-time (JIT) production systems have revolutionized the way products are designed and manufactured. These technological and strategic advances have changed the role of human operators in the manufacturing environment. The highly specialized work force of the low-tech manufacturing system has evolved into the multi-skilled work force of the high-tech manufacturing system. Among the multiple tasks that an operator is expected to perform in advance manufacturing systems (AMS) are job scheduling, inventory planning, machine set-up, problem-solving and quality inspection.

Throughout this evolution, human sensory detection capabilities have been a vital but often ignored component of the quality inspection task. Although automation is often employed to construct and assemble products within AMS, most inspections and quality checks are still done by human operators due to the inherent problems in machine vision and decision-making. While humans remain responsible for inspection, it has been widely accepted that the quality inspection task performed by humans is prone to error. Some studies indicate human inspectors typically find only ~80% of the defects. Despite the contributions of human factors research to the understanding of human performance in the quality inspection task, the manufacturing trend has been to design quality schemes that compensate for poor inspector performance instead of trying to improve it (Drury 1992).

2. A VISUAL QUALITY INSPECTION TAXONOMY

Quality is usually defined as fitness for use, or the extent to which a product meets the consumer requirements. Inspection is the act of measuring or examining carefully the quality of a product. Sensory inspections (performed by means of the human senses to assess a product's qualitative characteristics) and physical inspections (performed by means of measuring devices to assess a product's quantitative characteristics) are the main types of quality inspections. The visual quality inspection task has been described as consisting of the following subtasks (Wang and Drury 1989): (1) orient the item, (2) search the item, (3) detect any defect, (4) recognize/classify the defect, (5) decide the status of the item, (6) dispatch the item, and (7) record the information about the item. These subtasks can be combined into two main components: search and decision-making. Thus, the simplest description of the visual quality inspection task is to: search, recognize a defect and make a decision on the part's acceptability within the quality limits.

2.1. The Visual Search Component

Visual search is a sequential process that proceeds as a series of fixations linked by eye movements and which terminates upon successful detection of a defect or the complete inspection of the unit. It has been shown that almost all of the information in a visual search is obtained during the fixations, which account for > 90% of the search time. In the inspector's field of view, a defect is only visible within a limited area referred to as the visual lobe. During a fixation the visual lobe is located around the central fixation point. The visual lobe size will be affected by the luminance of the object inspected, the contrast between the object and the defect on the object, the defect size and the distance of the defect from the inspector's eyes. Megaw and Richardson (1979) conducted eye movement studies of inspectors and concluded that inspectors do not follow a simple pattern or strategy in searching an object. They observed that while a very random appearing search pattern was used for the inspection of complex units (e.g. circuit boards), a more systematic search pattern was used for the inspection of simpler ones (e.g. knitwear).

In addition to the lobe size and the search strategy, the time available for the inspection will affect human performance in the visual search component of the inspection. The more time the inspector has to search, the better the chances are of finding the defect. For the visual search subtask, the best inspectors are those who use fewer, longer fixations, as compared with those with a larger number of shorter duration fixations.

2.2. Decision-making Component

Given its strict relevance to decision-making, signal detection theory (SDT) has been used to explain the decision-making component of the quality control inspection task. In a quality inspection context, SDT proposes that the human, functioning as a defect detection device, builds up in the neural system two distributions of activity: one relating to the probability of accepting a unit, the other to the probability of rejecting it. The degree of separation of these two distributions' means is a measure of the inspector's discriminability of the defects (d¢). The criterion level (b), which is the ratio of the two ordinates of the curves at a given level X_c, delineates the boundary between accepting and rejecting a unit, and in doing so takes in some good units to be rejected and some faulty units to be accepted. Inspectors make a correct decision either by accepting a good unit (correct rejection) or by rejecting an unacceptable unit (hit). They fail either by not detecting a rejectable defect (miss) or by falsely reporting the presence of a rejectable defect (false alarm).

Both decision-making performance measures (d¢ and ß) are derived from the hit rate and the false alarm rate. The pure decision-making component of the quality inspection task can be measured by concentrating on tasks that require no search. The general conclusion of quality inspection studies reviewed by Drury and Fox (1975) is that "the decision-making component is among those rare tasks where a human being behaves like a rational economic decision maker, balancing the costs and payoffs involved to arrive at an optimum performance." As a normative model, SDT defines the optimal criterion ($ß_{opt}$) used by the ideal observer to optimize economic gains. Based on the values of a hit and a correct rejection, and the costs of a miss and a false alarm, $ß_{obs}$ of a rational observer can be calculated and compared with that of the theoretical ideal observer ($ß_{opt}$). After comparing

ß$_{obs}$ with ß$_{opt}$, it has been generally found that rational observers do fairly well at optimizing their winnings. However, they tend not to set extremely low or high criterion, even in situations where these strategies would lead to optimal performance.

3. SIGNAL DETECTION THEORY MODELS OF INSPECTION

SDT was first used to model the decision-making performance of the quality inspection task by Wallack and Adams in 1969. After using SDT to study the performance of industrial electronics inspectors in a visual, subject-paced task, they concluded that SDT performance measures (d¢, ß) were more useful than the other available measures. Wallack and Adams concluded that SDT is useful because in addition to relating performance to payoff, it also indicates the magnitude and the direction of improvement required. Although not all of the research using SDT to study the decision-making performance has been conducted in an industrial inspection context, its findings have been beneficial in understanding human quality inspector performance.

More recently another measure of criterion level or decision-making response bias called index *c* has been developed. The main difference between ß and index *c* is in the way these bias indices locate the criterion (X$_c$). The likelihood ratio measure, ß, locates X$_c$ by the ratio of the heights of the SDT distributions (N, SN), while index *c* locates X$_c$ by its distance from the intersection of the two distributions. The range of *c* is, therefore, the same as that of d¢, although zero is at the center rather than an endpoint. This parametric index is considered to be more effective than ß over a full range of sensitivity in recognition memory experiments and in vigilance. Unlike index *c*, ß has a tendency to produce extremely high values for conservative observers resulting in a scale that does not produce equal intervals. Index *c* is generally more sensitive than ß to the non-perceptual manipulations of signal probability, payoff and probability shifts. Also, ß is less effective than *c* at differentiating variations in response bias tendencies when sensitivity approached chance.

In most if not all of the research using SDT, the quality inspection task has been characterized as a vigilance situation in which the inspector's sole task is to examine a stream of products to detect and remove the defective ones. This characterization is no longer consistent with the reality of the operator's responsibilities in modern manufacturing, also known as advanced manufacturing systems (AMS). The quality inspection task in AMS is no longer a specialized task; instead, it is one of multiple dissimilar tasks conducted by a highly skilled operator. Recent research indicated that the performance of the operator in the quality inspection task while multitasking in an AMS will be determined by the interaction between the number of different types of defects that can be presented at the same time in the inspected parts and multitasking (Pesante-Santana 1997).

3.1. Changes Over Time in the SDT Parameters

Based on the review of 12 studies conducted between 1969 and 1975, Swets (1977) indicated that all 12 experiments showed an increasingly strict criterion (ß) over time when the signal-to-noise ratio was low; eight of the experiments showed a constant sensitivity (d¢) over time. In the four studies in which d¢ did not remain constant, it was found to decrease by 20%. An increase in ß (conservative criterion) represents a decrease in signals detected as well as a decrease in false alarm errors. In general, the performance changes over time have been characterized by a shift in the subject's response criterion (ß).

3.2. Determinants of Decision-making Performance

It is often reported that the decision-making performance is affected by the payoff matrix, knowledge of results (KR) and signal ratios (fraction of defective). Conservative or stringent decision-making should be expected if the operator knows that the cost of an incorrect decision may be a disciplinary action, a monetary penalty or job termination. A similar conservative decision should be expected if the process defective fraction is low (signal ratios), or if the operator receives information (knowledge of results) about specific defects or process conditions that needs to be detected for the benefit of the production area. When none of the conditions previously mentioned are not present a lax decision-making performance should be expected from the operator. However, it is important to recognize that a key element in the decision-making performance is the training. The operator's decision-making performance will not be consistent with the scenarios described in this section if s/he does not have the appropriate training on quality inspection.

4. QUALITY INSPECTION TRAINING

Training is essential to improving the decision-making performance of human operators. Many authors have reported that deficiency in knowledge (education or training) and or feedback (knowledge of results) is a cause of errors consistently made by inspectors. Drury (1992) listed five techniques that have proven effective in training for inspection: (1) cueing, (2) feedback, (3) active training, (4) progressive parts and (5) develop schema.

5. HYBRID INSPECTION SYSTEMS

Zero-defect products and shorter lead-time production are vital for the survival and success of modern manufacturing businesses in the current highly competitive world class manufacturing environment. According to Drury and Sinclair (1983), "this can often be achieved only by 100% inspection, which is known to be unreliable when performed by humans." This dilemma prompted an industry movement towards automated inspection systems. The advent of microprocessor-based automated inspection devices, at prices competitive with human inspection, called for a human factors reassessment of the human–machine function allocation possibilities in quality control (Drury and Sinclair 1983). Gramopadhye *et al.* (1992) proposed a framework for function allocation in inspection. They recommended accuracy, speed, flexibility and reliability as the performance criterion for the inspection system.

One of the first efforts to reassess the human–machine function allocation possibilities in quality control was conducted by Drury and Sinclair (1983). They compared the performance of experienced inspectors and a prototype optical/microprocessor inspection device. The main findings were: (1) neither the human nor the automated systems achieved an outstanding performance and (2) the automated system was better at locating the defects (search) but could not classify them as acceptable or rejectable (decision-making) as well as the human inspectors.

Using the accuracy and speed performance criterion measures combined with the false alarm rate and the hit rate, Hou *et al.* (1993) calculated a cost-based evaluation function. This cost-based evaluation function coupled with the Drury and Sinclair (1983) findings led to the conclusion that allocating the search function to machines and the decision-making function to humans results in better performance than pure human or pure machine inspection. This computer–search/human–decision-making system is known as a hybrid inspection system (HIS). The idea behind HIS is to capitalize on the machine speed and precision to scan the inspection unit, and on the decision-making ability exhibited by humans.

6. QUALITY INSPECTION TASK LOAD

Quality inspection tasks that impose a sustained load on working memory (to recall what the quality acceptability criterion looks like) will demand the continuous supply of processing resources. Parasuraman (1979) conducted an experiment using a successive-discrimination task (which imposed a memory load) in which the signal was specified as the decrease in the intensity of a flashing light. The signals were presented irregularly at a mean rate of two signals min^{-1}, and the event rate was 30 events min^{-1}. The duration of the task was 45 min. He concluded that the performance in such areas of vigilance application as radar monitoring and industrial quality inspection could be adversely affected when the operator has to discriminate a signal from a standard represented in memory and when the event rate is high. This performance decrement may result either from signal-data limits (weak signal in noise), or memory-data limits (quality of stored representation of the standard in delayed comparison memory tasks).

Like many other tasks, quality inspection has been identified as having an inverted-U-shaped relationship between task demand and performance level. The inverted-U theory states that for a given task there is an optimal level of workload or demand that yields the highest level of performance. A departure in either direction from the optimal level of work will result in a performance decrement. While most of the results of vigilance research support the right-hand side of the inverted-U theory (overload) there is a lack of support for the left-hand branch (assertion that the task performance level can be improved by increasing the load). McGrath (1965) conducted one of the first experiments that supported the left-hand side in a vigilance task. After comparing easy and hard visual monitoring tasks conducted concurrently he concluded that the presence of the hard task facilitated performance on the easy one. Wiener *et al.* (1984) conducted an experiment in which a control group performed a vigilance task (the signal was the decrease in distance between two dots presented on a computer screen), and a second group performed a one-dimensional compensatory tracking task in addition to the vigilance task. They found that the performance of the second group (vigilance and tracking tasks) in terms of signal detection exceeded the performance of the control group (vigilance task only). They concluded that these research results provided support for the facilitating effect of increasing the task load (left side of the inverted-U).

Some researchers describe the quality inspection task as being intrinsically boring. According to them, this explains why it is often the case that mild stress will increase the performance in

terms of detection and response time. However, Wickens (1992) has indicated that vigilance tasks with working memory loads are susceptible to interference from concurrent tasks. The results of a recent study on the effects of multitasking on the decision-making component of the quality inspection task were consistent with the information presented herein (Pesante-Santana 1997). The performance of the operator in the quality inspection task while multitasking in an AMS will be determined not only by the number of different types of defects that can be presented at a time in the inspected parts, but also by the mental processing resources required to meet the demand imposed by the multiple independent tasks and the memorized quality criterion. The best performance will be obtained when the additional tasks' load minimizes the monotony of the quality inspection task without interfering with the processing resources needed for the memorized quality criterion.

7. RECOMMENDATIONS

Whenever human operators perform the quality inspection a certain degree of error should be expected. However, some ideas oriented to minimize the quality inspection errors are:

- Identify the inspector with the best performance and understand his/her inspection strategy.
- Provide off-line training/practice on the actual task to be conducted using the strategy of the best inspector.
- Provide continuous feedback to the operators on their performance.
- Assure an appropriate task allocation to avoid unwanted task loading.
- Use the appropriate payoff matrix to improve the operator's performance.
- Let the proper process information (KR) flow among the operators.

REFERENCES

DRURY, C.G., 1992, Inspection performance. In G. Salvendy (ed.) *Handbook of Industrial Engineering* (New York: Wiley), 2282-2314.

DRURY, C.G. and FOX, J.G., 1975, *Human Reliability in Quality Control* (New York: Taylor & Francis).

DRURY, C.G. and SINCLAIR, M.A., 1983, Human and machine performance in an inspection task. *Human Factors*, 25, 391–9.

GRAMOPADHYE, A.K., DRURY, C.G., SHARIT, J. and SUDIT, M., 1992, A framework for function allocation in inspection. In P. Brödner and W. Karwowski (eds) *Ergonomics of Hybrid Automated Systems III* (Amsterdam: Elsevier), 249–57.

HOU, T., LIN, L. and DRURY, C.G., 1993, An empirical study of hybrid inspection systems and allocation of inspection functions. *International Journal of Human Factors in Manufacturing*, 3, 351–67.

MCGRATH, J.J., 1965, Performance sharing in an audiovisual vigilance task. *Human Factors*, 7, 141–53.

MEGAW, E.D. and RICHARDSON, J., 1979, Eye movements and industrial inspectors. *Applied Ergonomics*, 10, 145–54.

PARASURAMAN, R., 1979, Memory load and event rate control sensitivity decrements in sustained attention. *Science*, 205, 925–7.

PESANTE-SANTANA, J.A., 1997, The effects of multitasking on quality inspection in advanced manufacturing systems. Doctoral dissertation submitted and approved at the Industrial and Systems Engineering Department, Virginia Polytechnic Institute and State University, Blacksburg.

SWETS, J.A., 1977, Signal detection theory applied to vigilance. In R.R. Mackie (ed.) *Vigilance: Theory, Operational Performance and Physiological Correlates* (New York: Plenum), 705–18.

WALLACK, P.M. and ADAMS, S.K., 1969, The utility of signal detection theory in the analysis of industrial inspector accuracy. *AIIE Transactions*, 1, 33–44.

WANG, M.J. and DRURY, C.G., 1989, A method of evaluating inspector's performance differences and job requirements. *Applied Ergonomics*, 20, 181–90.

WIENER, E.L., CURRY, R.E. and FAUSTINA, M.L., 1984, Vigilance and task load: in search of the inverted U. *Human Factors*, 26, 215–22.

WICKENS, C.D., 1992, *Engineering Psychology and Human Performance* (New York: HarperCollins).

Quality Management, Continuous Improvement and Total Quality Management

K. Zink

Universität of Kaiserslautern, PO Box 3049, D-67653 Kaiserslautern, Germany

1. BASIC PRINCIPLES AND DEFINITIONS

When talking about quality management, continuous improvement, and total quality management, it is necessary to define these terms and to show where they originate. But first we have to clarify what we mean when using the term "quality".

1.1 Quality

The word "quality" is derived from the Latin "qualitas", meaning "of what kind or sort". As quality is a word in every day use there are many meanings. Related to quality assurance activities, quality was often described as the quality of a result (e.g. a product). So the dimensions of quality could be described by features like performance, reliability, maintainability, and safety. During recent decades this more technical and objective understanding of quality has had to be completed by a subjectively perceived quality, and so other dimensions like features of the product, aesthetics or perceived quality have had to be added. Apart from the quality of a product, the quality of services is gaining more and more importance, adding dimensions such as responsiveness, courtesy or communication. This development is reflected in the latest version of ISO 8402 from 1995 defining quality as "the totality of characteristics of an entity that bear on its ability to satisfy related and implied needs".

The term "entity" includes not only products but also processes, organizations or persons. Products in the context of ISO 8402 (DIN 1995, pp. 3, 4) are described as "the result of activities or processes and can be tangible or intangible, or a combination thereof".

At this point it is already obvious that there is also a broader scope in the ISO (e.g. including processes and services), but does fulfilling stated and implied requirements contribute to long-term customer satisfaction? Many authors in this field do not believe so. They support the idea of an "attractive quality" that pleases the customer, because it is unexpected. Exceeding requirements only makes sense if customers are offered more or better solutions for their problems than they actually expect — and this leads to additional benefits for them.

In the context of total quality management an even broader understanding of quality arises.

Associating quality with normative aspects, working conditions must be addressed as a further element of a comprehensive quality understanding. This includes participation, technical, methodical or social training on the job, cooperation within and across departments, and providing adequate operating resources. Such task orientation of quality has been known for a long time as "quality of working life".

A comprehensive quality understanding also involves an organization's attitude towards the environment. Therefore, a quality of external relationships can be defined that relates to social responsibilities — for example, environmental protection

DIN EN ISO 9000

Quality Management and Quality Assurance Standards - Guidelines for Selection and Application

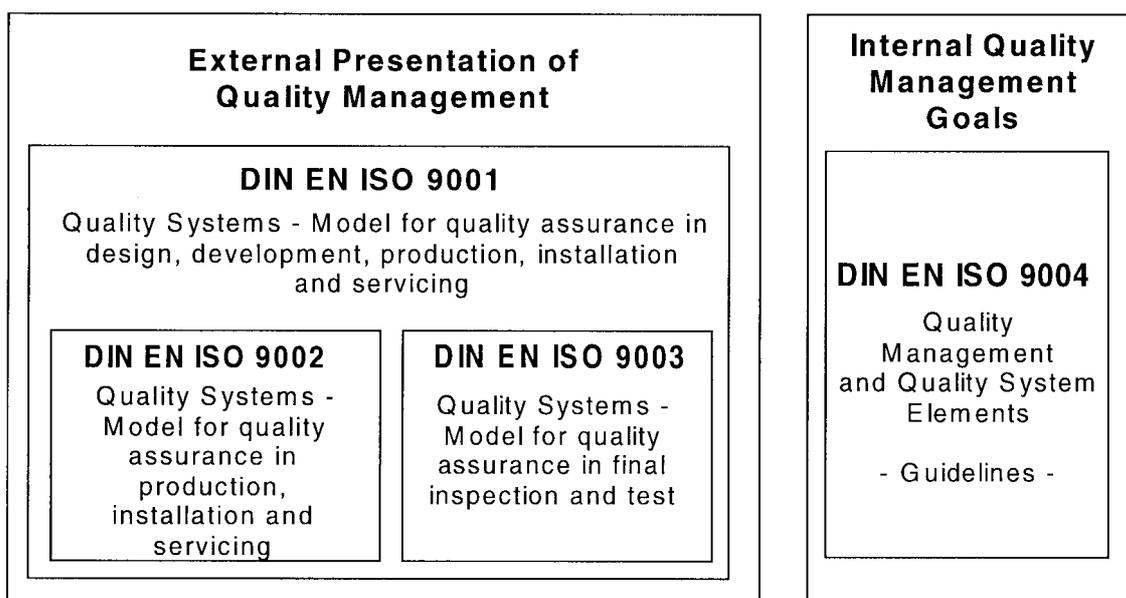| **External Presentation of Quality Management**	**Internal Quality Management Goals**
DIN EN ISO 9001 Quality Systems - Model for quality assurance in design, development, production, installation and servicing	**DIN EN ISO 9004** Quality Management and Quality System Elements - Guidelines -
DIN EN ISO 9002 Quality Systems - Model for quality assurance in production, installation and servicing **DIN EN ISO 9003** Quality Systems - Model for quality assurance in final inspection and test	

Figure 1. The DIN EN ISO 9000 series of standards

or an active role in the community. Environmental protection has an impact on products (waste management), processes (emissions and consumption of resources) and working conditions (work-related diseases).

Obviously, the elements presented here are not independent of each other. While the links between product (or service) and process quality are generally accepted, some still find it difficult to integrate the quality of working life or of external relationships into this concept.

The development of quality understanding described above can also be seen in the criteria for national or international quality awards. They started with a broader understanding of quality which in the meantime has been redefined as "business or organizational excellence".

1.2 Quality management

It is obvious that the definition of quality management is closely related to the definition of quality. Thinking of quality as a result of a process leads to a specific quality assurance which is mainly achieved by testing. Going back into history, the start of this approach is very much connected with the beginning of industrialization, especially Taylor's concept of scientific management which led to a separation of "thinking" and "doing" and brought quality assurance tasks to superiors or special departments. This development was enhanced by introducing Statistical Process Control (SPC) which was a consequence of mass production — and required relevant qualifications. In Japan during the 1960s and in the US during the 1970s, employee-oriented approaches arose. In Japan, the quality (control) circle as a means to include employees in (quality) problem solving was invented; in the US "Zero Defect Programs" tried to motivate people to do things right the first time. In the following years and up to the present time, quality has been seen as a management task too — and consequently quality assurance has developed into quality management. Coming back again to ISO definitions (ISO 8402), quality management is understood as including both quality control (as "operational means to fulfil the quality requirements") and quality assurance (as aiming at "providing confidence in this fulfillment both within the organization and externally to customers and authorities"), as well as the additional concepts of quality policy, quality planning, and quality improvement. quality management operates throughout the quality system. These concepts can be extended to all parts of an organization. "The quality system consists of the 'organizational structure, procedures, processes, and resources needed to implement quality management" (ISO 8402, 1995, p. 17). Quality management and quality assurance standards are described in the ISO 9000 series, for example. The following figure gives a general overview:

As mentioned before, the descriptions of quality management depend on the definition of quality.

1.3 Total quality management

Searching for the origin of total quality management (TQM), one has to refer to several sources. On the one hand, there is the Japanese way of "company-wide quality control" (Ishihawa 1985) or the US approaches of "total quality control" (Feigenbaum 1956). Whereas the content of TQM was already being developed and described in these early approaches, the term "total quality

management" appeared later — especially in the context of quality awards. In the meantime, even ISO has defined TQM as a "global management strategy" with "the participation of all members of the organization for the benefit of the organization itself, its members, its customers and society as a whole" (ISO 8402, 1995, p. 4). In practice TQM is mostly understood and evaluated on the basis of international quality awards. These awards have been created to promote the idea of a holistic quality understanding, which in the meantime (as mentioned above) was redefined as "business excellence" and also applied for example to non-profit organizations (NPOs) around the globe (Asia, America, Australia, and Europe) where such awards can be found. One of the latest developments is shown in the Figure 2.

As can be seen in figure 3, the European Quality Award (and also, for some years, the US Award) gives the same weighting to "enablers" and "results":

1.4 Continuous improvement

The idea of continuous improvement is not new at all, so that all the approaches that have been adopted for many years by both internal and external consultants driving the continuous

Australian Quality Award
1999 Criteria

1.0	**Leadership and innovation**	**(180)**
1.1	Strategic direction	60
1.2	Organisational culture	40
1.3	Leadership throughout the organisation	40
1.4	Environmental and community contribution	40
2.0	**Strategy and planning process**	**(100)**
2.1	Understanding the business environment	40
2.2	The planning process	30
2.3	Resources and assets	30
3.0	**Data, information and knowledge**	**(100)**
3.1	Collection and interpretation of data and information	30
3.2	Integration and use of information for decision making	40
3.3	Creation and management of knowledge	30
4.0	**People**	**(160)**
4.1	Involvement and commitment	60
4.2	Effectiveness and development	50
4.3	Health, safety and well-being	50
5.0	**Customer and market focus**	**(150)**
5.1	Knowledge of customers and markets	60
5.2	Customer relationship management	50
5.3	Customer perception of value	40
6.0	**Processes, products and services**	**(160)**
6.1	Innovation process	40
6.2	Supplier relationships	50
6.3	Management and improvement of processes	30
6.4	Quality of products and services	40
7.0	**Business Results**	**(150)**
7.1	Indicators of success	100
7.2	Indicators of sustainability	50

Figure 2. Australian Quality Award.

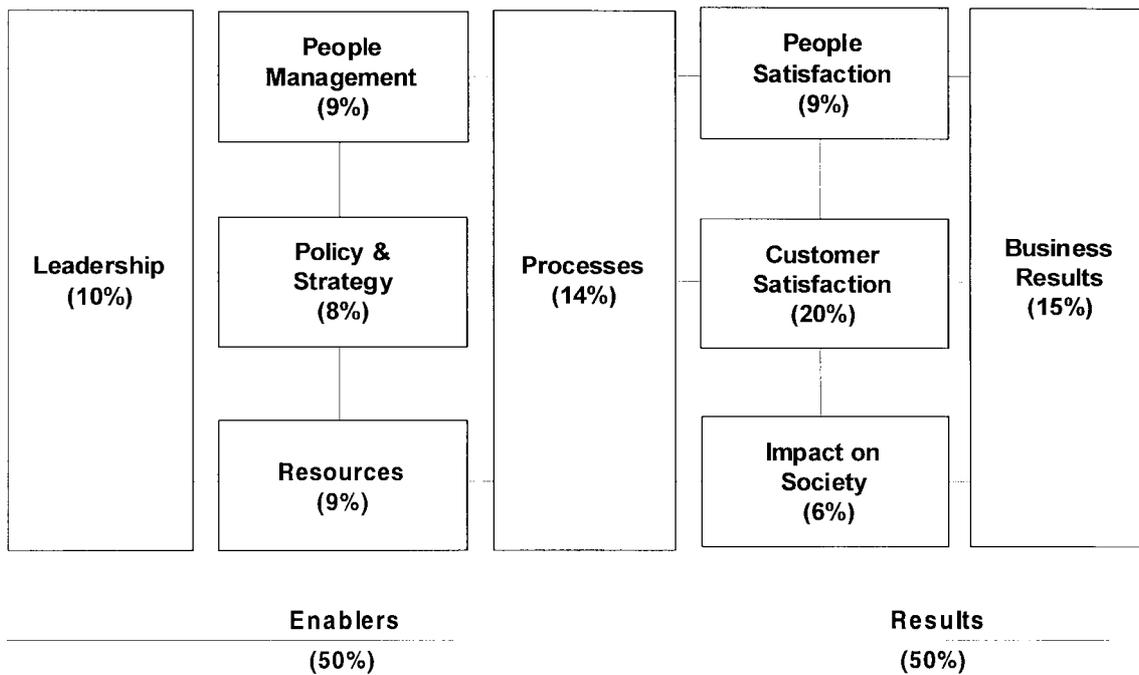| Leadership (10%) | People Management (9%) | Processes (14%) | People Satisfaction (9%) | Business Results (15%) |
| Policy & Strategy (8%) | | Customer Satisfaction (20%) |
| Resources (9%) | | Impact on Society (6%) |

Enablers
(50%)

Results
(50%)

Figure 3. The European Model for Business Excellence (European Quality Award)

improvement process will have to be included here. However, this process was, on the one hand, driven by experts (e.g. coming from industrial engineering) and, on the other hand, caused by the occurrence of specific weak points. After the World War II, instead of this externally driven improvement process, which was only partially "continuous", there were different routes of development in Japan and the Western countries. In the early 1950s and '60s, the Japanese discovered that employees are the real experts concerning problems connected to their work(place). They therefore introduced the so-called quality control circles (QCCs), which have been a huge success. Ishikawa especially saw QCC as one of the most important parts of company-wide quality control. Seeing this success, Western countries (i.e. Western Europe and the US) tried to introduce the idea of employee participation. However, although a success at first, the quality circles suffered after some time because, unlike in Japan, they were not part of a broader approach; thus many of the "quality circles" (as they were named in the West) were abandoned after two or three years, and "lean production" or "lean management" teamwork and continuous improvement became again topics for discussion. At this time the Japanese idea of "KAIZEN" was promoted, especially by Imai, who proposed the advantages of innovation in small steps using different concepts of continuous improvement (KAIZEN). As a result, many of the former quality circles have been revitalized by this concept. Imai gave the idea of continuous improvement a broader base by proposing three different target groups: individuals, teams, and management, and this is illustrated in figure 4 (Imai 1984).

2. THE COMMON DENOMINATOR AND INTERDEPENDENCIES

In comparing the three concepts, we find one common denominator — namely, the improvement of quality — but, as explained before, the definition of quality differs. While the former quality circles and (especially) "total quality management" refer

to a broad definition of quality management or quality assurance, the quality of products or services was, for many years, a central point. But here, too, the foreseeable developments show a more extended approach.

Based on this understanding, although total quality management may deliver the frame, quality management and continuous improvement are also part of TQM. In addition, quality management itself includes the idea of "quality improvement" (ISO 8402, 1995, p. 18), and in Imais's book KAIZEN, the term "total quality control" is the most used one. Therefore, there is no doubt that all three concepts are interdependent. Nevertheless a hierarchical approach, using TQM as the frame, is the most structured one.

As mentioned in defining total quality management , business excellence models play a predominant role as a basis for quality awards in (industrial) practice (even in NPOs). Using these models normally includes the tasks of self-assessment and benchmarking. The aim of these approaches is to find "strengths" and "areas for improvement". Carrying out a yearly self-assessment therefore leads to a structured identification of topics for improvement. Using this tool as (another) base for strategy and operative planning ensures that improvement becomes continuous (in contradiction to past methods). Since all the international award schemes include "process" as a central element, quality management systems are also covered.

Comparing the US Malcolm Baldrige National Quality Award Categories, for instance, or items of 1989 with those of 1999, it becomes obvious that the understanding of quality is now "business or organizational excellence" or, as termed in the Baldrige National Quality Program 1988, performance excellence.

3. TOPIC RELEVANCE FOR ERGONOMICS AND HUMAN FACTORS

As shown before, the field of quality has undergone a transition

	Management-Oriented KAIZEN	Team-Oriented KAIZEN	Individual Oriented KAIZEN
Tools	Seven statistical tools Seven new tools Professional skills	Seven statistical tools Seven new tools	Common sense Seven statistical tools
Involves	Managers and professionals	QC-circle (group) members	Everybody
Target	Focus on systems and procedures	Within the same workshop	Within one's own work area
Cycle (Period)	Lasts for the duration of the project	Requires four or five months to complete	Any time
Achievements	As many as management chooses	Two or three per year	Many
Supporting system	Line and staff project team	Small-group activities QC circles Suggestion system	Suggestion system
Implementation cost	Mostly small investments	Mostly inexpensive	Inexpensive
Result	New system and facility improvement	Improved work procedure Revision of standard	On-the-spot improvement
Booster	Improvement in managerial performance	Morale improvement Participation Learning experience	Morale improvement KAIZEN awareness Self-development
Direction	Gradual and visible improvement Marked upgrading of current status	Gradual and visible improvement	Gradual and visible improvement

Figure 4: The three segments of KAIZEN

from inspection of quality assurance to quality management and continuous improvement. The latest development may be characterized by total quality approaches. Generalizing this process we see more and more holistic concepts instead of fragmented ones, and this includes a change in attitudes. While, previously, quality was a primary task of (inspection) specialists, nowadays, top managers are dealing with the topic. Looking at ergonomics and human factors one can also note substantial changes in scope and issues of interest.

Although, in the past, micro-ergonomic approaches have been accompanied predominantly by the belief that ergonomic interventions only cause costs, we now have a broader conception of ergonomics. For some years organizational design and management or macro-ergonomics have been accepted, including participatory ergonomics, and outside industry new concepts such as "community ergonomics" have been defined. Thus, it can be said that not only is there a development in the (total) quality field to include more and more "new" human factors issues such as the "quality of working life" based on broad stakeholder

concepts, but also a "rediscovery" of the relationship between working conditions and manufacturing quality.

The extension of the types of issues in ergonomics/human factors include areas such as (cross-functional) processes, participation (also in continuous improvements) or quality of life in communities, which are very much related to quality topics like process management, continuous improvement or excellence councils as an approach to regional development.

The inclusion of "people management" issues in all quality awards — for example, participation, working conditions, care for people and people satisfaction — are promoting these topics to top management level. In this context, studies such as the one carried out at Harvard University entitled "The human equation: building profits by putting people first" (Pfeffer 1998) are very helpful.

Returning to the basics of quality management and ergonomics, these have several factors in common. For instance, both approaches are based on:
- prevention
- measurement

1315

● continuous improvement

The best results can be gained by participatory approaches — and both topics are only successful if they are management driven.

REFERENCES

Australian Quality Council, 1998, *Australian Business Excellence Framework 99* (St Leonards).

DEMING, W.E., 1982, *Quality, Productivity and Competitive Position* (Cambridge).

Deutsche Institut für Normung (DIN), 1995, DIN EN ISO 8402. 1995-08, Berlin.

EKLUND, J., 1997, Ergonomics, quality and continuous improvement: conceptual and empirical relationships in an industrial context, *Ergonomics*, 40(10), 982–1001.

European Foundation for quality management, 1997, *Self Assessment 1998: Guidelines for Companies* (Brussels).

FEIGENBAUM, A.V., 1956, Total Quality Control, *Harvard Business Review*, Nov.–Dec., 93–101.

IMAI, M., 1984, *KAIZEN: The Key to Japan's success* (New York), pp. 26—7.

ISHIKAWA, K., 1985, *What is Total Quality Control?* (New York).

JURAN, J.M., 1979, Basic Concepts. In Juran, J.M., Gryna, F.M., and Bingham, R.S. (eds.) *Quality Control Handbook, 3* (York/London/Singapore), 2.1–2.24.

JUSE (ed.), 1989, *QC Circle Koryo: General Principles of the QC Circle* (Tokyo: QC Circle Headquarters).

National Institute of Standards and Technology, 1998, *Malcolm Baldridge National Quality Award: 1999 Criteria for Performance Excellence* (Gaithersburg).

PFEFFER, J., 1998, *The Human Equation: Building Profit by Putting People First* (Boston).

ZINK, K.J., 1998, *Total Quality Management as a Holistic Management Concept* (Berlin/Heidelberg).

Quality of Life and Usability Engineering

E. A. Averboukh

Interdisciplinary Centre for Quality Life and Usability Studies (LUSI-Centre), University of Kassel, Germany

1. INTRODUCTION

Growing and rapidly changing demands towards higher quality of life along with unprecedented rapid technological evolution require to redefine and/or revisit and bridge diverse engineering approaches to manage these changes in a human- and environment-friendly way.

Quality of life can be defined as an extent to which humans achieve their social, economic and ecological goals, and satisfy their needs using technological resources for performing certain functions and tasks in diverse environments and contexts.

Humans and technological resources that they use form an interactive *system* where the behavior of each interacting component affects the behavior and efficiency of interaction of the whole "set."

Hence *quality of life engineering* can be defined as an interdisciplinary process of interactive human–machine–environment *system engineering* within cost and time constraints. It covers design, production and maintenance of trustworthy systems that are *compatible* with humans and with environment and *support user's goals and behavior*. This process ensures that *human's needs are satisfied* and individual *human's resources are enhanced* throughout a system's entire *life cycle*.

Humans interact with technologies or technical systems via so-called *user interface*. This interaction is constrained by their individual sensory, cognitive and biomechanical capabilities. As technologies become more and more intelligent and computer-aided, user interfaces are often called human–computer interfaces (HCI) and provide visual, audio, tactile, etc. communication modalities for presenting information to the human users and for accepting user's requests/inputs. So-called *usability* of technologies and of their HCI, i.e. of their front-end to the users depends on interface suitability to human users, to their capabilities and task performance in specific environments and contexts. Hence usability is one of the most critical quality of life factors and is defined as effectiveness and efficiency of interaction and subjective satisfaction of human users (ISO 1995).

By analogy, *usability engineering* can be regarded as user- and task-centered design, evaluation, production and maintenance of user- or human–computer interfaces within the entire system's life cycle and within cost and time constraints.

Main engineering efforts shift to the early design and evaluation activities and especially to the so-called architecting phase in order to model and to test mutual performance of human users and technical systems and to avoid and/or eliminate usability problems and mistakes before starting detailed design and product development.

2. CRITERIA AND METRICS FOR EVALUATING USABILITY AND QUALITY OF LIFE

Following *usability criteria* drive the design, development and evaluation of interactive systems and their interfaces:

1. *Effectiveness* of user's task performance or so-called *usefulness* of the system, i.e. the possibility for target end-users to achieve their task goal and to perform desired function(s) with the help of technical system in target environments and contexts within given time constraint.

2. *Efficiency* in terms of how easy (or how sophisticated) it is for target users to identify or to recognize systems' role and functionalities, to learn and to use them, to remember over time, etc. in typical task-situations.

   For example, following relative time-estimates may be applied as a metrics for evaluating efficiency:
   - time for error-free task-performance under the given context (t_1/t_0);
   - time for coping with problems or errors (t_2/t_0); and
   - time for recovering from errors (t_3/t_0),

   where $(t_1/t_0) + (t_2/t_0) + (t_3/t_0) = 1$, t_0 is the time to solve the task or to learn task-interaction.

   Efficiency can be also estimated in terms of so-called task-complexity, e.g. as number of task-performance errors, etc. Only rough estimates can be achieved although, as it is difficult to define unambiguously the boundaries of task-situation, as well as to simulate real-life task-scenarios in laboratory's environment, etc. When measurements of the metrics are performed manually or semi-automatically, as it is usually made in industrial usability labs, only qualitative evaluation of efficiency is feasible.

3. Subjective user's *satisfaction* and/or enjoyment, i.e. *user's sensations* when using the product or system.

   Usability criteria apparently transform into the superior quality of life criteria if analyzed in the broader then just ease-of-use and ease-of-learn contexts, i.e. for a brighter range of human activities covering producing, buying, owning, using, presenting, selling, decomposing, recycling the product, etc.

   Human sensations like, for example, feeling of comfort or discomfort can be considered as integrated quality of life criteria. They have to be evaluated in general, i.e. regarding the product or interactive system as a whole, as well as in relation to the specific human user's functions and activities from user's and from the third party's points of view. For example, sensations of driving the motorcycle have to be evaluated from the driver's point of view, i.e. when accelerating, braking, handling, etc., and from the third parties point of view, e.g., humans who are driving cars or walking along the same road, etc.

   Activity-related comfort subcriteria have to be further decomposed regarding human sensory systems (vision, audio, tactile, etc.) involved in these activities, relevant thermal, posture and movement comfort, etc.

   Following metrics are used for evaluating subjective satisfaction or sensations on-line, i.e. during the activity performance and/or off-line, i.e. before and/or after performing certain activities:
   - direct physiological measurements, such as blood pressure, urine composition, heart frequency, etc.;

- indirect measurements of cognitive load, stress, fatigue, etc., such as eye movement characteristics, sweating, temperature distribution over the target points or areas on the human body, etc.; and
- indirect measurements of sensations by monitoring and pattern-recognition of face mimics or expressions, wording, so-called "body language," etc.

All mentioned above quality of life and usability characteristics are subjective. They depend on culture of human users and on their environment in general, as well as on the user's gender, their individual physical, psychological, physiological profiles, cognitive level and professional background, as well as on their linguistical level, i.e. ability to express their wishes and preferences as well as ability to read, etc. These characteristics are not static, but have their own dynamics and evolution. Transformations reflect users learning and mutual adaptation processes when applying target technology, on the one side, and permanent and quite natural human desires to "want more and better," i.e. to intensify positive sensations and to prevent boredom, on the other side.

Availability of non-contact and preferably on-line measurement methods, task-integration and evaluation of the information acquired, its further processing and interpretation into the product's design features and forecast of their future transformations become critical for the success of the discussed engineering processes. Advanced error-tolerant data-modeling, -reduction and -management approaches and relevant computer-aided engineering platforms are, hence, of high demand and determine overall efficiency of quality of life engineering.

3. FLEXIBLE SYSTEM ARCHITECTURES FOR QUALITY OF LIFE AND USABILITY

3.1. Model-based Adaptive User Interfaces

Quality of life and usability engineering activities start from the very early design phases, particularly from analysis and architecting of products, systems and their user interfaces.

Architecting driven by human goals, needs and wishes is a top-down iterative approach and is one of the central system engineering/design activities.

Architectures of systems and user interfaces have to be flexible enough to support easy and time- and cost-effective adaptation to the evolution of the interactive system as a whole, i.e. to the changing requirements of human users, of their functions,

Figure 1. Advanced model-based adaptive User Interface.

activities and organizational structures, of the environmental conditions and enabling technologies, as well as of individual capabilities and preferences of human users.

The functional structure of advanced adaptive user interfaces is illustrated in figure 1.

Three following models are imbedded to control traditional user interface functions which manage dialogue and information presentation and exchange in a user-friendly way:

1. The Human User model (HM) which comprises and tracks information on human goals and objectives, functions and activities, human user behavior and interaction with the technical system via user interface in different situations and contexts, human resources or capabilities available, i.e.

- "long-term" user characteristics such as, e.g., demographic, linguistic, psychological, physical disabilities, certain aspects of professional level and skills which do not change within system life-cycle or change rather slowly, and
- "short-term" user characteristics such as, e.g., physical, physiological and cognitive characteristics which depend on human activities and tasks, their duration and workload, environment and situation, etc.

and generic and task-specific metrics for evaluating quality of life and usability of system for end-users.

2. The Technical System model (TM) which tracks information on technical goals like production, safety, etc., technical functions and system performance, system status in specific environments and situations, technical status and resources available, generic and task-specific metrics for evaluating system performance (e.g. time, costs, safety, etc.) and extent to which the mentioned above goals can be achieved.

3. The User Interaction model (UIM) which monitors interaction, bridges User- and Technical System Models at different levels of abstraction and from diverse view points, interpret them into relevant interactive modalities for information presentation and managing dialogue between human users and technical systems.

3.2. Evolutionary User-centered System Architecting

Top-down evolutionary and iterative modeling and architecting processes are schematically shown in Figure 2. Models embedded into the user interface reflect different views on system's architecture, i.e. so-called

- operational view (objectives, functions, organization, activities, and information-exchange requirements);
- system view (human capabilities- and enabling technology-dependent requirements and constraints); and
- technical view (requirements of multiplatform- and network interconnections among all subsystems and set of interface standards) (Levis 1998).

Diverse views on architecture of interactive systems have different dynamic of evolution. Enabling technologies, for example, are developing very rapidly. They determine and constraint currently achievable efficiency of supporting user's behavior, as well as it's future transformation when technology is upgraded. Hence, early design activities phase, i.e. function analysis, requirements definition and particularly architecting have to be performed differently, i.e. incrementally. Design, development and evaluation of the limited number of most critical functions using currently available enabling technologies, i.e.
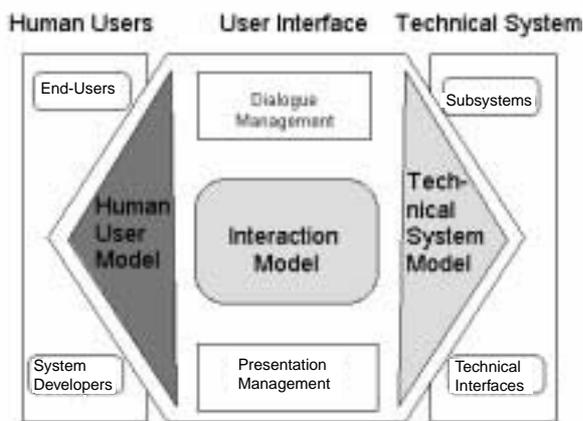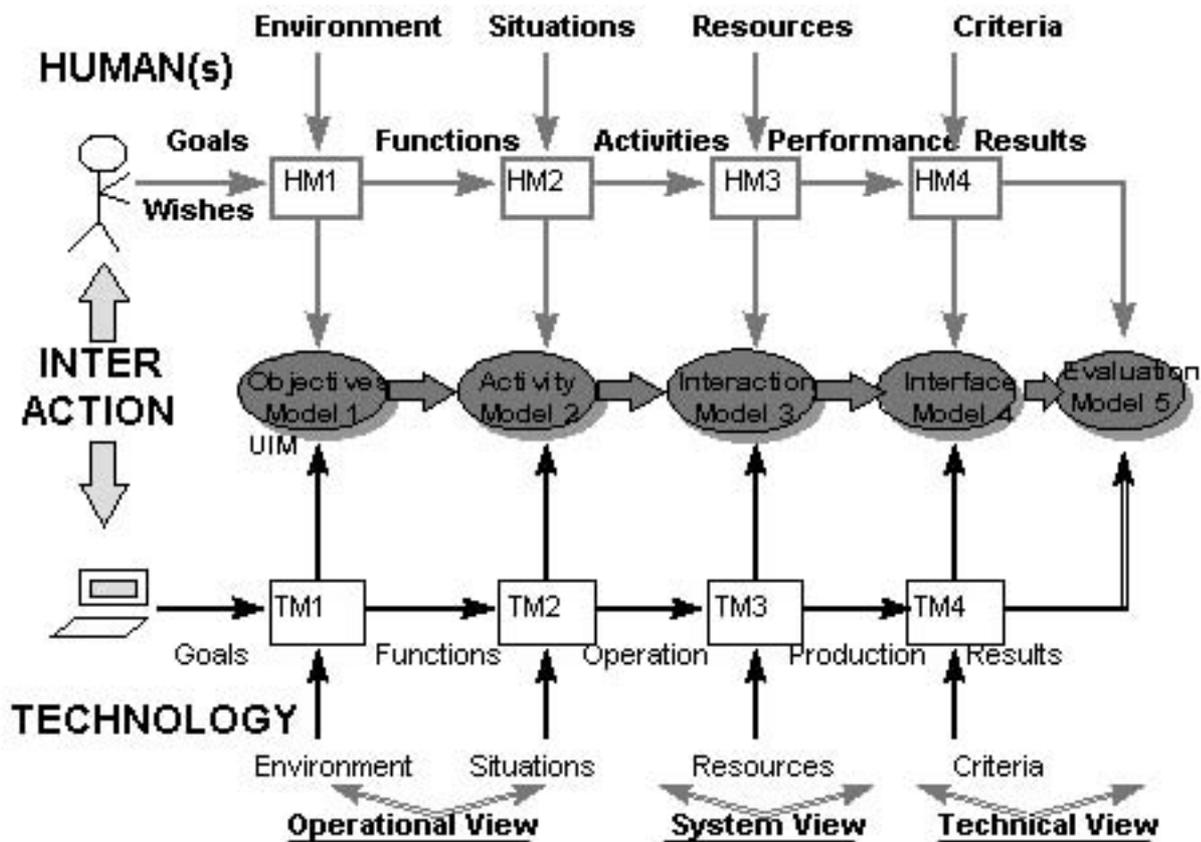
Figure 2. Evolutionary top-down multi-view system architecting and modelling.

"develop a little, test a little" from different architectural views becomes the only feasible way of user-centered engineering of interactive systems. That is why decomposed hierarchical structures of the human behavior models, technical and interaction models have to be considered in adaptive system architectures (Figure 2).

Such decomposition enhances flexibility, minimizes redesign and optimizes adaptation efforts and costs. Joint executable or so-called evaluation-model provides possibility to evaluate mutual compatibility and joint dynamic performance according to diverse quality of life and usability criteria (Venda and Venda 1995, Levis 1998).

Such modular hierarchical structures of the models provide also a basis and an efficient frame for system's life-long information- and usability-knowledge management.

4. ADVANCED USABILITY ENGINEERING LIFE–CYCLE

Usability engineering life cycle is an iterative process that comprises user- and task-driven design and development activities, as well as evaluation or so-called usability testing activities. They are focused on incremental architecting, modeling, prototyping and testing of the mentioned above submodels TM_i, HM_i and IUM_i, $i = 1,5$ (Figure 2) for different categories of target system's users from diverse architectural views. End-user participation from the very early design stages becomes crucial for system's usability.

Relevant usability testing methods have to be applied at different phases of the engineering life-cycle particularly

depending on system's overall goals and time- and cost constraints (Figure 3).

One of the first design activities which determine system's usability is devising it's purpose, selecting specific view point, analyzing human user's and supporting technical functions and building relevant functional hierarchy according to the following principles (Venda and Venda 1995):

- functions at the same level of hierarchy have to be exhaustive and not overlapping;
- they have to be formulated as a "verb–noun" phrase using the words familiar to end-users; and
- repetition of subfunctions at different level of the hierarchical tree is not allowed.

It is especially important to follow these rules when designing human computer interfaces using object-oriented programming languages.

Such top-down functional analysis and selecting of significant system's and interface functionalities via the usability test "Value of Functionality" is indispensable for focusing design and development priorities in a user-centered way. Incremental and hierarchical engineering approach is schematically presented in the Figure 3 as a L-approach to emphasize the above mentioned philosophy "develop a little, test a little." It differs from the traditional V- or waterfall system engineering approaches, where first all possible functions and relevant requirements are specified, designed, developed and only then evaluated.

The reliability and effectiveness of the results of usability testing for discovering usability problems depend on the number of test persons and evaluators involved, representativeness of test
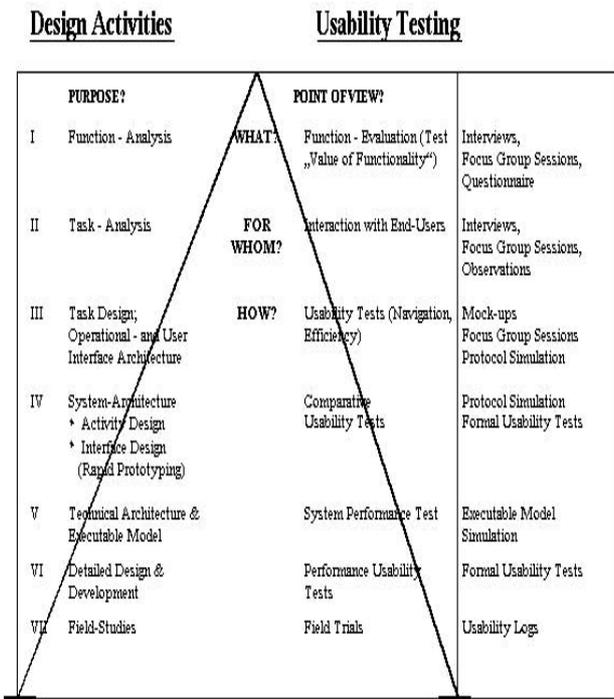
Figure 3. Λ-approach to usability engineering of interactive systems.

persons regarding expected user's population, number and representativeness of test-situations and scenarios, and fidelity of the prototypes, etc.

"Ideal" test with a representative sample of at least 20 test persons and three independent evaluators can reveal majority of main usability problems with probability > 70–75% when testing rather low fidelity prototypes (Nielsen 1993). Such exhaustive usability testing is rarely feasible in industrial practice and is often substituted by so-called low-cost or discount usability engineering methods, such as, for example, heuristic evaluation by single experts.

The following set of usability heuristics is typically used for evaluation (Nielsen 1993, Gerhardt-Powals 1996):

1. Minimize workload
   - locate displays and controls in easily reachable and visible zones;
   - elect locations depending on user's tasks, activities and expected motions;
   - provide enough free space for task performance;
   - automate unwanted workload (free cognitive resources for high-priority tasks; and
   - eliminate mental calculations, comparisons, etc.).
2. Improve visibility
   - locate most important and frequently used information in so-called primary areas or windows; and
   - minimize parallax, reflections, etc.
3. Support recognition
   - group data in consistently meaningful ways;
   - reduce uncertainty: display data in a manner that is clear and obvious;
   - use names and/or symbols that are conceptually related to function (context-dependent) and can be recognized without additional help or information;

- use the same names and/or symbols and their groups in the same locations for the same functions.
4. Support interpretation
   - present status and changing of status of interface elements or groups of elements;
   - use coding (form, color, size, length of lines, location, distance, graphical images, etc.);
   - provide multiple coding where appropriate;
   - fuse data: reduce cognitive load by bringing together lower level data into a higher summation;
   - allow user to detect and to interpret seldom situations and/or slight changes which may be crucial for system behavior by presenting critical values;
   - present not only data but also system processes which are important; and
   - provide meaningful aids to interpretation, i.e. use a familiar framework, everyday words, metaphors, etc.
5. Improve navigation
   - display data according to the task flow and to the eye movement or reading stereotypes (e.g. from top to bottom and from left to right); and
   - provide shortcuts that suit to specific task-flows, use-scenario, behavioral stereotypes or which support rapid access to the task-critical information.
6. Limit recall memory
7. Reduce information
   - reduce presented information only to the needed by the user at a given time;
   - allow users to remain focused on critical data; and
   - exclude extraneous information that is not relevant to the current tasks.
8. Optimize user control
   - limit data-driven tasks; reduce time spent assimilating raw data;
   - user controls the dialogue; and
   - provide feedback.
9. Prevent errors
   - reduce uncertainty;
   - consistency and standards; and
   - help users recognize, diagnose and recover from errors.

5. COMPUTER–AIDED PLATFORMS AND FUTURE TRENDS

5.1. Current Practice

Current industrial usability engineering practices show their rather low efficiency. The main reasons are usually following:

- significant underestimating and often lack-of-competence in performing user- and task-oriented system and user interface design activities;
- miscommunication between so-called usability testers or human factors professionals and system designers or developers, e.g. software engineers; and
- lack of integrated cross-disciplinary computer-aided platforms for consistent, mutually balanced and systematic multi-view performing of the above mentioned L-usability engineering activities (Figure 3), etc.

All this consequently devaluate the efforts of human factors

professionals and related investments in corporate usability engineering programs.

First, using expensive usability labs, performing formal usability tests or heuristic evaluations, etc. without being able under the given time and cost constraints systematically to consider obtained test results in system's redesign, does not contribute to better usability and quality of life.

Second, low cost usability testing methods often do not provide reliable enough basis for making right user-centered design decisions.

Third, many companies especially who are working in the same area perform similar low-cost usability tests instead of cooperative performing explicit and informative tests for typical interaction tasks.

Hence main research efforts have to be focused on turning quality of life and usability *studies* into relevant *engineering* approaches and *computer-aided engineering platforms.*

5.2. Future Trends

Systematic approach and notations for multi-view task-oriented architecting of flexible information systems are proposed by Levis (1999). Further enhancement of this technology with user-interface design and testing activities and with related knowledge-management tools is in progress.

Currently available Kansei Engineering methods and software tools allow to model human user's wishes and preferences and link them with static design images of relevant products. Further evolution of these technologies particularly towards dynamic or interaction modeling (Figure 2) has to be considered.

Internet technologies win recognition as a virtual platform especially for intercultural quality of life and usability engineering activities. Hence developing Internet-based databanks for different application areas, where explicitly tested (modular) models of human behavior and interaction with target technologies, i.e. HM_i, TM_i, UIM_i, $i = 1,5$ for different application areas, diverse target user groups, environments, etc. are available as a main data-patterns, will apparently challenge and drive further research and development in the field. Further standardization in this area will help to save time, efforts and costs and speed up developing human-friendly technologies and improving quality of life.

REFERENCES

GERHARDT-POWALS, J., 1996, Cognitive engineering principles for human–computer performance. *International Journal of Human–Computer Interaction*, 8, 189–211.

ISO 9241, 1995, *Ergonomic Requirements for Office Work with Visual Display Terminals, Part II: Guidance on Usability.* Draft International Standard ISO CD 9241-11.

LEVIS, A.H., 1998, Adaptation in decision making organizations. In *Proceedings of the 4th International Command and Control Research and Technology Symposium,* Nasby Park, Stockholm, Sweden, September.

VENDA, V.F. and VENDA, YU.V., 1995, *Dynamics in Ergonomics, Psychology, and Decisions: Introduction to Ergodynamics* (Norwood: Ablex).

NIELSEN, J., 1993, *Usability Engineering* (Boston: Academic Press).

Rapid Macroergonomic Redesign

A. Majchrzak

Marshall School of Business, University of Southern California, Los Angeles, CA 90089-1421, USA

1. THE CASE AND REQUIREMENTS FOR INFORMATION TECHNOLOGY

Critical to the success of many organizations today is an ability to rapidly redesign as their market and customer needs change (e.g. Goldman *et al.* 1995). Not only must the specific human factors of a particular workstation be redesigned rapidly, but the macroergonomics of an organization needs to be rapidly redesigned as well. Macroergonomics is the study and design of jobs, organizational strategy, organizational structure, incentive systems, and training programs, in conjunction with the technology (Hendrick 1987). Evolving from sociotechnical systems (STS) theory (Cherns 1976, 1987, Trist and Murray 1993, Taylor and Felten 1993), macroergonomics is focused on designing these features of the organization so that human skills and abilities are effectively used to achieve personal and organizational goals.

Both macroergonomic and STS design have been traditionally implemented largely as a manual process whereby a design team, composed of a cross-functional mix of members of the organization, become engaged in their own effort at identifying alternative macroergonomic designs. This manual process involves interviewing all members of the organization, collecting objective data about problems in the business process, discussing implications of the data and interview results for technical and organizational redesign, and then persuading each other and management about alternative redesign options. Information technology has not generally played a large role in this process, but see Butera and Schael (1997) for an exception.

As organizations find that they need to speed up the process of macroergonomic redesign, however, the need for information technology becomes more apparent and pressing. In particular, information technology can be useful to an STS-oriented redesign process in a variety of ways. Information technology can hep in a number of ways, outlined under the following headings. Thes headings can be used as requirements of a decision support system for organizational designers. Section 2 onwards looks at TOP Modeler, a tool that meets these requirements.

1.1. Levelling the Knowledge Field

Any redesign should involve a range of stakeholders from shop-floor workers to managers, from line to staff. However, each stakeholder in the design process represents a different level of expertise and different abilities to argue confidently in front of others. For example, it is not unusual to find workers with less than a high school graduate degree unable to convince qualified engineers about problems of a particular technical approach. Information technology can provide decision support capabilities to all stakeholders to help them analyze solutions in a way that overcomes their differences in status, language, experience, and articulation.

1.2. Encouraging rapid Knowledge-based What-if

In considering an option for redesign, the implications of that option for all aspects of the organization should be identified and analyzed. This often takes time and information. Computer-based tools can speed up this process and provide instantaneous feedback, encouraging the consideration of more options and a genuine what-if approach to options not previously considered. Visualization of alternatives can be made much simpler and more realistic with computerized tools.

1.3. Conducting a Comprehensive Assessment of the Organization

The open-systems nature of STS designs means that all aspects of the organization should be considered during any design effort. Invariably, something "falls through the cracks" such as when the culture or incentive structures are overlooked until later in the design. Waiting to consider these factors until later, however, creates the risk that too many other factors will have become constrained, thus limiting the alternatives which can be considered. A knowledge-based tool can ensure that the comprehensive range of factors are always kept in front of the STS design team to avoid overlooking important issues.

1.4. Comparing the Current Organization to Known Best Practices

To encourage the integration of "best practices" into organizational analysis and redesign, design consultants often encourage design teams to visit best-practice sites or otherwise obtain an early understanding of best practices. It is difficult to keep these best practices in front of the designers as the design process unfolds and design decisions are made. A computer-based tool with a knowledge base of best practices can perform this function by allowing users to compare their design options against best practices and immediately receive feedback about gaps.

1.5. Seducing people into a Sociotechnical Systems Orientation

People who are in the greatest need of practising and benefiting from an STS orientation to redesign are often the ones least likely to attend an STS seminar or consciously decide to adopt an STS-oriented approach to organizational redesign. A computer-based tool can help with this learning by making the process of organizational analysis so engaging that users may not be aware they are even learning a new approach to redesign. That is, just as young students are not necessarily aware they are learning valuable reading skills when playing with certain computer games, organizational designers do not have to know they are learning valuable STS skills when performing an organizational analysis on the computer.

2. TOP MODELER

TOP Modeler is a computer-based tool to help designers *align* the strategic, organizational design, technological, and human resource features of an organization (TOP stands for technology, organization, and people). The theory behind TOP Modeler is that an organization will be able to achieve its desired business strategies only when the organizational design, technology, and human resources are aligned with each other and with that strategy. Often people improve one aspect of their organization (such

as a reengineered workflow or technology or organizational chart) without paying adequate attention to other aspects of their organization. The best-intentioned changes will fail to achieve successful improvements unless all necessary changes are made. However, some changes are more important than others, and it can be hard to find out which changes are most critical to achieve a systemic change in the organization.

TOP Modeler contains a knowledge base of systemic relationships among a variety of different features of a manufacturing organization. The knowledge base was developed with a $10 million, five-year investment of the US Air Force ManTech program by an industry-based consortium of the National Center for Manufacturing Sciences, Digital Equipment Corporation, Texas Instruments, Hewlett-Packard, Hughes, General Motors, and the University of Southern California. After development, the knowledge base was validated on one hundred companies. Users of the software include anyone involved in improving their organization: engineers designing a new technology, managers restructuring their organizations, business process re-engineers, and direct labor workers responsible for the continuous improvement of their manufacturing operations.

2.1. Defining a Business Strategy

To begin the analysis process, the design team must agree on a unit of analysis: What is the "organization" being redesigned? Sometimes design teams are confused and believe they are designing an entire factory when they have actually only been asked to redesign a particular department. It is important to obtain agreement on the scope of the organization being designed. For the knowledge base, the scope is immaterial; it has been structured for organizations as small as workgroups and as large as factories (with less than 200 workers).

Having defined a unit of analysis, users select a set of business strategies for their organization. There are three components of an organization's business strategy that must be established: business objectives, process variance control strategies, and organizational values. Sometimes these strategies are determined by management levels above the organization being redesigned. And sometimes the organization under redesign is given some flexibility in establishing its own strategies.

2.1.1. Business objectives

Business objectives are defined as measurable goals that the organization would like to achieve in the near future. The objectives are phrased as continuous improvement objectives rather than objectives which reach a static state; this is because competitors will drive any organization to improve continuously. Two types of business objectives have been identified: (a) those intended to reduce cost while simultaneously improving customer satisfaction — minimizing throughput time, maximizing quality, and maximizing changeover flexibility (reduced setup time); and (b) those intended to enhance the organization's future adaptability — achieving employee teaming, maximizing design for manufacturability, maximizing production process flexibility, and maximizing product development flexibility. Definitions for each of these objectives were developed by the industry consortium and are available at the click of a mouse www.topintegration.com.

2.1.2. Process variance control strategies

Process variance control strategies are the second component of an organization's business strategy. Process variances are technical variations (planned or unplanned) in the production workflow that create uncertainty in the processing of materials. In TOP Modeler there are 18 variances (e.g., quality of incoming material and employee turnover), selected by asking hundreds of production managers to list the process variances that most directly impact their ability to meet production goals. For any of the variances relevant to an organization, the organization must define a control strategy: Will the organization work to proactively eliminate the variance? Will the organization work to effectively cope with the effects of the variance on production? Or will the organization simply do nothing and let the effects of the variance propagate unfettered (the default strategy)?

2.1.3. Organizational values

Organizational values describe the preferences or underlying beliefs the managers have about how employees (management included) should behave. Such values describe the degree of collaboration, risk-taking, and continuous improvement expected of employees. After selecting the three different components of the business strategy, the tool helps users determine whether the strategic components are aligned with each other. For example, an organizational value of no risk-taking is inherently in conflict with a business objective of product development flexibility.

2.2. Describing the As-is Organization

With the business strategy described and aligned, the next step in a TOP Modeler alignment process is to determine the extent to which the current as-is organization is designed to achieve that business strategy. This is done by first describing the as-is organization and then comparing the as-is organization to the ideal organization needed to achieve the business strategy. To describe the current as-is organization, the knowledge base contains a comprehensive list of 11 feature sets for an organization. The 11 feature sets can be viewed as questions about what the current organization looks like:

1. What information resources do the workers in the organization have access to?
2. What do the production process characteristics of the organization look like?
3. For which decisions, if any, are workers empowered?
4. What are the employee values being demonstrated?
5. To what degree is customer involvement encouraged?
6. Which skills do employees have?
7. What does the reporting structure look like?
8. What norms does management actively encourage?
9. What tasks and activities do employees perform in their jobs?
10. What do the general technology characteristics look like in the organization?
11. What are the performance measures and rewards for employees?

For each feature set there are specific features for describing the organization. In total there are over 300 specific features that have been identified for describing the current as-is organization.

2.3. Comparing the As-is Organization to an Ideal

When a user selects a business strategy, this has the effect of defining an "ideal" organizational profile, which the organization should adhere to if the business strategy is to be achieved. Ideal profiles in TOP Modeler are specific to the different business strategies and were developed based on a variety of sources (including an extensive review of empirical academic studies, company best practices, the practical experience of the industry consortium involved in developing the knowledge base, a scientific validation survey of 80 companies, and 20 cases of closely observed use of the knowledge base).

When a user has described both the business strategy for the organization and the current as-is state of the organization, this as-is state can be compared against the ideal. The results will indicate gaps between the two; with gaps interpreted as opportunities for improvement. For example, if the business strategy required a particular empowerment feature and the current as-is organization did not provide that degree of empowerment, then a gap in empowerment would be identified. TOP Modeler helps to quickly identify each gap, either interactively or as lists of improvement opportunities. The emphasis in TOP Modeler is to provide a comprehensive list of improvement opportunities, not just selected opportunities, to ensure the design team does not allow anything to "fall through the cracks."

Once a comprehensive set of improvement opportunities is identified, STS tenets suggest that priorities must be set by the user in order to ensure user commitment to the final implementation plan. Users may set priorities based on management dictate or detailed cost-benefit analyses. Although the tool does not set the priorities, it can provide support by informing the user about which business strategies are affected by each gap, as well as allowing designers to quickly model the proposed new organization to determine whether gaps would be closed, given a proposed prioritization.

3. LESSONS LEARNED FROM DESIGNERS USING TOP MODELER

Thus far, TOP Modeler (and its precursors Action, TOP Integrator, and Hitop) has been used in over 50 cases of organizational and/or technological change. The companies that have used it have ranged from very small companies to very large companies, located in the US, Brazil, and Switzerland. Here are some of the lessons learned.

3.1. IT Can Have Unforeseen Uses in Macroergonomic Design

3.1.1. Strategic plans for small companies

Used by government-sponsored manufacturing consultants (e.g., Switzerland's CCSO), it helped small companies to develop strategic plans for restructuring. In one case TOP helped the consultant to understand that the company's initial strategic plan was unlikely to succeed until management agreed to reduce the amount of variation it allowed in its process.

3.1.2. Persuading clients not to relocate

Used by the large software vendor EDS, it helped a client manufacturing company decide not to relocate its plant from one foreign country to another, because it looked too expensive to close the gaps created by the move.

3.1.3. Joint ventures between large companies

Used by the large manufacturer General Motors (GM), it helped decide whether a joint venture plant was ready to be opened. GM decided on delaying the opening because TOP helped to reveal differences of opinion in how to manage the workforce.

3.1.4. Better good practice for a small manufacturer

Used by the small manufacturer Scantron, it showed whether their good practices needed improving. TOP helped them to discover that while they did indeed have many best practices, they needed to involve the workforce more closely with the suppliers and customers, so that's what they did.

3.1.5. Changing technology in a large manufacturer

Used by the large manufacturer Hewlett-Packard, it helped identify the workforce and organizational changes needed for their new production technology to operate correctly. The result was a substantial improvement in ramp-up time when the new product and production process were introduced.

3.1.6. Improving teamwork in a maintenance crew

Used to redesign a maintenance crew at Texas Instruments, it revealed how the envisioned team approach needed several important improvements before startup.

3.1.7. Pinpointing inconsistencies in manufacturing objectives

Used by a strategic planning committee at a large manufacturing company, it identified areas of misalignment among elements of a new strategic plan (in this case between quality and throughput time).

3.1.8. Revealing the implications of manager-to-manager agreements

Used by a divisional manager at a manufacturing plant, it enabled him to verify the business strategy given him by his group manager. TOP showed he had agreed to deliver on new product development without having authority over the necessary people, skills, and other resources. He went back to his group manager to renegotiate these resources.

3.1.9. Unforeseen uses: some conclusions

These examples suggest a diversity of applications for computer-based STS design tools. Some users find its versatility frustrating since they are more familiar with simulation packages that provide a finite and well-defined set of functionalities. Others recognize that opportunities for injecting a knowledge-based approach into macroergonomic design can seldom be known or specified in advance; they place a high value on versatility. Valuing the versatility seems to be important to the ultimate successful use of information technology in macroergonomic design. Those who tend to value the versatility appear to have the following characteristics:

- Deep understanding of their own organization (including its strategy and operation).
- Basic understanding of the principles of organizational design and human behavior.
- Successful at translating externally derived information about possible process improvements (from books, consultants, and site visits) into the organization.

- Responsibility and authority to suggest process improvements.
- Good group process and presentation skills.

I have also found that successful users engage others in the organization in discussions about alternative ways to use and diffuse the tool. These discussions are often quite value-laden. For example, one manager refused to allow workers unsupervised access to TOP Modeler because, in his words, "they might begin to believe that they know more about organizational design than I know and question my authority. I don't want that; that could lead to chaos." This suggests that having discussions early on about the role of information technology in the macroergonomic design process will help to clarify expectations so these latent values do not come up later to kill the pilot experiences.

3.2. Macroergonomic Designers Don't Like to Admit IT Helped Them

I found several examples where discussions among design team members about macroergonomic design issues proved contentious and frustrating without the use of a knowledge base tool. I would then introduce TOP Modeler to help users reformulate their strategy in greater detail, the discussion usually became more productive, leading to a quick consensus and an action plan. Yet, when asked if TOP Modeler helped them, the design team said no.

Upon further examination, I found that some users believe the knowledge base of TOP Modeler cannot publicly provide any value to them since it would then suggest that they could not perform their own work adequately without it; and no one who is proud of their ability to redesign organizations wants to believe in any dependence or help from a machine! If left unmanaged, this opinion of not attributing value to the knowledge base is detrimental to its further use, especially if an organization is in the piloting stage to determine whether company-wide use should be considered.

I found two remedies to this situation. First, people who believe that they are expert organizational designers are particularly threatened by the tool and will sabotage it unless they have a particular role to play in its use or deployment. Possible roles include integrating the tool into an existing organizational change process which is controlled by the experts, helping experts understand how the tool can be used to more quickly and widely disseminate their own personal knowledge bases, and more clearly identifying the boundaries between the expert's organizational change efforts and those of the users of the tool.

The second remedy is to agree, at the outset of a pilot effort, on the basis for evaluating the success of the information technology. I have used a variety of different metrics for success. One such metric is, Did the user learn anything? In a systematic study of 30 users, I found that 80% reported learning something about TOP integration or about their own organization from using TOP Modeler. Another such metric is, Did the user learn what they had expected to learn? I found that only 50% reported yes to this. Yet another metric is, Could the user generate action plans from its use? In that same study, I found 100% of the users were able to generate action plans as a result of using the tool. A final metric I used was, Were the action plans implemented? In cases where I was told the outcome, I found that about half reported

implementing the action plans. It is clear that the choice of metric will affect how successful the pilot appears to be.

3.3. Who is Responsible for Making Macroergonomic Decisions?

TOP Modeler was designed as a decision support system which would (a) remind users of the comprehensive array of TOP features when redesigning their organizations and (b) provide them with data about the nature of systemic relationships among features so that decisions about organizational redesigns would not be limited to single functional units, single sociotechnical elements, or single interventions. From this perspective, the ideal user would be a person who interacted with the knowledge base, trying out different ideas about business strategies and alternative future designs of the organization, eventually to draw conclusions about the trade-offs for different design options and to decide on a particular design.

I found this ideal user describes a very limited set of the people who are likely to actually use the tool. Most users are not given the time or authority to consider alternative macroergonomic designs, in part because macroergonomics is an interdisciplinary issue that touches on many different functional responsibilities. Instead, they see their role as one of considering a specific macroergonomic design when there is a specific problem with the current design. As a result, such users are focused on fixing a specific problem in the least costly, most advantageous fashion. With this perspective, they are pressed for time and need quick results; they do not appreciate a knowledge base that reminds them of the comprehensive and systemic nature of the problem. Nor do they appreciate a knowledge base that has no cost figures associated with it (it is impossible to generate cost figures for a generic knowledge base).

This means there is often an immediate mismatch between the user's needs to quickly solve a problem and the tool's functionalities to help users solve systemic problems. I found about half of the users to effectively manage the mismatch by using the tool as a check on their decision-making process (e.g., as they decided on their business objectives, they would use the tool to make sure the objectives were not in conflict). The other half of users, however, found the tool extremely frustrating because, in one user's words, "I want it to tell me what to do. Don't just list my problems. Tell me which one to fix first." This suggests that an additional requirement needs to be added to the requirements for STS-based information technology described at the beginning of this article: when users are unwilling or unable to accept decision-making responsibility for their macroergonomic design process, only an expert system tool may be of value.

In the spirit of STS — based on the assumption that there are many different ways to achieve an effective outcome and that for users to be committed to an action, they must create it themselves — it is questionable whether an expert system would actually prove useful. However, not enough research has been done to find out what happens when expert systems are given to those who want them. Do the users implement the expert system's suggestions or do they find a different excuse?

4. SUMMARY

This article began with several requirements for the use of information technology in macroergonomic design. To indicate that

these requirements were feasible, it went on to describe an IT-based tool that fits most of them. Finally, it described some lessons learned from using the tool in macroergonomic design. These lessons clearly indicate there is a role for information technology.

REFERENCES

BUTERA, F. and SCHAEL, T, 1997, The convergence toward a new sociotechnical system approach. In Salvendy, Smith, and Koubek (eds) *Advances in Human Factors/Ergonomics 21A: Design of Computing Systems* (Amsterdam: Elsevier), pp. 201-4.

CHERNS, A., 1976, Principles of sociotechnical design. *Human Relations*, 29, 783-92.

CHERNS, A., 1987, Principles of sociotechnical design revisited. *Human Relations*, 40, 153-62.

GOLDMAN, S.L., NAGEL, R.N., and PREISS, K., 1995, *Agile Competitors and Virtual Organizations* (New York: Van Nostrand).

HENDRICK, H.W., 1987, Organizational design. In Salvendy (ed.) *Handbook of Human Factors* (New York: Wiley), pp. 470-94.

TAYLOR, J.C. and FELTEN, D.F., 1993, *Performance by Design* (Englewood Cliffs NJ: Prentice Hall).

TRIST, E. and MURRAY, H., 1993, *The Social Engagement of Social Science: A Tavistock Anthology* (Philadelphia PA: University of Pennsylvania Press).

Risk Management

B. Zimolong and G. Elke

Department of Psychology, Work and Organizational Psychology, Ruhr University Bochum, Germany

1. MANAGEMENT PRINCIPLES

Management as a function comprises all processes and functions resulting from the division of labor in an organization such as planning, organization, leadership and supervision. In most organizations more or less formalized management systems serve to structure, develop and direct business processes. Systems differ with respect to branches, nature of business, company size and Human Factors such as culture and policy. As firms grow in size management systems gain complexity and become difficult to use, thus resulting in domain-specific systems such as management of occupational safety, health, loss control, environmental resources, quality or personnel.

Risks include all aspects of accidental losses that may lead to any wastage of the organization's, society's and environmental resources. These resources cover personnel, materials, machinery, procedures, products, money and natural resources: soil, water, energy, and natural areas. Losses may result from the presence of potential harm to one or more elements of the system, either because of the interactions with other elements inside the system or with the environment outside the system. Risk is the measuring stick for this potential, which may be defined as the probability that harm will become manifest within a certain period.

Risk management can be understood as an approach to reintegrate the domain-specific management systems into an integrative management concept. The essence of risk management is to control risks, i.e. to prepare, protect and preserve the resources of the enterprise, the society and the environment. Risk control strategies may be classified into four main areas: risk avoidance, risk retention, risk transfer and risk reduction.

The system elements to be managed include among others:
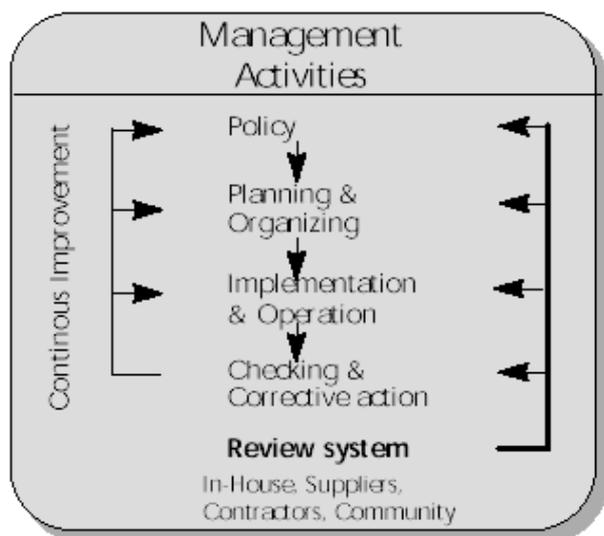
Figure 1. Management control cycle system based on the environmental management approach in ISO 14001 and BS standard 8800.

- Health and safety of employees, suppliers, contractors, customers and residents of the community (e.g. improvement of public health and safety).
- Reliability and safety of products and services; of materials, equipment, work systems and plants; of transport of hazardous goods.
- Integrated pollution control, radiation protection, waste minimization, recycling and waste disposal.
- Sustainable management of natural resources (soil, water, natural areas and coastal zones), reduction in the consumption of non-renewable energy.

Sustainable development means the improvement in the quality of life that does not impair the ability of the ecosystem to maintain life. Managing for sustainability is predominantly based on the principles of intergenerational and intra-general equity as well as social and ecological balance (Hutchinson and Hutchinson 1997).

Many of the features of risk management are indistinguishable from the sound management practices advocated by proponents of quality and business excellence. Quality and environmental management systems have been introduced as international standards prepared by the International Standards Organization (ISO) in Geneva. The European Community (EC) Eco-Management and Audit Scheme (EMAS) agreed by the EC states in 1992 was set up as a voluntary certification scheme for the assessment of a company's environmental impact. The British Environmental Standard BS 7750 and EMAS contributed to the development of the ISO 14000 series "Environmental Management Systems" standards. Initiatives to launch an international standard on "Occupational Health and Safety" (OH&S) management systems based on the British Standard BS 8800 (1996) have been delayed. BS 8800 shares common management principles with the ISO 9000 "Quality management" series and the ISO 14000 "Environmental Management" standards. In many countries, national guidelines give guidance on OH&S management systems.

The management approach of the ISO standards are based on generic management principles that are derived from different theoretical and organizational perspectives. The elements of the systems are considered to present "best practices" of successful enterprises. They are designed to be used by organizations of all sizes and regardless of the nature of their activities. The key elements of a generic management system are integrated in the management control cycle outlined in Figure 1. The cycle is based on ISO 14000 (environmental management) and on standard BS 8800 designed for an OH&S management system. The key elements from BS 8800 are set out below:

- *Effective OH&S management* stands for developing, coordinating and controlling a continuous improvement process by setting and adjusting OH&S standards. The corporate policy is summarized in a vision, containing perspectives for the future and providing an idea of identification for all members of the organization. Policy and strategy can only become reality if they are translated into a stringent planning process. Both internal and external assessment methods can be used for the evaluation of a strategy's effectivity and efficiency: This needs to be done systematically through regular reviews of performance based on data from monitoring activities and from audits of the OH&S management system.

- *Formulation of an OH&S policy* addresses the preservation and development of physical and human resources and reductions in financial losses and liabilities. The policy provides guidance on the allocation of responsibilities and the organization of people, of resources, communication and documentation. It influences design and operation of working systems, the design and delivery of products and services and the control and disposal of waste. OH&S policy should be aligned with people management policy to secure the commitment, involvement and wellbeing of employees, of suppliers, contractors and customers.

- *OH&S planning* requires a comprehensive organizational approach that emphasizes prevention and involves risk identification, evaluation and control. Hazards are identified and risks assessed and controlled according to a systematic plan before anyone or anything could be adversely affected. Pro-active planning particularly adopts procedures to set OH&S objectives, devise and implement plans to meet them and to monitor both the implementation and effectiveness of these plans. Reactive planning means that measures are only considered after the occurrence of incidents such as loss-damages, accidents or deficient safety performance.

- *Implementation and operation.* Designing OH&S structures concerns the divisions of responsibility and distribution of formal authority, the creation of hierarchical or lean structures, the degree of self-regulation of work groups and units and the formal relations between groups and leaders. Establishing and maintaining control is central to all management functions including OH&S. The allocation of OH&S responsibilities to line managers and work groups serves as an important tool to foster the integration of OH&S into the daily work activities with specialists acting as advisers.

- *Checking and corrective actions* are the final steps in the OH&S management control cycle and part of the feedback loop needed to enable the organization to maintain and develop its ability to control successfully risks. Both qualitative and quantitative measures provide information on the effectiveness of the OH&S system. Learning from experience is supported through performance reviews and independent audits.

- *Management review* is a periodic status review of the OH&S management system and considers the overall performance of the OH&S management system; of individual elements and the findings of audits. The review identifies the actions that need to be taken to adjust any deviations.

2. RISK MANAGEMENT SYSTEM

The above principles seem to suggest that science and industry have reasonable models of how safe and reliable organizations work. This is not the case. The standards only presume that the targets of health, safety, quality and environmental protection are achieved by doing the right things within a reproducible structure, they do not offer guidelines based on scientific research and company's "best practices" which allow to evaluate if the things were done right.

A holistic concept of risk management requires the system to be integrated into the processes of an organization, e.g. integration of health, safety and environmental (HSE) activities
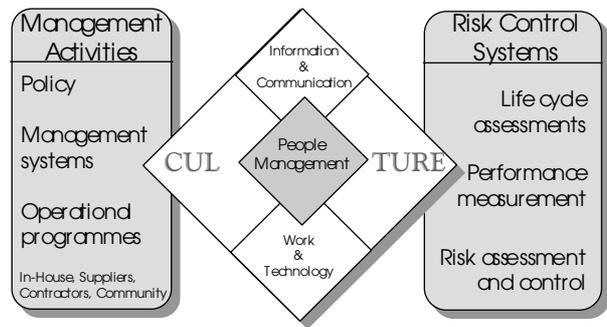
Figure 2. Holistic risk management system.
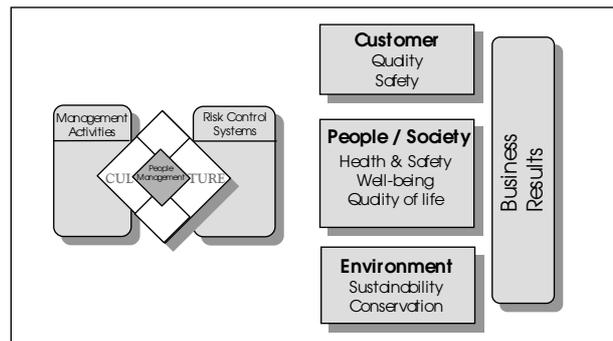
Figure 3. Multidimensional targets of risk management.

into the daily routines of managers, supervisors and employees and of HSE standards and processes into the life cycle of products, services and work systems. Finally human resource management principles have to be applied to foster long-term commitment and involvement of the workforce to HSE, including suppliers, contractors and partner firms.

The Holistic Risk Management System (HRMS) as suggested by Zimolong (1997) has been extended to comply with the management control cycle of the standards of BS 8800 and ISO 14000 (Figure 2). HRMS covers three domains: health, safety and environment. The system addresses the key elements of management as people management, management of information and communication, (re)design of work and technology and creation of an HSE supporting culture. Generic management activities include those of the management control cycle of the standards as described above. The elements of the risk control system to be managed are first of all the risks associated with the life cycle of systems, products and services; second, the process of continuous risk performance measurement as part of the information and communication management; and third, risk assessment and risk control. The outcomes of the risk management process as seen in Figure 3 are assessed with regard to multidimensional targets, which include reliability, quality, safety, health and ecological sustainability. They are given equivalent weight at improving business results. These goals are, however, seldom compatible, often contradictory, may raise internal and external conflicts and have to be solved on company's or society's level, if at all.

3. KEY ELEMENTS OF RISK MANAGEMENT
3.1. People Management and Culture
The implementation of HSE vision and policy into missions and operational programs is a company-wide task and not only the

duty of each individual to comply with standards, regulations and laws. Corporate HSE policy must aim at prevention and integrate the targets of health, safety, reliability and sustainability into business performance, including systems, processes, working conditions, external relationships to business partners and the public. Implementing vision and corporate policy is helped by the creation of a positive culture that secures involvement and participation at all decision levels. Managers take full responsibility for controlling all those factors that could lead to ill health, injury, loss or damage. Corporate HSE culture becomes visible in the way complaints are dealt with, suggestions are treated, errors, mishaps and incidents are blamed for (guilt-free atmosphere) and especially in how problems are solved in daily activities.

Effective management of HSE requires the support and commitment of the employees. The knowledge and experience of the workforce is a valuable resource in the development and operation of the risk management system. Structural arrangements for HSE consultations and representation provide effective means for involving the workforce, for example, HSE committees. Participative risk assessments and controls promote the active involvement of the workforce at the operational level. Implementing an effective HSE people management system requires appropriate structural conditions. Among those to be considered are:

- integrating HSE goals into corporate goals;
- assessing management and employees with regard to multidimensional targets, including HSE objectives; and
- planning and improving people resources, which includes the regular and systematic examination of all processes within personnel management as well as their adaptation to changes in general conditions.

Management of an organization's human resources includes such activities as personnel planning, recruitment, placement, development, performance appraisal, training and competency assessment, counseling and guiding of individuals and groups (leadership), payment concepts, back-to-work programs and rehabilitation at the workplace. Such personnel systems include valid selection procedures for hiring and promotion purposes, performance appraisal and review systems to ensure that the person is measured on the right goals and standards and receives accurate feedback, effective training procedures for the development process and labor relations that are conducive to employee motivation.

Personnel development of people's capabilities means in-company training focused on the company related targets, closing the gap between the actual state of the art and the target state through qualitative personnel planning and supporting and realizing career planning with personnel-related targets. Contractors, suppliers and temporary workers have to be included in the training program according to the level of risk to which they may be exposed or could cause.

With respect to HSE a major purpose of the performance appraisal process is to modify behavior, to feed back information to the employee for counseling and development purposes so that the person will start doing or continue doing the activities critical to effectively performing on the job. The feedback must lead to the setting of and commitment to specific HSE goals. Goal setting affects performance by directing the attention and actions of individuals and/or groups, mobilizing efforts, increasing persistence and by motivating the search for appropriate performance strategies. Setting difficult, yet achievable goals and providing performance feedback, is widely accepted as to be one of the most powerful behavioral modification techniques. Performance feedback is thought to fulfill several functions: as a reinforcer when it conveys success, as a punisher when it conveys failure and as an antecedent when it prompts or cues the conditions under which responses will be reinforced and/or punished. Perhaps it is because it can operate in all these ways at once that feedback has been found to be an especially powerful behavior modification technique (Locke and Latham 1990).

External and internal rewards play a significant role in getting people to accept goals and motivating them to maintain goal-relevant behavior in the long-term. The key reward in organizational settings is probably not feedback but rather the consequences to which feedback leads such as recognition, praise, raises, financial rewards and promotions, privileges, material or monetary rewards, which can be embedded within compensation or performance appraisal systems and preferred activities or assignments. Internal, self administered rewards that can occur following high performance include a sense of achievement based on attaining a certain level of excellence, pride in accomplishment and feelings of success and efficacy.

According to a German study (Zimolong 1999), the most frequently documented as well as practiced personnel systems in OH&S are performance appraisal, leadership, training and reward systems for managers and supervisors. In a successful HSE management system at least two personnel systems are combined with leadership, mostly appraisal and reward systems. Companies striving to get better do not yield typical patterns of personnel systems in their specific transition phase, however, they use the most systems, whereas poor companies use the fewest systems. Which system they prefer depends on the organizational and cultural characteristics of the company. Organizations differ fundamentally in how they control the compliance with the application of personnel systems. Successful companies have deployed monitoring systems. If deviations from standards have been observed, necessary corrective actions are undertaken by the personnel department or by line managers to remedy any deficiencies.

The deployment of corporate targets to the level of teams and individual employees is an essential step towards the achievement of HSE. Management by Objectives (MbO) in a cooperative fashion meets one of the most important requirements of HSE, that of involving all employees. Goal agreement concepts on the individual level cannot be efficient unless there is a functioning system of regular feedback and if consequences are to be expected in the case of goal deviation.

Whereas participatory approaches were pursued in the past mainly under the aspect of the quality of working life, it is the optimum use of human resources nowadays that is emphasized in people empowerment, making it a predominantly an economic target. Teams usually have a higher degree of empowerment. Not only do they accomplish their regular tasks, but also they deal with all processes concerning their unit. This means that they integrate organizational, planning, supervision and improvement tasks, which were traditionally performed by supervisors or special functions/departments. Transient team concepts with

problem-solving involvement include both quality circles and project teams. Forms of permanent teamwork with different scopes of problem-solving and decision-making autonomy have been introduced not only in the manufacturing area, but also in the development, sales and administrative areas.

3.2. Information and Communication

Risk management means coordinating and controlling a continuous improvement process through formulating, setting and adjusting objectives and standards, monitoring and reviewing deviations from standards and providing remedial actions. From this point of view, risk management is a communication concept related to person, group and unit levels. It includes the idea of a complex network of interrelationships at the intragroup, intergroup and even interorganizational levels. In practice, communication and information between people, groups and units is difficult to organize and maintain due to physical, social and psychological barriers. Organizations are open systems and linked in multiple ways with the social and political environment. Successful organizations establish and maintain arrangements for identification and reception of relevant HSE information from outside the organization. This may include new legislation, or at least amendments, information and developments in HSE (best practices) and information necessary for the identification of hazards and control of risks, for example from customers or suppliers.

There is a variety of systems and procedures which ensure the communication of necessary information to all people in the organization who need it. Written statements of HSE policy, documented performance standards, reporting systems and newsletters are quite a few examples. Direct forms are oral presentation, face-to-face discussions at consultative meetings, health and safety tours, team briefings as well as monthly or weekly "tool box talks" on health and safety issues. The new information and communication technologies play an important part to support the processes, for example intra- and Internet communications via e-mail, electronic boards and management information systems.

Organizations need sufficient knowledge, skills and experience to identify and manage HSE risks effectively. Documentation is the process of assembling and retaining HSE knowledge. Accessibility of data is a critical factor to use the knowledge, for example, key documents, such as working procedures, records and instructions should be accessible at the point of use. Other options are:

- Training managers to a sufficient level of competence and keep them up-to-date with developments in HSE.
- Employing appropriate HSE professionals as part of the management team.
- Hiring external specialists support where in-house expertise and/or resources are insufficient to meet the organization's needs.

3.3. Design of Work and Technology

The outcomes of safety, health and well-being depend to a significant degree on the installation and maintenance of safe and reliable technology, as well as on ergonomic and psychosocial aspects of work design. General topics and methods of systems safety engineering are concerned with guarding energy sources,

design and redesign of machinery, equipment and processes, application of environmental standards and establishment of inspection systems, such as statutory engineering inspections of pressure vessels, cranes and lifting machines, or electrical installations (Ridley 1994).

Ergonomic and psychosocial aspects of work are potential contributors to the health and well-being of employees and organizations. Their health effects can also be regarded as contributors to work motivation, or work performance. From an ergonomic perspective, the design of the environment, workstations, tasks, work organizations and the tools or technology should reasonably accommodate the employees' capacities, dimensions, strengths and skills. Particularly, it should accommodate the human sensory capabilities, support the decision-making processes and adapt to the human performance abilities. Error-prone designs place demands on performance that exceed capabilities of the user, violate the user's expectancies based on his or her past experience and make the task unnecessarily difficult, unpleasant and error-likely. One example are back injuries: high physical energy expenditure, no variation in work movements and postures, prolonged standing, sitting or stooping, or repetitive work can contribute to the development of musculoskeletal problems, low-back pain and back disorders.

Critical aspects of the success of ergonomic interventions for providing health and psychosocial benefits are a strong management support for the ergonomic program. Further aspects are: the involvement of managers, supervisors and employees, the willingness to participate in defining problems, proposing solutions and improving work practices; and the application of engineering improvement to reduce biomechanical and physiological loads.

There are various characteristics of working that have generally been shown to have negative physical and/or psychological consequences, for example, machine-paced work, a lack of task control, high job demands, shiftwork, time pressure and poor supervisory relations. A person normally copes with transactional periods of stress by either altering the situation or controlling his or hers reactions. Problems arise when work conditions are in conflict with human capacities and expectations over a long period and when coping fails.

The extent of the negative consequences varies from individual to individual depending on the perceived threat, individual constitution and coping mechanisms. Tension, boredom, worry, anxiety and irritability are inevitably some of the first indicators of strain. Emotional stress reactions are quite normal responses. Short-term stress reactions may include increases in blood pressure, adverse mood states and job dissatisfaction. Depression and apathy are later symptoms. Long-term stress reactions may even accumulate to cardiovascular diseases and upper extremity disorders (Kalimo et al. 1997).

Job control and social support are beneficial factors for well-being and job satisfaction. The person can have control over various job demands, such as the task itself, pacing of the work, work scheduling, the physical environment, decision-making, other people or mobility. When job control is high, the other job demands tend to have less potential adverse effects on health. Social support is thought to exert a protective function during conditions of stress, i.e. to "buffer" a person against the harmful effects of the social environment. Evidence for the beneficial effect

of job control and social support is conclusive enough to promote better job designs and interpersonal interaction at the workplace.

The criteria in job design deal with the physical work environment, compensation systems, institutional rights and decisions, job content, internal and external social relations and career development. According to quality of working life principles the following characteristics of job content are of main interest: variety of tasks and task identity; feedback from the job; perceived contribution to product or service; challenges and opportunities to use one's own skills; and individual autonomy. Depending on the degree of autonomy, the team design approach allows members to regulate their work activities by themselves. People get a better idea of the significance of their work and create greater identification with the finished product or service. If team members rotate among a variety of subtasks and cross-train on different operations, the team can become more flexible. Teams with heterogeneous background also allow for synergistic combinations of ideas and abilities not possible with individuals working alone and such teams have generally shown higher satisfaction, better involvement and superior performance, especially when task requirements are diverse.

In any redesign process there are tradeoffs among specific improvements and achieving the best "overall" job design solution. There is no perfect job design that provides complete psychological satisfaction and health for all employees and maximizes the outcomes of the organization. Making jobs more mentally demanding increases the likelihood of achieving people's goal of satisfaction and motivation, but may decrease the chances of reaching the organization's goals of reduced training, staffing costs, decent wages and error-free products and services. Which tradeoffs will be made, depends on the outcomes the organization prefers to maximize.

4. RISK CONTROL SYSTEMS

4.1. Life Cycle Assessments

Production is cyclical in nature. Raw materials are taken from the earth, processed, used and returned to the earth as waste. Life cycle assessment (LCA) measures the continual process of extraction, use, disposal and extraction. The LCA, commonly referred to as "cradle-to-grave" assessment, examines all aspects of environmental impacts associated with the life of a product or service from the extraction of the raw materials through the pre-production process to the distribution and final disposal. LCA is an integral part of an ecological auditing procedure and provides a systematic framework for impact analysis and the continual updating of procedures. LCA analyses all environmental impacts related to product development and permits the comparative analysis of the environmental impacts of similar products. Assessed is the product rather than the system, as in EMS, but does not include the impacts of the processes involved, such as energy usage. LCA cannot be restrained to just one site or indeed to one company, it requires significant cooperation down the supply chain to produce a product LCA. Since all areas of production have environmental impacts, each stage needs to be traced back to the source to fulfill the requirements of the LCA process. This includes the life cycles of other embedded systems, for example of capital equipment, work systems and goods.

The term system is used to clearly describe defined activities, processes and equipment during the lifetime of the system. The overall lifecycle of a product may comprise the following phases: conception and planning; design and engineering; use/operation; modification and maintenance; demolition and disposal. With regard to OH&S the system is to work the way it was planned and nobody should be harmed by an accident, toxic substance or malfunction. OH&S and ecological management share common objectives. For example, the replacement of hazardous raw materials and the auditing of all substances used in the production process is one measure to improve OH&S as well as one way to minimize waste.

Risk control systems are the basis for ensuring that adequate performance measurement systems and both proactive and reactive control activities are provided and maintained during the operational lifecycle. At the input stage, the aim is to minimize hazards and risks entering the organization. At the process stage, the focus is on containing risks associated with the process. At the output stage, controls minimize the export of risks off-site that may arise from work activities, products and services (HSE 1997). Particularly the system safety concept emphasizes that hazards built into a technology or activity and the preventive measures to eliminate and control them are largely conditioned by the decisions made in the planning and design phase of the activity.

In all stages of the life cycle of a system performance standards are important to control the risks. They cover the physical resources (i.e. workplaces, materials and substances, plants), the human resources (recruitment, selection of personnel and contracting organizations) and the information related to health and safety, risk control and positive health and safety culture. From a systems point of view, unsafe acts and unsafe conditions are substandard practices and substandard conditions, i.e. deviations from an accepted standard or practice. A vast number of substandard conditions involve poor ergonomic design of machine, equipment and the work environment. It is essential to consider these practices and conditions only as symptoms, which point to the latent failures (Wagenaar *et al.* 1994) or basic causes behind the symptoms. The origins of deviations from a standard are deficiencies in management and organization. Detailed lists that cover different factors such as work environment, individuals doing their job and management are provided in Wagenaar *et al.* (1994).

A holistic risk management considers all phases of the life cycle of the plant or facility it exploits and the goods and services it delivers. Each phase is an activity that must be managed according to the formulated policy in its own right and which must feed forward into subsequent phases and have a feedback loop to facilitate organizational learning processes. This approach requires the development and deployment of a performance measurement system, based on a combination of both qualitative and quantitative as well as proactive and reactive monitoring data.

4.2. Performance Measurement

Information needs vary at different levels and in different parts of an organization. For example, senior staff need key performance indicators to confirm that the HSE system is working effectively. At the operational level many other performance indicators may be necessary to monitor implementation and effectiveness of risk controls. Large organizations have developed a system where annual measurement summaries are reported upwards to senior

staff. Both proactive and reactive monitoring data are used as outcome indicators to determine whether objectives are achieved. Proactive monitoring data include:

- number of staff suggestions for HSE improvements;
- staff perceptions of management commitment to HSE (e.g. from an in-house survey);
- frequency of HSE audits;
- quality of health surveillance reports; and
- workplace exposure levels.

Reactive monitoring data include:

- number of unsafe acts observed on a safety tours;
- unsafe conditions examined on a "walk through" basis;
- near misses reported from staff on a monthly basis; and
- lost-time accidents sampled from healthcare station.

Selecting appropriate outcome indicators depends on the chosen objectives. Proactive monitoring is used to check compliance with the organization's HSE activities, for example to confirm by signature that the weekly "tool box" talks have been attended by all members of a unit. Reactive monitoring is used to investigate, analyze and record HSE management system failures — including incidents and accidents. Measuring performance against predetermined standards is based on a proactive examination of both hardware (premises, plant, substances) and software (people, procedures and systems, including individual behavior). Deviations from standard reveal when and where action is needed to improve performance.

In addition to routine monitoring of HSE performance, there is a need for periodic audits that enable a deeper and more critical appraisal of all the elements of the risk management system. Concerning nature and extent of an audit, it may cover the whole risk management system or part of it; technical matters concerning plant, equipment or process; corporate processes or functional units. A validation audit is intended to control the effectiveness of the risk management system, a compliance audit verifies whether the organization or part of it is complying with its own standards and procedures.

4.3. Risk Assessment and Control

Risk assessment and control is part of the proactive HSE planning approach that emphasizes prevention. The main purpose of risk assessment is to determine whether planned or existing controls are adequate.

The following steps are necessary to carry out a risk assessment:

- identify hazards;
- estimate the risk from each hazard, e.g. the likelihood and severity of harm;
- decide if the risk is tolerable; and
- prepare risk control action plan (if necessary).

A hazard is a source of potential harm or damage or a situation with potential for harm or damage. The harmful effects of health hazards, such as hearing loss, cancer, liver damage, silicone are regarded as illnesses. However, back pains may result from improperly designed chairs, headaches from poor ergonomic layout of VDT work place. This exceeds the traditional view on accidents. Consequently, controls are not always identical: the prevention of contact or its reduction to a level where no harm is done is valid only for hazardous or toxic materials, whereas illnesses resulting from poor design require ergonomic standards,

planning and sometimes complete re-installation of the working system.

In the simplest case, hazards can be identified by observation, comparing the circumstances with the legal standards and guidance. In more complex cases, measurements such as air sampling or examining the methods of machine operation may be necessary to identify the presence of hazards presented by chemicals or machinery.

Competent people with practical knowledge of the work activities can carry out risk assessments, preferably with colleagues from another part of the organization who may have greater objectivity. A risk assessment based on a participative approach provides a further opportunity to integrate employees into the activities, use their knowledge, develop a shared perception of hazards and risk, empower responsibility and accountability.

Assessing risks is necessary to identify their relative importance and to obtain information about their extent and nature. This will help to identify where the major efforts in prevention and control should be placed and to make decisions on the adequacy of control measures. It is generally not necessary to make precise numerical calculations of risk. Complex methods for quantified risk assessment are normally only required where the consequences of failure could be catastrophic. Risk assessments in major hazard industries, for example, in the chemical or nuclear industry, may be required by legal standards and guidance. Hazard and operability studies (HAZOPS) and hazard analysis systems such as event or fault tree analysis may be applied when planning a new system or major changes of an existing system.

5. RISK MANAGEMENT CHALLENGES

The past two decades have witnessed a significant transformation in how firms are structured. Tall organizations with many management levels have become flatter, competitors that have adopted a modular organizational structure have gained market shares. Organizational de-layering and the rise of smaller, often entrepreneur-based firms gives self-management new meaning covering personal self-management, self-leading teams and semi-autonomous units. Companies and public services adopt cooperative forms of work at a very fast pace. Teleworks provides flexibility in both working hours and the location of work and allows employees to cultivate tailored life-styles while working a full-time job. These "boundaryless" organizations, e.g. organizations whose membership, departmental identity and job responsibility are flexible create new challenges for risk management particularly for people management.

The traditional approach to managing people focuses on selection, training, performance appraisal and compensation for individuals in specific jobs. It also presumes a hierarchy of control rather than horizontal workflow sequences. When organizations are restructured around teamwork instead of individual performance, different forms of team autonomy and HSE responsibilities are emerging.

Risk management systems in small, medium and modular organizations must rely on the participation of members, on self-management, personal and team responsibility. Selection, performance appraisal and reward policies are the most likely candidates for change. Contingent pay and peer pressure

generated by teams are emerging as substitutes for both managerial influence and internalized member commitment. HSE criteria, rules, procedures and achievements have to be reformulated and integrated into the systems. De-layered organizations also need to develop combined systems, particularly, a recruitment system which includes participatory concepts of job analysis and assessment and a team management system focusing on performance appraisal, compensation, rewards and benefits and personal career development. Systems are based on peer reviews instead of managerial appraisal and on team based output measures instead of individual performance. This requires an intensive qualification process for the employees and new cognitive and social skills to run the systems.

The rewards themselves are changing. Promotions and formal status have become less important and are replaced by lateral moves, presented as career-building assignments Workers often perceive training as a reward, providing self-actualization and the motivation to learn: career development with increased responsibility, autonomy and likelihood of advancement; and psychosocial benefits, including increased confidence, new friendships and better well-being in non-working life.

REFERENCES

BRITISH STANDARDS INSTITUTION, 1996, *BS 8000: Guide to Occupational Health & Management System* (London: BSI).

HEALTH AND SAFETY EXECUTIVE, 1997, *Successful Health and Safety Management: HS(G)65* (London: HSE).

HUTCHINSON, A. and HUTCHINSON, F., 1997, *Environmental Business Management* (London: McGraw-Hill).

KALIMO, R., LINDSTRÖM, K. and SMITH, M. J., 1997, Psychosocial approach in occupational health. In G. Salvendy (ed.) *Handbook of Human Factors and Ergonomics* (New York: Wiley), 1059–84.

LOCKE, E. A. and LATHAM, G. P., 1990, *A Theory of Goal Setting and Task Performance* (Englewood Cliffs: Prentice Hall).

RIDLEY, J., 1994, *Safety at Work* (4th ed.) (Oxford: Butterworth-Heinemann).

WAGENAAR, W. A., GROENEWEG, J., HUDSON, P. T. and REASON, J. T., 1994, Promoting safety in the oil industry. *Ergonomics*, 37, 1999–2013.

ZIMOLONG, B., 1997, Occupational risk management. In G. Salvendy (ed.) *Handbook of Human Factors and Ergonomics* (New York: Wiley), 989–1020.

ZIMOLONG, B., 1999, *Elaboration and Test of New OSH Concepts for New Types of Work Organization* (Bilbao: Agency for Safety and Health at Work).

The Role of Ergonomist in a Design Engineering Environment

M. Bellemare[†], D. Imbeau[†], J.-G. Richard[†] N. Perron[‡] and S. Gauthier[‡]

†Institut de recherche en santé et sécurité du travail du Québec, Montréal, Québec H3A 3C2, Canada

‡Alcan CP1500 Jonquière, Québec, Canada, G7S 4L2

1. INTRODUCTION

To be an ergonomist in an engineering design environment may be assumed in many ways. Traditionally, ergonomists have been asked for inputs at the final stage of design processes to help designers fix some details to ensure that the final product will fit the human being. More recently, with new methods in engineering new ways of intervention are developing. Moreover, ergonomic interventions now cover many domains in the design field: technical devices, either hardware or software, work organization and training. Ergonomists are now being asked to participate in teams designing whole plants. A model of ergonomic interventions that defines the ergonomist as an effective player in the design process is presented here.

The ergonomist may be an "actor" in the design process, that is, s/he may not only be a resource person called when advice is required but also can actively take part in the design team so as to influence the design output as well as the design process itself. This kind of contribution rests on the consideration that designing is not only a technical process but also a social one (Bucciarelli 1988). Besides being a recognized expert in human characteristics and work activities analysis (de Keyser 1991), the ergonomist can then achieve a complete intervention through structured interactions with project staff and future users. Two conditions seem to be required for such an implication. First, the ergonomist must have a good understanding of the project itself: from what it stems, what are the stakes of the project, what is planned, what is the calendar, what are the methods to be used to reach the results. Second, the ergonomist must analyze the structures in the project: who are the key actors, for example, who is the sponsor (i.e. the payer and often the main decision-maker), who is the project manager, who are the stakeholders. Such knowledge allows the construction of the ergonomic intervention with the different actors. Even if the ergonomist masters a particular field of competence, s/he must have a good understanding of these aspects to provide the right input at the right time. Also, the ergonomist must have some knowledge of the interactions between the different actors to be able to give advice in a form that helps the project manager negotiate the different demands generated by the ergonomic intervention. This may for instance, be the case with suppliers when a fixed budget has been set.

The engineering design environment should include two structures through which the ergonomist will act (see Figure 1). The first is a steering committee composed of the stakeholders, project manager and sponsor. The second is one or more users group (composed of maintenance and production workers, su-
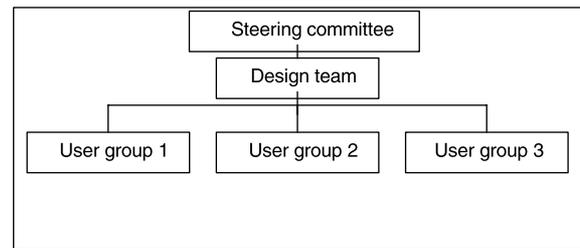
Figure 1. Structures the ergonomist has to interact with in a design project.

pervisors) working on specific topics which the sponsor or one of its representatives will lead from the very beginning until the end of the design process and even after when necessary. The ergonomist will support this group in analyses of the project's ergonomics aspects.

How the ergonomist may contribute to the different steps of the design process will now be examined. Through structures described above and using the concept of "future work activity" developed by Daniellou (1987) the intervention of the ergonomist takes place through the following design process stages: (1) defining the problem, the opportunity; (2) choosing the concept; (3) conducting preliminary studies; (4) conducting detailed studies; (5) construction; and (6) start-up.

One must keep in mind that these stages often overlap extensively in real world projects and that the designing process is an iterative one: many loops backwards in the temporal flow can be observed. For instance, in an overhead crane project that spanned over 1 year, the last month before start-up while the new building structure was being completed, the operator cabin was still at the preliminary studies stage. This example illustrates the short delays an engineering design team is often faced with, particularly when it comes to workstation design.

2. DEFINING THE PROBLEM, THE OPPORTUNITY

Even though the ergonomist might be called at the very beginning of a project, the project's objectives are often already defined. In a longer-term project, however, the objectives may change. Hence, the ergonomist must keep in touch with the stakeholders to be aware of these changes as soon as they occur.

Most of the time, design projects are aimed at solving existing problems. The analysis of the problem will have an impact on its solution. It is important to ensure that the problem is analyzed systematically to identify its true determinants. Often, users will dictate a solution to a problem that has not been subjected to a systematic analysis. The designer may then be locked in a process where the output will not meet users' expectations despite the strong feeling users may have that "this is the solution," Many engineering projects consist in undoing what has been poorly designed because of incomplete analyses. Moreover, as suggested by Schön (1983), "setting" a problem which needs a technical answer is not a technical problem.

Another type of projects arises from an opportunity offered in the company life. For example, an acid unloading station is re-designed in an aluminum plant because of a change in train wagons configuration. This opportunity offers a possibility of

improving working conditions. For example, going back to accident reports of the past years, new objectives stemming from a health and safety point of view may be added (e.g. reduction in manual materials handling work, fewer awkward postures, improved safety). This task of enriching project objectives can be done with the help of a work group composed of production and maintenance workers and proposed to the steering committee as part of the project.

The ergonomist can also assist users in defining their problem from the very beginning of the design process, on the basis of fieldwork done at the installations to be changed. Field observations are very useful when investigating an existing situation. In these observations the ergonomist will try to determine what in the work environment can be harmful to workers and also what in the work activities needs to be improved from a production point of view. The work group can then validate what has been found by the ergonomist and also provides pertinent information about situations that rarely occur but which are critical to installation efficiency. Analysis of a problem will often show that health and safety issues are associated with system dysfunctions and that correcting the latter will improve the former. In particular cases, however, such as with paced production lines, designing for a productivity increase can yield a worsening in working conditions. The role of the ergonomist here is to document the consequences on workers health and safety of the design choices and to ensure that these are thoroughly discussed by the stakeholders. As opposed to the ensuing design process stages where engineers are leading, this first step should be lead by the sponsor and users.

3. CHOOSING THE CONCEPT

Once the problem space is correctly defined, the following step is to determine which concept will provide the most satisfying answer to the objectives set previously. An approach used by most designers is to visit reference sites. The project manager will indeed try to find existing facilities where the setup is close to what is expected in the new design. The ergonomist may add to the technical data obtained by engineers during such visits by completing with work activities data (e.g. kind of problems users have in the existing set up). From his/her knowledge of the situation to be changed, the ergonomist will also try to anticipate future events that could occur in the system under development. One key issue here is for the ergonomist to be aware of the sources of variability in the existing situation such as variability related to operators, to raw materials, to products, to the environment, etc. and to anticipate how these sources of variability will be taken into account in the system under development. Preparing the visit with users, visiting and debriefing with them will usually provide interesting ergonomics input and help in establishing advantages and disadvantages of a concept from the users point of view.

4. CONDUCTING PRELIMINARY STUDIES

Once the concept is chosen, studies are conducted by engineers to determine the feasibility of the project and to quantify the resources needed. At this point, the resources required to ensure the objectives pertaining to the ergonomics aspects must be anticipated. It can take the form of a special budget for conducting simulations on prototypes, or testing workstation concepts

through mock-ups, or having workers participate in meetings and visits. A design alternative can also be evaluated from an ergonomics point of view (e.g. given this particular concept, what will be the activities of the production worker, of the maintenance staff). The role of the ergonomist at this stage is generally underestimated. Since at this stage the structural aspects of the project are decided upon, it will often be difficult to change these later without significantly increasing the project costs (engineering and overall costs). Therefore, if major difficulties from the users point of view have surfaced, these may be attenuated by revising alternatives or design choices. Once this stage is completed the project is ready for budget approval and the main suppliers who will become active players in the next steps can be known.

5. CONDUCTING DETAILED STUDIES

For the project manager, the aim of this fourth stage is to come up with precise plans that will render possible the realization of the installation. In most engineering environments, it is at this stage that the need for an ergonomist is striking: information on conveyor heights, control panel design, seat types, all of these issues need urgent answers. There are, of course, textbooks, reference manuals and software products that can help designers. But most of these data need interpretation and this can only be done by an ergonomist as long as s/he can anticipate what the work activities will be.

This is where it pays to have the ergonomist in the project from the very start: s/he can provide better advice and help the design team in working out the details of the installation with the different suppliers. One way of helping designers in the completion of plans is to simulate with different supports (plans, mock-ups, software products, prototypes) the anticipated workers activities. Scenarios can be constructed from data collected in the existing situations taking into account the future user characteristics, the regular production and maintenance task as well as the related critical tasks. The main goal of simulation is to make forecasts on the future work activities: are the working conditions acceptable for most of the population and from an operations point of view, are the objectives met. Modifications can still be made at this stage without adding undue expenses to the initial cost. This is also the stage where the word "iteration" takes all of its meaning; if the forecast on the workers future activities is poor, designers will modify their proposal which can again be tested and so on until the design is deemed acceptable. Discussions on the acceptability of the proposals will often oppose different professional fields: production, health, safety, maintenance may not necessarily point towards the same solution. The final word always results from compromise. These discussions are important for the remainder of the design process. For example, in a discussion about a joystick, the engineer in charge of the project may suggest to choose one with a high resistance to guarantee that it will not be activated accidentally, while the ergonomist may suggest the force to be as low as possible to prevent musculoskeletal strain. On the other hand, maintenance people will be preoccupied by frequently replacing the joystick's springs. The final decision may reflect all of these considerations; however, every actor involved in the discussion then knows what to expect from the selected design in terms of consequences for the work activities and for the function s/he represents in the plant.

6. CONSTRUCTION

The ergonomist is usually not involved at the construction stage. But from experience, many changes are made on the spot, particularly when building in an existing plant where there are many constraints. Some attention must be paid not to introduce undue risks for future users. One of the rewards related to an early involvement of the ergonomist in the design process is that the work done in the previous stages with the other design actors, including workers from user groups, makes them more attentive to what is being done on the construction site.

7. START–UP

Start-up is a critical period where workers learn to use their new workstations and tools as well as living in a new environment. The training program that has been developed (with a likely contribution of the ergonomist) in the previous stages will be put into practice. As part of the ergonomist's intervention at this stage, an analysis of the work activities and particularly of the problems faced by users during the start-up, can help fine tune the new system and improve the training program. The start-up stage is also a good time for evaluating the ergonomic intervention with the design actors and for modifying accordingly the strategies and methods for a future project.

Whatever type of involvement in a design environment, the ergonomist can enhance the quality of an engineering project. The earlier his or her involvement, the better the results. Even if an early involvement of the ergonomist is not always achievable, one must keep in mind that an engineering project is a re-definition/definition of future work situations. At each stage of the design process, decisions are taken that will affect the way future work will be done. The stakes underlying these decisions require for the ergonomist to act not only in providing information and advice on the end results, but also in keeping alive interactions between stakeholders, sponsors, users and the project manager.

ACKNOWLEDGEMENTS

REFERENCES

BUCCIARELLI, L.L., 1988, An ethnographic perspective on engineering design. *Design Studies*, 9, 159–68.

DANIELLOU, F., 1987, Les modalités d'une ergonomie de conception. Introduction dans la conduite des projets industriels. *Cahiers de notes documentaires*, no. 129, 517–23.

DE KEYSER, V., 1991, Work analysis in French language ergonomics: origins and current research trends. *Ergonomics*, 34, 653–69.

SCHÖN, D.A., 1983, *The Reflexive Practitioner: How Professionals Think in Action* (New York: HarperCollins).

Safety Culture

A. I. Glendon

Department of Applied Psychology, Griffith University Gold Coast, Queensland 9726, Australia

1. INTRODUCTION

The term "safety culture" arose from the Chernobyl nuclear disaster in 1986, in which cause was attributed to a breakdown in the organization's safety culture (IAEA 1986). The concept was heralded as a substantive issue in official inquiry reports into disasters at King's Cross Underground station in London in 1987 and on the North Sea platform *Piper Alpha* in 1988. The term rapidly gained currency within the safety management lexicon. The notion of safety culture arose from the more inclusive concept of organizational culture as well as the related concepts of organizational climate and safety climate. Organizational culture is often held to be critical to an organization's success or failure. Safety culture is frequently identified, for example by disaster inquiries, as being fundamental to an organization's ability to manage safety-related aspects of its operations – successfully or otherwise. The assumed link between safety culture and major disasters was subsequently extended to include accidents of all types.

It is generally agreed that culture operates at different levels and through various mechanisms. Cultural features are complex, shared characteristics of a group dynamic relating to a system (e.g. group, community, nation, religion). They include beliefs, values, attitudes, opinions, motivations, meanings, ideas, expectations, linguistic features, specific actions, rituals, ceremonies, symbols, general behaviors and associated norms. More tangible aspects include buildings, uniforms, documents, liveries, logos, equipment and design features. Within the culture of an organization or other definable system, subcultures may be identified in respect of groups whose characteristics are sufficiently distinct from those of the mainstream culture or from other subcultures.

2. ORGANIZATIONAL CULTURE

Two contrasting perspectives on organizational culture dominate the literature as well as managerial and professional practice within organizations (Waring and Glendon 1998). These have been described as functionalist and interpretive. Functionalist approaches assume that organizational culture exists as an ideal to which organizations should aspire so that it can, and should be, manipulated to serve corporate interests. The notion that organizational culture has as its prime function to support management strategies and systems is premised on the assumption that it can be reduced to relatively simple models of prediction and control. This approach primarily aligns organizational culture as a support for managerial ideology, goals and strategy, in extreme cases involving managerial use of "culture" to coerce and control.

However, many aspects of culture are not reducible to cause–effect quantitative relationships, and ideological use of culture as a weapon in organizational struggles reveals a powerful unitarist bias. Interpretive approaches to organizational culture provide an alternative approach to this concept and assume that organizational culture is an emergent complex phenomenon of social groupings and serves as the prime medium for all members of an organization to interpret their collective identity, beliefs and behaviors. Thus, organizational culture is not owned by any single group but is rather a unique creation of all the organization's members.

From assumptions characterizing interpretive approaches to organizational culture, it follows that managerial attempts to manipulate culture, for example in seeking to drive rapid organizational change, are likely to fail because of the application of an inadequate model of processes that they are attempting to manipulate. An analogous point, in respect of organizations seeking to enhance safety culture as the "philosopher's stone" to improving health and safety, is made by Cox and Flin (1998).

In practice, many organizations display elements of both approaches. For example, through rigorous adoption of formalized risk management practices, an organization might be seen as invoking functionalist aspects of its safety culture. Simultaneously, a more interpretive side may be revealed by individual and group commitment to open-ended learning from past mistakes, such as those leading to accidents. This could be achieved through communication and discovery processes that involve a developing identity for organization members.

3. FROM ORGANIZATIONAL CULTURE TO SAFETY CULTURE

Contrasting perspectives on organizational culture inform directions that organizations take in respect of safety culture. Functionalist assumptions may be appropriate for components of safety culture that can be shown to follow reliable cause–effect type models, for example as represented by readily quantifiable relationships. An interpretive approach may be subsumed within an essentially functionalist framework as a management tactic for maintaining control. For example, the dominant safety paradigm for an organization may be a legislative/standards driven formalized risk assessment system. Within this may be incorporated employee consultation or involvement of various types, which may use various methods to enhance or augment the mainstream system. This differs from a system that is premised by trusting safety to staff at all levels and which is tolerant of diverse and heterogeneous contributions.

To understand the nature of major accidents within organizations and systems, it is essential to invoke the notion of culture (Reason 1997). The notion that diversity and flexibility are essential components of an organizational culture that contributes positively to safety is championed by several researchers in the field (see, for example, contributors to Cox and Flin 1998). This challenges the functionalist assumption that a monolithic culture across an entire organization is desirable, and is premised upon the certainty of an unknown future in which accidents or disasters will differ from those previously experienced and also be difficult to imagine, but which could nevertheless occur. A number of authors caution that seeking to use components of safety culture as a "defence in depth" against major accidents, perhaps through a "total safety culture" ideal, is doomed to long-term failure through organizational complacency as well as lack of diversity and flexibility of response.

4. DEFINING SAFETY CULTURE

A range of meanings has been attached to safety culture, three of which are summarized by the UK Institution of Occupational Safety and Health (1994):

- those aspects of culture that affect safety (Waring and Glendon 1998);
- shared attitudes, values, beliefs and practices concerning safety and the necessity for effective controls;
- the product of individual and group values, attitudes, competencies and patterns of behavior that determine the commitment to, and the style and proficiency of, an organization's safety programs (cf. UK HSC 1993).

The latter two definitions are premised upon a functionalist perspective. This approach implies that safety culture is conflict-free and aligned with the objectives of a controlling function. The first of the three meanings reflects a more interpretive approach, also espoused by Glendon and McKenna (1995), implying a looser definition around a set of principles that is at least pluralist in orientation.

HSC (1993) proposes as a definition of safety culture, "the product of individual and group values, attitudes, perceptions, competencies, and patterns of behavior that determine the commitment to, and the style and proficiency of, an organization's health and safety management." It argues that organizations with a positive safety culture are characterized by communications founded on mutual trust, by shared perceptions of the importance of safety and by confidence in the efficacy of preventive measures. From their review, the HSC concluded that effective health and safety provision depends as much on organizational culture as upon specific attention to health and safety matters.

While definitions of culture and safety culture abound, partly because the theoretical underpinning for safety culture is weak, the concept remains problematic. For example, there is little or no theoretical basis for determining what makes an organization's safety culture "positive," "healthy," "strong," "excellent" or merely "good" – or the antonyms of these frequently used descriptors. Reason (1998) argues that a safe culture is one that is informed and just, being based upon problem solving rather than indiscriminate blame apportionment, and includes reporting as an essential element. Reporting of mistakes and violations is likely to be encouraged when people feel that the organization trusts them and shows evidence of responding in a problem solving manner that rewards their behavior.

5. CULTURE AND CLIMATE

Confusion continues between use of the terms "culture" and "climate" and they are often used interchangeably. While there is a strong relationship between them, organizational climate refers essentially to the perceived quality of an organization's internal environment in which employee attitudes and perceptions feature prominently. Climate is regarded as a more superficial concept than culture, being descriptive of important aspects of the current state of an organization. Culture is often seen as being long-term and strategic, while climate is perceived as short-term and tactical. Scaled dimensional measures are the most popular means of measuring organizational climate and many of these have been devised. Dimensions typically assessed include autonomy, cohesion, trust, pressure, support, recognition, fairness and innovation.
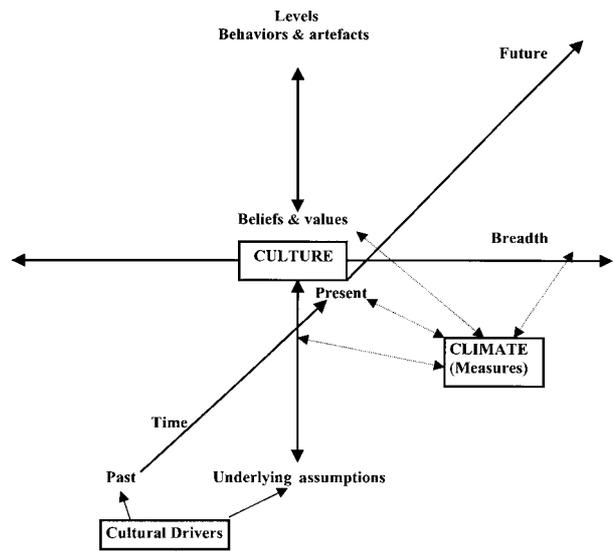
Figure 1. Organizational culture and organizational climate (Glendon and Stanton 1999).

A typical three-level classification of organizational culture embodies relatively accessible, intermediate and deep levels. The most accessible level refers to observable behaviors and associated norms. The intermediate level includes attitudes and perceptions, which are not directly observable, but which may either be inferred from behaviors or assessed through questioning. At the deepest level are core values, which are much less readily assessed. Other key dimensions of organizational culture that have been identified include depth, breadth, progression, strength, pervasiveness, direction and localization (Waring and Glendon 1998).

Some authors consider climate to be one manifestation of culture. Figure 1 represents one view of the relationship between organizational culture and climate. Here organizational culture is represented at three different levels as well as having breadth and time dimensions. Organizational climate measures can access certain components of the dimensions of organizational culture across a limited range. For example, those relating to member attitudes, beliefs and perceptions as well as cultural breadth at the time that the survey is undertaken and perhaps a little in the past as well. However, as organizational culture change is generally taken to occur over a period of years, the time frame for assessing culture must reflect this important time dimension.

6. SAFETY CLIMATE

Contemporaneously with the derivation of safety culture from organizational culture, another important influence was from the associated term "safety climate," which comes from a more empirical tradition. Some researchers distinguish between safety culture and safety climate. The prime research method for investigating safety climate is the questionnaire, completed by sufficient numbers of employees to enable statistical analysis to reduce a large number of items to a small number of dimensions. Typical safety climate dimensions include management attitudes to safety, safety training, worker involvement and employee risk perception.

"Safety culture" is generally taken to be more embracing than

"safety climate," although the two terms have similar meanings. "Culture" implies a notion of residing within an organization. "Climate" has more passive connotations, reflecting attitudes and perceptions of organization members to both internal (e.g. management actions) and external (e.g. economic) influences. Exclusively psychometric (e.g. questionnaire survey) approaches have proliferated as measures of safety climate (cf. Cox and Flin 1998, Hale *et al* 1999), but have restricted range in respect of representing measures of safety culture.

7. MEASURING SAFETY CULTURE

Several empirical studies have sought and found associations between many ascribed individual components of safety culture, safety climate and other measures. However, convincing evidence of causal links between global measures of safety culture or even safety climate and safety performance outcomes has so far proved elusive. This has led to a degree of skepticism among some Occupational Health and Safety (OHS) professionals in respect of whether these terms might obscure more concrete safety performance measures, such as accident outcomes. It remains a possibility that attitudinal and cultural aspects of safety are not particularly strongly related to behavioral and environmental components. However, a definitive answer to this issue must await more fundamental advances in measurement of these concepts.

Ways in which safety culture and safety climate have been measured and explored include case studies, comparative studies and psychometric approaches. Hitherto most empirical studies in this field, mainly using questionnaires, have measured dimensions of safety climate. Studies using triangulated methodologies to investigate broader, deeper and historically derived safety culture features have been undertaken and it is in this area that methodological advances are most required.

Studies that model various aspects of safety climate and safety culture are beginning to appear (see, for example, contributors to Cox and Flin 1998, Hale *et al* 1999), although where questionnaires are the only data gathering instrument, common method variance and restriction of range are generic methodological problems. A few studies seek to validate the safety culture or safety climate measure used by correlating features of it with accident or other outcome measures (e.g. protective clothing or equipment use). However, as behavioral measures these have limited utility. Despite the possibility of discriminating between high and low accident rate organizations on the basis of a range of features that may be broadly ascribed to the culture or climate pool of elements, using safety culture components as a predictor of either large-scale disasters or accident/incident rates remains problematic. The two types of outcomes typically have distinct etiological patterns.

Typically a functionalist perspective views culture as a route to identifying, assessing, controlling and monitoring apparently measurable features of an organization's internal environment. In contrast, an interpretive approach would seek data from multiple sources with a prime aim of advancing learning through increasing understanding. Measuring safety culture is problematic. Ethnographic approaches, for example involving participant observation, action research and grounded theory, while possibly the most valid from an interpretive perspective, are also costly and time consuming. Before determining how safety culture can be measured, there needs to be some agreement as to what it is,

although a grounded theory approach would incorporate this issue within its analytic fabric.

A triangulated methodology is required. One component of this methodology might be some form of audit, which follows from a functionalist risk management perspective. The Human Factors subgroup of the UK Advisory Committee on the Safety of Nuclear Installations (now NUSAC) (HSC 1993) present a framework for measuring safety culture, comprising questions on leadership, management style, communication, policy, planning, risk assessment, risk management, perceptions, stress, training, monitoring and regulation. In 1998 the UK HSE released a tool for measuring safety climate that was partly based on the HSC (1993) report. Cooper (1998) proposes a methodology for changing safety culture that incorporates risk assessments, audits, training, climate surveys and behavior change programs.

Because organizational culture, or some aspect of it such as safety culture, exists at different levels and across several dimensions, complex measures are required to assess it. Questionnaires or similar tools are inadequate for measuring all aspects of organizational culture, including safety culture components. Measurement of safety culture must be based upon an adequate multifaceted model and seek to measure at all levels from individual psychological to executive strategic. Approaches that have been used to assess and analyze organizational culture include soft systems methodology, organizational climate surveys supported by triangulated methods and grid-group analysis. Other methods that have been used include repertory grids, "twenty statements test" and group discussions. In many high reliability organizations, conventional safety audits and accident analyses that are coded for human and organizational factors can be incorporated within a triangulated methodology. A more comprehensive analysis might also require checklists (to capture data on artifacts) and activity analysis (for information on behavior patterns).

From a variety of sources, key dimensions of safety culture for assessment include:
- extent of trust and shared concern for safety among groups within the organization;
- a variety of perceptions and other aspects of culture among subgroups;
- organizational learning, including reflection on practice and feedback systems;
- norms and rules permitting flexibility in dealing with all types of safety issues;
- top management commitment, support and resource allocation;
- soundness of safety policy and applied safety management techniques;
- continuous motivation of all staff;
- safety training provision;
- fostering a problem solving and not a blaming approach to safety issues; and
- persistence of purpose.

At least of equal importance with measurement is the development of theoretical models that adequately express the mechanisms by which safety culture or safety climate might influence individual, group and organizational behaviors that are relevant to safety outcomes. A combination of ethnographic and field experimental approaches could help to identify and

clarify these mechanisms by building adequate theoretical models for empirical testing. This is likely to be a fruitful approach to developing a useful safety culture concept in the future.

8. CONCLUSIONS

Debate on safety culture and safety climate is at an early developmental stage. Within organizational practices, there is some convergence between theoretically distinct functionalist and interpretive approaches. Maintaining a theoretical distinction between the overlapping concepts of safety culture and safety climate is useful, even though in practice they may be treated as partially fused components of a safety management system.

As Pidgeon (1998) and others observe, a major current barrier to improving comprehension of safety culture is an adequate theoretical model that can drive data gathering and analysis to facilitate better understanding. Such a model is likely to include reference to safety culture dimensions, different parties' objectives, un/desired outcomes, processes and procedures, and internal and external contexts, as well as incorporating both functionalist and interpretive perspectives.

Although important, safety culture remains an enigma with several variations. This is due to the complexities of large organizations and the larger overlapping systems within which they reside. A central paradox of safety culture is that by seeking to use it as a rigid functional means to control hazards to health, safety and organizational integrity, more serious outcome scenarios that result from multiple but very rare combinations of events, requiring more flexible organizational responses, may be overlooked.

REFERENCES

COOPER, M.D., 1998, *Improving safety culture: A Practical Guide* (Chichester: Wiley).

COX, S. and FLIN, R., 1998, "Safety Culture" special issue. *Work and Stress*, 12.

GLENDON, A.I. and McKENNA, E.F., 1995, *Human Safety and Risk Management* (London: Chapman & Hall).

HALE, A.R., GLENDON, A.I. and STANTON, N.A., 1999, "Safety Culture and Safety Climate" special issue. *Safety Science*.

HEALTH AND SAFETY COMMISSION, 1993, *Organising for Safety*. Third Report, Human Factors Study Group, Advisory Committee on the Safety of Nuclear Installations (London: HMSO).

INSTITUTION OF OCCUPATIONAL SAFETY AND HEALTH, 1994, *Policy Statement on Safety Culture* (Leicester: IOSH).

INTERNATIONAL ATOMIC ENERGY AGENCY, 1986, *Summary Report on the Post-accident Review Meeting on the Chernobyl Accident.* International Safety Advisory Group, Safety Series 75-INSAG-1 (Vienna: IAEA).

PIDGEON, N., 1998, Safety culture: key theoretical issues. *Work and Stress*, 12, 202–16.

REASON, J., 1997, *Managing the Risks of Organisational Accidents* (Aldershot: Ashgate).

REASON, J., 1998, Achieving a safe culture: theory and practice. *Work and Stress*, 12, 293–306.

WARING, A.E. and GLENDON, A.I., 1998, *Managing Risk: Critical Issues for Survival and Success into the 21st Century* (London: ITBP).

Safety, Ergonomics and Total Quality Management

A. Leino and M.K. Mattila

Occupational Safety Engineering, Tampere University of Technology, PO Box 541, FIN-33101 Tampere, Finland

1. INTRODUCTION

Ergonomics is a product quality criterion from the viewpoint of total quality management (TQM). An effective quality system should be designed to satisfy stated or implied needs and expectations of customers while serving to protect the organization's interests (EN ISO 9004-1: 1994). Today the manufacturing personnel in an industrial context are considered to be internal customers, and the end-users are considered to be external customers. All customers have needs and expectations, including ergonomics and quality, regarding their interaction with products (Eklund 1997).

Management of quality, occupational safety, health and ergonomics have many overlapping elements. On the one hand the origins of quality, safety or ergonomic problems may derive from the same faults, and on the other hand handling of one specific problem, for example a quality problem, may unintentionally cause problems in other areas, such as occupational safety (Zwetsloot 1994).

Modern safety management emphasizes the pro-active approach in planning, organization and measurement. These emphases are identical to those of modern quality management. Therefore, occupational safety and health (OSH) and ergonomic issues can be effectively and economically integrated in the total quality management system (Mattila 1997).

A quality system based on the ISO 9001 or ISO 9004 standards is a valuable management resource in the control of quality. Also a management system for environmental issues, introduced in the ISO 14000, standard is widely implemented. management systems for occupational health and safety have been produced e.g. in the BS 8800 standard.

Integration of these systems has been recommended (Zwetsloot 1994) where companies implement two or more management systems simultaneously. A well-structured integrated system combines diverse management systems with quality management, improves the company's quality of performance and saves time, work and money.

2. ERGONOMICS IN TQM

The TQM system introduces four assumptions as starting points: improved quality is profitable, people want to do a high-quality job, all the parts of the organization are highly interdependent and quality is the ultimate responsibility of top management (Hackman, Wageman 1995). In addition to these assumptions, the basic principles and foundations of TQM are customer focus, process orientation, employee participation, management involvement and continuous improvement (Eklund 1997).

Table 1. Management of safety and ergonomics as part of the quality system in accordance with the standard ISO 9001 (1994)

ISO 9001 Quality standard	Integrating safety and ergonomic issues
4.1 management responsibility	— occupational safety and ergonomic issues are included in company quality strategy and quality policy — occupational safety and ergonomic objectives are included in business activity objectives (for example job satisfaction)
4.1.2 Organization	— occupational safety and ergonomic issues are included in responsibilities and authorities at each level within the organization — occupational safety and job satisfaction are measured regularly — a management representative has been nominated for occupational safety and ergonomic issues — enough resources have been allocated to maintain and continuously improve safety and ergonomics — management review includes occupational safety and ergonomic issues
4.2 Quality system	— safety and ergonomics management issues are included in a company's quality system — occupational safety and ergonomic issues are considered in working instructions and courses of actions
4.3 Contract review	— contract review includes occupational safety and ergonomic issues — product safety requirements are recognized — technical product documents and instructions correspond to customers' requirements — the working and enviromental conditions of products are examined
4.4 Design control	— safety and ergonomic aspects are considered in product design — customers' special ergonomic demands are considered in the product design process — satisfactory technical documents and instructions for use are delivered with the product to make sure that the product is used in a right, ergonomic and safe manner
4.5 Document and data control	— document and data control covers also occupational safety and ergonomics, courses of action and work instructions
4.6 Purchasing	— purchasing and subcontracts share the same product safety requirements as a company's own design and production — production purchasing takes occupational safety and ergonomic requirements into account — subcontractor audits include occupational safety
4.9 Process control	— risk and ergonomic assessment is included in continuous evaluation and follow-up of working conditions — safety and ergonomic aspects are also included in condition assessment and preventive maintenance of machines and devices — occupational safety and ergonomic aspects are taken into account in working instructions and courses of action — employees participate in developing work ergonomics

ISO 9001 Quality standard	Integrating safety and ergonomic issues
4.14 Corrective and preventive action	– sickness leaves and accidents are analysed – job satisfaction is observed – ergonomics and order of work environments are observed – the efficiency and effectiveness of corrective and preventive occupational safety and ergonomics are observed
4.15 Handling, storage,	– insruments for lifting and their ergonomics are included in regular condition assessment – occupational safety and ergonomic aspects are considered especially when stocking and preserving critical products – occupational safety and ergonomic issues are considered during packaging, loading and unloading a product
4.16 Control of quality records	– occupational safety records are included in quality records
4.17 Internal quality audits	– safety and ergonomic issues are included in qualtiy audits – audits are conducted by persons who are trained to assess quality, safety and ergonomic issues or an expert group conducts the audits
4.18 Training	– occupational safety and ergonomic issues are included in training in working methods – special safety and ergonomic training is arranged and a register is kept of these trained employees – enough employees should be trained for first aid and their preparedness is maintained by training – employees' preparedness to act in cases of urgency and failures is developed by regular training and exercises
4.19 Servicing	– occupational safety and ergonomic issues are taken into account when working in customers' premises during commissioning, mounting and service – safe and ergonomic commissioning, mounting and maintenance are taken into account in the technical product documents and instructions

TQM regards the entire business process from supplier to customer. Therefore, it must consider all elements of a company's processes and dimensions of quality, including ergonomics (Luczak *et al.* 1998). It has been proved that ergonomic work conditions support product quality. There is a clear relationship between poor work conditions, such as insufficient lighting, noise and vibration, and effects on human performance, including error frequency and other unwanted deviations.

Integration of OSH and ergonomics management into the TQM system has proved to be efficient and possible in practice. Connections between safety and ergonomics management and quality management are described in Table 1.

3. MODEL FOR IMPLEMENTING INTEGRATED MANAGEMENT SYSTEM

A model for implementing a safety and ergonomics oriented TQM system for small and medium-sized companies has been developed and tested. The model integrates occupational health and safety and environmental issues in the TQM system. The phases of the model are described in Figure 1.

One of the prerequirements is that it is ensured that the top management has a clear vision of the development when the process is starting, before the development work has progressed. This can be achieved by adopting the basic principles and definitions of TQM, ergonomics and safety. The objectives of the programme have to be unambiguous in order to achieve an enduring commitment from the top management.

The model significantly emphasizes the importance of the analysis of the current situation, and the implementation of risk analysis and assessment is crucial. The review consists of the following phases:
- survey of customer satisfaction.
- survey of job satisfaction.
- identification of main processes and responsibilities.
- assessment of occupational safety.
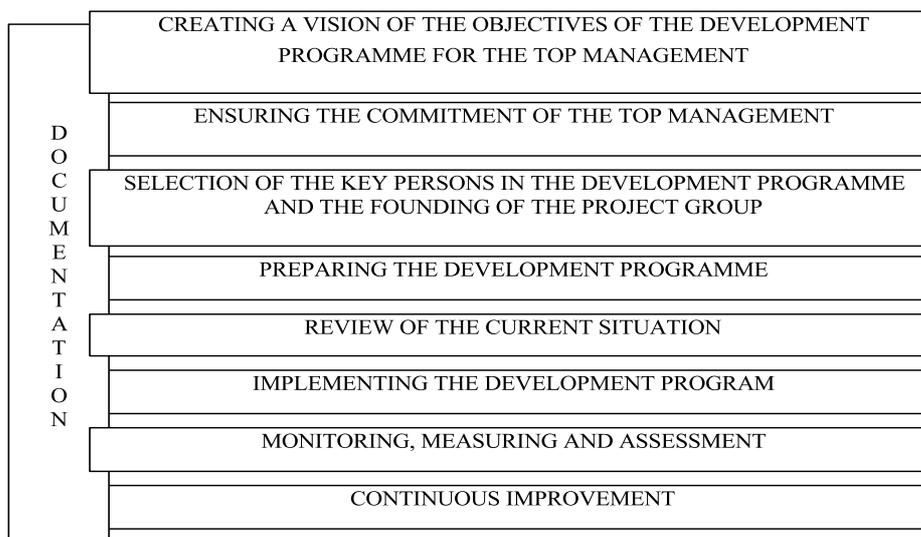
Figure 1. Operations model for implementing occupational safety oriented TQM system

- assessment of work environment, including ergonomics.
- assessment of environmental aspects.

Based on the analysis and risk assessment, the areas for improvements are selected. Responsibilities will be allocated, schedules for implementation documented and the structure of monitoring, measuring and assessing defined. The documentation of the integrated management system is completed. Continuous improvement concerns the whole management system, including ergonomics.

4. RECOMMENDATIONS

1. Total quality management, TQM, regards the entire business process from supplier to customer. Therefore, it must consider all elements of a company's processes and dimensions of quality, including ergonomics.

2. The integration of management systems for quality, environment, occupational safety and health and ergonomics has become actual where companies implement two or more management systems simultaneously.

3. The integration of management systems has proved to be beneficial. Both modern safety and quality management emphasize the pro-active approach in planning, organization and measurement. An integrated system combines diverse management systems with quality management, improves the company's performance and saves time, work and money.

4. System integration is a continuous process which needs to be evaluated at regular intervals. The Malcolm Baldridge National Quality Award is an assessment method that can be used for monitoring and reviewing the integrated TQM system. The assessment criteria could be developed to cover also ergonomic issues.

5. It has been proved that ergonomic work conditions support product quality as well as human performance.

REFERENCES

EKLUND, J., 1997, Ergonomics, quality and continuous improvement – conceptual and empirical relationships in an industrial context. *Ergonomics*, 40, 982–1001.

HACKMAN, J. R., WAGEMAN, R. 1995. Total quality management; empirical, conceptual, and practical issues. *Administrative Science Quarterly*, 40.

INTERNATIONAL STANDARDIZATION ORGANIZATION, 1994, *ISO 9004-1, Quality management and quality assurance standards – Part 1. Guidelines for selection and use* (Geneva: ISO).

LUCZAK, H., KRINGS, K., GRYGLEWSKI, S., STAWOWY, G., 1998, Ergonomics and TQM. In Karwowski and Salvendy (eds), *Ergonomics in Manufacturing* (Society of Manufacturing Engineers).

MATTILA, M., 1997, Integrated and Safety Oriented TQM System for SMEs. *Neue Ansätze für innovative Produktionsprozesse. Kongressbeiträge des 1. Kasseler Kolloquiums vom 17. bis 19. September 1997*. Universität Gesamthochschule Kassel.

ZWETSLOOT, G. 1994. Joint Management of Working Conditions, Environment and Quality. *Search of synergy and organizational learning* (Amsterdam: Dutch Institute for the Working Environment).

Self-managed Work Teams

J. R. Wilson[†] and C. Whittington[‡]

†Institute for Occupational Ergonomics, University of Nottingham, Nottingham, NG7 2RD, UK

‡Holly Bank, Church Lane, Farndon, nr Chester CH3 6QD, UK

1. INTRODUCTION

The notion of teams is implicit and explicit within manufacturing and other industries. There are, for instance, design, development, core or steering teams. Increasingly companies are concerned with teams in a different sense, those that are formed within the organization of production, as an improved way to re-design work. These teams are termed here as self-managed work teams, although the same or very similar concept may be described instead as self-regulating, autonomous, semi-autonomous or self-directing. These distinctions have emerged to help discriminate between teams with varying levels of autonomy and with different types of internal or external leadership and supervision. However, all of the definitions embrace the concept of a work group that has the opportunity for greater independence of decision-making than is traditionally available, although the boundaries of this decision-making, and the degree of mutual cooperation and support that are required, may vary.

2. WHAT ARE SELF-MANAGED TEAMS?

Self-managed teams differ from the more traditional concept of the work group in that the team, rather than the first-line supervisor, controls many of the critical management processes of planning, organizing, prioritizing and staffing. Although the degree of autonomy exercised may vary, the team will normally have responsibility for a complete area of work. Activities within the team are normally interdependent and team members are generally multi-skilled so as to enable interchangeability between different activities. In the view of Rothwell (1993), teams can be classified along scales of formal/informal and permanent/temporary; the self-managed work team (SMT) is formal and permanent.

Teams at work are often not defined in the literature, but where definitions are given there is some agreement. Hackman (1990: 4–5) believes that organizational work groups are those with attributes of being "real" groups (that is, they are intact social systems with boundaries, member interdependence and differentiated member roles), with one or more tasks to perform with collective responsibility, and who operate within an organization or a larger social system. Katzenbach and Smith (1993) define teams as comprising a small number of people possessing complementary skills who are committed to a common purpose, and who are mutually accountable for goals and approach. In terms of manufacturing industry, self-managed teams (or self-regulating work groups) usually have 4–12 members (although as many as 20–30 might be found) with an optimum often specified as about six or eight; personnel specifications will usually include self-organizing or interpersonal skills as well as technical ones, and in the more sophisticated implementations decision-making and problem solving skills also.

SMT vary with respect to two main dimensions: the extent to which managerial authority is delegated (autonomy) and the degree to which roles within the group are interchangeable (multi-skilling). Neither autonomy nor multi-skilling will occur overnight and new teams will need considerable help in making the transition to full self-management. In practice, many teams are likely to move along a continuum, and the role of the supervisor will gradually evolve from controller (stage 1), to team leader (stage 2), then to team coordinator (stage 3) and boundary manager (stage 4), and finally to team resource (stage 5). This progression is mirrored in the development of the team members who move from individuals who are told what to do, to working together as a group, assuming shared responsibility and finally running all the day-to-day operations and being accountable for their own work.

Where a team will sit on this continuum could be determined both by planning (e.g. how much management is willing to give up in terms of control) and also by events (e.g. the speed and enthusiasm with which the team acquires new skills). Therefore, the rate of development of teams, and the extent to which they actually reach the different levels of maturity, will vary considerably.

2. HISTORY AND EARLY DEVELOPMENTS

Interest in the idea of teamworking in manufacturing and other industries, and particularly in what became known as autonomous (later semi-autonomous) work groups, grew out of the work by a number of researchers at the Tavistock Institute in London. Researchers such as Trist, Bamforth and Emery developed a socio-technical systems theory and consequent principles for group work design, from their experiences during World War II and afterwards (e.g. Trist and Bamforth 1951; and see below). The founding group was composed of psychologists, sociologists and anthropologists. The basis behind socio-technical design is the recognition that the social system in any new implementation of work will have as great an influence on its success as the technical and economic systems; by recognizing the interaction of these subsystems and jointly optimizing them, there is an opportunity to improve both work performance and quality of work life for individuals. By extension, a further tenet is that the work organization is a whole open system, influenced by what we would know these days as the complete supply chain.

The first experiments in the UK in group work design took place in the mining industry (the Longwall Mining Study) and afterwards spread into Australia, Canada, Norway, India and Sweden among other countries. The basic principles lying behind these implementations of (semi-)autonomous work groups meant they were:

- Groups of workers with high levels of self-management engaged in cooperative completion of whole identifiable parts of the process.
- Independent of external control for long periods.
- Groups that could formulate goals decide when, where and how to work, distribute tasks internally, decide membership and decide on leadership.
- Given increased responsibilities, often including training, target setting, quality and maintenance.

There was a considerable debate on variations in autonomy at around this time, particularly on the types and degrees of autonomy. For instance, a group could have work method

autonomy (discretion regarding work methods and procedures), work schedule autonomy (the extent to which people can control when they do things) or work criteria autonomy (whether people can choose the criteria to evaluate their own performance). Also the degree of autonomy could vary on a spectrum, from the group examining alternatives and selecting in consultation with management through to the group both developing the alternatives and making a final selection among them on its own.

Some of the most famous cases in group working in the 1960s and 1970s took place in Scandinavia. Along with experiments in a wire drawing and paper mill in the Norwegian Industrial Democracy program, and further experiments in a fertilizer plant, the best-known initiatives were in the Swedish motor industry, and in particular at Saab–Scania (Södertälje) and Volvo (Kalmar and elsewhere). Both these developments involved motor vehicle assembly; the changes involved elements of what we would call today product based manufacture, and also included initiatives to improve the work environment as well. The principal observable outcome was to increase the work cycle time, from ~1 to 30 min (e.g. Valery 1974, Agurén *et al.* 1976).

Although many evaluations of the Saab and Volvo experiments at the time were positive, in terms both of social and technical measures, there have been criticisms ever since. These vary from suggestions at one end of the spectrum that they were one-off developments that could never be repeated because of the scale of effort required, through to suggestions at the other end of the spectrum that really the managers were merely playing with the situation and the changes were not particularly radical. Nonetheless, the developments have had a profound impact on work re-design for a period and have been re-visited more recently as work team design for the 1990s has re-emerged.

3. SOCIO–TECHNICAL SYSTEMS AND THEORY

One important outcome of the work of the 1960s and 1970s has been the socio-technical systems design principles bequeathed to later researchers and investigators. One listing of these appears in Cherns (1987), as follows.

- Principle 1: Compatibility
  The design process must be compatible with its objectives.
- Principle 2: Minimal Critical Specification
  Negatively — "no more should be specified than is absolutely essential." Positively — "state what has to be done but not how to do it."
- Principle 3: Variance Control
  Variances should be controlled as close to their source or origin as possible.
- Principle 4: Boundary Location
  Organization and function boundaries should be drawn with care to allow the sharing of information, knowledge and learning.
- Principle 5: Information Flow
  Information needed to take actions should be provided to those who require it and when they require it.
- Principle 6: Power and Authority
  Groups should exercise the degree of power and authority needed in order to accept responsibility for their performance.
- Principle 7: The Multifunctional Principle

It is beneficial for an organization to have employees with knowledge of a wide range of tasks.
- Principle 8: Support Congruence
  Systems of social support should reinforce required behavior and be consistent with the work design philosophy.
- Principle 9: Transitional Organization
  Planning and design are required for the transition from a traditional to a new organization.
- Principle 10: Incompletion
  Systems design is an interactive and continuous process.

To these 10 principles one must add the basic principles of group and teamwork design, and indeed good job design generally. The objective of organization design should be to provide a high quality of work life for employees and team members as well as providing for the successful performance of the organization. To achieve this the jobs within the teams should:
- be demanding and varied;
- provide the ability to learn on the job;
- provide for an area of decision-making;
- give a degree of social support and recognition;
- relate work life to social life; and
- give a feeling that the job leads to a desirable future.

4. ADVANTAGES AND PROBLEMS FOR SELF-MANAGED TEAMS

During the 1990s, teamwork in manufacturing, as in other businesses, has become the organizational form of choice. This has largely been for business and production reasons rather than being directly aimed at improving the quality of work life. Nonetheless in many industrial and service sectors teamworking is now regarded almost as a panacea and as the only practical way of handling a complex, technical and rapidly changing commercial environment. It is misleading, though, to assume that SMT are applicable to every work situation or that effective teams are easy to establish, and this is reflected in the extent to which potential benefits appear to have (or, in fact, have not!) been actually realized in practice. While many company presentations appear to show significant overall benefits to the organization, the academic literature suggests less impressive results.

Advantages of work organization in teams for companies and individuals are well documented. They help meet many of the principles of socio-technical systems theory, and provide a convenient mechanism for job rotation or enrichment and for increases in job characteristics such as variety, involvement, significance and feedback. They parallel new production methods (e.g. cells) and imperatives (e.g. involvement, responsibility, high competence), and are less vulnerable to the uncertainty possible with individual-based jobs and isolated workstations and processes. Demands at work must be balanced by improved control (over timing, method and boundary), social interaction and responsibility; teamwork can aid all of this and especially can give effective boundary control, timing control, support from colleagues and rational allocation of demands and responsibilities.

There are, though, several problems with a team-based organization of which one must beware. Inter- and intra-group conflict, dislike of groups (just as of participation or increased autonomy generally) shown by significant minorities of workers, harmful effects on supervisory and technical support jobs,

increases in self or group paced stress, and difficulties in establishing equitable and workable performance measurement and payment systems, have all been found in different organizations at different times.

In recent years, the emphasis for manufacturing teams has been on achieving high performance, sometimes to the exclusion of quality of working life considerations. For example, the introduction of teams in many cases is driven by a restructuring, downsizing or process re-engineering exercise. In this context, messages about improving the quality of working life and enhancing motivation and commitment can have rather a rather hollow ring. If badly handled, the resulting tension between employee needs and business needs can mean that the conditions for fully implementing self-managed teams are unlikely to be met. This is reflected in the figures, which are often quoted for the proportion of teamworking initiatives failing to meet expectations. For example, figures quoted in discussions at a recent conference on SMT suggested that two-thirds of programs aimed at the introduction of such teams achieve only moderate success or else fail.

5. DESIGN OF TEAMS

Less prominent in the literature than the generally favorable comments about teams and the "vision" required is detailed advice on *how* to set-up and implement teamworking and *how* to evaluate it. This is a serious hurdle for human-centered systems generally, where various approaches to work re-design, which are not mutually exclusive and are of varying degrees of formality, have been found, including:

- top-down framework of all issues and their interactions, change at a higher level supporting change at a lower one;
- bottom-up, *ad hoc* or sequential change, each decision throwing up the need to make several more;
- skills-based changes, identifying and making explicit the existing and potential shopfloor skills;
- following a job design model at individual or group work level, with the other areas of human-centered systems interfaces, tasks and environment, support systems) dealt with as convenient;
- use of a decision support system to assist engineers, designers and managers at plant level to re-design work;
- contemporary approaches to allocation of function, allowing for under- as well as over-load, in a task-oriented approach; and
- use of a task force, which in change to SMT means that the process and content of the change are fundamentally linked.

Across the literature, conditions and management strategies for changes to teamworking have been defined, propositions laid down for types of team, content for different levels of autonomy specified, suggestions about team leadership, size of teams and payment systems given, and decision aiding tools and decision support systems proposed. Analytical frameworks do exist which may be used in selection and training of team members, for instance Belbin's (1981, 1993) team roles.

However, a detailed program covering all implementation issues likely to be faced by a company introducing group work is hard to find. Published advice is more at the level of desirable concepts (such as shared vision, values and conviction,

management faith, common knowledge base), rather than the direct operational guidance needed for implementing work teams while managing all the other direct and tangential issues involved. One must face the fact that such guidance may not exist because it is impossible to generalize from specific cases, and that content of any change must be set in the light of prevailing circumstances. This would make the mechanisms used to support change to teamwork even more crucial (Wilson and Grey Taylor 1995).

On team evaluation also there is a relatively inadequate literature to give guidance. As well as assessing technical performance, organizational infrastructure (e.g. absenteeism) and work attitude outcomes, we need to evaluate the very group structure and its workings, in terms of how well the team is operating as a group. Measures are beginning to emerge but some of the better initiatives in this area may possibly do well to learn from related work with the armed forces or other such teams (e.g. Brannick *et al.* 1997).

6. CONCLUSIONS

Recommendations for a group or team structure appear so often in ergonomics, occupational psychology, management or health safety literature, that "teams" are in danger of being seen as a panacea for all hitherto intractable problems of work organization. It sometimes seems that "teams" are turned to by companies both as a first (easy) answer and are also clutched at as the last chance when other solutions have not succeeded. They are, in fact, a way of organizing manufacturing activities that require considerable planning, communication, flexibility, insight and luck, but which present a model for work which can satisfy a number of needs of both employers and employees.

REFERENCES

AGURÉN, S., HANSSON, R. and KARLSSON, K.G., 1976, *The Volvo Kalmar Plant – The Impact of Design on Work Organization* (Stockholm: Rationalisation Council).

BELBIN, R.M., 1981, *Management Teams: Why They Succeed or Fail* (Oxford: Butterworth-Heinemann).

BELBIN, R.M., 1993, *Team Roles at Work* (Oxford: Butterworth-Heinemann).

BRANNICK, M.T., SALAS, E. and PRINCE, C., 1997, *Team Performance Assessment and Measurement* (Mahwah: Lawrence Erlbaum).

CHERNS, A., 1987, Principles of sociotechnical design revisited. *Human Relations*, 40, 153–62.

HACKMAN, R.J. (ed.), 1990, *Groups That Work (And Those That Don't): Creating Conditions for Effective Team-Work* (Oxford: Jossey-Bass).

KATZENBACH, J. R. and SMITH, D. K., 1998, *The Wisdom of Teams* (New York: McGraw Hill).

ROTHWELL, S., 1993, Team-building, involvement and empowerment. *Manager Update*, 5, 19–31.

TRIST, E.L. and BAMFORTH, K.W., 1951, Some social and psychological consequences of the Longwall method of coal getting. *Human Relations*, 4, 3–38.

VALERY, N., 1974, Importing the lessons of Swedish workers. *New Scientist*, 62(892), 27–8.

WILSON, J.R. and GREY TAYLOR, S.M., 1995, Simultaneous engineering for self directed work teams implementation: a case study in the electronics industry. *International Journal of Industrial Ergonomics*, 16, 353–65.

Service Quality and Ergonomics

C. G. Drury and A.-C. Chen

Department of Industrial Engineering, University at Buffalo, Buffalo, NY 14260, USA

1. INTRODUCTION

A service industry is one that produces intangible goods for customers, in contrast with manufacturing and agricultural industries whose products are physical objects. The world's nations have developed from largely agricultural economies through manufacturing (following the industrial revolution) to service-based. A developed country will now have service as its largest employment and revenue sector, followed by manufacturing with agriculture a distant third. In addition, our primary industries depend on service companies, for example, to transport our food from producer to consumer, or to raise the capital required to expand manufacturing.

Service industries are becoming global in the same way as are manufacturing industries. An organization seeks to balance the need for local knowledge and direct customer involvement against tactical advantages of different locations around the world. Thus, it is often found that a call to an airline is answered from a rural location, remote from airline activities, where commuting costs for workers are less and possibly so are wage rates for the organization. This globalization is currently limited by the language knowledge needed for direct customer interaction, but it is not limited by communications availability. A service agent can interact as well with data sources from anywhere on the planet, in what appears to the customer to be real-time. There are obvious implications for both organizational and worker well-being in this globalization process.

With these trends in mind, the definitions of some of the unique characteristics of the service sector are given. Examined too is how service quality can be defined and measured, and the ergonomic implications of service quality are presented

2. SERVICE DEFINED

Service can be defined as the production of largely intangible items for the benefit of customers. Items are intangible in the sense that they are only secondarily evidenced by a physical item or commodity. Thus, a visit to a physician may involve paper records, X-rays and medicines but the objective of the visit is to diagnose, give a prognosis and develop a course of treatment for the customer. Service typically represents a change in the status of a customer or their belongings. Thus, a car is washed, a customer is transported to work, funds are moved to or from a customer's bank account or a telephone conversation takes place. Hence, the key commodity in service is information rather than a physical product. Note, however, that a large fraction of manufacturing activities is in fact services, such as buying, information management, personnel, marketing or sales.

Another key to service is that the customer is involved much more directly than is the case for manufacturing or agricultural operations. A customer interacts personally, or via communications media, with an agent (human or machine) of the service company, rather than being at the end of a long chain of operations. Thus, many service employees see customers continually, rather than only occasionally. In the transactions between customer and service employee, the outcome is decided by their joint actions. At the start of a transaction it is quite possible that neither will have a clear idea of either the objective of the transaction or its outcome.

This closeness of customer contact and interaction is a key dimension of service activity. Across the whole spectrum of service organizations it can range from high, for example interacting with an architect for the design of a home extension, to low, as is a customer placing a coin into a vending machine for a beverage. Haywood-Farmer (1988) developed a general classification for service industries based on three dimensions:

1. Degree of customer interaction.
2. Degree of labor intensity.
3. Degree of customization.

Customer interaction is greater in some industries, such as travel agencies, where the customer is an active participant in the service provision. Labor intensity refers to the relative abundance of personnel in some service industries, such as teaching low values imply a more automated service. Customization incorporates the notion of differentiation between the needs of different customers. A fast-food restaurant has few options on its menu compared with a full-service restaurant, allowing customers less freedom to choose according to their different tastes.

With these dimensions go three attributes of the service, each of more or less importance depending on the industry:

1. Physical facilities, procedures.
2. Personal interaction.
3. Professional judgement.

In industries with low customer interaction, low labor intensity and low customization, the physical facilities and highly developed procedural rules are what dominate a customer's judgement of service quality. Personal interaction is important for service transactions with high labor intensity, such as beauty parlors or barbershops. Professional judgement distinguishes services with a high degree of customization and customer interaction, where a highly differentiated outcome is needed.

These classification schemes become useful when the ideas of quality and quality management are applied to service industries. Service quality will now be examined before explaining its impact on ergonomics.

3. SERVICE QUALITY

Quality is fitness for use of either physical or intangible goods. The Quality Movement has been reviewed elsewhere in this encyclopaedia, so service quality will only be concentrated on. Service quality has been defined and measured from three different perspectives: performance observations, customer perceptions and service failures.

Service performance observations are the most straightforward definitions for service quality and, more importantly, they are easily measured. They can be further categorized into either temporal or behavioral measures. Temporal measures, which are widely used in various service settings because of their intrinsic simplicity and measurability, quantify service performance as customer's waiting time (e.g. time waited

for service responses or time spent in processing services) or promptness. On the other hand, behavioral measures are qualitative performance observations. Service personnel's greetings, verbal responses and even non-verbal gestures (e.g. smiling, bowing or handling customer's belongings) are frequently considered as behavioral performance measures. These performance observations have arisen from the direct transfer of the quality experiences from manufacturing settings to service settings. The applicability of statistical process control (SPC) techniques is emphasized. Most importantly, service performance measures are meant to be objective, explicit and able to be monitored on-line and internally. However, the meaningfulness of these measures to service quality improvement or even to service quality is in some doubt.

The second type of service quality measure comes from the customer's perspective, i.e. the quality of service is determined directly by the customer. The well-known Gap model and SERVQUAL (Parasuraman *et al.* 1988) defined service quality as the gap between customer's expectations and customer's perceptions of service. Although this service quality definition has been debated, the emphasis on a user perspective with customers as the final judge of service quality is well accepted. However, the major shortcoming of this definition is that the information from the customers is external and may be difficult to relate directly to service quality planning.

The third approach for defining and measuring service quality focuses on error-free service. That is, service quality is assured as a reduction in service defects or failures and/or the presence of recovery modes for service failures. Service recovery strategies such as unconditional service guarantee are promoted as the corrective approach to ensure service quality. Critical incident techniques (CIT) are widely utilized to assess and collect the information regarding service failures. Fail-safe design (Chase and Stewart 1993) has also been promoted as a preventive approach towards error-free service. One major advantage of using these failure-based approaches is that service quality is described in terms of the tasks and the context of the interactions between service personnel and customers. Therefore, the information collected is more activity- or context-oriented and thus more meaningful to quality assurance or quality improvement. Caution should of course be exercised because of reporter- or context-biases in interpreting service failure data.

4. THE HUMAN ROLE IN SERVICE QUALITY

If quality is seen as the control of errors, then examples of human factors in service quality are relatively plentiful. In the medical profession there have been several studies of error, how errors arise, how they propagate through the system and how they can be controlled. Bogner (1994) provides a number of such examples and more recent work has applied human error analysis techniques to the blood supply. Similarly in transportation, error studies go back to the dawn of human factors in the Second World War, and forward to studies of automation-induced pilot error in modern airlines. Recently, errors in aviation maintenance, another service industry, have had similar attention.

These error studies, although certainly in the service sector, do not address some of the characteristics more typical of the 'higher' ends of the service taxonomy, where human professional judgement during direct customer interaction plays an important

role. In one sense they are closer to our usual examples from manufacturing. Within the service quality community, human factors/ergonomics has often been confined to the measurement of customer attitudes, opinion and satisfaction. There is clearly a role for ergonomics in a set of tasks with these characteristics:

There is direct interaction between the operator and the ultimate customer in many service applications, such as retail or banking. Most interactions between customers and operators are less routine than those between physical processes and operators. Each operator/customer interaction is in a sense unique.

The error measures available, or even possible, may not be as amenable to SPC techniques. For example, it is possible to plot control charts of the number of errors per hundred or thousand or million interactions, but can these errors really be added together? What action does the SPC chart indicate when a significant increase in error rate is discovered? In one sense, if there are enough errors to count, the service system has already failed its customers.

Service may have unique characteristics, but many human factors techniques should be directly applicable. An example is the model of customer/operator interaction in service quality proposed by Chen and Drury (1997). They used examples of service providers who must handle a wide variety of customer requests, for example bank officers or travel agents. Both of these operations are undergoing rapid computer conversion, so that there are often two parallel customer/operator models: with and without a human agent. In both cases, the ultimate aim of the customer is to access data (available through the computer, print media or via networks of contacts) to solve a unique problem, such as obtaining a loan, making an investment or booking a vacation. In both of these the customer's objective is to obtain the service required, but which the customer may be currently unable to articulate. It is the task of the agent in Figure 1 or the computer system in Figure 2 to make the customer's needs explicit enough to access the correct data base and to search it for the solution that best matches the customer's needs.

As an example, a bank customer may wish to invest $1000 for 3 months, but much interaction would be required to determine whether the customer's investment strategy should be characterized as aggressive (e.g. maximum return on investment) or risk-averse (e.g. minimum probability of monetary loss). The customer may not even use such terms. To complicate matters further, the customer may be willing to invest a somewhat

Figure 1. Customer service direct.
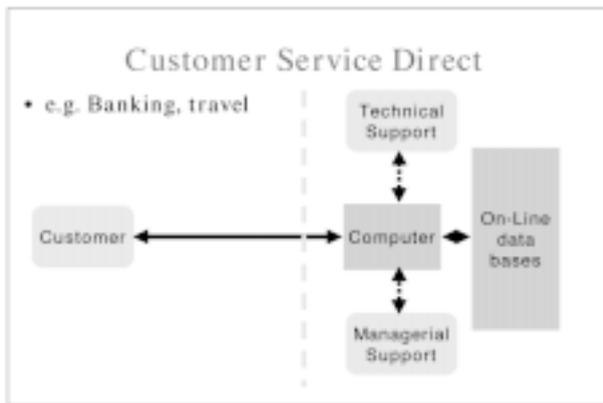
Figure 2. Customer service via agent.

different amount (e.g. if a particularly attractive investment had a $1200 threshold) or to change the time scale slightly.

With a human agent (Figure 1) this explanation of alternatives may be an interesting and feasible, if time-consuming, exercise. But where the allocation of function has been changed to an entirely computer-mediated transaction (Figure 2) many differences arise. For example, a computer program may be able to ask pertinent questions but would be unlikely to be useful to a customer who had little concept of elasticity and risk. Thus, changing the function allocation would likely change the type of customer who could be served. This particular service could end up with quite different customers choosing the alternative system designs represented by Figures 1 and 2, with the less informed and less computer literate customers migrating to a service that may have a higher price.

The implications of using models such as Figures 1 and 2 to study service quality is that they remain focussed on the central transaction between customer and system. Thus, human factors/ ergonomics techniques have to be used to understand the essence of the service. In this way it can be understood *why* errors arise rather than just counting and charting them. This allows one to move beyond statistical process control techniques and to develop a true application of ergonomics to service quality.

REFERENCES

BOGNER, J., 1994, *Human Error in Medicine* (Hillsdale: Lawrence Erlbaum Associates).

CHASE, R. and STEWART, D., 1993, Fail-safeing services. In Scheuing and Christopher (Eds) *The Service Quality Handbook* (San Francisco: AMACOM), pp. 347–57.

CHEN, A.-C. and DRURY, C.G., 1997, Human errors in customer service: a research framework. In *Proceedings of the 13th Triennial Congress of the International Ergonomics Association '97*, Tampere, Finland, 3, 141–3.

HAYWOOD-FARMER, J., 1998, A conceptual model of service quality. *IJOPM*, 8, 19–29.

PARAMSURAMAN, A., ZEITHAML, V. A. and BERRY, L. L., 1988, SERVQUAL: a multiple-item scale for measuring consumer perceptions of service quality. *Journal of Retailing*, 64, 12–37.

Shift Work

K. Kogi

The Institute for Science of Labour, 2-8-14, Sugao, Miyamae, Kawasaki, Kanagawa 213, Japan

1. REASONS FOR SHIFT WORK

Shift work is done by two or more groups of workers who alternate their work shifts assigned at different times of a given day. A shift refers to hours of the day in which a worker or a group of workers is scheduled to be in the workplace. This enables the undertaking to operate longer than the usual period of a work day, extending the business to night or early morning hours. The total period of time covered by the shifts can be either continuous for 24 hours, or discontinuous for less than 24 hours a day.

In the modern industrial society, shift work spreads to various sections of industry. While the emphasis may depend on different economic structures, the reasons for employing shift work are similar in different countries:

(a) Shift work is necessitated by technological requirements for continuous plant operation, as in iron and steel mills, metallurgical factories, chemical and other continuous processes.

(b) Shift work is used in many manufacturing, mining, and construction undertakings for economic reasons in order to profitably use costly or automated equipment. This profitability results from the rapid depreciation of expensive equipment, increased competitiveness due to reduced overall costs per unit of production, and increased flexibility in meeting changes in demand.

(c) Shift work is widely used for public services attending people during night and early morning hours. Often, 24-hour operation is needed as in communication services, public transport, power generation, health and welfare services, hotels and security. Though not necessarily for continuous 24 hours, radio and television stations, car service stations, entertainment centers and restaurants also require shift work. Shift work is increasingly used in computerized services and local authorities.

(d) Natural environmental factors or the interaction with natural resources may require shift work as in the case of farming work, livestock raising, plantations, fishing, food processing, and others.

(e) Social factors may be mentioned where business or service operations are extended for social reasons, such as reduced normal hours of work resulting in the adoption of shift work.

Thus the spread of shift work largely depends on the industrialization and urbanization of the national economy. As industrialization proceeds, about 10–20% or even more of the workforce may be engaged in different forms of shift work. A series of international labor standards are relevant to shift work. The Night Work Convention (No. 171) adopted in 1990 by the International Labour Organization (ILO) has broadened the scope and applies to both genders and nearly all occupations, stipulating the provision of safety and health services, maternity protection, resting periods, social services, consultation on work schedules, and others. The European directive 93/104/EC also confirmed the need for regulating measures of shift work for improving safety and health protection.

2. SHIFT WORK SCHEDULES

Shift work schedules may be organized in a large variety of ways. A shift system may include day work plus one or more shifts worked outside these normal day-work hours. Thus the number of shifts per day may be two, three or more. These shifts may be shorter or longer than, or of the same duration as normal day-work hours, sometimes overlapping with them. Major problems of shift systems arise from the extension of business hours that result in a phase-displacement of the sleep period and therefore substantial changes in the daily life of shift workers. These changes create many problems compared with a minor extension of day-duty hours toward the evening that may not necessarily hamper the sleep period.

Most shift systems can be classified according to the four basic features:

(a) whether the crews of different shifts rotate or work permanently on a particular shift;

(b) whether any shift extends into hours that would normally be spent asleep (i.e. night hours);

(c) whether the shifts cover the 24 hours of the day; and

(d) whether the shifts are worked throughout the week or with a free weekend.

As shown in table 1, shift systems are first categorized according to whether they rotate (rotating systems) or not (e.g. permanent night work), and then into two groups, those involving night work and those without night work. The latter is represented by "double dayshift" systems worked by morning and afternoon shifts and normally without weekend work. Those night-working systems are either manned for 24 hours a day (24-hour systems) or for less than 24 hours a day (discontinuous systems). A typical discontinuous system involving night work is one in which day and night shifts are alternated with operations suspended for a few hours between them and during the weekend. Twenty-four-hour systems can be further divided into continuous systems, worked 7 days a week, and semi-continuous systems that do not involve weekend work.

The number of shifts per day may be two or three in typical shift systems. Today, most discontinuous systems have two shifts, either morning and afternoon shifts (as in the case of "double day-shift" systems) or day and night shifts. Twenty-four-hour systems have two, three, or more shifts, the length of the shifts varying much between different systems. Two 12-hour shifts are frequently seen, whereas the two shifts may differ in length, for example, comprising an 8-hour shift and a 16-hour shift. Typical three-shift systems have three 8-hour shifts, but the actual shift length may vary, sometimes overlapping with each other for the reason of work duty changes. There can be four-shift systems consisting of 6-hour shifts or split-shift systems. Permanent shift systems are generally less frequent, although countries such as the United States engage many permanent night workers in addition to rotating shift systems.

There are a large number of irregular shift systems in which shifts of various starting and finishing times are worked by workers to cover the required production or services demands. These irregular systems are common for transportation, telecommunication, and services. Further, it should be noted that

Table 1. Classification of different shift schedules

Systems	(a) Rotated?	(b) Night-involving	(c) 24-hour business	(d) Weekend work	Number of straight shifts
Permanent shifts	no	mostly yes	yes or no	yes or no	various
Double-day shifts	usually yes	no	no	usually no	usually 5
Discontinuous systems involving night work	usually yes	yes	no	usually no	up to 5–6
Semi-continuous systems	usually yes	yes	yes	no	up to 5–6
Continuous systems	usually yes	yes	yes	yes	1–5
Irregular systems	mostly yes	yes or no	yes or no	yes or no	various

mixed or hybrid systems are not uncommon. An increasing tendency of shift work is to have systems comprising different shift schedules; for example, continuous three-shift schedules combined with weekend 12-hour shifts.

Increasing flexibility in organizing shift systems is an emerging feature for many shift systems. This relates to the need for extending business hours despite the general trend for shorter hours of work per worker as well as to the need for accommodating workers' preferences for flexible working time. Flexibility is achieved by varying shift hours or by allowing workers to change their own shift-rotating schedules depending on their preferences as well as by combining different shift schedules. There are many innovative patterns for increasing flexibility. Flexitime allowing individual workers a choice of the day's working hours can also be introduced to shift systems if workers can change their starting and finishing times by making arrangements among those working at the same worksite. Compressed work weeks (e.g. three 12-hour shifts per week) or combinations of different shift systems may be used for increasing flexible manning of various shifts. Part-time work and job-sharing schemes may also be used. Numerous variations of flexible shift schedules are increasingly applied, including reduction in the frequency of unpopular night shifts, more time off in weekends, additional working time reductions or additional holidays, opportunities to exchange shifts as well as preferential treatment of time off or vacations for particular workers. Advantages and problems in adopting flexible work schedules for the enterprise and for the worker are listed in table 2.

Shift schedules with increased flexibility are advantageous for more competitive business operation and for responding to workers' preferences. Economic and manning advantages for the enterprise are significant. This contrasts with often remarkable disadvantages for workers who may have to work longer shifts or suffer from irregular schedules and sometimes increasing strains in social and family life. The need to reorganize job content as a result of increasing flexibility is also obvious. This is because it is often necessary to redistribute job duties among shift workers when more flexible work schedules are introduced, such as reorganizing night work tasks, improving communication, or applying multi-skilled jobs. Thus it is important to redesign work tasks themselves in realizing flexible shift systems.

3. ASSESSMENT OF SHIFT WORKERS' PROBLEMS

The problems of shift workers relate to both the phase-displacement of their work–sleep periods and adverse negative working conditions that may be combined with shift work. The effects of the phase-displacement lead to disruptions in physiological "circadian" rhythms. Human biological rhythms are adapted to daily work–sleep schedules and cannot be easily inverted even if we change the phase of these work–sleep schedules. These internal rhythms are called "circadian rhythms" as the rhythms have an approximate cycle period of 24 hours and manifest day-and-night changes in functional levels of various parameters reflecting the "internal clock". Circadian rhythms are found in many physiological functions, such as body temperature, blood pressure, heart rate, blood concentrations of proteins, electrolytes, hormones, etc., as well as cerebral functions. Some of the internal rhythms can be phase-displaced relatively easily in accordance with the change in work–sleep periods, but many others cannot be so easily phase-displaced. It is well known that, in the case of "jet lag" after travelling by jet planes to a place with a time difference of several hours or more, people suffer from disturbed sleep, feeling unwell, and reduced work capacity. Such jet-lag symptoms disappear within a few days as the internal circadian rhythms are adapted to the new time zone. In the case of shift workers, however, similar "shift-lag" symptoms persist as long as the night-work–day-sleep schedule continues.

As a result of this incomplete reversal of circadian rhythms, the shift workers may be subjected to problems in work, sleep, safety, and health, in addition to other effects associated with shift work and with disrupted social and family life. These problems are best shown by the stress and strain relationships of shift workers as shown in table 3.

Shift work, in particular night work, can have negative

Table 2. Advantages and problems in adopting flexible shift schedules

	Advantages	Problems
For the enterprise	Longer business hours	Higher administrative costs
	Adjusting to fluctuating demands	More difficult supervision
	Improved labour productivity	More welfare services
	Improved recruitment	Increase in training costs
For the worker	Shorter working hours	Fatigue by longer shifts
	More time for personal needs	Strains in family life
	Increase in voluntary options	Separate social life
	More responsible jobs	Temptation to a second job

Table 3. Stress and strain relationship of shiftworkers

Stress --->	Intervening variables --->	Strain --->	Potential effects
- Phase-displacement of work/sleep hours	- individual characteristics - job-related factors	Sleep changes Fatigue	- Inefficiency - Unsafe work
- Unfavourable working, organizational, and social conditions	- environmental conditions - domestic circumstances	Less capacity Complaints	- Impaired health - Reduced well-being

impacts on health and well-being of workers. Deterioration of health can be manifested in disturbances of sleeping and eating habits, and, in the long run, in more severe health disorders. Since reaction time, computation and problem-solving performance are significantly affected, declining during night hours often with sleepiness, one serious consequence of fatigue and sleep deficit by shift work is the increased risk of accidents. Chronic health disorders prevalently deal with gastrointestinal (colitis, gastroduodenitis, peptic ulcer), neuropsychic (chronic fatigue, anxiety, depression) and cardiovascular (hypertension, ischemic heart diseases, cerebrovascular disorders) functions. In addition, reports show that shift and night work may have more specific adverse effects on women's health both in relation to their hormonal and reproductive functions and their family roles. Many shift workers leave shift work after a short time because of serious disturbances, while those remaining in shift work show different levels of adaptation and tolerance. The effects of such stress conditions can vary widely among shift workers in relation to many intervening variables, as well as working life situations.

4. DESIGN OF SHIFT SYSTEMS

The design of shift systems is a complex issue, often requiring special skills. We should note that there are no ideal systems simply computed using given criteria. This is because a variety of organizational and social aspects must be taken into account. Not only the nature of work and manning demands, but also safety and health effects, social needs, workers' preferences, absences and fluctuating business must be considered. Usually, a basic shift pattern is negotiated between managers and workers to agree on modified patterns meeting the local needs. This cautious approach is recommendable.

The design process greatly differs between different categories of shift systems shown in table 1. The shift patterns agreed on may be relatively straightforward in the case of permanent night and evening shifts or weekly rotated systems without weekend work. In the case of continuous or irregular systems, however, the design process is complex and varies considerably on account of local customs and differences in preceding systems. Normally, options for shift design relate to a number of factors: permanent or rotating shifts, direction and rate of rotation, length of work shifts, number of consecutive shifts, number of consecutive days off, shift changing times as well as special provisions and rules for work shift use. One prominent problem arises from the number of consecutive night shifts and that of days off, both varying greatly according to industries and local customs. Longer periods of consecutive night shifts have been common in some countries such as the United States, while rapidly rotating systems with 3–4, or less, consecutive work days in the same shifts are increasingly popular in many other countries. But even in the latter systems, actual shift patterns differ greatly. Some useful suggestions for shift pattern design are given in table 4.

The night shift is the most disturbing because of disruptions in circadian rhythms, sleep difficulties, and well-being problems. Thus, minimizing the use of night shifts should be the primary concern. Reduction in overall working hours and avoidance of excessive overtime work are necessary for shift workers considering their special burdens. Wherever possible, night work should be avoided or minimized. Except for established local lifestyles favoring permanent night workers, the use of permanent night shifts is not recommendable for the majority of shift workers. For rotating shift patterns, the suggestions in the table are generally advisable.

Because there is no "optimal shift systems", managers and workers should find a reasonable compromise between business demands and workers' needs. Many guidelines and manuals are available. There are many innovative shift patterns and the advantages of these new patterns should be referred to. Recent examples include rapidly rotating (two or three consecutive night shifts) five-team three-shift systems, systems with free weekends, systems with

Table 4. Useful suggestions for designing better shift systems

(a) Nature of work	The nature of work and workload suitable for extended hours.
(b) Use of night shifts	The use of night shifts minimized as much as possible.
(c) Overtime	Overtime not added with arrangements for cover of absentees.
(d) Consecutive night shifts	Only 1–3 night shifts in succession wherever possible.
(e) Between-shift intervals	Short intervals between two shifts (e.g. 7,10 hours) avoided.
(f) Free weekends	Some free weekends with at least two consecutive days off.
(g) Length of shift	Depending on physical/mental load, shorter for night shifts.
(h) Shift rotation	Forward rotation (morning to afternoon to night) preferred.
(i) Shift change times	Allowing individuals some flexibility.
(j) Good sleep periods	Good bedroom quality, securing longer sleeps and naps.
(k) Eating/physical fitness	Eating at regular times, keeping fitness with medical checks.
(l) Social and family life	Sharing household work, planning time off positively.

Table 5. Some examples of rotating shift schedules

System	Cycle period	Notation
Discontinuous system		
2-team 2-shift	14 days	11111HH22222HH
2-team 2-shift	14 days	11111HH33333HH
Semi-continuous system		
3-team 3-shift	21 days	11111HH22222HH33333HH
4-team 3-shift	28 days	112233HHHH1122H33HH11H2233HHH
7-team 3-shift	49 days	111111H33H22HH333HHHH22333HH2 2233HHH11111H1H222HH
Continuous system		
4-team 3-shift	8 days	112233HH
	28 days	1122333HH1122233HH1112233HHH (2-2-3 system)
	28 days	112222H333HH1122H3333HH111HH
5-team 3-shift	10 days	112233HHHH
	15 days	111222333HHHHHH
	35 days	1122333HH1122233HH1112233HHHH HHHHHH
	35 days	11111hh3333HHH222HH1DHHHH33N HHH222H
	or	1133HHH222H11D33HHHHHHHH1133N HHH222H
6-team 3-shift	42 days	112233HddddddHH33HHH1D2233HHH Hddd22NHH111HH
4-team 12-hour -shift	4 days	DNHH
	8 days	DDHHNNHH
	56 days	HNNHHDDDHHNNHHDHHNNHHHDD HHNN HDDHHNNNHHDDHHNHHDDHHHNN HHDD

NOTE: Sequence of shifts for one of the shift crews indicated by 1: morning shift 2: afternoon shift, 3: night shift, D: 12 hour day shift, N: 12-hour night shift, d: 8-hour training, H: day-off.

reduced staffing of night shifts as well as continuous systems involving day-work periods taken in turn by different shift teams.

Rapidly rotating schedules have advantages compared with slow or weekly rotation. Faster rotation tends to disrupt the circadian rhythms in a lesser degree because the original day-work–night-sleep style is quickly resumed before fatigue and sleep deficiencies accumulate. Field studies show that the majority of shift workers favor quicker rotation after a trial period. Rapid rotation seems to facilitate more opportunities of participation in family life and more regular social contacts. To confirm the long-term effects of rapid rotation, detailed studies may be necessary. Some basic examples of weekly and rapidly rotating schedules are indicated in table 5.

Other features of shift systems including the length of shifts, between-shift intervals and shift-changing times also influence both physiological adjustment and social life of shift workers. Forward rotation is favored as it can avoid short intervals and easier to adapt. While avoidance of lengthy shifts and too short between-shift intervals is generally favored, they occur in many shift schedules. For example, 12-hour shifts are frequently adopted as they lead to less commuting frequencies and more days off. It is argued that 12-hour shifts, as in the case of four-team two-shift systems in continuous operations, may be applicable if the work is not physically or mentally strenuous. As one of disadvantages of 12-hour shifts is increased fatigue, single night shifts or a maximum of two night shifts seem recommendable. Another argument for 12-hour shifts is that these

Table 6. Examples of support for shift workers

(a) Better work schedules	Consultation about shift schedules, increasing flexibility
(b) Keeping capacity at work	Ergonomic work design, fail-safe operations
(c) Securing restful conditions	Flexible schedules, resting facilities, housing
(d) Support for family care	Child care facilities, child care leave, social services
(e) Facilitating coping measures	Group work, training on coping
(f) Support for health	Health counselling, health check, access to daywork
(g) Support for career planning	Retraining schemes, educational leave, early retirement

shifts can increase flexibility in shift patterns. Examples include combination of 12-hour shifts with three-shift schedules or the use of part-timers for week-end 12-hour shifts. It is necessary to examine the use of flexible working time arrangements in assessing shift schedules.

5. MANAGING SHIFT WORK AND SUPPORT FOR SHIFT WORKERS

Improving shift patterns and providing various support measures for shift workers are both important in managing shift work. Examples of support measures for shift workers are listed in table 6.

Safety and health measures must be given a priority, including safe-work methods, good teamwork, rest periods, better work environment, health checks, counseling, and providing access to day work for workers suffering from ill-health due to shift work. Support for good housing conditions and for family care is important. Maternity protection and support for women shift workers having family and other burdens are very important. Reduction of sleep problems, food intake, physical fitness, and support for taking various coping measures are also important. Further, individual day-work–night-sleep patterns are important in supporting their adjustment to shift schedules, including "morning" and "evening" activity patterns. In general, support for sleep and coping strategies of individual workers in their different daily life situations is essential. While many compensation schemes exist for shift work, comprehensive approaches should be taken to meet varying needs of shift workers.

REFERENCES

ÅKERSTEDT, T., 1996, *Wide Awake at Odd Hours: Shift Work, Time Zones and Burning the Midnight Oil* (Stockholm: Swedish Council for Work Life Research).

CORLETT, E.N., QUEINNEC, Y. and PAOLI, P., 1988, *Adapting Shift work Arrangements* (Dublin: European Foundation for the Improvement of Living and Working Conditions).

COSTA, G., 1996, The impact of shift and night work on health. *Applied Ergonomics*, 27, 9–16.

FOLKARD, S. and MONK, T.H., 1985, *Hours of Work: Temporal Factors in Work-Scheduling* (Chichester: Wiley).

HÄRMÄ, M., 1993, Individual differences in tolerance to shift work: A review. *Ergonomics*, 36, 101–9.

INTERNATIONAL LABOUR OFFICE, 1990, The hours we work: new work schedules in policy and practice. *Conditions of Work Digest*, vol. 9, no. 2 (Geneva: ILO).

KNAUTH, P., 1993, The design of shift systems. *Ergonomics*, 36, 15–28.

KNAUTH, P., 1997, Hours of work. In *Encyclopaedia of Occupational Health and Safety* (Geneva: International Labour Office), pp. 43.2–43.15.

KNUTSSON, A., 1989, Shift work and coronary heart disease. *Scandinavian Journal of Social Medicine*, 544, 1–36.

KOGI, K.,1995, Increasing flexibility in shift work arrangements. *Work Stress*, 9, 211–18.

KOGI, K., and THURMAN J.E., 1993, Trends in approaches to night and shiftwork and new international standards. *Ergonomics*, 36, 3–13.

MONK, T., and FOLKARD, S., 1992, *Making Shift Work Tolerable* (London: Taylor & Francis).

RUTENFRQANZ, J., 1982, Occupational health measures for night and shiftworkers. *Journal of Human Ergology*, 1, Supplement, 67–86.

WEDDERBURN, A., 1991, Guidelines for shiftworkers. *Bulletin of European Shiftwork Topics*, 3, 3–56.

WEDDERBURN, A., 1998, Continuous shift systems. *Bulletin of European Studies on Time*, 11, 3–46.

Shiftwork and Sleep

Professor D. I. Tepas

Industrial Psychology, Connecticut Transportation Institute, University of Connecticut, 179 Middle Turnpike, Box U-202 Storrs, CT 06269-5202, USA

1. INTRODUCTION

With regard to hours of work, the term *shift* is defined as the hours of the day that a worker is required to be in the workplace. Given this definition, most workers are shiftworkers. Work schedules take many forms. A worker may have *permanent* hours, with their shift at the same time of day every workday. Many workers have *rotating* hours, with the time of day they work changing in a scheduled way. Some workers are employed on *irregular* hours, with shift starting times and/or durations varying in an erratic or unpredictable way.

The work schedule categories for individual workers, noted in the previous paragraph, may or may not be combined to form the *work schedule system* for a group of individuals operating within an organization. These operations may be *continuous*, employing workers around the clock, seven days a week. In many cases these are *discontinuous*, not requiring around the clock work and/or weekend operations. Emergency and/or military activities often require *sustained* operations, requiring workers to perform as long as they can. Given these variations in operations and the forms work shifts hours take, an immense variety of work schedule systems are possible.

Work schedule systems requiring night work, unpredictable work hours, long or short workdays, and weekend duty are often perceived as being non-traditional, unusual, infrequent, undesirable, uncommon, and unneeded work. Although some of these adjectives accurately describe many existing work schedule systems, these generalizations are not warranted and many are associated with service requirements. For example, the requirement for night work in the healthcare industry precedes the factory system and the industrial revolution. The humanitarian value of emergency search-and-rescue operations is widely accepted. Weekend work to monitor nuclear operations is mandatory. In fact, it would appear that the majority of US workers are now employed on what some would term non-standard work schedules (Presser 1995). In the contemporary world, the question is not if these alternative work schedule systems are to be used, but rather how and where these systems should be deployed.

2. SLEEP

At first glance, the importance of sleep to the ergonomic study of shiftwork is not obvious. Some of the factors contributing to this impression: many work systems do not require night work; nearly everyone has experienced acute but short-term exposure to sleep deprivation without having a catastrophic problem; for many people, sleep appears to be a behavior they can control within what they perceive of as reasonable limits; and, many human factors people assume that time-of-day effects are the same as time-on-task effects. Research in the last 20 years has demonstrated that: sleep deprivation and/or time-of-day factors can lead to catastrophic events; there are few (if any) successful countermeasures to avoid falling asleep when one is sleepy; and, time-of-day and time-on-task effects interact and sometimes mask

each other. Numerous studies, by many investigators using a variety of methods, have reported that sleep complaints are elevated in shiftworkers. But what does this mean?

People are diurnal beings. since they tend to be active during the day and they do have a physiology driven by an internal biological clock. These variations in physiology and behavior are said to be *circadian*, given their cycle time of about 24 h. Research has shown that it is generally difficult to make quick and large changes in the timing of biological clock circadian variation. Data from jet-lag observations, sustained operations, and individuals in social isolation tend to confirmed this finding. These studies demonstrate that over a short-term period attempts to be active when the internal biological clock says sleep result in degradation of performance, changes in mood, and disturbed sleep. Under these acute conditions, performance is often driven by biological factors.

The impact of long-term chronic exposure may be quite different. Shift workers subject to chronic exposure to night or rotating shiftwork have biological, social and cultural problems (Tepas, Paley and Popkin 1997). With regard to sleep, these problems are manifest as a *reduction* in sleep length (Tepas and Carvalhais 1990), rather than as a sleep disturbance (Mahan, Carvalhais and Queen 1990). Equally important is the conclusion that social factors may outweigh biological ones (Monk 1989). Thus, evaluation of the chronic impact of a bad work schedule must consider not only the impact of sleep deprivation but also the workers perceptions and preferences.

A study of experienced shiftworkers demonstrates the role of chronic sleep deprivation and worker perceptions (Tepas and Mahan 1989). This was a study of 324 hourly workers on permanent discontinuous shifts. These workers were selected from a larger pool of workers to produced three matched groups, with 108 workers in each group. The three groups were matched on the basis of age, shift tenure, gender, and industry. The three groups were: D/D, workers who were employed on the day shift and most preferred to the day shift; N/D, workers who were employed on the night shift and most preferred the day shift; and N/N, workers who were employed on the night shift and most preferred the night shift.

Figure 1 shows the mean sleep length for workers in each of these three matched groups on workdays and non-workdays. For all three groups, non-workday sleep length was significantly longer than workday sleep length, but the differences between the three groups were not significant on non-workdays. On workdays, the D/D group slept significantly longer that the other two groups. These data demonstrate a chronic reduction in the quantity of sleep on workdays, with night shiftworkers showing a significantly greater impact. There was no evidence in this data base to support the notion that workers who like night work sleep longer or fully make-up for lost sleep on their days-off.

Shift workers often complain about the quality of their sleep, even though their polysomnographic sleep recordings suggest a reduction in sleep length is the major factor, rather than a disturbance of sleep (Tepas, Walsh, Moss and Armstrong 1981). These reductions in sleep length are most often associated with night shiftwork and they do *not* disappear with continued exposure to night shiftwork. It is important to note that the reduction in sleep length associated with night shiftwork is also present in workers who prefer night work and do not report many complaints about the quality of their sleep. Figure 2 shows
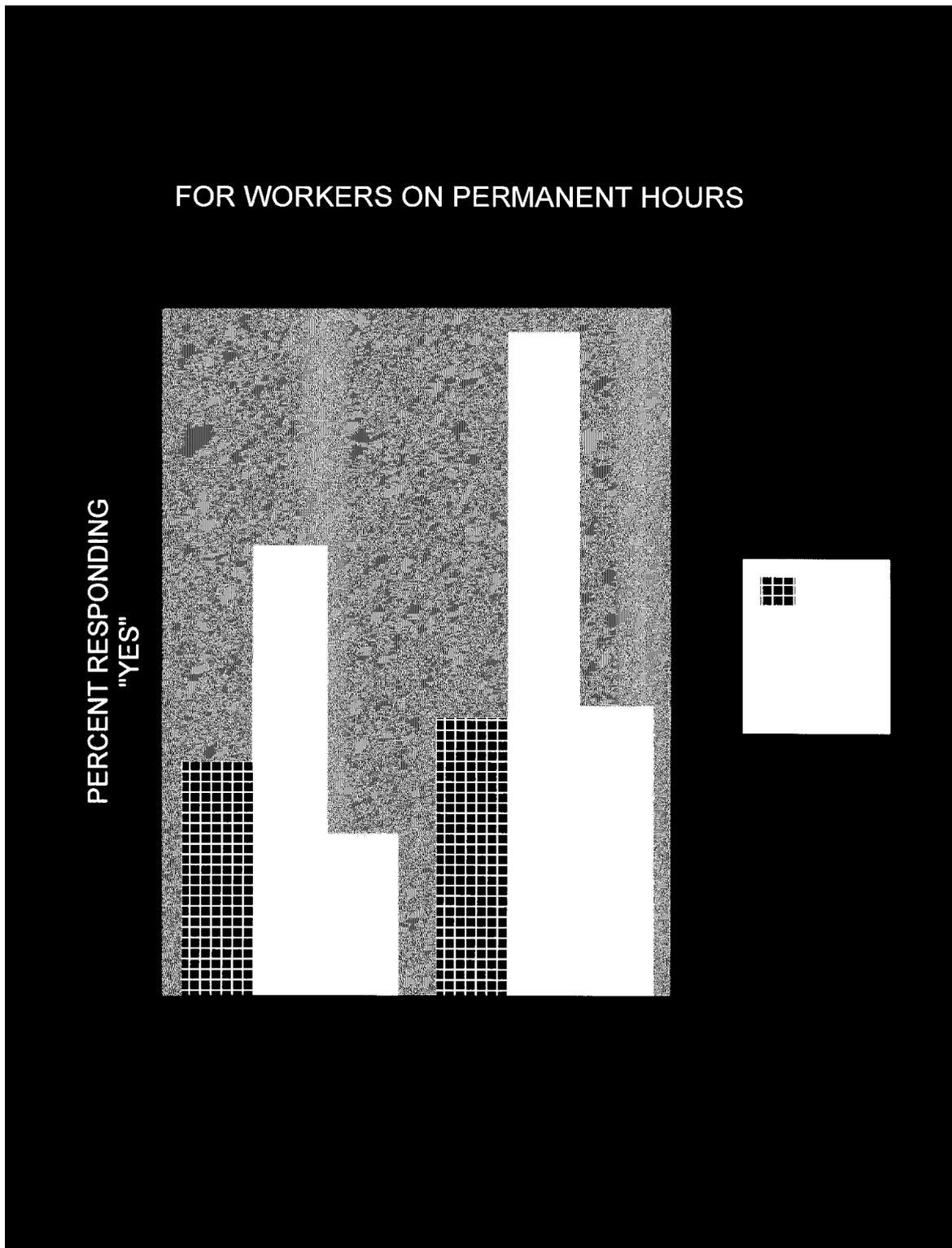
FIGURE 1. Mean reported workday and non-workday sleep length in minutes for workers on permanent hours.
[*Note*: **Data from three matched groups, with 108 workers in each group: D/D, workers who were employed on the day shift and most preferred to work the day shift; N/D, workers who were employed on the night shift and most preferred to work the day shift; and, N/N, workers who were employed on the night shift and most preferred to work the night shift. Graphs plotted from data reported by Tepas and Mahan (1989).]**

the response of these three groups to two sleep quality questions: "Do you often have difficulty falling asleep?" and "Do you often have difficulty staying asleep?". The results obtained for the two questions are quite similar in each case. Workers in the N/D group most often reported that they had sleep problems. These data suggest that complaints by night shiftworkers about the quality of their sleep are related to their work shift preferences, perhaps more than to working the night shift per se.
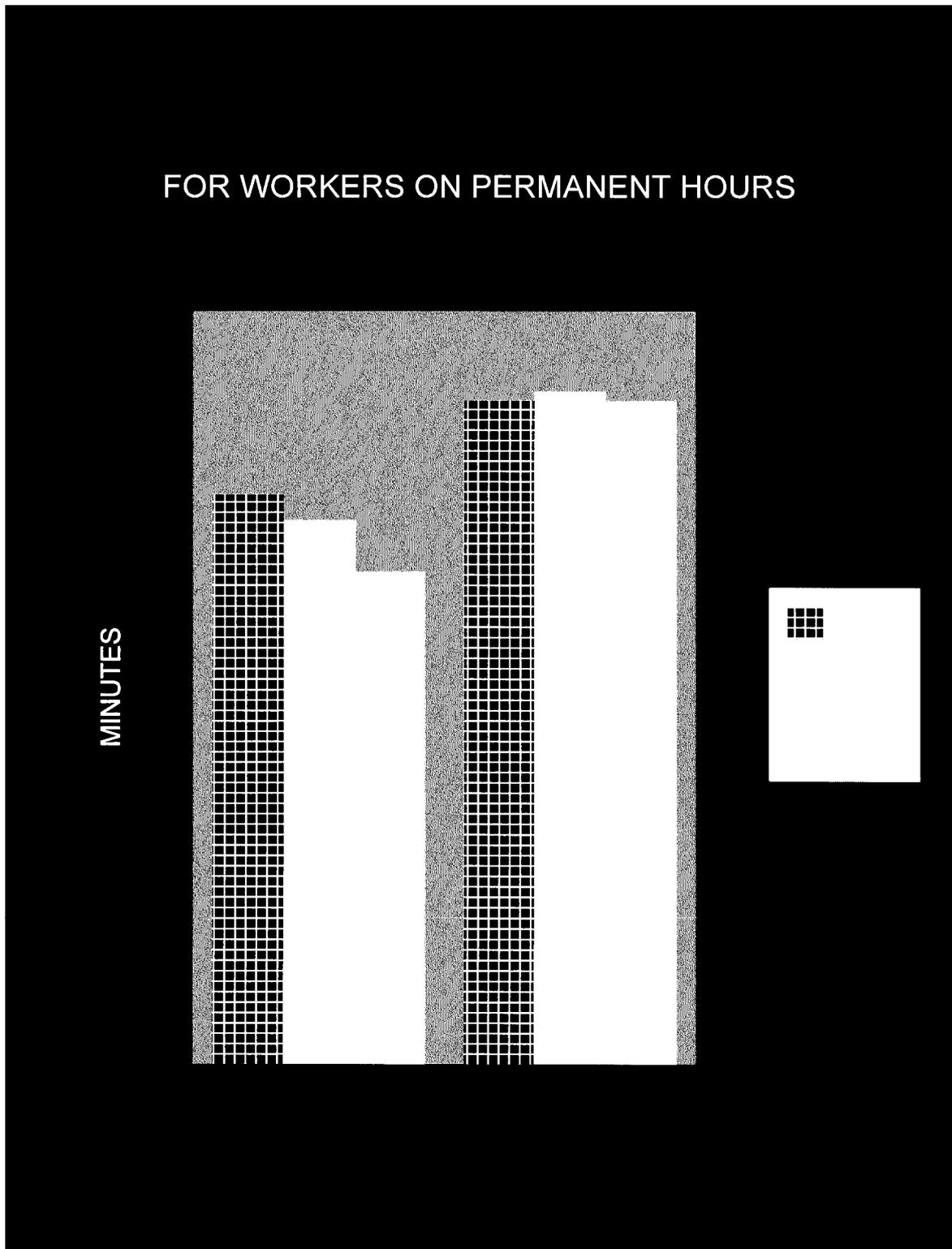
FIGURE 2. Percentage of workers responding "Yes" to two sleep quality questions: "Do you often have difficulty falling asleep?" and "Do you often have difficulty staying asleep?".
[*Note*: Data from three matched groups, with 108 workers in each group: D/D, workers who were employed on the day shift and most preferred to work the day shift; N/D, workers who were employed on the night shift and most preferred to work the day shift; and, N/N, workers who were employed on the night shift and most preferred to work the night shift. Graphs plotted from data reported by Tepas and Mahan (1989).]

When chronic exposure to a work schedule system is to be assessed, sleep length data are often a benchmark which can be used to assess the merits of a specific system. *Sleep data provide the ergonomic investigator of work shift systems with a valuable tool.* However, this does *not* mean that ergonomic analysis need only focus on estimates of physiological systems or chronobiology. Chronic reductions in sleep are associated with health and safety hazards, but the usability of a given work schedule system depends on many variables. The state and interaction of multiple physiological, social and cultural variables must be considered. In most cases, if not all, intervention efforts call for a macroergonomic approach to work schedule system design.

3. ERGONOMIC DESIGN CONSIDERATIONS

Given the conclusion that multiple interacting variables determine the impact of a given shift schedule, it should not be surprising that work schedule systems take many forms. In fact, thousands of different work schedule systems are being used. In practice, managers often behave as if they think work schedule systems take only a few forms. They assume that one or two of these forms can be labeled by the expert as the best schedule available. However, the contemporary consensus among work schedule experts is quite different. Research data leads work schedule experts to assume that there are *no* universal work schedule solutions for shiftwork which can be applied to most workplaces or all workers. Some dimensions which should be considered in designing a system are: worker age and gender; job task load; household size and duties; worker preferences and experience; current employment conditions; and, assessments of stress, fatigue, health and safety data.

Just as it is reasonable to conclude that there is no single variable which determines shiftworker maladjustment or adjustment, it is also reasonable to expect that multiple work schedule solutions may all be acceptable for a given workplace. Most ergonomic design efforts consider and evaluate alternative designs, and this is also true when one evaluates work schedules. A good work schedule systems design or evaluation effort includes a consideration and evaluation of alternative work schedule forms. Some of the options which should be considered are: whether the hours of work should be permanent or rotating; the direction of rotation, if there is to be any; the rate of rotation, if there is to be any; the length of workdays; what days of the week are to be non-workdays; the time of day shifts start, and the temporal regularity of shifts.

4. WORK SHIFT USABILITY TESTING

The macroergonomic focus for work shift design places an emphasis on the development of assessment tools, rather than a search for universal solutions. The array of tools used to schedule, design and/or evaluate work is referred to as *workware*, just as the term software refers to an array of tools used to program computer hardware to perform tasks and/or evaluate hardware. *Work shift usability testing* is one form of workware. Just as a computer software program requires software usability testing, a good work schedule systems requires work shift usability testing. Like other forms of usability testing, work shift usability testing makes the assumption that reliable and valid testing with real users is required.

Recommended approaches to work shift usability testing are detailed in Tepas (in press). These are new and evolving technologies whose use will become more common as around-the-clock operations continue to expand. For the present, it is reasonable to assume that many managers are not aware of the range of alternative work schedules available, or the need for a professional systems approach to the selection and evaluation of work schedule systems (Tepas 1994). The definitions and notations used for many shiftwork terms and categories often vary in the literature, in popular use, and from nation to nation (Tepas, Paley and Popkin 1997). Good ergonomic practice should stress the importance of clear definitions, good notation systems, and the quantitative assessment of existing and new installations. Unlike some ergonomic design problems, work shift usability testing does require a macroergonomic approach.

REFERENCES

MAHAN, R.P., CARVALHAIS, A.B., and QUEEN, S.E., 1990, Sleep reduction in nightshift workers: Is it sleep deprivation or a sleep disturbance disorder? *Perceptual and Motor Skills*, 70, 723–30.

MONK, T.H., 1989, Social factors can outweigh biological ones in determining night shift safety. *Human Factors*, 31, 721–4.

PRESSER, J.B., 1995, Job, family and gender: Determinants of nonstandard work schedules among employed Americans in 1991. *Demography*, 32, 577–98.

TEPAS, D.I., 1994, Technological innovation and the management of alertness and fatigue in the workplace. *Human Performance*, 7, 165–80.

TEPAS, D.I. In press, Work shift usability testing. In Karwowski and Marras (eds.) *The Industrial Ergonomics Handbook* (Boca Raton: CRC Press).

TEPAS, D.I. and CARVALHAIS, A.B., 1990, Sleep patterns in shiftworkers. In Scott (ed.) *Shift work* (Philadelphia: Hanley and Belfus). pp. 199–208.

TEPAS, D.I. and MAHAN, R.P., 1989, The many meanings of sleep. *Work and Stress*, 3, 93–102.

TEPAS, D.I., PALEY, M.J. and POPKIN, S.M., 1997, Work schedules and sustained performance. In Salvendy (ed.) *Handbook of Human Factors and Ergonomics, Second Edition.* (New York: Wiley), pp. 1021–58.

TEPAS, D.I., WALSH, J.K., MOSS, P.D. and ARMSTRONG, D., 1991, Polysomnographic correlates of shiftworker performance in the laboratory. In Reinberg, Vieux and Andlauer (eds) *Night and Shift Work: Biological and Social Aspects.* (Oxford: Pergamon). pp. 179–86.

Shiftwork Health Consequences

G. Costa

Department of Public Health, University of Verona, I-37134 Verona, Italy

1. INTRODUCTION

Shiftwork is increasing in modern society as an important tool of flexibility of the work organization. It enables one to cope with round-the-clock activities in relation to technological constraints, productive demands, social services and leisure activities. According to some recent statistics ~20% of workers in industrialized countries are involved in some form of shiftwork including nightwork.

There are several types of shift systems that differ according to their main characteristics such as: permanent or rotating; continuous or discontinuous (interruption on weekend or on Sunday); with or without nightwork; regularity/irregularity and length of the shift cycles, duration of the shifts (e.g. from 6 to 12 h); number of crews; start and finishing times of shifts; speed (fast, slow) and direction (clockwise or counterclockwise) of shift rotation.

Shiftwork, in particular that including nightwork, can have a negative impact on workers' health due to a mismatch between psychophysiological functions and environmental synchronizers, which can cause significant disruptions of biological rhythms, social and family life. This can have adverse effects on performance efficiency, health and social relations, and lead to increased risk of errors and accidents, morbidity and absenteeism, with consequent high economic and social costs both for the individual and the society.

2. INTERFERENCE ON PSYCHOPHYSIOLOGICAL FUNCTIONS

On shiftwork, and particularly on nightwork, workers are compelled to alter their normal sleep–wake cycle according to the changed activity and rest periods. This interferes with the regular oscillation of bodily functions ("circadian rhythms"), which are linked to the normal diurnal pattern of human life showing, in general, higher levels during the day (ergotrophic phase) and lower levels during the night (trophotropic phase). This rhythmicity is controlled by a "body clock," located in the suprachiasmatic nuclei, and influenced by environmental "synchronizers," in particular activity, sleep and light exposure.

The perturbation caused by shiftwork is evidenced by a flattening of the amplitude and a shift of the acrophase (peak of the function) of the biological rhythms, which can be more or less pronounced according to the number of successive nightshifts worked and the forward (i.e. morning–afternoon–night) or backward (i.e. afternoon–morning–night) rotation of the duty periods. Consequently, a de-synchronization among the partial adjustments of the different functions is observed, also due to the "masking" influence of the physical activity, which has more effect on the functions having a prevalent exogenous component (e.g. heart rate) than those with a strong endogenous control (e.g. body temperature, cortisol).

In relation to this perturbation, workers complain to a greater or lesser extent of a syndrome similar to "jet lag," characterized by fatigue, sleepiness, insomnia, dyspepsia, irritability and impaired mental agility.

3. INTERFERENCE ON SLEEP AND MENTAL PERFORMANCE

Shiftworkers have to change sleep times and strategies according to the duty periods; consequently, both sleep length and quality can be affected considerably according to the variable retiring and rising times on the different shifts.

On nightshifts in particular, the subsequent diurnal sleep is perturbed both for physiological reasons, being difficult to fall asleep and sleep for a long period when sleep starts during the rising phase of the body temperature rhythm, and for unfavorable environmental conditions (light and noises in particular) or domestic commitments (mealtimes, child care): sleep is shorter, frequently interrupted and altered in its stage sequencing.

A reduction of hours of sleep is generally recorded also on morning shifts, due to the early start time of the duty period, which causes a cut off of the last part of the sleep, richer of REM phase.

In the long run, sleep perturbations and deprivations, besides having influence on de-synchronization of circadian rhythms and on "shift–lag" syndrome, can yield persistent and severe disturbances of sleep itself, chronic fatigue and psychoneurotic syndromes (such as chronic anxiety or depression), which often require treatment by hypnotic or psychotropic drugs.

Sleep disturbances, chronic fatigue and oscillatory fluctuations of vigilance and performance can be also important contributing factors of the "human error" and consequent accidents. It is worth mentioning that shift and nightwork and sustained operation have been claimed as important causal factors in many work accidents and in some tragic events occurred at night, e.g. the Three Mile Island and Chernobyl nuclear accidents, the Bophal disaster, the Exxon Valdez shipwreck, as well as the Challenger space-shuttle explosion.

Moreover, sleepiness due to a too early starting hour on morning shifts or in case of prolonged duty periods, can favor higher frequencies of errors and accidents also during the dayshifts, as it has been reported in train and bus drivers, and long-haul pilots.

However, the epidemiological studies on accidents in industrial shiftworkers are not all in agreement on this matter; in fact, this has to be evaluated taking into account the complex interactions between the chronobiological and the organizational factors, such as environmental conditions (e.g. lighting), job content, fluctuations in work load, time pressure, long working hours (e.g. 12-h shifts), number of workers and supervision levels.

4. DIGESTIVE AND CARDIOVASCULAR DISORDERS

Shift- and nightwork interfere also with meal times and content, thus favoring digestive disorders. These are due to displacement between meal times and phases of gastrointestinal secretion and motility, as well as to changes in food quality (e.g. cold prepackaged food), too short work pauses for meals, nibbling and, sometimes, increased intake of caffeinated drinks, tobacco smoking and alcohol consumption.

Disturbances of appetite, dyspepsia, heartburn, abdominal pains, constipation, grumbling and flatulence are frequently complained by shiftworkers (20–70 compared with 10–25% of dayworkers), who may also develop more serious illnesses such as chronic gastritis, gastroduodenitis, peptic ulcer and colitis. According to many epidemiological studies peptic ulcer shows an incidence 2–5 times higher, and a shorter latency from starting work and diagnosis, among shiftworkers with nightshifts opposed to dayworkers; on the other hand, there is a significant decrease of complaints and disorders after transfer from shiftwork to daywork.

Shiftwork can also be a risk factor for cardiovascular diseases, in particular coronary heart diseases. As a stressor, it can cause a neurovegetative activation with increased secretion of stress hormones and consequent effects on blood pressure, heart rate, thrombotic processes, lipid and glucose metabolism; these can derive also from interference on compensatory mechanisms connected to more stressful living conditions and more risky life styles (i.e. smoking, diet, sleep disturbances).

In recent years some epidemiological studies have drawn the attention on this topic, showing a prevalence of some risk factors for cardiovascular diseases in groups of apparently healthy shiftworkers, and a higher incidence of complaints related to angina pectoris and hypertension in some groups of shiftworkers. Other studies have reported an increased risk for myocardial infarction in occupations with high proportions of shiftworkers, and a higher morbidity for cardiovascular and ischemic heart diseases with increasing age and shiftwork experience.

5. INTERFERENCE ON WOMEN'S REPRODUCTIVE AND PARENTAL FUNCTION

Women can be more vulnerable to shift- and nightwork in relation both to their more complex circadian and infradian (menstrual cycle) hormonal rhythms and to extra demands related to family life and domestic commitments.

Disorders of the menstrual cycle and reproductive function have been reported in many groups of women shiftworkers, leading to a higher incidence of menstrual pains, abortion and interference on fetal development, such as premature births and low birthweight.

Besides, women shiftworkers (those married with small children, in particular) can have more difficulties in combining their irregular working schedules with the additional domestic duties, thus suffering more for sleep troubles and chronic fatigue than their male colleagues.

This can support the view that women shiftworkers, without being discriminated in terms of parity, should have more protection, i.e. exemption from nightwork when pregnant, and possibility of transfer to daywork during the first 2–3 years of age of their children.

6. TOXICOLOGICAL RISK AND DRUG EFFECTIVENESS

Shiftwork can influence the risk of intoxication due to chemical substances in relation to both the circadian fluctuation of biological susceptibility to xenobiotics, and a de-synchronization of the mechanisms of detoxification. This has been proved by studies of experimental chronotoxicology, which show a circadian rhythmicity of effectiveness of some toxic compounds given at different times of the day, and variations of the susceptibility after changes of the light–dark regimen. A suggestive evidence in humans came from the Bhopal disaster (1984), where surprisingly none of the nightworkers died from the vapors of methyl isociante, while thousands of inhabitants of the near villages died in their sleep; at the same time thousands of cattle died, while small nocturnal active rats were observed to be scurrying around the corpses and carcasses.

Therefore, risk assessment should consider the circadian fluctuations of physiological responses (in terms of absorption, metabolism, excretion) in relation to shift- and nightwork: some studies have already evidenced a circadian pattern of excretion of some toxic substances or metabolites, which can be used for a better strategy of protection of the workers.

On the other hand, these aspects should also be considered in case of regular drug assumption for certain illnesses (e.g. diabetes, epilepsy, hypertension, asthma, hormonal and sleep disorders), as the persistence on shift and nightwork can interfere significantly with the efficacy of the pharmaceutical treatment, which requires a precise timing of administration and a stable life regimen.

7. EFFECTS ON FAMILY AND SOCIAL LIFE

Persons engaged in shift- and nightwork are frequently out of phase with the society and can face greater difficulties in their social lives since most family and social activities are arranged according to the day-oriented rhythms of the general population. Consequently shiftwork can lead to some social marginalization due to the interference between the workers' time budgets (working hours, commuting and leisure times) and the complex organization of social activities (on daily and weekly basis mainly), particularly when these refer to groups of persons and require regular contacts.

Shiftwork may also interfere with the coordination of family timetables according to family composition (i.e. number and age of children, cohabiting persons), personal duties (i.e. school, housework), availability of public services (i.e. shop hours and transportations). "Time pressure" is a constant condition among those who have a high family burden (e.g. women with small children), and this can have a negative influence on marital relationships, parental roles and children's education.

Such family and social difficulties are often the main cause of maladaptation to shiftwork, since they have also a clear influence on the development of the psychosomatic disorders already mentioned.

On the other hand, shiftwork may consent a more flexible use of daytime hours to meet particular needs or preferences (i.e. access to public services, to study, second job, solitary hobbies) or in case of persons who give a higher priority to family and domestic duties than to personal leisure (e.g. women with children).

8. FACTORS INFLUENCING SHIFTWORK EFFECTS ON HEALTH

There is a high interindividual variability on tolerance to shiftwork. According to several studies, ~15–20% of workers are forced to leave shiftwork in a short time because of health disorders, whereas 5–10% do not report any complain during their working life; the majority withstand shiftwork with different levels

of maladaptation, showing different times and severity of manifestation in terms of troubles or diseases.

Many factors concerning individual, social and working conditions can interact and influence both short-term adjustment and long-term adaptation.

As concerns the individual aspects, the more important ones refer to age, gender, physical fitness, behavior and personality traits, For example, aging can be associated with a progressive intolerance due to reduced psychophysical fitness, decreased restorative properties of sleep, and proneness to internal de-synchronization of circadian rhythms. Besides, subjects having the characteristics of "morningness" generally face more difficulties in short-term adjustment to nightwork compared with the "evening" types. Moreover, people who present high levels of neuroticism, or the characteristics of rigidity of sleeping habits and lower ability to overcome drowsiness, have more difficulties in their adaptation to irregular work schedules. On the other hand, good physical fitness and a strong commitment to shiftwork could favor a better coping.

Family situations (e.g. marital status, number and age of children, partner's (shift)work, housing) and social conditions (e.g. socio-economic status, labor market, leisure activities, social services, commuting times) can also play an important role on short- and long-term tolerance.

Moreover, working conditions (e.g. work sector, work environment, work load, job characteristics, income level, qualification, job satisfaction, career opportunities, human relations) and, in particular, the organization of the working hours (e.g. shift schedules, number of crews, rest pauses, night duties, free weekends, etc.) have a crucial influence both on the biological adjustment and on the psychophysical conditions.

All these factors can affect tolerance to shift and nightwork according to the specific circumstances, and their interactions can give rise to not only possible additive or multiplicative, but also to subtractive effects. Consequently, health complaints or disorders can manifest with various degrees of severity and in different periods of worker's life, thus making difficult to compare the effective harmfulness of shiftwork in different groups and individuals.

This is also influenced by the process of self-selection that often occurs among shiftworkers (the "healthy worker effect"); in fact, many epidemiological studies examined groups of workers where a more or less relevant part had left shiftwork because of health troubles or social problems.

Also reports on absenteeism can be sometimes misleading. Although it is a rather indirect indicator of health, it can be affected by the shift scheduling (i.e. amount of nightwork, speed of shift rotation, overtime, shift start and finishing times) in association to other factors connected to work organization, socio-economic conditions and individual characteristics. However, shiftworkers may show a lower absenteeism rate, despite of a higher frequency of troubles and illnesses, both because of a higher solidarity towards colleagues (since an unexpected absence may cause more problems for shift handovers than for normal daywork) and for a higher threshold in the perception, labeling and reporting of complaints, that shiftworkers often accept as "part of the job."

Therefore, proper intervention measures should be undertaken to prevent negative effects on health and well-being. They should deal with work organization, social support, medical control, education and counseling for better coping strategies with shift and nightwork (see the chapter on "Prevention and compensation of shiftwork effects").

REFERENCES

AKERSTEDT, T., 1996, *Wide Awake at Odd Hours* (Stockholm: Swedish Council for Work Life Research).

COLQUHOUN, W.P., COSTA, G., FOLKARD, S. and KNAUTH, P., 1996, *Shiftwork: Problems and Solutions* (Frankfurt aM: Peter Lang).

COSTA, G., 1996, The impact of shift and nightwork on health. *Applied Ergonomics*, 27, 9–16.

FOLKARD, S. and MONK, T. (eds), 1985, *Hours of Work: Temporal Factors in Work-Scheduling* (Chichester: John Wiley).

INTERNATIONAL LABOUR OFFICE, 1990, *Night Work Report V(1)* (Geneva: ILO).

SCOTT, A.J. (ed), 1990, Shiftwork. *Occupational Medicine: State of Art Review*, 5.

WATERHOUSE, J.M., FOLKARD, S. and MINORS, D.S., 1992, *Shiftwork, Health and Safety: An Overview of the Scientific Literature 1978–1990* (London: HMSO).

Shiftwork Stress

L. Smith

Shiftwork and Safety Research Group, School of Psychology, University of Leeds, Leeds LS2 9JT, UK

1. INTRODUCTION

1.1. Shiftwork

Shiftwork is widespread in all industrialized and developing countries. It is a way of organizing working time that can be very disruptive to the lives of shiftworkers and their employers. Shiftwork often requires people to work regularly at night and this can result in direct disruption to the human biological clock, impairment to sleep and mood, reduced alertness and increased fatigue, and interference to social and family life. There may also be consequences in terms of employee performance, safety, productivity, moonlighting, absence for sickness and turnover (Colquhoun *et al.* 1996). This article highlights the issue of shiftwork as a stress factor. To accomplish this, models of shiftwork stress are presented that indicate different perspectives to the application of this issue to working life and identify practical implications with respect to shiftworker's lives.

1.2. Shiftwork as a Stress Factor

Cooper and Marshall (1976) identified a number of sources of stress at work within a model of organizational and personal disease development. They included: factors intrinsic to the job such as work over-/underload, time pressures, shiftwork, physical working conditions and repetitive work; role-based stress such as work role ambiguity, work role conflict and levels of responsibility; interpersonal relationships with subordinates, colleagues and superiors; career development factors such as lack of job security, under/over promotion thwarted ambition and organizational structure and climate issues, including office politics, communications, participation in decision-making and the organization of work and organizational trust. Job conditions are a major stress factor and a major part of a person's working conditions is, of course, their work schedule. Cooper and Marshall (1976) recognized this fact and included shiftwork as an element of their model. The following section presents a number of shiftwork-related models of stress.

2. MODELS OF SHIFTWORK STRESS

A number of models of shiftwork stress are founded on stimulus-based theories of stress that focus on environmental factors which cause a response in the person, i.e. strain. Shiftwork stress models using this approach describe shiftwork as an environmental factor that causes strain that manifests itself as health problems (e.g. Akerstedt and Froberg 1976, Rutenfranz *et al.* 1981, Knutsson 1989). Transactional models have also been used. These models describe stress as a product of the ongoing interaction between the person and her/his environment. Stress occurs when demands tax or exceed the resources available to a person. It is an approach that emphasizes the role of cognitive appraisal, personality and behavior in the stress process. A number of shiftwork stress models has adopted this perspective (e.g. Haider *et al.* 1982, Monk 1988, Olsson *et al.* 1990).

A number of these models have been proposed, such as a model with three components: objective stress, subjective strain and intervening variables. The stress of shiftwork is seen as the work rota-imposed requirement to alter the timing of activity and sleeping relative to the normal phases of circadian (24-h) rhythms in physiological processes and performance. Being out of phase can lead to strain in terms of increased health and sleep complaints and lowered well-being, but the pathway between stress and strain may be mediated by intervening variables such as family situation, personality, housing/locality and organizational factors.

Another model (Akerstedt and Froberg 1976) proposed two separate causal links between shiftwork and health: first, disruption to circadian rhythms; and second, disruption to social rhythms. Shiftwork hours conflict with physiological, family and social rhythms, and results in the establishment of new circadian and social rhythms. These interact to cause stress and health problems. Severe disturbance to social rhythms is also believed to contribute to psychological stress.

Knutsson (1989) suggested a shiftwork-health model. The model was based on cross-sectional, prospective and longitudinal cohort studies on shiftwork with heart disease as the health measure. The data provided evidence of a dose–response relationship between years of shiftwork experience and heart disease, and increased levels of heart disease risk factors in shiftworkers compared with day workers. Three causal paths to disease are suggested: first, disturbed physiological rhythms; second, behavioral changes; and third, disturbed social rhythms. For this model disruption to body rhythms and social rhythms have separate effects. Disturbed circadian rhythms lead to poor sleep, increased susceptibility to illness and internal desynchronization, all of which lead to disease. Disrupted social rhythms result in the adoption of new behaviors (often palliative, or inappropriate coping behaviors such as increased smoking, alcohol intake or drug abuse) in response to the interference with social life and stress. These, in turn, are a risk to health.

This type of model moves away from a stimulus–response conceptualization with a direct link between shiftwork and poorer health. Instead it considers that the personality, thoughts and behaviors of shiftworkers will moderate the stress experience (Haider *et al.* 1982). The model suggests that the balance between three spheres of activity is central to adaptation to shiftwork schedules, namely: sleep behavior, family life and attitudes to shiftwork.

These domains interact and exist in a dynamic equilibrium that functions in a regulatory capacity. For example, a conflict can exist because more time with the family may be gained at the expense of obtaining less sleep. Difficulties with sleep and family relationships may then lead to negative attitudes towards shiftwork that in turn result in poor sleep. Destabilization to the "dynamic equilibrium" between the three domains is caused by the different activity pattern when working shifts. Resulting conflicts may then affect physiological systems and cause impairment to health. In addition to this, shiftworkers may engage in more "risky" coping behaviors that contribute to impairment to health, e.g. increased smoking, caffeine, alcohol, or drug use and changed eating habits. However, the model recognizes that the destabilization process may be modified by personality factors. Coping efforts form explicit components of other models (e.g. Monk 1988). The ability to cope with shiftwork is determined

by the interaction and direct effects of three factors: the biological clock/body rhythms, sleep, and social and domestic factors. For example, the body clock has to be reset when working shifts. This results in sleep complaints, malaise and digestive problems. Sleep is influenced by both the body clock and external factors such as daytime noise. The shiftworker's routine conflicts with timing of family and community activity. So, all three areas are interdependent. A successful adapter to shiftworking is viewed as someone who applies multifaceted strategies that affect aspects of the three most influential factors.

Shiftwork experience is suggested to moderate the destabilization process. Three phases are suggested: first (0–5 years), there is the adaptation phase where the quality of sleep, family life, social activity and work demands have a strong influence on health; second (5–20 years), there is the sensitization phase in which the primary predictors of health will be job satisfaction, attitudes towards shiftwork and the family situation; and third (20+ years), family and social situations are suggested to remain stable or improve but sleep behavior linked to aging and the development of risky behaviors will have the greatest adverse effect on health.

Olsson *et al.* (1990) offered an approach that viewed shiftwork as only one of many occupational stressors that may lead to poor health outcomes. Shiftworkers are viewed as engaging in a two-way relationship with their environment. Threats to well-being are cognitively appraised. There is also the appraisal of a personal capability to deal with or manage the potential threats. Coping strategies are adopted in response to these judgements. Coping can alter the environment directly and/or it can change the appraisal process to alter the stressfulness of the situation. Problem-focused coping may involve dealing with sleep problems by learning a relaxation technique, or adopting a strategy of napping to promote alertness, or even to changing to permanent daywork. Emotion-focused coping, on the other hand, refers to the shiftworker expressing their feelings about their difficult situation, or using relaxation techniques and mental strategies to help reduce anxiety or unwanted thoughts. Thus, the thoughts and behaviors of the shiftworkers play a central part in determining their stress experiences when working shifts.

More recently Barton *et al.* (1995) recognized that the problems faced by individual shiftworkers are both complex and multifaceted. Their model of shiftwork stress is a representation of the way in which problem areas relate to the features of shift systems and to one another. The model offers a conceptual framework that helps guide research efforts to examine the impact of different shift systems on individuals and organizations. Shift systems and their features such as shift timing and duration are seen to impact on health and safety. This influence may be relatively direct but there can also be indirect effects resulting from the influence on the whole of a person's life. Interrelated disruptions of biological rhythms, sleep, and family and social life may be moderated by individual differences as well as situational differences such as domestic circumstances or commuting time. The acute effects of shiftwork, such as when an individual returns to a block of night shifts, may also be moderated by the more chronic effects on mental health. In the long-term, physical health may be impaired by chronic impact on, for example, psychological well-being while safety may be compromised by the acute negative impact of shiftwork on mood and cognitive performance.

3. CONCLUSIONS AND IMPLICATIONS

Theoretical and research-based models of how shiftwork impacts upon the lives of shiftworkers are useful for a number of reasons. For example, Taylor *et al.* (1997) noted that the models help to identify possible interventions that will minimize the negative impact of shiftwork by (1) providing an understanding of the potential interactions between variables considered to influence the experience of problems when working shifts, (2) helping to understand the development of shiftwork problems and identifying possible strategies to combat shiftwork disruption and (3) helping to identify research questions that may form the basis of future study.

Some models are explicitly based on empirical data. For example, Haider *et al.* (1981) formulated their model from the results of tests on cross-sectional data. Causal relationships were inferred from a complex pattern of associations. The models presented provide an overview of the shiftwork research field by organizing existing knowledge about the development of shiftwork-related problems. They can be considered to be both theory-driven and theory building as they have helped the formulation of general hypotheses that could add to the understanding of shiftwork effects and ways of intervening.

Stress theory has been used to help describe the shiftwork–health relationship through the application of stimulus-based and transactional models of stress. These conceptualizations tend to influence suggestions for strategies aimed at reducing shiftwork problems. In the case of stimulus-based models that assume that the work environment causes ill health, the basic premise is to remove and/or improve detrimental work conditions. That is, work stress simply 'happens' to the shiftworker and removing the source of the stress will result in the prevention of health problems. Transactional models tend to advocate educational and stress management training approaches to improve coping skills. This is because they recognize the role played by the individual's thoughts and behaviors in the experience of work stress. It should be noted, however, that one negative aspect of adopting a transactional rather than a stimulus-based approach is that the responsibility for the employee's health is transferred, at least in part, to the shiftworker. It is possible that in future there will have to be a re-emphasis of the impact of shiftwork as a risk factor and the shiftwork environment. This would result in an increased resort to providing better shift systems and using other interventions at the level of the organization (e.g. installing bright lights) if the problems of shiftwork are to be reduced for the largest numbers of people who regularly work shifts.

REFERENCES

AKERSTEDT, T. and FROBERG, J., 1976, Shiftwork and health: interdisciplinary aspects. In Rentos, P.G. and Shepard, R. D. (Eds) *Shiftwork and Health* (Washington: US Department of Health Education and Welfare), pp. 179–97.

BARTON, J., SPELTEN, E., TOTTERDELL, P., FOLKARD, S., SMITH, L. and COSTA, G., 1995, The standard shiftwork index: a battery of questionnaires for assessing shiftwork-related problems. *Work and Stress*, 9, 44–30.

COLQUHOUN, W. P., COSTA, G., FOLKARD, S. and KNAUTH, P., 1996, *Shiftwork: problems and solutions* (Frankfurt am Main: Peter Lang).

COOPER, C. and MARSHALL, J., 1976, Occupational sources of stress: a review of the literature relating to coronary heart disease and mental ill health. *Journal of Occupational Psychology*, 49, 11–28.

HAIDER, M., KUNDI, M. and KOLLER, M., 1982, Methodological issues and problems in shiftwork research. In Johnson, L. C., Tepas, D., Colquhoun, W. P. and Colligan, M. (Eds) *Biological Rhythms, Sleep and Shiftwork* (New York: Medical & Scientific Books), pp. 145–63.

KNUTSSON, A., 1989, Shiftwork and coronary heart disease. PhD thesis, National Institute for Psychosocial Factors and Health, Karolinska Institute, Stockholm.

MONK, T. H., 1988, Coping with the stress of shiftwork. *Work and Stress*, 2, 169–72.

OLSSON, K., KANDOLIN, I. and KAUPPININ-TOROPAINEN, K., 1990, Stress and coping strategies of three-shift workers. *La Travail Humain*, 53, 175–88.

RUTENFRANZ, J., KNAUTH, P. and ANGERSBACH, D., 1981, Shiftwork research issues. In Johnson, L. C., Tepas, D., Colquhoun, W. P. and Colligan, M. (Eds) *Biological Rhythms, Sleep and Shiftwork* (New York: Medical & Scientific Books), pp. 165–95.

TAYLOR, E., BRINER, R. B. and FOLKARD, S., 1997, Models of shiftwork and health: an examination of the influence of stress on shiftwork theory. *Human Factors*, 39, 67–82.

Situational Awareness Issues at Work

M. A. Vidulich

Crew System Interface Division, US Air Force Research Laboratory, AFRL/HECP, 2255 H Street, Wright-Patterson Air Force Base, OH 45433-7022, USA

1. INTRODUCTION

Situational awareness is a term that arose out of the fighter pilot community to explain the outcomes of air combat training exercises. In some cases, the more experienced pilot in the more capable aircraft might unexpectedly lose an engagement. In debriefings, these events were often explained by asserting that on this specific occasion the less experienced pilot had achieved better situational awareness and used it to win (Waddell 1979). With a user community already convinced that situational awareness was a valuable concept in human–system performance, it was natural that the human factors research community would follow suit and investigate situational awareness. Throughout the 1980s and 1990s, situational awareness emerged as a very active human factors research area. For example, a bibliography of situational awareness was published in 1994 that listed and reviewed > 230 situational awareness citations available at that time (Vidulich *et al*. 1994).

2. DEFINITIONS OF SITUATIONAL AWARENESS

Not surprisingly, one of the first great issues in situational awareness research was the matter of defining situational awareness. For example, Dominguez (in Vidulich *et al*. 1994) found and analyzed 15 different definitions of situational awareness. He consolidated the different definitions into a proposed consensus definition of situational awareness as (Vidulich *et al*. 1994: 11) "Continuous extraction of environmental information, integration of this information with previous knowledge to form a coherent mental picture, and the use of that picture in directing further perception and anticipating future events". Probably the most popular definition of situational (or "situation") awareness was advanced by Endsley (1995a: 36), "Situation awareness is the perception of the elements in the environment within a volume of time and space, the comprehension of their meaning, and the projection of their status in the near future". Both of these definitions and most others emphasize the operator's perception of the elements of the current situation and their integration with knowledge to form an adaptive understanding of what is going on and to maximize the opportunity to choose an appropriate course of action. The situational awareness concept has demonstrated great vitality in inspiring research into selection of individuals predisposed towards developing good situational awareness, training procedures to maximize operator situational awareness, and interface design to aid situational awareness (for examples, see Vidulich *et al*. 1994). In all of these proposed applications of the situational awareness concept, an underlying theme is the need for situational awareness measurement. After all, the evaluation of any intervention designed to improve situational awareness requires that some means for assessing situational awareness exist.

3. MEASUREMENT OF SITUATIONAL AWARENESS

A recent analysis of the sensitivity of situational awareness metrics in 65 studies of interface manipulations (Vidulich, 2000) found that three major approaches to situational awareness measurement have emerged: performance-based, memory probe and subjective ratings. The results of all three of these measurement approaches are instructive in understanding the importance of the situational awareness concept and the role that it may play in human–system analysis.

4. PERFORMANCE–BASED MEASUREMENT

Measuring operator performance is not usually considered a direct measurement of situational awareness. However, a core assumption behind the interest in situational awareness is that in most circumstances better situational awareness will be associated with better performance. Almost all of the papers in the Vidulich (2000) review included some measure of performance. In about one-half of the studies the performance measure was included in addition to some specified situational awareness metric. However, in the other half of the studies, the performance measure was the basis for inferring any changes in situational awareness. Overall, performance measurement was the most common type of measure used in the database. Regardless of whether the experimenter used performance as a measure of situational awareness, changes in interfaces that were intended to improve situational awareness tended to improve performance. This finding demonstrates the potential practical value of the situational awareness concept applied in system design.

4.2. Memory Probe Measures

If situational awareness is regarded as the information in awareness, then explicit measures are the most direct method of assessing situational awareness. Two main categories of explicit measures can be identified: retrospective event recall and concurrent memory probes. Retrospective event recall occurs after the operator has completed a simulated mission. Concurrent memory probes usually involve randomly stopping the simulation sometime prior to its completion to assess the contents of awareness at that moment. For the most part, retrospective and concurrent approaches share the same general strengths and weaknesses. The greatest strength is face validity. In any complex goal-directed task, conscious decision-making would be expected to be important. Explicit measures assume that such conscious data are reportable. The best known approach for collecting this type of data is the situation awareness global assessment technique (SAGAT; Endsley 1995b). SAGAT, and very similar procedures, have been widely used in system evaluations.

Vidulich (2000) found that the sensitivity of memory probe measures of situational awareness was strongly influenced by the breadth of probe questions used in the evaluation. Evaluations that used a very broad set of memory probe questions were generally sensitive to the interface manipulations. On the other hand, evaluations that used a very small set of focused questions (or even one question) tended to be insensitive to interface manipulations.

4.3. Subjective Ratings

Implementing subjective ratings as a measure of situational awareness is simple; operators are merely asked to rate their experience of situational awareness after completing a task. Early attempts at subjective situational awareness ratings used a straightforward unidimensional situational awareness scale. Although this was simple, this approach also tended to be insensitive. Currently the most popular and validated subjective metric is the situation awareness rating technique (SART) developed by UK researchers at Farnborough. SART has been demonstrated to be sensitive to many manipulations of situational awareness in laboratory and high-fidelity simulator evaluations. In fact, SART was the most reliable measure of situational awareness examined in the Vidulich (in press) review.

4.4. Situational Awareness as a Meta-measure

As described above, manipulations of interface design that were undertaken to improve situation awareness tended to improve task performance. This suggests that situation awareness researchers are justified in believing that the situation awareness construct represents an important phenomenon in human–machine interaction. As Selcon *et al.* (1996) have suggested, it might be worthwhile to consider situation awareness a "meta-measure" in system evaluation. These researchers pointed out that it is seldom possible to evaluate all of the mission relevant uses of a display. Thus, a metric approach is needed that provides a generalized link to likely future performance in a wide variety of settings. In other words, rather than focusing on task-specific performance, it might be preferable to examine meta-measures that encapsulate the cognitive reaction to using a given interface. One proposed meta-measure was mental workload. Presumably, if a task can be performed with one interface while experiencing less mental workload, this should provide for more robust transfer to using that interface in more challenging conditions. Another possible meta-measure is situational awareness. Presumably the interface that provides a better understanding of the situation will provide the better chance of reacting appropriately to that situation. The fact that the Vidulich (2000) database suggested

that manipulations designed to improve situational awareness generally improved performance in a wide variety of tasks and by a wide variety of measures is consistent with the suggestion that situational awareness measurement could provide a useful meta-measure of system interface design.

5. CONCLUSIONS

The main points concerning the concept of situational awareness and its measurement can be summarized as follows:

- As a component of human factors research, situational awareness has displayed considerable vitality since the 1980s.
- Sensitive situational awareness measurement tools have been developed and demonstrated.
- System design changes intended to improve situational awareness also tend to improve system performance.
- In system design evaluations, situational awareness can serve as a meta-measure that will test overall design quality in a fairly generic fashion.

REFERENCES

ENDSLEY, M. R., 1995a, Toward a theory of situation awareness in dynamic systems. *Human Factors*, 95, 32–64.

ENDSLEY, M. R., 1995b, Measurement of situation awareness in dynamic systems. *Human Factors*, 95, 65–84.

SELCON, S. J., HARDIMAN, T. D., CROFT, D. G. and ENDSLEY, M. R., 1996, A test-battery approach to cognitive engineering: to meta-measure or not to meta-measure, that is the question! In *Proceedings of the Human Factors and Ergonomics Society 40th Annual Meeting*, 228–32.

VIDULICH, M. A., 2000, Testing the sensitivity of situation awareness measures in interface evaluations. In Endsley, M.R. and Garland, D. J. (Eds) *Situation Awareness Analysis and Measurement* (Mahwah: Lawrence Erlbaum Associates).

VIDULICH, M., DOMINGUEZ, C., VOGEL, E. and MCMILLAN, G., 1994, Situation awareness: papers and annotated bibliography. Technical Report AL/CF-TR-1994-0085 (Wright-Patterson: Armstrong Laboratory, Wright-Patterson Air Force Base).

WADDELL, D., 1979, Situational awareness. *USAF Fighter Weapons Review*, Winter, 3–6.

Sociotechnical Systems Analysis

B. M. Kleiner

1. INTRODUCTION

While many valid and potentially useful work process evaluation/ improvement methods exist and are in use, a general framework which has strong theoretical support and which integrates ergonomic interface design and function allocation and other macro-ergonomic tools is described. The following methodology is detailed in Hendrick and Kleiner (1999). It has been developed based on the writings of Emery and Trist (1978), Taylor and Felton (1993), Clegg *et al.* (1989) and the authors' own experience with large-scale organizational evaluation and change in academia, industry and government (Kleiner 1996, Kleiner and Drury 1999) with specific emphasis on integrating socio-technical systems theoretical propositions and prescriptions with micro-ergonomics.

This methodology can be used to conduct both formative and summative evaluation of production, non-production, linear and non-linear work processes. Although the methodology can be used by an inside or outside consultant, it is best used in a participatory ergonomics framework.

2. PHASES OF WORK SYSTEM PROCESS ANALYSIS AND DESIGN

2.1. Phase 1 — Environmental and Organizational Design Subsystems: Initial Scanning

Achieving a valid organization/environment fit may be the most important step to ensure future work system success. Mission, vision and principles (or MVP) provide definition or identity to the work system. Once defined, a gap between what the organization claims are its defining characteristics and its actual organizational behavior is identified. To assess the severity of this gap, the formal company identity statements are documented. Gaps in MVP can be inferred from what the work system actually produces in terms of products, bi-products, services, attitudes and behavior relative to what it formally declares as its outputs.

System scanning is the process of defining the workplace in systems terms including defining relevant boundaries. This step is aided by the construction of an input/output diagram, in which inputs, outputs, suppliers, customers, processes, internal controls and feedback mechanisms are identified for the work system. The system scan also establishes initial "boundaries" or domains of responsibility. As described by Emery and Trist (1978), there are throughput, territorial, social and time boundaries to consider.

In the environmental scan, the organization's subenvironments and the principal stakeholders within these subenvironments are identified. Their expectations for the organization are identified and evaluated. Problems are viewed as opportunities for process or interface improvement. The work system itself can be changed to conform with external expectations or the work system can seek to change the expectations of the environment to be consistent with its own direction. The ergonomist designs or redesigns interfaces among the organizational system and relevant subenvironments to improve communication, information exchange and decision-making. Rather than delay to address organizational design until all subsystems have been analyzed, it is useful develop preliminary organizational design hypotheses at this stage based upon the environmental scan and iterate later. By referring to the empirical models of the external environment, initial optimal levels of complexity (both differentiation and integration), centralization and formalization can be proposed.

2.2. Phase 2 — Technical Subsystem Analysis: Define Production System Type and Performance Expectations

As described above, it is important to identify the work system's production type since the type of production system can help determine optimal levels of complexity, centralization and formalization. The system scan performed in the previous phase should help in this regard and the ergonomist can consult the production models discussed previously.

In this context, the key performance criteria related to the organization's purpose and technical processes are identified. First and foremost, this requires a determination of success factors for products and services, but may also include performance measures at other points in the organization's system, especially if decision-making is important to work process improvement. As described in Kleiner's (1997) framework adapted from Sink and Tuttle (1989), specific standardized performance criteria guide the selection of specific measures which relate to different parts of the work process. Measures can be subjective, as in the case of self-reports, or measures can be objective, observed from performance.

Once the type of production system has been identified and the empirical production models consulted, the organizational design hypotheses generated in the previous phase should be supported or modified until the personnel subsystem can be thoroughly analyzed as well. From the Clegg *et al.* (1989) function allocation methodology, this is also the appropriate time to specify system level objectives. Requirements specifications can be developed, including ergonomic requirements. Also included are system design preferences for complexity, centralization and formalization.

2.3. Phase 3 — Flowchart the Technical Work Process and Identify Unit Operations

Unit operations are groupings of conversion steps that together form a complete piece of work and are bounded from other steps by territorial, technological, or temporal boundaries. They are exemplified by natural breaks in the process (i.e. boundaries determined by state changes (transformation) or actual changes in the raw material's form or location (input) or storage of material). For each unit operation or department the purpose/ objectives, inputs, transformations, and outputs can be defined (see Phase 1).

The current workflow of the transformation process (i.e. conversion of inputs to outputs) should be flow charted, including material flows, workstations and physical as well as informal or imagined boundaries. In linear systems such as most production

systems, the output of one step is the input of the next. In non-linear systems such as many service or knowledge work environments, steps may occur in parallel or may be recursive. Unit operations are identified. Also identified at this stage are the functions and subfunctions (i.e. tasks) of the system (Clegg *et al.* 1989). The purpose of this step is to assess improvement opportunities and coordination problems posed by technical design or the facility. Identifying the workflow before proceeding with detailed task analysis can provide meaningful context in which to analyze tasks. Once the current flow is charted, the ergonomist can proceed with a task analysis for the work process functions and tasks.

2.4. Phase 4 — Collect Variance Data

A variance is an unexpected or unwanted deviation from standard operating conditions or specifications or norms. As Deming and Shewhart suggested, variances can be assigned special or common causes, the former being abnormal causes and the latter expected system variation from normal operations. Special variances need to be tackled first to get the work process in control, at which time common variation can be tackled for overall system improvement. For the ergonomist, identifying variances at the process level as well as the task level can add important contextual information for job and task redesign. By using the flowchart of the current process and the detailed task analysis that corresponds to the flow chart, the ergonomist can identify variances. Deviations in raw material are called *input variances* and deviations related to the process itself during normal operations are called *throughput variances*. These can both be identified at this stage. Differentiating between types of variances helps determine how to control the variances.

2.5. Phase 5 — Construct Variance Matrix

Key variances are those variances that significantly impact performance criteria and/or may interact with other variances thereby having a "multiplying" effect. The purpose of this step is to display the interrelationships among variances in the transformation work process to determine which ones affect which others. The variances should be listed in the order in which they occur down the vertical *y*-axis and across the horizontal *x*-axis. The unit operations (groupings) can be indicated and each column represents a single variance. The ergonomist can read down each column to see if this variance causes other variances. Each cell then represents the relationship between two variances. An empty cell implies two variances are unrelated. The ergonomist can also estimate the severity of variances by using a Likert-type rating scale. Severity would be determined on the basis of whether a variance or combination of variances significantly affect performance.

A variance is considered *key* then if it significantly affects quantity of production, quality of production, operating costs (utilities, raw material, overtime, etc.), social costs (dissatisfaction, health, safety, etc.), or if it has numerous relationships with other variances (matrix). Typically, consistent with the Pareto Principle, only 10–20% of the variances are significant determinants of the quality, quantity or cost of product.

2.6. Phase 6 — Personnel Subsystem Analysis:

Construct Key Variance Control Table and Role Network

The purpose of this step is to discover how existing variances are currently controlled and whether personnel responsible for variance control require additional support. The Key Variance Control Table includes: the unit operation in which variance is controlled or corrected; who is responsible; what control activities are currently undertaken; what interfaces, tools or technologies are needed to support control; and what communication, information, special skills or knowledge are needed to support control.

Role analysis addresses who interacts with whom, about what and how effective these relationships are. This relates to technical production and is important because it determines level of work system flexibility. A *role network* is a map of relationships indicating who communicates with the focal role. First, the role responsible for controlling key variances is identified. Although multiple roles may exist which satisfy this criterion, there is often a single role without which the system could not function. With the focal role identified within a circle, other roles can be identified and placed on the diagram in relation to the focal role. Based upon the frequency and importance of a given relationship or interaction, line length can be varied, where a shorter line represents more or closer interactions. Finally, arrows can be added to indicate the nature of the communication in the interaction. A one-way arrow indicates one-way communication and a two-way arrow suggest two-way interaction. Two one-way arrows in opposite directions indicate asynchronous (different time) communication patterns. To show the content of the interactions between the focal role and other roles and an evaluation of the presence or absence of a set of functional relationships for functional requirements, the following are indicated: the short-term goal of controlling variance; adapting to short-term fluctuations; integrating activities to manage internal conflicts and promote smooth interactions among people and tasks; and ensuring long-term. Also the presence or absence of particular relationships is determined by describing the work process functions in terms of five types of relationships: vertical, equal, cross-boundary, outside or non-social. The effectiveness of the relationships identified in the role network can also be evaluated. At this juncture the organizational design hypotheses can be tested against the detailed analysis of variance and variance control. The role analysis and variance control table may suggest for example, a need to increase or decrease formalization or centralization.

2.7. Phase 7 — Function Allocation and Joint Design

Having previously specified system objectives, requirements and functions it is now appropriate systematically to allocate functions and tasks to human and machine or computer. It is helpful to review the environmental scan data to check for any subenvironment constraints (e.g. political, economic, etc.) before making any mandatory allocations (Clegg *et al.* 1989). Next, provisional allocations can be made to the human(s), machine(s), both or neither. In the latter case, a return to developing requirements specifications previously is called for using four groups of criteria: technical feasibility; health and safety;

operational requirements (i.e. physical, informational, performance); and function characteristics (i.e. criticality, unpredictability, psychological). See Kleiner (1998) for a review of macro-ergonomic directions and issues in function allocation.

Technical changes are made to at best, prevent or at worst, control key variances. Human-centered design of the following may be needed to support operators as they attempt to prevent or control key variances: interfaces, information systems to provide feedback; job aids; process control tools; more flexible technology; redesign work station or handling system; or integrating mechanisms.

After considering human-centered system changes in the previous step, it is time to address supporting the human directly by addressing knowledge and/or skill requirements of key variances and any selection issues that may be apparent. In the variance control table, who controls variances and the tasks performed to control these variances were identified. At this stage, it is suggested that personnel system changes to prevent or control key variances be identified. This may entail specific skill or knowledge sets that can be acquired through technical training, formal courses, workshops or distance learning.

At this point in the process, organizational design hypotheses have been generated and iteratively adjusted as new analyses are performed. Next, take the specifications for organizational design levels of complexity, centralization and formalization and produce specific structures. Depending upon the level of work system process analysis, this may require design/redesign at the organizational level or at the group/team level or at both levels. The former level is exemplified by choices between functional structures, product structures, geographic structures or matrix structures while the latter may specify a functional team, cross-functional team, self-managed or high performance work team.

2.8. Phase 8 — Roles and Responsibilities Perceptions

It is important to identify how human operators perceive the roles documented in the variance control table, especially if the table was initially constructed by those who do not occupy the roles identified. Through interviews, role occupants can participate in an analysis of their perceptions of their roles. Using the previously constructed table, expected roles, perceived roles any gaps can be identified. Gaps can be managed through training and selection as well as technological support. In essence, two role networks are operating, the one needed and the one perceived. Any variation between the two can be reduced through participatory ergonomics, training, communication, interface design or tool design.

2.9. Phase 9 — Design/Redesign Support Subsystems and Interfaces

Now that the work process has been analyzed and jointly designed, other internal organizational support subsystems may require redesign. Of particular interest are the management system, and reward system and maintenance system. In all cases, the goal is to determine the extent to whether a given subsystem impacts the socio-technical production system; to determine the nature of the variance; determine the extent to which the variance is controlled; and determine the extent to which tasks should be taken into account in redesign of operating roles in the supporting subsystem units.

According to the Clegg *et al.* (1989) method of function allocation, individual and cumulative allocations made on a provisional basis can be further evaluated against: requirements specifications (including the scenarios developed earlier); resources available at the time of implementation (including human and financial); and the sum total outcome. In addition to a check of function allocation, interfaces among subsystems should be checked and redesigned at this juncture.

Especially at the team and individual levels of work, the internal physical environment should be ergonomically adjusted if necessary to promote effectiveness. Investigating the technical and personnel variance analyses, an assessment can be made about whether there are physical environmental changes that will promote improvement. These changes might include changes to temperature, lighting, humidity, noise control/hearing protection, etc.

2.10. Phase 10 — Implement and Improve

Finally, it is desired to implement the work system changes prescribed, design interfaces and allocate functions. Since in most cases, the ergonomics team will not carry the direct authority to implement the changes suggested by the analysis, proposals with recommendations for change may be required for presentation within the formal organizational structure. Based on the proposal feedback, modifications and an iteration to an earlier analysis or design stage may be necessitated.

Once the proposal for change is accepted and implementation begins, regular reviews of progress are required. To compliment the weekly formative evaluations performed by the implementation team, semi-annual summative evaluations should be performed by an objective outside party.

REFERENCES

CLEGG, C., RAVDEN, S., CORBERTT, M. and JOHNSON, S., 1989, Allocating functions in computer integrated manufacturing: A review and new method. *Behavior and Information Technology*, 8, 175–90.

EMERY, F. E. and TRIST, E. L., 1978, Analytical model for sociotechnical systems. In Pasmore and Sherwood (eds), *Sociotechnical Systems: A Sourcebook* (LaJolla: University Associates), 120–33.

HENDRICK, H.W. and KLEINER, B.M., 1999, *Macroergonomics: An Introduction to Work System Design* (Santa Monica: Human Factors and Ergonomics Society) (in press).

KLEINER, B.M., 1996, Macroergonomics lessons learned from large-scale change efforts industry, government, and academia. In Brown and Hendrick (eds), *Human Factors in Organizational Design and Management* (Amsterdam: North-Holland), 483–8.

KLEINER, B.M., 1998, Macroergonomic directions in function. In Vink and Koningsveld (eds), *Human Factors in Organizational Design and Management VI* (Amsterdam: Elsevier).

KLEINER, B.M., 1997, An integrative framework for measuring and evaluating information management performance. *International Journal of Computers and Industrial Engineering*, 32, 545–55.

KLEINER, B.M. and DRURY, C.G., 1999, Large-scale regional economic development: Macroergonomics in theory and practice. *Human Factors and Ergonomics in Manufacturing* (in press).

SINK, D.S. and TUTTLE, T.C., 1989, *Planning and Measurement in Your Organization of the Future* (Norcross: Industrial Engineering and Management Press).

TAYLOR, J.C. and FELTON, D.F., 1993, *Performance by Design* (Englewood Cliffs: Prentice-Hall).

Socio-technical Theory

R. J. Badham†, C. W. Clegg‡ and T. Wall‡

†Department of Management and BHP Institute for Steel Processing and Products, University of Wollongong, NSW 2522, Australia

‡Institute of Work Psychology, University of Sheffield, Sheffield S10 2TN, UK

1. INTRODUCTION

Socio-technical theory has a long tradition stemming from its explicit theoretical origins in the 1950s. It was initially established in the classical work of the Tavistock Institute in the British coal mining industry, soon to be followed by the work of Rice and colleagues in the textile industry in India, and by Emery Thorsrud and others in the Norwegian Industrial Democracy program. As a general approach to the analysis and design of organizational structures, it has promoted a philosophy of participatory democracy, an open socio-technical systems conceptual framework, and an action research methodology. Its key ideas are now embodied in a wide range of concepts and practices employed in a number of countries by researchers and practitioners working as university academics and researchers as well as external and internal consultants to organizations (van Einjatten 1993, Taylor and Felten 1993).

In the 1950s the Tavistock Institute was involved in a number of projects investigating how "longwall mining" had mechanized and enlarged the scale of coal mining operations, and replaced traditional group-based methods of work by specialized and fragmented manual jobs coordinated and controlled by an external supervisor. The change in work had broken down the previous social structure and led to a catalogue of individual, organizational, social and performance problems. Yet, in a number of sites, the Tavistock researchers discovered that different forms of work organization had been introduced. Increasing mechanization had also been accompanied by the creation of relatively autonomous groups interchanging roles and shifts and regulating their affairs with a minimum of supervision. High levels of personal commitment, low absenteeism, infrequent accidents and high productivity accompanied cooperation between task groups.

The Tavistock researchers, who had a strong background in trade unionism and work psychology, provided a particular interpretation of this experience. They were impressed by the dramatic effects of new technology, concerned about the breakdown of traditional group work, but also excited by the potential of the naturally occurring experiments in new forms of work. The last of these was seen as providing a model for combining higher levels of mechanization with new forms of group cohesion, self-regulation and participation in decisions about work arrangements. For this reason, the book which overviewed the Tavistock mining studies was subtitled "The Loss, Rediscovery and Transformation of a Work Tradition" (Trist 1981).

The commitment of the Tavistock researchers to improving working conditions and enhancing the ability of workers to control their own work organization has remained an enduring part of the socio-technical legacy. This has been reflected in frequent advocacy of autonomous or self-regulating work groups. It has also found expression in recommendations ranging from individual job enrichment to institutionalized co-determination and industrial democracy. The ultimate failure of many of the experiments in the UK coal industry to take hold was seen by some Tavistock researchers as a sign that socio-technical re-design had to address the "extended social field" of forces at a broader societal level. The Norwegian Industrial Democracy program in the 1960s, involving four major socio-technical field experiments in direct worker participation, was given more chance of success as the social setting was seen to be more conducive to successful and enduring outcomes (Trist 1981).

The open socio-technical systems framework that emerged from the Tavistock Institute during the 1950s and 1960s was strongly influenced by von Bertalanffy's (1950) "Open systems in physics and biology." The Tavistock researchers developed the framework in an attempt to reflect on and generalize their ideas about the interdependence of technology and work organization, the existence of choices between different forms of work organization, and their commitment to group self-regulation. The key features of the open socio-technical systems framework were that work or production operations should be seen as:

- *Systems* with interdependent parts.
- *Open* systems adapting to and pursuing goals in external environments.
- Open *socio-technical* systems possessing an internal environment made up of separate but interdependent technical and social subsystems.
- Open socio-technical systems with *equifinality*, i.e. in which system goals can be achieved by different means. This recognized the existence of organizational choice in the type of work organization that could complement any given technology.
- Open socio-technical systems in which performance depends on *jointly optimizing* the technical and social subsystems, i.e. where neither the technical nor social subsystems are optimized at the expense of the other.

The adoption of the open socio-technical systems framework has led to socio-technical theory being characterized as a representative of both *organic* approaches to organizations and *contingency* approaches to organisational design. In contrast to mechanistic models of organizations that underlie formal administrative and Tayloristic approaches, organizations are viewed as *purposeful* systems, with effective organizations possessing a social system aligned to that purpose. In opposition to universalistic theories presenting *one best way* of organisational design, the understanding of organizations as open systems also implies that the optimum socio-technical design is *contingent* on the purposes and environment of the organization.

The open socio-technical systems framework has also provided the basis for more detailed analysis and re-design of systems based on specific characterizations of the technical and social subsystems and the relations between them. At the core of technical analysis and re-design lies a view of technology as more than machines and equipment. The technical subsystem is seen as a conversion process transforming system inputs into outputs. This conversion process proceeds through a number of stages or "unit operations," each of which is subject to unpredictable "variances" (deviations from predicted or required output) which affect completion of the task. These variances, and causes such as machine breakdowns, etc. have to be controlled if system goals are to be achieved. The need to control variances is the central

link point between the technical and the social system, as people within the social system are required to assist in the control of variances. This is summarized in the now classic "variance control chart" (Taylor and Felten 1993).

The social subsystem is regarded as more than a set of technical control tasks that have to be performed by people. These technical tasks are combined into individual jobs and the responsibilities assigned to groups. In any analysis and re-design of the social subsystem, these jobs and, equally importantly, the broader social roles of which they are a part are examined for their implications for the technical subsystem *and* the extent to which they enhance or reduce the quality of working life for the individuals and groups involved in production. It is the commitment to this social analysis and the quality of working life that underlies the celebrated comment in 1975 by one of socio-technical theory's founders, Herbst, "the product of work is people" (Trist 1981, Davis and Taylor (eds) 1979).

These key systems concepts have been differently interpreted by alternative branches of socio-technical theory, and have undergone substantial development since their origin in the 1950s. Among the more recent developments are the integral socio-technical model of de Sitter and The Netherlands school, influenced by the systems theory of Luhmann (van Einjatten 1993, Mathews (ed.) 1997) and the GAIL model employed by a number of US researchers and practitioners (Taylor and Felten 1993). Some socio-technical writers have also sought to broaden the unilinear technical analysis into a broader multidimensional analysis more suitable for knowledge work and services (Barko and Pasmore (eds) 1986).

The socio-technical approach has been described as a *practical paradigm* by van Beinum (van Einjatten 1992). It is well known as a set of re-design principles, partially embedded within the open socio-technical system concepts but also incorporating a set of design heuristics and social values. The classical formulation of socio-technical design principles was articulated by Cherns in 1976, and further refined in 1987 as the following (Cherns 1987):

- *Compatibility* — the process of design should be compatible with the design objectives (i.e., processes should be highly participative).
- *Minimal critical specification* — while objectives should be specified, the means of achieving them should not be.
- *Variance control* — variances should be controlled at source (and should not be exported across boundaries).
- *Boundary control* — boundaries should not be drawn so as to impede sharing of information, knowledge or learning.
- *Information flow* — information should be provided to those who require it when they require it.
- *Power and authority* — those who need equipment, materials, or other resources to carry out their responsibilities should have access to them and authority to command them.
- *The multifunctional principle* — individuals and teams should take on multiple roles to increase their response repertoires.
- *Support congruence* — supporting systems and subsystems need to be congruent. (e.g. planning, payment systems and career systems).
- *Transitional organization* — periods of transition require planning and design, and transitional organizations may be different from the old and the new systems, and are themselves subject to socio-technical design.

- *Incompletion* — re-design is continuous and is the function of self-regulating teams.

Chern's principles provide criteria that can be used to guide the design of individual jobs, group work, technology, work processes, organizational structure and the design process. These principles have been taken up and further developed by socio-technical and other researchers and practitioners. *Individual job design criteria* for promoting worker motivation and performance have been further developed, particularly in Anglo-Saxon and continental European work psychology, and are now a widely accepted component of work re-design (van Einjatten 1993). *Group work principles* have been further extended by work and social psychologists and have been developed in a wide variety of methods for analyzing group work and establishing semi-autonomous, self-regulating or self-managing work teams (Ulich and Weber 1996). Methodologies for *the re-design of work processes* have been further extended in different socio-technical traditions, and have been taken up and developed by both Total Quality Management and Business Process Re-engineering philosophies. (Taylor and Felten 1996). A broader commitment to *organizational re-design* has been emphasized by many branches of socio-technical theory, in particular the Dutch socio-technical school, as a prerequisite for effective and enduring work re-design (Mathews (ed.) 1997). Since Burns and Stalker's classic contrast between "mechanistic" and "organismic" organizational structures in 1961, socio-technical practitioners have promoted different versions of organic structures, finding its most recent expression in advocates of "high performance" and "excellent" organizations (Buchanan and McCalman 1989).

While socio-technical theory has traditionally failed to intervene effectively in or develop adequate design criteria for the *re-design of technology*, understood as machinery and equipment, socio-technical researchers have attempted to remedy this situation by developing criteria for human–machine interface design and the appropriate allocation of functions between technology and people (Grote *et al*. 1995, Clegg *et al*. 1996). Influenced by the rise and diffusion of information technology, socio-technical researchers have collaborated with ergonomists, human–machine interface specialists, and branches of work psychology, to help develop technology design criteria inspired by socio-technical design principles. These have been well represented in the "human centered systems" and "computer-supported cooperative work" movements (Badham (ed.) 1991). Finally, the *process of re-design* has received considerable development. Socio-technical consultants have systematized and refined different stage models for socio-technical re-design, extending Emery's original nine-step method, with substantially different models regularly applied by US socio-technical consultants, and the Dutch and Swiss socio-technical schools (van Eijnatten 1993, Grote *et al*. 1995). There has also been substantial development in attempts to provide more effective methods for participative involvement in design. This has been strongly influenced by Emery's participative design approach, involving such techniques as "deep slice" re-design groups and "search" conferences. These have been developed into systematic re-design packages by socio-technical practitioners in the US and elsewhere (Greenbaum and Kyng (eds) 1991).

As with any influential theoretical tradition, socio-technical theory has a number of in-built ambiguities and tensions, giving rise to different paths of development, and sometimes differences

of opinion between advocates. Socio-technical theorists have differed widely in their degree of commitment to a systems perspective, the systems models that they deploy and, particularly, in the degree to which they have emphasized the contingency and choice implied by the open systems model. While socio-technical theory has been widely associated with a universalistic prescription for highly autonomous and self-regulating group work, a number of socio-technical theorists have emphasized that this prescription contradicts the principles of system openness, organizational choice and equifinality. The nature and level of participation advocated by different socio-technical theorists and practitioners has also varied widely. While some emphasize participation in work re-design, others are far more committed to industrial democracy and broader changes in corporate governance and societal structures. For some practitioners, socio-technical interventions are strongly client driven and expert assisted processes while, for others, it is a far more participatory process driven and "owned" by the direct users or recipients of its results (Greenbaum and Kyng (eds) 1991, van Eijnatten 1993). Finally, alternative views on action research have informed different socio-technical theorists and practitioners. For some, this is a question of the type of action, with answers ranging from more expert driven views of "action science" as a test of knowledge, to more participatory views of research as a tool for promoting the interests of disadvantaged groups in the workforce. For others, the central concern is the degree to which socio-technical theory has managed to combine research excellence with practical effectiveness. Different critics have condemned either an observed decline in in-depth research on the actual nature of work practices or an increasing academicization of the discipline at the cost of effective interventions in organizational change processes.

Socio-technical theory is now in a somewhat anomalous state. On the one hand, it has been subjected to severe criticism on a number of grounds (Barko and Pasmore (eds) 1986, Mathews (ed.) 1997). These include: an excessive focus on shopfloor work re-design and neglect of broader organizational and strategic change; an over-preoccupation with social and technical issues and lack of attention to hard economic issues and realities; a general failure to establish enduring successful exemplars and long term proven results; an inability to effectively address and intervene in technological change despite its proclaimed technical as well as social orientation; a perceived lack of impact on the development, implementation and use of the vast array of new information and communication technologies which are revolutionizing (or have the potential to revolutionize) many organizations; a simplistic identification of autonomous work groups as a universal solution for work re-design; and a general neglect of power and politics in its concepts and practices. It has also failed to develop the kind of cult managerial following and sets of tools characteristic of total quality management and business process re-engineering movements, though some might express relief at this. There is also, arguably, an excessive degree of fragmentation and dispute amongst its supporters, with socio-technical advocates often preferring to classify their work in different terms as, variously, self-managing or self-directed work team philosophies high performance work system design, participative design, self-design or rapid design (US and Australia), integral renewal (The Netherlands) or MTO (Mensch Technik

Organization) (Switzerland). A recent attempt to reinvigorate socio-technical theory as an international STS2, network has so far not succeeded in overcoming this self-admitted fragmentation.

On the other hand, it can be argued that many of socio-technical theory's key ideas and concepts have been widely diffused, so much so that they have become part of the conventional wisdom of many disciplines, consultants and managers. In the face of rapidly changing markets and innovations, there is now a far greater commitment to organizational development and flexibility, teamwork, organizational decentralization, continuous innovation and learning. Arguably, at least in the US, The Netherlands and Switzerland, socio-technical practitioners have been influential in explicitly promoting socio-technical ideas and concepts in industry as well as academia.

It is clear that socio-technical researchers and practitioners have not succeeded in establishing themselves as an intellectual discipline or as a profession and set of institutionalized practices with authority over socio-technical activities in organizations. There would not even be common agreement on whether such a goal would be desirable. Yet, the central ideas of the socio-technical tradition have an enduring relevance, in large part because they continue to be honored in theory and to some extent in practice. The challenge for both academic disciplines and organizations is to work further on overcoming the technology-social divide, increase worker or user participation in change, and take advantage of technical and organizational choices to promote increased productivity and the humanization of work. In addressing these issues, some heed should be given to the recent verbal reflections of Lou Davis at the 1997 HCI Conference on his past work as one of the founders of socio-technical theory in the US, "Our main mistake was that we under-estimated the role of power."

REFERENCES

BADHAM, R. (ed.), 1991, Special issue: work, organization and culture. *AI and Society: Journal of Human Centred Systems and Machine Intelligence*, 5.

BARKO, W. and PASMORE, W. (eds), 1986, Special issue: sociotechnical systems. *Journal of Applied Behavioural Science*, 22.

BUCHANAN, D. and McCALMAN, J., 1989, *High Performance Work Systems: The Digital Experience* (London: Routledge).

CHERNS, A., 1987, Principles of socio-technical design revisited. *Human Relations*, 40, 153–62.

CLEGG, C.W., COLEMAN, P., HORNBY, P., MacLAREN, R., ROBSON, J. CAREY, N. and SYMON, G., 1996, Tools to incorporate some psychological and organizational issues during the development of computer-based systems. *Ergonomics*, 39, 482–511.

DAVIS, L.E. and TAYLOR, J.C. (eds), 1979, *Design of Jobs* (Santa Monica: Goodyear).

EIJNATTEN, F. VAN, 1993, *The Paradigm that Changed the Workplace* (Stockholm, Arbetslivcentrum).

GREENBAUM, J. and KYNG, M. (eds), 1991, *Design at Work: Cooperative Design of Computer Systems* (Hillsdale: Lawrence Erlbaum Associates).

GROTE, G., WEIK, S., WAEFLER, T. and ZOELCH, M., 1995, Criteria for the complementary allocation of functions in automated work systems and their use in simultaneous engineering projects. *International Journal of Industrial Ergonomics*, 16, 367–82.

MATHEWS, J. (ed.), 1997, Special Issue: Sociotechnical Theory. *Human Relations*, 50.

TAYLOR, J. and FELTEN, D., 1993, *Performance by Design: Socio-technical Systems in North America* (Englewood Cliffs: Prentice Hall).

TRIST, E., 1981, *The Evolution of Sociotechnical Systems: A Conceptual Framework and an Action Research Program* (Ontario: Ontario Quality of Working Life Centre).

ULICH, E. and WEBER, W.G., 1996, Dimensions, criteria and evaluation of work group autonomy. In M.A. West (ed.) *Handbook of Work Group Psychology* (Chichester: Wiley).

VON BERTALANFFY, L., 1950, The theory of open systems in physics and biology. *Science*, 3, 1–10.

Systems Approach to Training

K. Henriksen, R. C. Glennon and R. E. Jones, Jr

Raytheon Systems Co., 1070 Parkridge Boulevard, Suite 200, Reston, Virginia 20191 USA

1. INTRODUCTION

In terms of its origin, the systems approach to training (SAT) is an adaptation of the systems engineering process, a methodology that rapidly evolved during World War II for the efficient development and deployment of new equipment and weapons systems. SAT is best viewed as a subsystem of a larger operational system. The larger operational system imposes certain requirements on the personnel who are charged with operating, maintaining and managing the system. The training subsystem is responsible for ensuring that assigned personnel can satisfy their job requirements, employing what has become known as SAT. Decisions regarding training development are, thus, based on an analysis of system requirements and subsystem relationships. Since the training subsystem is frequently under development at the same time as the primary operational system, practical considerations regarding schedule, resources and costs have significant influence. With the build-up of the US defense industry during the 1950s and 1960s, the requirement for an efficient and cost-effective way to train large numbers of recruits who became the operators and maintainers of the new systems was apparent.

An early conceptualization of the systems approach to training is generally credited to Miller (1953) who presciently took into account such present-day concerns as training device design for individuals and teams, task allocation, task procedures design, human factors engineering, job aids, behavioral analysis and research, performance evaluations, personnel selection and collective training. A discussion of some of the concurrent educational influences to the systems approach to training – programmed instruction, task analysis, the behavioral objectives movement, self-paced instruction, mastery learning, individualized instruction, criterion-referenced testing – can be found in Risner (1987). By 1976, a five-volume set of prescriptive procedures entitled *Interservice Procedures for Instructional Systems Development*, developed at Florida State University, was introduced for use by the armed forces whereby the outputs of one phase served as inputs to the next phase (Bronson *et al.* 1976). Most practitioners use the terms "systems approach to training" and "instructional systems development" interchangeably (as will be the tendency in the present paper) since both provide an orderly and systematic approach to developing instructional materials by integrating the processes of analysis, design, development, implementation and evaluation. The major distinction between the two models appears to be the point of entry to the analysis phase. Analysis in the SAT model will most likely start at the broader mission and collective task level, while with the Instructional Systems Development (ISD) model, it typically starts at the individual task level.

Rather than signaling a lack of acceptance, the fact that various versions of the SAT process have been adopted by the military services, by private industry and by a broad cross-section of government agencies attests to its versatility and robustness. Although the process has a behavioral flavor with its emphasis on observable outcomes, it is derived from neither educational theory nor pedagogic method (Meister 1985). It is simply a very pragmatic approach for ensuring the job-relevance of training materials in an efficient and cost-effective manner. When first introduced, it represented something akin to a paradigm shift in the way knowledge is presented. Rather than starting with generalized scholarly knowledge which traditionally gets organized into a textbook and is relayed to students by teachers under the assumption that knowledge for its own sake is good, the SAT approach starts with an organizational need, for which training has been identified as a viable solution, and then proceeds to deliver a body of tailored instruction to designated students who have a job requirement for the specified skill or knowledge (Branson and Grow 1987). Once the SAT creators formalized the steps of the process in the form of a set of guidebook procedures, it was assumed that lay personnel who were placed by their services in the role of trainers could develop the training materials at relatively low cost. Although subject-matter experts are utilized for their technical input, they typically do not develop nor deliver the instruction.

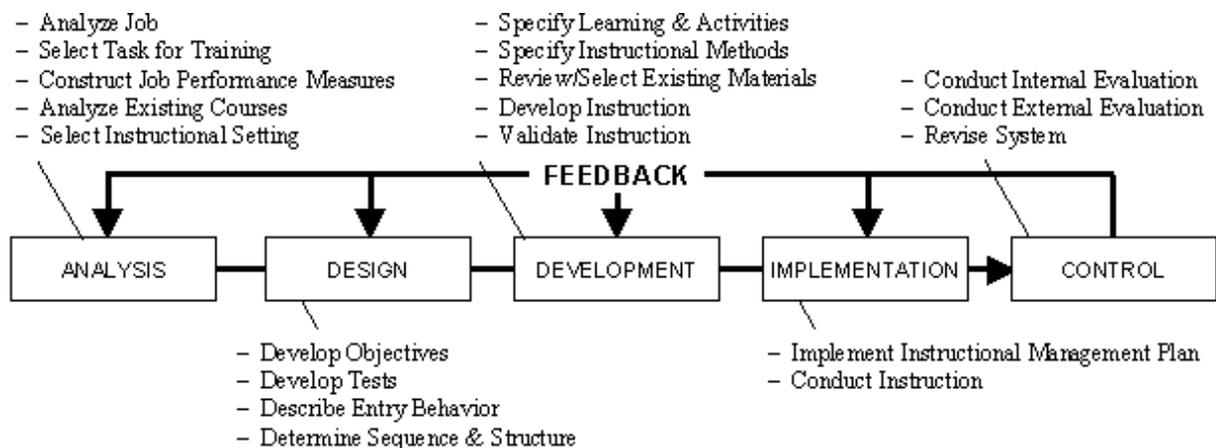Figure 1 shows the earlier SAT model that was adopted by

Figure 1. **Phases and steps of the systems approach to training.**

the US armed forces. It has served as a prototype of the numerous versions spawned during the past 25 years. Each of the five phases of the model is discussed in turn.

2. ANALYSIS

The Analysis phase provides the foundation upon which successful materials are based. To ensure that instructional materials are based on job requirements, a job/task analysis is performed. In brief, a job/task analysis establishes what constitutes adequate on-the-job performance. It answers the questions: what tasks, performed in what manner, under what conditions, in response to what cues, to what standards of performance, make up the job? Not all tasks uncovered during the job analysis are candidates for training. Only those tasks and functions that job holders need to know (and do not already know) for the successful performance of their jobs are earmarked for training. Tasks that are performed infrequently or information that is "nice to know" but not essential is not included. To collect this information, three data collection methods are typically used: review of existing documentation (e.g. job descriptions, operator manuals); structured interviews with job incumbents, supervisors and subject-matter experts; and, where possible, observation of on-the-job performance. Job performance measures that serve as indicators of successful task performance are developed next. Each task selected for training has a job performance measure associated with it. Since the development of training materials is both time-consuming and labor-intensive, careful review of existing training materials and documentation helps the training developer avoid duplication of effort and may provide insight into what has worked well with similar tasks in similar operations. With respect to the instructional setting, ideally training is conducted in a setting that will maximize the student's ability to acquire and retain the instruction; however, logistical or administrative considerations frequently influence the venue for training.

3. DESIGN

The development of training objectives is the first step of the Design phase. Subsequent steps in the SAT process depend heavily on carefully prepared objectives. If the training objectives are poorly defined or couched in terms too vague, the instructional program will lack focus. Foremost, a well-stated objective specifies the desired outcomes in observable and measurable terms. It identifies what the student is to do, the conditions under which the task is performed, and the standard of performance that must be achieved. Such objectives aid the instructional developer by specifying the exact performances and knowledge the training is expected to produce. When desired outcomes are specified in observable and measurable terms, it becomes readily apparent when the instructional methods have succeeded, when they have failed, and when they need further development to increase their effectiveness. Criterion test items provide students with immediate confirmation about whether they are performing the specified terminal behavior called for by the course objectives. These tests are based on job performance measures that represent the best approximations of task performance, given the practical considerations of ease of administration and acceptable fidelity. Test items can be used as pretests for placing students at an appropriate instructional level, as within-course progress tests and for confirmation of mastery of objectives at the end of an instructional unit. To describe entry behavior, an entry-level test is devised and administered to assess initial assumptions about what learning objectives students know and do not know. Where the learning of an objective is dependent upon the prior learning of another objective, careful consideration is given to the way learning objectives are sequenced and organized in groupings.

4. DEVELOPMENT

The Development phase starts with a classification of learning objectives by learning category and subcategory to identify learning guidelines necessary for optimal learning to take place. A media selection process determines how the instruction is to be presented, taking into account learning categories, media characteristics, training setting criteria and costs. Compared with other decisions, the selection of media has a considerable impact on resources, development schedule and costs. Instructional management plans for the appropriate management and allocation of resources are developed. Existing materials are analyzed for appropriateness and new materials are developed for learning objectives not covered by existing materials. Specific development activities, of course, depend upon the media under development. Each medium has its own unique features and trade-offs that need to be considered. For example, multiple paths through lesson modules are one of the features that set interactive computer-based forms of instruction apart from more traditional media. Whenever multiple paths are open to students, the number of sequences that development personnel must undertake to develop increases considerably. It is, thus, necessary to arrive at a reasonable balance between all the possible paths that a developer might advocate for adding interest and individuality to the courseware and what is actually possible given the very real constraints of time, resources and budget. Once the materials have been developed, they undergo validation or formative evaluation where small-group tryouts are conducted with representative users for the sole purpose of improving the quality of the instructional materials.

5. IMPLEMENTATION

The Implementation phase affords the opportunity to observe and evaluate the instructional materials under actual training conditions. An implementation plan is developed which includes all facets of training management and administration, personnel roles and responsibilities, training system description, initialization procedures and schedule, equipment storage and utilization, student grading and evaluation, training materials management, facilities utilization, training of instructors and contingency planning. Sometimes called operational tryouts, the first implementation of the instruction provides an opportunity to evaluate procedures for the combined and integrated functioning of instructional materials, instructors (if required), actual students, equipment and facilities. Feedback from this larger and more representative sample in the job or training setting allows some of the less obvious problems to be uncovered, thus ensuring a more comprehensive quality control process.

6. CONTROL

The last phase, Control, refers to evaluation and revision of the instructional materials. The training materials may have been successfully implemented, but how well do they hold up across

repeated offerings of the program? Internal evaluation addresses the effectiveness of the instructional materials in terms of the extent to which students can master the training objectives. Is it more difficult to reach criterion-level performance on some of the units of instruction compared with others? External evaluation, on the other hand, assesses the extent to which the training program prepares students for on-the-job performance. Despite the care and resources that go into developing instructional materials, it is possible that graduates fall short of satisfying on-the-job requirements. Sometimes on-the-job requirements change in a short period of time. If graduates of the program do not need what they were taught, or need instruction they did not receive, this information needs to be fed back to the maintainers of the training system. Because of the methodological difficulties of experimentally demonstrating transfer of skills and knowledge to the job setting, most external evaluations, when conducted, rely on interview and questionnaire techniques.

7. ISSUES AND CHALLENGES

7.1. Isolation of Critical Steps

Several issues deserve comment. Any practitioner of the SAT process will quickly point out the distinction between the ideal and real. Because the SAT process was originally developed for use by individuals who were not trained neither in the subject-matter nor in the development of instruction, it was highly prescriptive in nature. Since training development is a labor-intensive activity, training developers are quick to bypass unnecessary steps. Organizational needs with respect to training also vary considerably. In short, plenty of adaptation and reinvention occur. When different organizations and federal agencies codify these adaptations and produce their own SAT models, one realizes there is no universal model that is applied unaltered across organizational settings. This gives rise to the question about what steps are of limited value and perhaps better treated as optional and what steps are essential or without which the integrity of the model would suffer to the extent to render its SAT character unidentifiable. Despite the emphasis placed on iterative evaluation in the SAT process, there is little in the way of research that attempts to isolate what the critical variables are and show their relationship to measures of effectiveness.

7.2. Importance of Front–end Analysis

While there might not be empirical support for the necessity of certain processes, training researchers and practitioners alike agree on the importance of the front-end analysis phase. Not everybody understands the importance of a thorough job/task analysis. Sometimes clients do the job/task analysis in-house and hand it off to an external training developer to do the design and development work. The job/task analysis is the foundation upon which training materials are developed. It can require an extensive effort. If it is not done completely and correctly at the outset, a careful following of the remaining steps of the model will be all for naught. An incorrect or incomplete front-end analysis should be redone immediately before design and development activities are undertaken. The penalties for not doing so include significant rework, slippage in the schedule and budget overruns.

7.3. Cognitive Task Analysis Procedures

While the SAT model has been applied successfully to jobs where most of the tasks are procedural and performed in a stable envi-

ronment, the application of the model to more complex tasks and dynamic environments, where the role of the individual is more of a problem solver or decision maker, is less convincing. In addition to performing operations, today's complex automated systems call upon operators to monitor, interpret and intervene in the case of abnormal system states. One of the ironies of automation, as observed by Bainbridge (1987), is that the most successful automated systems with the rarest need for manual intervention require the greatest investment in operator training. The correction of unusual abnormal system states depends on more than simple "how-to-do-it" procedural sequences. Cognitive task analysis procedures have emerged during the past couple of decades which look at skilled performance as a composite of (1) declarative or domain knowledge (knowledge of the system or device), (2) procedural (how-to-do-it) knowledge, and (3) strategic (how-to-decide-what-to-do-and-when) knowledge (Gott 1989). While experts can coordinate these multiple types of knowledge quite efficiently, novice operators are likely to experience gaps in declarative knowledge structures and in knowing when to deploy specific pieces of knowledge. As a way of capturing the tacit knowledge of experts and the cognitive processes involved in the deployment of that knowledge, Hall *et al*. (1990) developed the PARI (Precursor, Action, Result and Interpretation) methodology. PARI is a verbal protocol approach for eliciting knowledge from technical experts working in complex problem-solving domains (e.g. avionics troubleshooting). By focusing on the knowledge structures and cognitive processes that underlie skilled observable performance, the approach is considered an extension of the earlier behavioral procedures. While cognitive task analysis techniques have commanded the attention of a large body of cognitive scientists, they also are intended as practical, analytic tools that can be used by non-scientists for training development (Hall *et al*. 1990). The new century, no doubt, will witness continued progress along these lines.

7.4. Opportunities for Innovation

Training managers and developers sometimes treat the SAT model as a panacea for developing effective instruction. Effective instruction depends on more than simply following the steps of the SAT process. By attending to form over substance, training managers may gain a false sense of security, yet be unaware of innovative techniques that can be employed for enhancing student motivation and involvement with the training materials. There is nothing about the SAT process that says instruction has to be repetitious and tedious, yet much of it is. At the same time, there is not much guidance in the SAT steps on how to develop truly engaging instruction. Sometimes organizations invest heavily in a given instructional technology (e.g. web-based training, video teleconferencing, a new electronic performance support system) and then see every training problem as requiring that particular technology. There is more than an element of truth in the old adage about giving someone a hammer and every problem becomes a nail that requires pounding. Contrary to feeling constrained by the SAT process or any given technology, training developers have the opportunity to give free vent to their imaginations and more fully to exploit the unique instructional qualities that each medium provides. Recent years have seen the development of new instructional technologies, methodologies and tools

for providing expert advice, for interacting with students on a tutorial basis, and for allowing students to explore problem domains where they can constructively enhance their knowledge structures for responding to situations not previously encountered. These developments have the potential to make instruction more responsive to student needs and more manageable. Starting with the SAT process as a useful point of departure, the challenge for training efforts in the new century will be both conceptual and pragmatic: to explore new ways of thinking about enhancing on-the-job performance and to seek new ways of maximizing the instructional effectiveness of the ever increasing array of training technology.

REFERENCES

BAINBRIDGE, L., 1987, The ironies of automation. In J. Rasmussen, K. Duncan and J. Leplat (eds) *New Technology and Human Error* (London: Wiley), pp. 271–83.

BRANSON, R.K., RAYNER, G.T., COX, J.L., FURMAN, J.P., KING, F.J. and HANNUM, W.H., 1976, *Interservice Procedures for Instructional Systems Development* (5 vols). TRADOC Pam 350-30 and NAVEDTRA 106A (Fort Monroe: US Training and Doctrine Command).

BRANSON, R.K. and GROW, G., 1987, Instructional systems development. In R. M. Gagne (ed.) *Instructional Technology: Foundations* (Hillsdale: Lawrence Erlbaum), pp. 397–428.

GOTT, S.P., 1989, Apprenticeship instruction for real-world tasks: the coordination of procedures, mental models and strategies. In E. Z. Rothkopf (ed.) *Review of Research in Education* (vol. 15) (Washington, DC: American Education Research Association), pp. 97–169.

HALL, E.P., GOTT, S.P. and POKORNY, R.A., 1995, *A Procedural Guide to Cognitive Task Analysis: The PARI Methodology*. AL/HR-TR-1995-0108 (San Antonio, TX: Armstrong Lab Brooks AFB, Human Resources Directorate).

MEISTER, D., 1985, *Behavioral Analysis and Measurement Methods* (New York: John Wiley).

MILLER, R.B., 1953, *Handbook on Training and Training Equipment Design*. Technical Report 53-136 (Dayton, OH: Wright-Patterson AFB, Wright Air Development Center).

RISNER, R.A., 1987, Instructional technology: a history. In R. M. Gagne (ed.) *Instructional Technology: Foundations* (Hillsdale: Lawrence Erlbaum), pp. 11–48.

Task Analytic Methodology for Design of an Aircraft Inspection Training Program

G. M. Nickles and A. K. Gramopadhye

Department of Industrial Engineering, Clemson University, Clemson University, Clemson, South Carolina, USA

1. INTRODUCTION

Aircraft inspection and maintenance is a complex system (Drury *et al.* 1990, Drury 1991) with many interrelated human and machine components. The linchpin of this system, however, is the human. Recognizing this, the Federal Aviation Administration (FAA), under the auspices of the National Plan for Aviation Human Factors, has pursued human factors research. In the maintenance area this research had focused on the aircraft maintenance technician (AMT). Since it is difficult to eliminate errors altogether, continuing emphasis must be placed on developing interventions to make inspection and maintenance more reliable and/or more error tolerant. Inspection is affected by a variety of geographically dispersed entities. These entities include large international carriers, regional and commuter airlines, repair and maintenance facilities, as well as the fixed-based operators associated with general aviation. An effective inspection is seen as a necessary prerequisite to public safety, so both inspection and maintenance procedures are regulated by the US federal government via the FAA. Investigators conducting this study found that while adherence to inspection procedures and protocols is relatively easy to monitor, tracking the efficacy of these procedures is not.

2. AIRCRAFT MAINTENANCE PROCESS

2.1 Maintenance Procedures

The maintenance process begins when a team that includes representatives from the FAA, aircraft manufacturers, and start-up operators schedule the maintenance for a particular aircraft. These schedules may be, and often are, later modified by individual carriers to suit their own scheduling requirements. These maintenance schedules are comprised of a variety of checks that must be conducted at various intervals. Such checks or inspections include flight line checks, overnight checks, and four different inspections of increasing thoroughness, A, B, C, and D; D is the most thorough and most time-consuming. In each of these inspections, the inspector checks both the routine and nonroutine maintenance of the aircraft. If a defect is discovered during one of these inspections, the necessary repairs are scheduled. Following these inspections, maintenance is scheduled to (1) repair known problems, (2) replace items because the prescribed amount of air time, number of cycles, or calendar time has elapsed, (3) repair previously documented defects (e.g., reports logged by pilot and crew, line inspection, or items deferred from previous maintenance), and (4) perform the scheduled repairs (those scheduled by the team including the FAA representatives).

2.2 Importance of Aircraft Inspection

In the context of an aging fleet, inspection takes an increasingly vital role. Scheduled repairs to an older fleet account for only 30% of all maintenance compared with the 60-80% in a newer fleet. This difference can be attributed to the increase in the number of age-related defects (Drury 1991). In such an environment the importance of inspection cannot be overemphasized. It is critical to perform these visual inspections effectively, efficiently, and consistently over time. Moreover, 90% of all inspection in aircraft maintenance is visual in nature and is conducted by inspectors, so inspector reliability is fundamental to an effective inspection.

2.3 Ergonomic Improvements in Aircraft Inspection

An analysis of the inspector's role in inspection has pointed to a number of issues, e.g., inspector-oriented issues, environmental design issues, workplace design issues (Drury *et al.* 1990, Lock and Strutt 1985). These issues have been continually addressed by the FAA as part of the National Plan for Human Factors in Aviation. Research conducted under this program has identified several ergonomic changes to the system and the inspector. System changes have included improved work control cards (Patel *et al.* 1994) and crew resource management interventions (Stelly and Taylor 1992). Inspector-oriented interventions are (1) selection and (2) training. Although individual differences in simulated aircraft inspection have been found (Thackray 1992), the correlations are typically too small to use as a basis for personnel selection. This article concentrates on training as an improvement strategy. Although aircraft inspection is used as a domain, some of the issues have broader applicability in training for industrial inspection.

3. OBJECTIVE: TRAINING FOR AIRCRAFT INSPECTION

As in any system that is highly dependent on human performance, efforts made to reduce human errors by identifying human-system mismatches can have an impact on the overall effectiveness and the efficiency of the system. Given the backdrop of the inspection system, the objective of this study was to use training as an intervention strategy to reduce inspection errors.

This article documents a task analytic methodology used in the development of a computer-based inspector training program, the automated system of self-paced instruction for specialized training (ASSIST) for aircraft inspectors. The ASSIST program is described in another article. The development of the inspection training program followed the classic training program development methodology. It began with a thorough analysis of the requirements and needs (goals) of the training program.

The next step was to establish the training group and identify the trainers and participants who would be involved. Following this, a detailed task analysis of the job was conducted to determine knowledge, skills, and abilities necessary for the job in order to specify the behavioral objectives of the training program. These objectives then formed the basis for evaluating the training program. After this, the next step was to define criteria measures against which inspectors will be trained and their performance measured, to meet the quality goals.

The abilities of the incoming trainees were compared to

requirements imposed by the task to determine gaps, and hence define the contents of a training program that will help close this gap and meet the defined criteria. At this stage, the appropriate training delivery system (i.e., the instructional technique), such as tutoring, on-the-job computer-aided instruction, had to be chosen. Once the training system is designed and developed, it goes through an evaluation to determine whether the program has met the ultimate goals.

The designer must choose criteria to be used for evaluation, identify a method and protocol for collecting evaluation data, and analyze the data to form conclusions about the training program's effectiveness. Figure 1 also shows the organizational inputs and objectives or goals at each step of the training program development phase. Up to this point, the steps in describing the training program have been sequential. But in reality many of them are iterative, with the designer fine-tuning the training program based on outputs obtained from later stages (as illustrated in Figure 1).

Thus training program design means translating content into a training program. Besides its content, the program is composed of the training methods, the training delivery systems, and the trainees. In the airline industry there is strict regulatory control over the trainees for inspection, as they are required to be licensed mechanics first. Hence the trainee is a relatively fixed input to the system.

4. DATA COLLECTION

The first stage of the assessment was to collect detailed information about the inspection process. A variety of data collection techniques were used to collect not only the process and procedural information and the idealized way of completing inspection tasks, but also information about the way tasks actually get accomplished. Here are the main information sources for collectiing task-descriptive data:

- *Observation and shadowing*: Data was collected by observing aircraft inspection and maintenance operations at various sites ranging from large international carriers to start-up and regional operators of general aviation. This involved watching various inspectors and maintenance technicians accomplishing various tasks over several shifts.
- *Interviewing*: Personal interviews conducted with inspectors, supervisors, aircraft maintenance technicians, lead mechanics, foremen, managers, planners, and other personnel associated with aircraft maintenance, were used
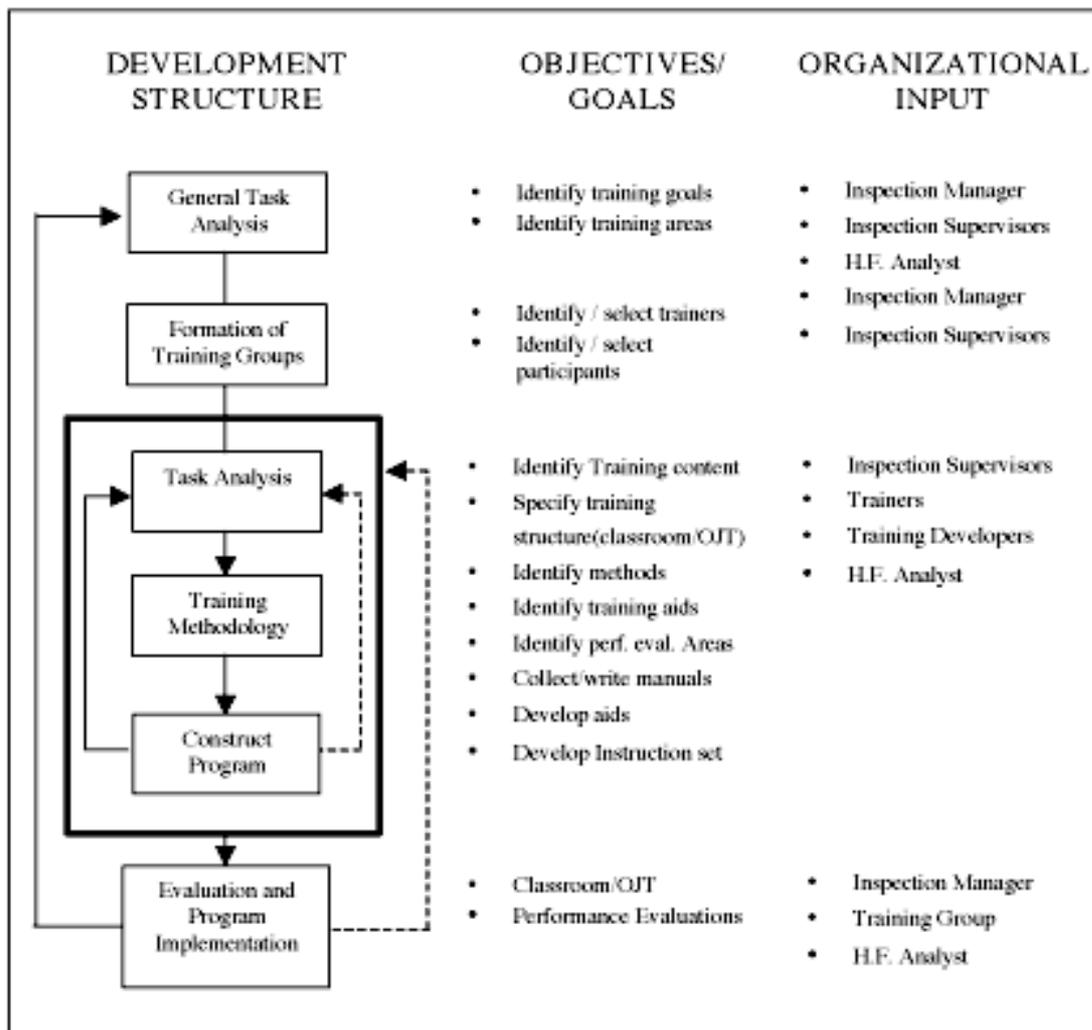
Figure 1. Model for training program development in commercial aviation

to collect data on aircraft maintenance tasks. Interviews with system participants at all levels helped investigators to collect data on the structure and functioning of the system as well as to collect data on the rare events such as system errors. This process was used to identify not only the prescribed way of working on the task but also the "quick and dirty" way that those tasks often get completed in reality.

- *Documentation*: Information on aircraft inspection and maintenance procedures was obtained through company-wide procedures, FAA mandated procedures (Federal Aviation Regulations, FARs), airworthiness directives, aircraft manufacturers' manuals and other documents.

Information from these sources was used to obtain a basic understanding of the tasks involved and to identify problem areas.

5. TASK DESCRIPTION

The information on the inspection process was represented in various formats: (1) an inspection flowchart, (2) a task description based on generic inspection, and (3) a task description form. Here are more detailed descriptions of each of these formats.

5.1 Inspection Flowchart

As a first step, a flowchart of inspection and maintenance activities was developed to illustrate the relationship between inspection and maintenance activities and the relationships among the personnel involved (Figure 2).
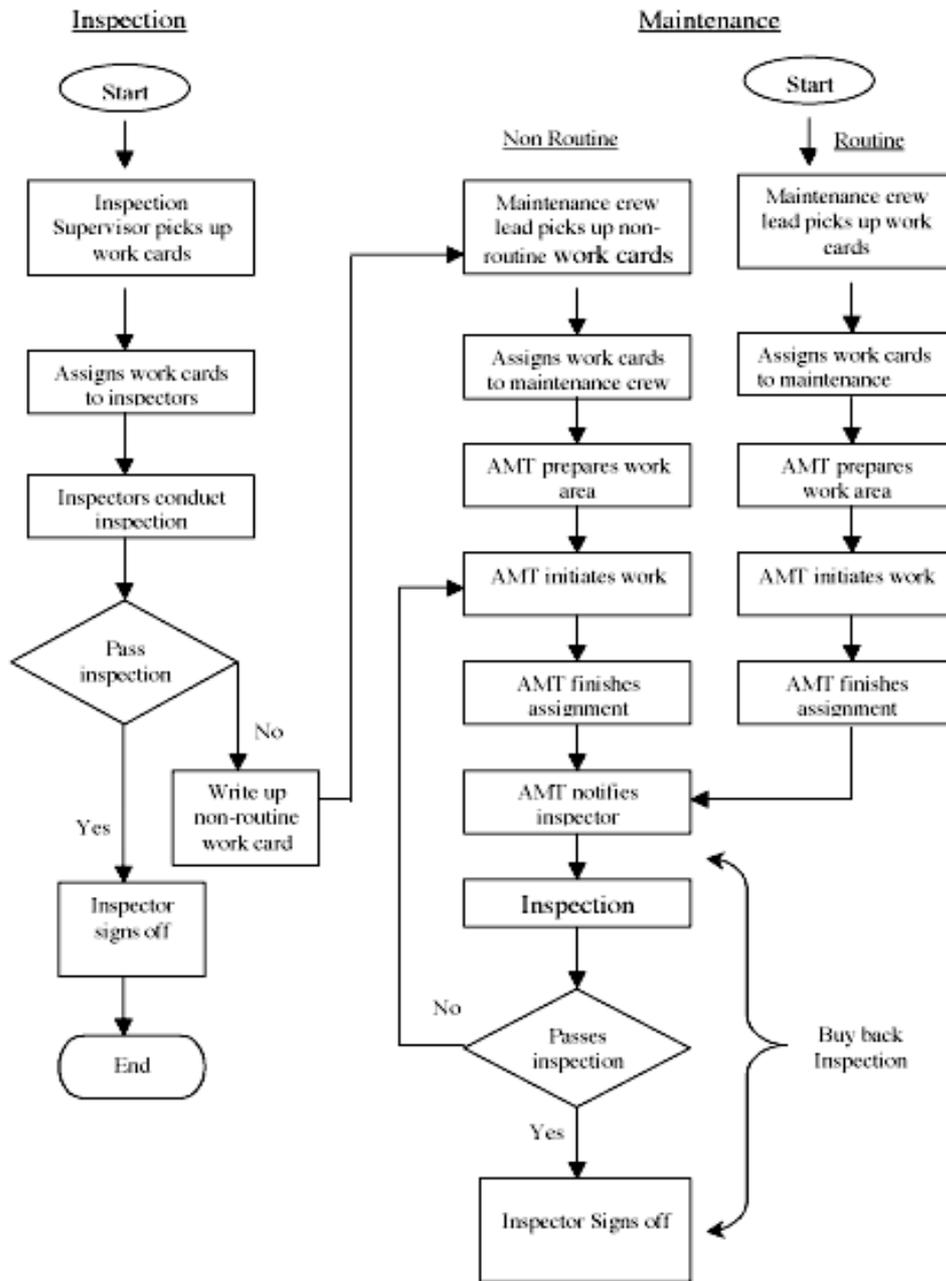
Figure 2. Aircraft inspection flowchart

5.2 Task description Based on Generic Inspection Activities

Although general task description approaches are widely available (Drury 1983; Kirwan and Ainsworth 1992), investigators in this case decided it would be advantageous to use an approach directly related to inspection. Literature describing human factors in inspections has produced the following generic list of inspection activities (Sinclair 1984): (1) present item to inspector, (2) search for flaws, (3) decide on rejection or acceptance of each flaw, and (4) respond by taking necessary action.

Table 1. Detailed breakdown of aircraft visual inspection by task step (Drury 1991)

Task Description	Visual Example	Task Step
1. Initiate	Get workcard. Read and understand area to be covered	1.1 Correct instructions written 1.2 ….
2. Access	Locate area on aircraft. Get into correct position	2.1 Locate area to 2.2 ….
3. Search	Move eyes across area systematically. Stop if any indication	3.1 Move to next lobe 3.2 …
4. Decision making	Examine indication against remembered standards (e.g. for dishing or corrosion)	4.1 Interpret indication 4.2 Access comparison standard 4.3 Access measuring equipment 4.4 Decide if it is a fault 4.5 Decide on action 4.6 Remember decision/action
5. Respond	Mark defect. Write up repair sheet or if no defect, return to search	5.1 Mark fault on aircraft 5.2 …
6. Repair	Repair area (drill out and repair rivet)	6.1 Repair fault
7. Buyback inspect	Visually inspect marked area	7.1 Initiate 7.2 ….

Note that not all steps may be required for all inspections; for example, some inspections may not necessitate search (e.g., color matching) and some may not necessitate decision making (e.g., absence of a rivet head on a lap splice). The unit of task description was the workcard (instructions outlining the specific inspection activity). A task was seen as continuing until a repair was completed and had passed airworthiness requirements. The workcard was the unit of work assigned to one particular inspector on one physical assignment, but the actual time required to complete work shown on a workcard is not consistent from workcard to workcard. Typically, inspectors were expected to complete work outlined by the workcard within one shift, but work might have to be continued across shifts. The work listed on each workcard could potentially require several inspections. Detailed task descriptions were obtained through site visits to various aircraft maintenance facilities (Table 1)

5.3 Task Description Form

Data collected through interviews and site visits was transcribed using a standard working form (Table 2), with a separate form for each of the five steps (initiate, access, search, decision making, and respond) in the generic task description. During the data collection phase, the human factors analysts remained with the inspectors, asking probing questions while these inspectors were actually at work. Following this step, human factors models of inspection performance and the functioning of individual subsystems were used to identify specific subsystems and potential human-system mismatches under the observations column in Table 2.

6. TASK ANALYSIS

Following this step, a detailed taxonomy of errors was developed from the failure modes of each task in aircraft inspection (table 3). This taxonomy, based on the failure modes and effects analysis (FMEA) approach, was developed due to the realization that a

Table 2. Task description form (Drury 1991)

Task descriptions	A	S	P	D	M	C	F	P	O	Observations
				Subsystems						
Search										
1.0 Inspect slat structure, wiring and installation										
1.1 Check wear on male duct of telescope shaft duct										
1.1.1 Check wear by moving it			X		X					• No prescribed force
1.2 Inspect slat well area for corrosion and cracks							X			• No standards for wear
1.2.1 Hold flashlight such that light falls perpendicular to the surface							X			• Holding flashlight for a long period at odd positions—strenuous
1.2.2 visually look for cracks and corrosion			X	X	X					
1.2.3 The visual indication confirmed by tactile			X	X	X					• Lack of information on type of cracks and figures
1.3 Look for play in slat actuator nut										

Attention: number of time-shared tasks   **Memory:** SSTS, working, long-term   **Senses:** visual, tactile, auditory
Perception: **Feedback:** quality, amount, timing   **Decision:** sensitivity, criterion, timing
Control: continuous, discrete   **Posture:** reading, forces, balances, extreme angles

Table 3. Task and error taxonomy for visual inspection (Drury 1991)

Task	Errors	Outcome
4.1 Interpret indication	Classify as wrong fault type	All indications located are correctly classified
4.2 Access comparison standard	Choose wrong comparison standards Comparison standard not available Comparison standard not correct Comparison incomplete Does not use comparison standard	correctly labeled as fault or no fault, and actions correctly planned for each indication
4.3 Decide if it is a fault	Type I error, false alarm Type II error, missed fault	
4.4 Decide on action	Choose wrong action Second opinion if not needed No second opinion if needed Call for buyback when not required Fail to call for required buyback	
4.5 Remember decision/action	Forget decision/action Fail to record decision/action	

Figure 3. Components of the ASSIST aircraft inspector training program

proactive approach to error control is needed to identify potential errors. Thus, the taxonomy was aimed at the phenotypes of error (Hollnagel 1989), i.e., the observed errors. Using the generic task description of the inspection system, the goal or outcome of each task was postulated as shown in Table 3 (third column). These outcomes then formed the basis for identifying the failure modes of each task and included operational error data gained form the observations of inspectors and from discussions with various aircraft maintenance personnel, which were collected over a period of two years. Later the frequencies of each kind of error were estimated, then the consequences of errors on system performance were deduced. The error taxonomy provided the analysts with a systematic framework to suggest appropriate

content for developing the content of the ASSIST training program.

The ASSIST training program specifically focused on the search and decision-making components of the inspection task. These appear to be determinants of inspection performance (Drury 1978, Sinclair 1984) and the two most critical tasks in aircraft inspection (Drury 1991, FAA 1993).

The overall structure of the ASSIST program is shown in Figure 3. As an example, table 4 shows how errors (column 5), identified from the error taxonomy, for each subtask of the decision-making task (column 1) were addressed by the specific modules of the ASSIST training program (columns 2, 3, and 4). Column 2 specifies the training content, column 3 outlines the method used for training, and column 4 specifies the training module within ASSIST.

Table 4. ASSIST program showing errors addressed for the decision task

Task	Content of assist	Method	Program Module	Error addressed from task analysis
4.1 Interpret indication	Present examples of defects and identify in simulator	Active and feedback	General module simulator	• Classify as wrong fault type
4.2 Access comparison standard	Use simulator to access information on defects, locations, and action	Active and feedback	General module simulator	• Choose wrong comparison standards • Comparison standard not available • Comparison standard not correct • Comparison incomplete • Does not use comparison standard
4.3 Decide if it is a fault	Use simulator with real defects and feedback	Progressive parts, active, and feedback	Simulator	• Type I error, false alarm • Type II error, missed fault
4.4 Decide on action	Complete NR card with feedback in correct way to fill out card	Active and feedback	Simulator	• Choose wrong action
4.5 Remember decision/action	Enter multiple defects and complete NR card with feedback	Active and feedback	Simulator	• Forget decision/action Fail to record decision/action

7. CONCLUSION

This article describes a step-by-step methodology to develop a task analytic training program. The research used existing knowledge of inspection training principles to design the ASSIST training program. Note that the development of ASSIST was driven by the requirements of the inspection environment rather than the ability of the research team to implement a specific ergonomic intervention or apply a specific technological solution.

ACKNOWLEDGMENTS

REFERENCES

DRURY, C.G., 1978, Integrating human factors in statistical quality control. *Human Factors*, 20, 561-70.

DRURY, C.G., 1983, Task analysis methods in industry. *Applied Ergonomics*, 14, 19-28.

DRURY, C.G., 1991, The maintenance technician in inspection. *Human Factors in Aviation Maintenance—Phase One Progress Report*, DOT/FAA/AM-91/16 (Washington DC: Office of Aviation Medicine), pp. 45-103.

DRURY, C.G., PRABHU, P., and GRAMOPADHYE, A.K., 1990, Task analysis of aircraft inspection activities: methods and findings. *Proceedings of the Human Factors Society 34th Annual Meeting*, pp. 1181-84.

FEDERAL AVIATION ADMINISTRATION (FAA), 1993, *Human Factors in Aviation Maintenance—Phase Two Progress Report*, DOT/FAA/AM-93/5 (Washington DC: Office of Aviation Medicine).

HOLLNAGEL, E., 1989, The phenotypes of erroneous actions: implications for HCI design. In Weir and Altry (eds) *Human Computer Interaction and Complex Systems* (London: Academic Press).

KIRWAN, B. and AINSWORTH, L.K., 1992, *A Guide to Task Analysis* (London: Taylor & Francis).

LOCK, M.W.B. and STRUTT, J.E., 1985, Reliability of in-service inspection of transport aircraft structures. Canadian Aviation Administration Paper 85013.

PATEL, S, DRURY, C.G., and LOFGREN, J., 1994, Design of workcards for aircraft inspection. *Applied Ergonomics*, 25, 283-93.

SINCLAIR, M., 1984, Ergonomics of quality control. Workshop document at the International Conference on Occupational Ergonomics, Toronto.

STELLY, J. Jr and TAYLOR, J., 1992, Crew coordination concepts for maintenance teams. *Proceedings of the Seventh FAA Meeting on Human Factors Issues in Aircraft Maintenance and Inspection*, Atlanta GA (Washington DC: FAA), pp. 173-206.

THACKRAY, R., 1992, *Human Factors Evaluation of the Work Environment of the Operators Engaged in Inspection and Repair of Aging Aircraft*, DOT/FAA/AM-92-3 (Washington DC: Office of Aviation Medicine).

Team Effectiveness and Competencies

J. A. Cannon–Bowers and E. Salas

Naval Air Warfare Center Training Systems Division, Department of Training Technology, Development Branch, Code 4961, 12350 Research Parkway, Orlando, FL 32826-3275, USA

1. INTRODUCTION

The recent rise in use of teams in modern industry can be traced to several sources. First, advances in technology have resulted in more complex systems that often require contributions from physically dispersed team members. Second, it has been recognized that team-based organizations have advantages in terms of employee motivation and involvement. Third, teams can often accomplish more than individuals and provide the redundancy required in high-risk environments. For these reasons, it has been important in recent years to examine the factors that contribute to team effectiveness. The purpose of this paper is to discuss some of these factors, in particular the competencies required for effective teamwork. Other papers in this volume describe additional factors that have been shown to have an impact on team performance (Bowers *et al.* 1999).

First, the knowledge, skill and attitude competencies that seem to be required for effective team performance are delineated. Next, the nature of competencies held by team members is described — whether they are specific or generic to the task and team. In conclusion, several measurement issues associated with assessing team effectiveness are discussed.

2. TEAM COMPETENCIES

As noted, the focus is on identifying and discussing the competencies required for effectiveness in a team. Note that team competencies can be thought to fall into two related categories (as originally outlined by Morgan *et al.* 1986). First, *taskwork* competencies refer to those knowledge, skills and attitudes (KSA) that a team member needs to accomplish his/her portion of the team task. Taskwork competencies are what are typically thought of as individual skills. However, taskwork competencies are a necessary but not sufficient condition for effective team performance. In other words, team members need to be proficient at their individual tasks, but in addition to this they require competencies at the team level. Hence, *teamwork* skills are required above and beyond taskwork skills if a team is to be most effective. Such teamwork competencies encompass not only both the team-level KSA required in teams but also the important teamwork processes that characterize effective teamwork. The following sections expand on the notion of teamwork competencies and how they are thought to affect team performance.

2.1. Types of Team Competencies

It has been argued that team competencies include all of those KSA that are necessary to perform effectively as a team member. According to Figure 1, this means that effective teams understand the underpinnings of effective teamwork and how these are expressed in the task at hand; are able to execute relevant
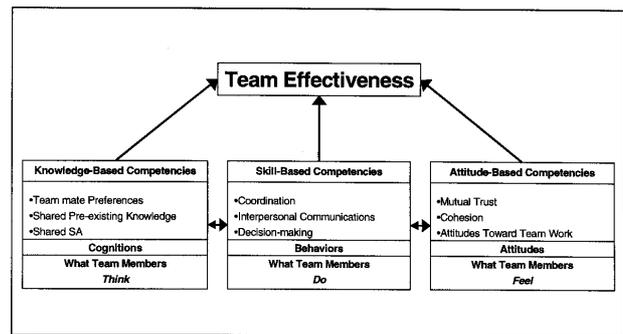
Figure 1. Competencies that affect team effectiveness.

teamwork skills when and how they are required; and have appropriate team-related attitudes that support their team performance. Each of these types of competencies will be detailed below.

2.1.1. Knowledge competencies

Knowledge competencies in teams express what team members need to know to be effective. This refers both to pre-existing knowledge that team members must bring to the team task, and also to knowledge that is constructed dynamically during the team's performance. Beginning with the pre-existing knowledge, several researchers have begun recently to study shared mental models or, more simply, shared knowledge in team members (for more detail on the subject, see Cannon-Bowers *et al.* 1995). Briefly, it has been suggested that when team members have a rich understanding of the task and its demands, and of the strengths, weaknesses, preferences, knowledge requirements and tendency of team mates, they are better able to coordinate their actions. This is particularly important in high workload situations where team members may not have much time to communicate or discuss potential task strategies. In such cases, effective teams appear to rely on an "implicit coordination" strategy where team members seem to be utilizing pre-existing knowledge about the task and each other to support performance.

In addition to shared pre-existing knowledge, team members must also be able to build dynamically a compatible picture of the situation to coordinate implicitly. This means that they must understand well how particular task cues are related to specific team strategies. This kind of knowledge in teams has been referred to as shared situational awareness (Salas *et al.* 1999).

Team members must also have knowledge about their mission, norms, roles and responsibilities. In addition, they require knowledge of teamwork skills, task sequencing and the team's relationship to the larger organization (Cannon-Bowers *et al.* 1995).

2.1.2. Skill competencies

Skill competencies reflect what team members must *do* to perform effectively. Recently, Cannon-Bowers *et al.* (1995) summarized the literature regarding teamwork skills. They found that > 130 labels had been given to team-related skills in the literature. They organized these into eight skill dimensions, including adaptability, which enables the teams to use information gathered from the environment and to adjust their strategies accordingly; shared situational awareness, which creates a common understanding of the task and situation; performance monitoring and feedback, which allows team members accurately to monitor each other's

performance and correct errors; leadership, which provides team members with the ability to direct, focus or coordinate activities; interpersonal skills, which enables members effectively to resolve conflicts and motivate each other; coordination skills, which enables team members to organize and use all the resources available; communication skills, which facilitates the exchange of information to be done clearly and accurately; and decision-making skills, which enables team members to gather and integrate information to execute the task.

2.1.3. Attitude competencies

Dick and Carey (1990) define attitudes as an internal state that influences an individual's choices or decisions to act in a certain way under particular circumstances. Team attitudes are important because they influence how team members will perform in the team. Some examples of attitudes relevant to teams include: (1) attitudes toward teamwork, which fosters the need to coordinate and communicate among team members; (2) mutual trust, which allows team mates to communicate their opinions without fear of being ridiculed; (3) collective orientation, which enables the disposition to receive and value inputs from others; and (4) cohesion, which creates the desire of team members to remain in the group. The importance of individual's attitudes to team success or failure has been investigated in a number of studies and has been show to strongly influence performance.

3. NATURE OF TEAM COMPETENCIES

According to Cannon-Bowers *et al.* (1995), team competencies (i.e. KSA) can be thought of as varying on two dimensions. First, they may be specific to a particular task at hand, or may be more general, thereby applying to a host of team tasks. For example, good communication skills might be general enough to have a positive impact on teamwork in many tasks. Hence, they would be considered *task-generic* teamwork competencies. On the other hand, task sequencing (i.e. knowing when to make contributions to the overall team task) is specific to the specific task at hand. This would be considered a *task-specific* teamwork competency.

Cannon-Bowers *et al.* (1995) also distinguish between *team-generic* and *team-specific* competencies. Team-generic competencies are those that remain the same regardless which specific team members are present. Conversely, team-specific competencies only apply to a specific set of team mates. For example, collective efficacy refers to the team's belief that it can accomplish the task at hand. By definition, this competency is team-specific because it will change depending on the team members being considered. On the other hand, collective orientation — an individual's propensity to be comfortable performing as part of a team — is not dependent on the particular team members being considered.

Figure 2 displays how Cannon-Bowers *et al.* (1995) postulate the types of teamwork competencies are related. Inspection of this figure reveals that the team specific/generic and task specific/generic categories can be combined to form four quadrants of competencies. Beginning in the upper right quadrant are *context-driven* competencies. These are specific to both the task and team. Moving to the upper right hand quadrant, the competencies here are labeled *team-contingent*. They are specific to the particular team members being considered, but generalize across tasks. In the lower left quadrant are *task-contingent* competencies that are

Relation To Task

	Specific	Generic
Specific	*Context-driven*	*Team-Contingent*
Generic	*Task Contingent*	*Transportable*

Relation To Team
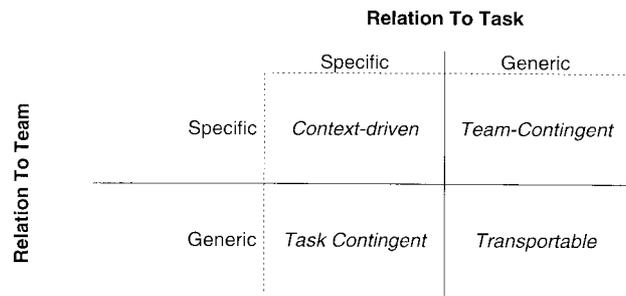
Figure 2. Types of team competencies.

specific to the task at hand, but applicable to a variety of team mates. Finally, in the lower right quadrant are *transportable* competencies that generally hold for any task or team.

To determine which type of competencies would be appropriate for any given team, Cannon-Bowers *et al.* (1995) offered some notions regarding how task and environmental characteristics drive the nature of team competencies required for effective performance. First, they hypothesized that high interdependency within a team task would require that team members to possess context-driven competencies. The rationale behind this proposition was that interdepency causes team members to be reliant on one another and better able to select team strategies in a smooth, coordinated manner. This means that team members have a rich understanding of the task and its demands, and also of the information needs and tendencies of their team mates. They also suggested that when the team's task environment is fairly stable, task-specific competencies would be required, but not necessarily team-specific competencies. This is because in stable tasks, the behavioral discretion of team members is low — that is, the task is done pretty much the same way by any team member. Hence, it is more important to understand the task in this case than it is to understand the idiosyncrasies of team mates performing it.

Another proposition offered by Cannon-Bowers *et al.* is that rapid turnover in a team also reduces the need for team members to hold team-specific competencies. Obviously, if the team's configuration changes rapidly, investing time in learning about team mates would be wasted. On the other hand, when the same set of team members interact together in a variety of different tasks, the opposite is true. That is, in this case, team-specific competencies should improve performance, while task-specific competencies may not be an option.

Cannon-Bowers *et al.* also argued that transportable team competencies, which are generic to both the task and the team, should be provided to team members who hold membership on multiple teams. Also, when the organization is moving to team-based organization, training transportable teamwork competencies may provide a viable precursor to more specific team training.

4. DELINEATING COMPETENCIES FOR TEAMS

Table 1 shows a fairly comprehensive listing of team competencies categorized in accordance with the scheme offered by Cannon-Bowers *et al.* (1995). It summarizes the results of numerous studies of team performance and should provide an excellent starting point for those interested in determining the KSA required for any particular team. These competencies can be used to

Table 1. Proposed competencies for teams.

Nature of team competency	Description of team competency	Knowledge	Skills	Attitudes
Context-driven	Team-specific; task-specific	Cue-strategy associations Task-specific team mate characteristics Team-specific role responsibilities Shared task models Team mission, objectives norms, resources	Task organization Mutual performance monitoring Shared problem-model development Flexibility Compensatory behavior Information exchange Dynamic reallocation of functions Mission analysis Task structuring Task interaction Motivation of others	Team orientation (morale) Collective efficacy Shared vision
Team-contingent	Team-specific; task-generic	Team mate characteristics Team mission, objectives, norms, resources Relationship to larger organization	Conflict resolution Motivation of others Information exchange Intrateam feedback Compensatory behavior Assertiveness Planning Flexibility Morale building Cooperation	Team cohesion Interpersonal relations Mutual trust
Task-contingent	Team-specific; task-specific	Task-specific role responsibilities Task sequencing Team role-interaction patterns Procedures for task accomplishment Accurate task models Accurate problem solving Boundary-spanning role Cue-strategy associations	Task structuring Mission analysis Mutual performance monitoring Compensatory behavior Information exchange Intrateam feedback Assertiveness Flexibility Planning Task interaction Situation awareness	Task-specific teamwork attitudes
Transportable	Team-generic; task-generic	Teamwork skills	Morale building Conflict resolution Information exchange Task motivation Cooperation Consulting with others Assertiveness	Collective orientation Belief in importance of teamwork

establish the selection requirements for team members, or as an indication of the type of team training that is required.

5. MEASURING TEAM EFFECTIVENESS

The discussion of team effectiveness cannot be complete without considering how team performance is measured. According to Cannon-Bowers and Salas (1997), team performance measures must fulfil three related concerns. First, they must describe the performance of interest. Owing to the dynamic nature of many team tasks, even this simplest level of measurement is complicated. The moment-to-moment changes in team performance — especially those involving communication — are sometimes difficult to capture. The next level of measurement involves comparing observed performance to a standard. This 'evaluative' level of measurement seeks to determine how the team's performance compares to some standard (whether it be absolute or normative). The final level of measurement that is needed to assess team performance, particularly in training, is at

the diagnostic level. Here the concern is to determine the causes of effective and ineffective performance. This is imperative as a basis to provide feedback and/or construct remedial activities.

To diagnose team performance, Cannon-Bowers and Salas (1997) argue that performance measures for teams must assess the processes involved in accomplishing the team task as well as the outcomes of team performance. Outcome measures are what are typically thought of when measuring performance; essentially, they describe whether the team's performance was successful. Outcome measures include task accuracy, timeliness, latency, quality, productivity and so forth. It is essential to assess team outcomes to get a full picture of how the team performed. On the other hand, process measures described *what* happened during the course of the team task. Moreover, they also help to describe *why* certain outcomes were achieved. They are essential if an understanding of why the team was effective is sought, and are a primary mechanism for diagnosing team performance as described above.

Cannon-Bowers and Salas (1997) also distinguish between measures at the team and individual levels. As noted above, team performance consists of competencies that both the individual and team levels. Hence, it is important that team performance is measured at both levels as well. When combined with the outcome versus process and descriptive, evaluative and diagnostic delineations described above, a simple matrix of performance measures for teams is formed. According to Figure 3, four categories of team performance measures — at the descriptive, evaluative and diagnostic levels — are necessary to provide a full picture of team performance.

Figure 3 indicates that within each quadrant a set of performance measures can be developed for any given task. For example, in the team process box, it is recommended that the team's degree of shared task models, understanding of cue/strategy associations and collective efficacy is assessed. In the individual process box (still referring to teamwork processes, but those held at the individual level), it is recommended that assertiveness, mutual performance monitoring and task procedures be measured. Taking this a step further, Cannon-Bowers and Salas (1997) offered suggestions about what type of measurement strategy would be most appropriate for each category. Beginning with the upper left quadrant (team process measures), they recommend that observational scales targeting teamwork processes are an effective means in assessing this aspect of performance. Such scales can be developed using subject matter

experts (SME) to guide delineation of critical behaviors that must occur in team members. SME can also be employed as experts as expert raters when the task is particularly complex. Another technique for assessing team process is content analysis. By reviewing what happened over the course of team performance — what decisions were made and by whom — an assessment of the team's process is (at least partially) possible. Such data can be collected by observers or through automated systems that capture various aspects team performance. Related to this, an analysis of the team's communications (called protocol analysis) is also a useful mechanism to assess team process.

Turning to individual process measurement, several techniques apply. First, policy capturing seeks statistically to determine the relationship between inputs (cues) and outputs (strategies or decisions) in and individual's decisions. Next, decision analyses — which are more qualitative in nature — can help to reveal how decisions were made and how these led to task action. Observational scales and protocol analyses can also be geared toward individual processes (as well as team processes as described above).

At the outcome level, several more traditional methods for assessing task performance may be useful. In addition to expert ratings and observational scales, critical incidents may be used, archival data (i.e. performance trends) and automated performance recording are all viable options (Cannon-Bowers and Salas 1997).

REFERENCES

BOWERS, C. A., JENTSCH, F. and MORGAN, B. B. Jr, 1999, Team performance (in this volume).

CANNON-BOWERS, J. A. and SALAS, E., 1997, Teamwork competencies: the intersection of team member knowledge, skills, and attitudes. In O'Neil, H. F. (Ed.) *Workforce Readiness: competencies and assessment* (Mahwah: Lawrence Erlbaum Associates), pp. 151–74.

CANNON-BOWERS, J. A., TANNENBAUM, S. I., SALAS, E. and VOLPE, C. E., 1995, Defining team competencies and establishing team training requirements. In Guzzo, R. and Salas, E. (eds) *Team Effectiveness and Decision Making in Organizations* (San Francisco: Jossey Bass), pp. 333–80.

DICK, W. and CAREY, L. 1990, *The Systematic Design of Instruction* (3rd ed.) (Glenview: Scott Foresman).

MORGAN, B. B. Jr, GLICKMAN, A. S., WOODARD, E. A., BLAIWES, A. S. and SALAS, E., 1986, *Measurement of team behaviors in a Navy environment*. Technical Report No. 86-014 (Orlando: Naval Training Systems Center).

SALAS, E., CANNON-BOWERS, J. A. and SMITH-JENTSCH, K. A., 1999, Principles and strategies for team training (in this volume).

	TEAM	INDIVIDUAL
PROCESS	• Shared Task Models • Cue/Strategy Associations • Task Organization • Compensatory Behavior • Collective Efficacy • Dynamic Reallocation of Function • Task Interaction	• Assertiveness • Information Exchange • Task-Specific Role Responsibilities • Procedures for Task Accomplishment • Cue/Strategy Associations • Mutual Performance Monitoring • Flexibility
OUTCOME	• Mission/Goal Accomplishment • Aggregate Latency • Error Propagation • Aggregate Accuracy	• Accuracy • Latency • Errors • Safety • Timeliness • Decision Biases

Figure 3. Framework for developing TPM in training.

Team Performance

C. Bowers, F. Jentsch and B. B. Morgan Jr

Department of Psychology, University of Central Florida, 4000 Central Florida Blvd, PO Box 161390, Orlando, FL 32817, USA

1. INTRODUCTION

The increasing complexity of modern technological systems, such as power plants, manufacturing factories or aircraft cockpits, has made it necessary for teams of workers to coordinate their actions to achieve successful performance. However, several notable failures, incidents and accidents in these systems suggest that even highly competent and trained workers are sometimes poorly suited to perform together as a team. Therefore, there has been a recent interest in assessing the nature of teams and the variables that impact team performance. Issues such as staffing, individual team member characteristics, and the design of workspace and workflow have been among the topics studied in team performance research.

2. DEFINITION OF TEAMS

The initial stages of team performance research were somewhat hampered by a debate concerning the definition of teams. Because the terms "teams" and "groups" were often used interchangeably, an early controversy arose concerning whether the large body of existing research on the performance of groups could be used to predict the performance of teams. The result of this debate has been that there appears to be some value in discriminating between teams and groups in more detail before generalizing from one to the other. Consequently, teams are currently thought to form a specific and identifiable subset of the various multi-person units that are described by the more general term "group".

Teams are differentiated from other collectives by three characteristics (Swezey and Salas 1992):

- To be part of a team, team members must work toward a shared, valued goal. While some groups are formed so that their members work toward a common goal, others are formed by affiliation (e.g. professional organizations), belief (e.g. political parties or religious groups), demographic characteristics (e.g. graduating classes in college), random assignment (e.g. juries in court cases) or purely by coincidence (e.g. audiences in a movie theatre). Thus, individuals may belong to the same group, but have widely differing goals. On the other hand, by definition, all members of a team work to accomplish the same goal.
- Team members work interdependently toward their common goal. That is, no individual member of the team can normally complete the entire team's task. Rather, the performance of team members is interdependent on the performance of other team members. In groups, on the other hand, members routinely have equivalent resources and information, and any individual can often represent the group through their actions. As a result, the size and member responsibilities for teams tend to be much more specifically defined than for groups.
- Team members must coordinate their activities. Given the requirement for interdependence, teams perform their work by sharing resources and sequencing their activities. Often,

the success of a team is determined by the ability of its members to coordinate effectively. Groups, on the other hand, have no such requirement. The content and timing of communication in groups is generally not dictated by the task. Consequently, the quality of their performance tends to be less correlated with coordination behaviors than in teams.

3. FACTORS AFFECTING TEAM PERFORMANCE

3.1. Team Effectiveness Model

Salas *et al*. (1992) have presented a model that describes many of the factors that have been shown to affect team performance. This input–process–outcome model is called the team effectiveness model (TEM). A number of input factors from this model are discussed below. Variables associated with team processes and outcomes are discussed within other entries in this encyclopaedia (for example, Cannon-Bowers *et al*. 1999).

3.2. Individual Team Member Characteristics

3.2.1. Task skills

Not surprisingly, teams whose members are more competent in their performance of individual tasks tend to demonstrate better team performance. It should be noted, however, that the correlation between individual team member skill competence and team performance is lower than one might expect. Further, the performance one would predict on the basis of individual team member skill competence is usually higher than that actually achieved by the team. This phenomenon has been described as 'process loss' (Steiner 1972).

3.2.2. Interpersonal skills

Because interdependence and coordination are required among team members, it seems reasonable that team members must possess interpersonal ("teamwork") skills in addition to their specific task skills. The research on interpersonal skills training for teams, however, is somewhat mixed. Research indicates that training that is specifically targeted at the coordination needs of the team within the context of its assigned task is more successful than general interpersonal skills training. For example, in the case of junior team members who struggle to communicate with senior team members, general social skills training (e.g. social mores, politeness) is less likely to improve team performance than targeted team skills training (e.g. assertiveness).

3.2.3. Attitudes

A relatively large body of research is devoted to individual attitudes toward teamwork and their impact on team performance. This research has shown that attitudes toward teamwork differ among individuals and that there is some relationship between these attitudes and team performance. This relationship has been the impetus for team training that focuses on improving team performance by improving team member attitudes toward working as a team. At the same time, however, the effectiveness of these interventions also hinges on the presence of adequate team member competence in the required task and interpersonal skills.

3.2.4. Personality characteristics

When individuals join teams, they tend to carry their long-term

patterns of behavior with them. Thus, it seems reasonable to predict that these personality styles and characteristics also affect team performance. Interestingly, the literature, in general, seems not to support this prediction. While differences in personality among team members are associated with differences in the processes used by teams to achieve their goals, they do not necessarily impact the outcome or result of the team's actions.

3.3. Team Characteristics

3.3.1. Team size

The size of a team is typically determined by its task demands. That is, the number of team members is a function of the resources demanded by the task. That said, it is tempting to form large teams to maximize the number of available resources. This strategy is especially effective in the domain of problem-solving if the goal is to achieve a very high-quality solution. However, it should be noted that increasing team size is associated with several drawbacks. First, it is more difficult for large teams to communicate and coordinate their actions. Thus, they are not usually effective in highly interdependent tasks. Second, large teams typically perform more slowly and make more errors. They are generally not recommended in critical, high time-pressure tasks. Third, increasing team size is associated with higher communication workload, perhaps leading to greater process loss.

3.3.2. Collective attitudes

Two collective attitudes that have received research attention are collective efficacy and team cohesion. Collective efficacy refers to the shared belief that the team can perform its tasks effectively and achieve its goals. Similar to effects reported with individuals, teams with higher collective efficacy seem to perform better than their low-efficacy counterparts. However, this conclusion is based largely on correlational research and requires additional study. Team cohesion refers to the degree to which team members prefer to remain in a specific team. This team attitude has been associated with several positive team outcomes, including lower absenteeism and reduced turnover. However, the relationship between cohesion and performance is less clear.

3.3.3. Team member homogeneity

It is reasonable to suggest that teams are more effective when their members are "compatible" along some number of traits. In general, research has shown that teams made up of more similar members usually require less time to coordinate and achieve higher performance in routine, proceduralized tasks. However, there might be an advantage for heterogeneous teams performing tasks that require creativity, such as dynamic problem-solving, new product development and the like.

3.4. Task Characteristics

- Task type — teams are employed to perform a variety of tasks, from relatively routine, proceduralized tasks (such as product assembly) to unpredictable problem-solving tasks (such as system diagnosis). While these tasks, in and of themselves, do not fully determine team performance, they are prone to interact with other team performance characteristics (as demonstrated above). For example, the nature of the task can interact with team size to influence team processes.

- Time pressure — as mentioned above, the need to communicate and coordinate among team members is an additional workload factor for teams. When confronted with time pressure, the team's choices are either to increase their speed or reduce their communication. Research has indicated that effective teams often possess the ability to maintain performance while reducing communication. That is, they can share resources in the absence of explicit communication. The mechanism underlying this pattern has been labeled "implicit coordination" (Kleinman and Serfaty 1989). Presumably, team members can coordinate implicitly because they possess a shared understanding of the task and can anticipate one another's needs. It should be noted that while this pattern might be effective for high time-pressure situations, it increases the risk of errors.

3.5. Work Characteristics

3.5.1. Work structure

Researchers have studied the impact of forming structured, hierarchical ("tall") teams versus somewhat less structured ("flat") teams. Generally speaking, the research has indicated that hierarchical teams tend to display performance disadvantages. This is probably due to the inability of members to choose subtasks that fit their personal strengths. It is interesting to note, however, that these performance differences tend to disappear under high time pressure. In fact, hierarchical teams appear to posses a slight advantage in high time-pressure situations.

3.5.2. Communication structure/modality

Team members in modern workplaces are often geographically distributed. This distance often is associated with communication structures that use a variety of modalities, from video to e-mail. There appear to be potential positive and negative effects of mediated (i.e. non-face-to-face) communication. Negative effects include increased time to perform, which might render this type of communication unsatisfactory for high time-pressure tasks. Furthermore, studies have suggested that team members might display more emotional behaviors in mediated communications, leading to reduced cohesion. However, advantages of mediated communication have also been reported. Namely, mediated communication often leads to an increased ability of all members to contribute. This, presumably, is associated with higher qualities of decision-making. Similarly, the opinions of junior members appear to be considered more fully in mediated communications.

3.6. Organizational and Situational Characteristics

3.6.1. Organizational support

One important environmental characteristic that affects the effectiveness of teams is the teamwork support provided by the organization. In many examples from industry, teams that were created to improve worker participation, job satisfaction and organizational performance have failed because on-going organizational support for teamwork was missing. For example, teams fail sometimes because their explicit goals are contrary to the goals of the larger unit in which they are supposed to work. Conversely, positive organizational support for teamwork can manifest itself in the form of organizational commitment towards participative decision-making throughout the organization,

mission statements that explicitly endorse teamwork, training initiatives that emphasize team skills, etc.

3.6.2. Supervision and reward

In addition to the general support provided to teams by the organization, the ways in which organizations supervise and reward teams have important implications for team processes and performance. Most team members have to balance individual task performances with efforts that are required to maintain intra-team communication and coordination. Research has shown that organizational feedback and rewards substantially affect the way in which team members create and maintain this balance. For example, team members who are rewarded on the basis of their individual performances are much more likely to emphasize the individual task aspects of their work and to neglect coordinating and communicating with other team members. Conversely, team members whose rewards are based on their attempts to coordinate with others are more likely to emphasize coordination and to neglect their individual tasks, even if that tends to hurt overall team performance. Therefore, organizational rewards should be structured in a way that mirrors the importance of individual and team tasks for overall organizational success.

4. CONCLUSIONS

Teams are created for a variety of reasons: because the work cannot (or should not) be completed by an individual, because organizations would like to take advantage of the diversity of resources that a number of individuals bring to a team, or because the use of teams increases redundancy and reduces the chance for (catastrophic) errors. In recent years, teams have show great promise for fulfilling these intents within a wide range of academic, industrial and military settings. However, it is erroneous to believe that selecting and training a group of competent individuals will necessarily result in effective team performance. As discussed above, a variety of factors converge to determine the success (or failure) of teams. Only by understanding the way in which these factors interact can organizations reap the full benefits of effective team performance.

REFERENCES

KLEINMAN, D. and SERFATY, D., 1989. Team performance assessment in distributed decison-making. In *Proceedings of the Symposium on Interactive Networked Simulations for Training*, 22–7.

SALAS, E., BOWERS, C. A. and CANNON-BOWERS, J. A., 1995. Military team research: 10 years of progress. *Military Psychology*, 7, 55–76.

SALAS, E., DICKINSON, T. L., CONVERSE, S. and TANNENBAUM, S. I., 1992. Toward an understanding of team performance and training. In Swezey, R. and Salas, E. (Eds) *Teams: their training and performance* (Northwood: Ablex).

STEINER, I. D., 1972. *Group Processes and Productivity* (New York: Academic Press).

SWEZEY, R. W. and SALAS, E., 1992. *Teams: their training and performance* (Northwood: Ablex).

Team Training

E. Salas, J. A. Cannon-Bowers, and K. A. Smith-Jentsch
Department of the Navy, Naval Air Warfare Center Training Systems
Division, 12350 Research Parkway, Orlando, FL 32826-3275, USA

1. INTRODUCTION

Teams have become a way of life in organizations. In fact, a recent survey conducted in US industries reported that 82% of businesses with 100 or more employees used teams in one way or another to accomplish their objectives. Teams are more popular than ever, and organizations are investing considerable resources in strengthening the communication and coordination mechanisms of their employees. Furthermore, many organizations are now designed around teams so that their very survival depends on effective teamwork.

This interest in teams by organizations has led to an explosion of team training research and practice. The aim of this research has been to address the question, What turns a team of experts into an expert team? That is, what developmental activities can enhance effective team functioning among highly interdependent members that have distributed expertise? In order to report the state of the art in team training, we outline what has been learned in answering this question. Specifically, we focus on team training principles and strategies that have been generated in recent years.

To begin our discussion, we define a team as "a distinguishable set of two or more people who interact, dynamically, interdependently, and adaptively toward a common and valued goal/

Figure 1. The structure of team training (Adapted from Salas and Cannon-Bowers 1997)

objective/mission, who have been assigned specific roles or functions to perform, and who have a limited life-span of membership" (Salas *et al.* 1992: 4). There are key concepts in this definition that bound the generalizability of this discussion. We focus particularly on teams that have distributed expertise, have high task interdependency and are organized hierarchically (i.e., there is a leader). Examples of this type of team include military command and control teams, medical emergency teams, some police and firefighting teams, cockpit crews, some sports teams, and some management teams.

2. WHAT IS TEAM TRAINING?

Team training is a set of tools, methods and content that together create an instructional strategy that focuses on enhancing teamwork. Figure 1 outlines the basic elements of team training. It reveals that team training is not a program or a place, as it has been popularly characterized. Team training interventions are theoretically based; built around sound instructional principles; based upon carefully crafted and designed instructional strategies; and focused on promoting the ease of coordination, communication, and synchronization among team members. Team training is more than just team building (i.e., where team members learn about each other) and individual training (i.e., where team members learn about their task). Team training is about designing and developing instructional strategiesæthrough tools, methods, team objectives, and competenciesæfor influencing team processes.

2.1. Tools

There are a number of useful tools available for designing effective team training strategies. For example, research has shown that an analysis of the coordination demands required for successful task performance can serve to focus the nature of practice and feedback. This coordination analysis requires that subject matter experts (SMEs) specify the criticality, frequency, and difficulty of performing certain team-related tasks. This is usually done via surveys or questionnaires. Team task analysis is also an available procedure. Although much work remains here to demonstrate its utility, this tool allows us to uncover the knowledge, skills, and attitudes (KSAs) necessary for effective teamwork given specific task requirements.

Performance measurement is the key to learning. Performance measurement tools embedded into training allow for learning to occur. Measurement tools for teams must consider a number of principles. These principles are readily adaptable/applicable to any team training environment. First, it is essential that any system of performance measurement has a strong theoretical foundation upon which to draw. There are many frameworks now available that can help guide the development process (Brannick *et al.* 1997). Second, a clear purpose for designing such a measurement system must be established upfront. This can be accomplished by first performing team/cognitive task analyses in order to tease out the criteria (i.e., KSAs) upon which the system should be based. Next the opportunities for measurement during training must be specified . This approach allows for specific and controlled learning opportunities and a manageable context to measure complex team performance. Any measurement system should also be aimed at not only determining what transpired but should also explain why it happened, in order to facilitate

learning. Therefore, process measures (i.e., to capture moment-to-moment actions and behaviors) must be designed and used. Usually these measures take the form of observation protocols or scales.

Additionally, team performance measurement systems should be capable of supporting the diagnosis of performance deficiencies so as to provide relevant remediation to team members. Another principle that should underlie the development of a team performance measurement tool is that it must not be cumbersome to use. Users should be able to rapidly collect, evaluate, and synthesize performance data so that immediate feedback can be provided to trainees. This can be realized by automating the system and by preparing instruction for the assessors.

A key element to the success or failure of any performance measurement system lies with the involvement of the user. Involving SMEs in the development process will help to ensure the end-product is accepted, useful, and effective. Finally, the tools and training necessary for instructors to consistently use the performance measurement system should be provided by way of instructor training packages as well as supporting materials (i.e., guide sheets, definitions, scenario events). Adherence to these principles will ensure the development of a strong team performance measurement tool to support team training.

Other tools to support team training include feedback and learning principles, which are used to guide and focus the design process. These are available in the literature from not only the human factors perspective but also the education and cognitive disciplines as well. Finally, task simulation and exercises are also tools for team training. These provide the environment or context that allows team members to practice the requisite KSAs.

2.2 Methods

Many methods are available to deliver instruction. They apply to both individual training and team training. There are information-based methods such as lectures, handbooks, slide presentations. These are widely used and effective when combined with other methods. There are also demonstration-based methods. These are more dynamic and stimulating. They include video demonstrations of effective and ineffective teamwork behaviors as well as multimedia systems that illustrate key principles and performance requirements in teams. Finally, there are practice-based methods. These are more sophisticated, costly, and interactive. They include simple to advanced exercises such as role-play exercises to high fidelity simulation environments (i.e., simulators) for teams to practice. These systems are visually rich and very interactive and engaging to the trainees. Moreover, we know simulators are effective training devices when appropriate instructional tools and features (e.g., performance measurement) are embedded in them (Salas *et al.* 1998).

2.2.1 Principles for enhancing the design of team training

Several principles can be used to guide the design and delivery of team training. We provide some representative examples (Cannon-Bowers and Salas 1998, Salas and Cannon-Bowers 1997, 1999). For example, team training must foster the development of team skills such as team leadership, adaptability, and compensatory behavior; link cue patterns to response strategies not only with regards to the task, but other team members as well;

provide team members with guided practice under realistic conditions; provide team members with motivational guidance and outline factors that hinder their motivation; make team members aware of the importance of effective communication; develop simulations that allow team members to experience different courses of action; facilitate team members' ability to learn about each other's roles and build realistic expectations about task requirements; diagnose team members' teamwork deficiencies with a combination of process measures and outcome measures; provide the necessary information to team members about effective teamwork, demonstrate those behaviors and actions, and allow for the practice with feedback on the requisite KSAs; and exercise team members with a variety of novel situations to build adaptive mechanisms.

2.3. Team Competencies

Team competencies are the heart of team training. The task context provides the required KSAs needed to successfully complete a team's tasks. KSAs in teams are quite different from those found for individuals.

Team competencies describe what team members need to know (i.e., knowledge), how they need to behave (i.e., skills), and what task-related attitudes (i.e., attitudes) they need to hold. Using the tools described earlier (i.e., team task analysis), team KSAs must be derived as a precursor to developing team training. In fact, the KSAs established for a team should provide the training objectives to be targeted in developing instruction.

Recently, researchers have focused attention on developing taxonomies of KSAs for different types of teams. In general, this work has highlighted the fact that team cognitionæthe knowledge componentæhas been overlooked in past research. In addition, results indicate that some teamwork skills are consistent across tasks, while others may be more generic. Likewise some team competencies may be specific to the particular team at hand, while others may hold regardless of team membership (Cannon-Bowers and Salas 1999, Cannon-Bowers *et al.* 1995).

2.4. Team Strategies: Examples

The methods, tools, and content (with the appropriate team training objectives) create a team training strategy. This strategy may take many forms and shapes, depending upon the needs, resources, and objectives of the training. Nowadays most team training strategies combine information, demonstration, and practice-based methods. Here are a few of them.

2.4.1 Team coordination training

Probably the most widely used in organizations, team coordination training is also known as crew resource management or aircrew coordination training. It has been used to train teams in a variety of settings, including aviation, medical, and military. Although the course content differs for each of these settings, there are certain commonalities inherent in all of these training programs. To begin with, this strategy concentrates on teaching team members the general components of teamwork in an effort to encourage it. Next, most of these programs use the general team performance or group dynamics literature as a framework. In addition, the training is usually delivered through a combination of information, demonstration (e.g., video examples), and practice-based methods (e.g., role plays) over a 2-5 day period.

2.4.2 Cross-training

Cross-training is a strategy in which team members are exposed to the basic tasks, duties, and responsibilities of the positions held by other members of the team. The intent of this strategy is to alleviate the decline in performance following personnel changes as well as to increase implicit coordination (i.e., being able to coordinate without the need to explicitly communicate). This is accomplished by the subsequent increase in the interpositional knowledge held by each team member following training (i.e., a strengthening of the team's shared mental models). The training is composed of cross-role information sharing (teammates, task, equipment, situation), enhanced understanding of interdependencies, roles and responsibilities, and cross-role practice and feedback. The training is usually delivered via lecture, multimedia systems, and/or roleplays. Research has shown that cross-training can be an effective means to enhance team performance. Among other things, cross-trained teams are better able to anticipate the information needs of their teammates, commit fewer errors, and display better team process quality; see Cannon-Bowers and Salas (1998) for a review of this and other team training strategies.

2.4.3 Team self-correction

Team self-correction, as defined by Blickensderfer *et al.* (1997), is a naturally occurring tendency for effective teams, following a performance episode, to review events, correct errors, discuss strategies, and plan for the future (e.g., debrief themselves). This team training strategy attempts to capture this tendency and train other teams how to engage in this process. The end result is a team that is able to correct its knowledge-based, skill-based, and attitude-based deficiencies without the aid of a formal instructor. This training is delivered through a combination of lecture, demonstration, practice, and feedback. The target of training is the team leader (if there is one) and team members themselves. Specifically, team members are taught how to observe their own performance, how to categorize their effective and ineffective behavior into a structured format, and how to use this information to give each other feedback (Cannon-Bowers and Salas 1998). When an instructor is present, he or she is taught how to probe the team members to come up with their own examples of effective and ineffective performance. This method of guided team self-correction has been demonstrated to improve team performance.

2.4.4 Event-based approach to training (EBAT)

EBAT is a general approach to the design of simulation-based exercises where training opportunities are systematically identified and introduced as "trigger" events in a training exercise. The purpose of doing this is to provide opportunities for students to exhibit specific team behaviors that are based on the team training objectives established through team task analysis. These behaviors are in turn captured, rated, and then used later for debriefing and feedback. This strategy has been successfully used in multiservice distributed military training environments as well as in aviation settings.

2.4.5 Assertiveness training.

In complex team environments, it is important that teams make use of all available information and resources. This means that the unique knowledge, skills, ideas, opinions, and observations of each team member must be exploited to arrive at the optimal decision. In fact, it is often the willingness of team members to voice concerns assertively that can help to avoid potentially life-threatening decisions (e.g., in law enforcement teams or aircrews). Hence efforts have been directed toward training team members to display effective task-related assertiveness. For example, team members may need to act assertively when providing feedback to other members, when stating opinions, when stating potential solutions, or when offering or providing assistance. Training to improve assertiveness should involve practice and feedback to ensure that assertiveness skills will be demonstrated (Smith-Jentsch *et al.* 1996).

REFERENCES

BLICKENSDERFER, E., CANNON-BOWERS, J.A., and SALAS, E., 1997, Training teams to self-correct: an empirical investigation. Paper presented at the 12th annual meeting of the Society for Industrial and Organizational Psychology, April, St. Louis MS.

BRANNICK, M.T., SALAS, E., and PRINCE, C. (eds), 1997, *Team performance assessment and measurement: theory, methods, and applications* (Mahwah NJ: Lawrence Erlbaum).

CANNON-BOWERS, J.A. and SALAS, E. (eds), 1998, *Making decisions under stress: implications for individual and team training* (Washington DC: APA Press).

CANNON-BOWERS, J.A. and SALAS, E A., 1999, Team effectiveness. In Karwowski *et al.* (eds) *International Encyclopedia of Ergonomics and Human Factors* (London: Taylor & Francis).

CANNON-BOWERS, J.A., TANNENBAUM, S.I., SALAS, E. and VOLPE, C.E., 1995, Defining team competencies and establishing team training requirements. In Guzzo and Salas, (eds) *Team effectiveness and decision making in organizations* (San Francisco: Jossey Bass), pp. 333–80.

SALAS, E. and CANNON-BOWERS, J.A., 1997, Methods, tools, and strategies for team training. In Quiñones and Ehrenstein (eds) *Training for a rapidly changing workplace: applications of psychological research* (Washington DC: APA Press), pp. 249–79.

SALAS, E. and CANNON-BOWERS, J.A., 1999, The anatomy of team training. In Tobias and Fletcher (eds) *Handbook on research in training* (New York: Macmillan).

SALAS, E., DICKINSON, T.L., CONVERSE, S.A., and TANNENBAUM, S.I., 1992, Toward an understanding of team performance and training. In Swezey and Salas (eds) *Teams: their training and performance* (Ablex), pp. 3–29.

SALAS, E., BOWERS, C.A., and RHODENIZER, L., 1998, It is not how much you have but how you use it: toward a rational use of simulation to support aviation training. *International Journal of Aviation Psychology*.

SMITH-JENTSCH, K.A., SALAS, E., and BAKER, D.P., 1996, Training team performance-related assertiveness. *Personnel Psychology*, 49, 909–36.

Team Training for Aircraft Maintenance

D.C. Kraus and A.K. Gramopadhye*

Galaxy Scientific Corporation, Tucker, Georgia, USA and *Industrial Engineering Department, Clemson University, Clemson, South Carolina, USA

1. INTRODUCTION

Research and development efforts related to human factors in aviation have often focused on the pilot and the cockpit. However, a recent study conducted in the USA has shown that 18% of the accidents could be attributed to maintenance-related issues (Phillips 1994). In response, the FAA has pursued maintenance-related human factors research. The Federal Aviation Administration (FAA) reports on human factors in aviation maintenance (Shepherd 1991, FAA 1993, 1995) have identified training as a primary intervention strategy to improving performance. However, this training has primarily focused on improving individual skills of the aircraft maintenance technicians (AMTs) This training has ranged from improving diagnostic skills for aircraft maintenance training (Johnson 1990) to acquiring and enhancing visual inspection skills to improve airframe structural inspection (Gramopadhye et al. 1992). A large effort in the past has concentrated on developing individual skills of AMTs; however, very little effort has been placed on developing team skills.

The task analyses of aircraft inspection and maintenance activities (Drury et al. 1990, Taylor 1990, FAA 1991, 1993) have revealed the aircraft inspection and maintenance system to be complex, requiring above average coordination, communication and cooperation between inspectors, maintenance personnel, supervisors and various other subsystems (planning, stores and shop) to be effective and efficient. A large portion of inspector and maintenance technician work is accomplished through teamwork. The challenge is to work autonomously but still be a part of the team. In a typical maintenance environment, first the inspector looks for defects and reports them. The maintenance personnel then repair the reported defects and work with the original inspector or the buyback inspector to ensure the job meets predefined standards. During the entire process, the inspectors and maintenance technicians work with their colleagues from the same shift and the next shift as well as personnel from planning, stores, etc., as part of a larger team to ensure the task gets completed (FAA 1991). Thus, in a typical maintenance environment, the technician has to learn to be a team member, communicating and coordinating the activities with other technicians and inspectors. Though the advantages of teamwork are widely recognized (Hackman 1990) in the airline industry, the work culture assigns responsibility for faulty work on individual AMTs rather than on the teams in which they work. The reasons for this could be the individual licensing process and personal liability, both of which often result in AMTs and their supervisors being less willing to share their knowledge and to work across shifts with less experienced or less skilled colleagues. The problem is further compounded since the more experienced inspectors and mechanics are retiring and being replaced by a much younger and less experienced workforce. Not only do the new AMTs lack knowledge or skills of the far more experienced AMTs they are replacing, but they are also not trained to work as a team member.

The FAA has continually addressed the earlier problem of the development of individual AMT skills. For example, the newly established FAR Part 66 (new AMT's certification requirements) specifically addresses the significantly technological advancements that have taken place in the aviation industry and the advancements in training and instructional methods that have arisen in the past decade. The FAA, through the Office of Aviation Medicine, has also funded efforts for the development of advanced training tools to train the AMTs of the future. These advanced tools include intelligent tutoring systems, embedded training, etc., which will be available to airframe and powerplant (A&P) training schools. It is anticipated that the application of these new training technologies will help reduce the gap between current AMT skills and those needed for the maintenance of advanced systems. However, the AMTs joining today's workforce are lacking in team skills. The current A&P school curriculum often encourages students to compete against one another, and often AMTs are not fully prepared for cooperative work in the future. To prepare student AMTs for the workplace, new ways have to be found to build students' technological, interpersonal and sociotechnical competence by incorporating team training and communication skills into their curriculum. This study explored the potential of advanced technology for team training within the aircraft maintenance environment.

The general objective of this research was to understand the role of team training and specifically that of computers for team training. The team training model (figure 1) for the aircraft maintenance environment, developed by Kraus et al. (1996), was used as a framework to develop a computer-based team training software, the aircraft maintenance team training (AMTT) software (Gramopadhye et al. 1995). The model served as the first step in understanding teamwork in aircraft inspection and maintenance operations, and illustrated the interaction between internal factors, external factors, the team process, training strategies and outcomes measures. Using the model and the AMTT software, a controlled study was designed and conducted to evaluate the effectiveness of computers for team training. The specific objectives of this research were (1) to evaluate the usefulness of computers in assisting AMTs in acquiring knowledge on team skills, (2) to determine whether a computer-based multimedia tool would be easy for AMTs to use, and (3) to determine whether computer-based team training was as effective as traditional instructor-based training.

2. COMPUTER–BASED TEAM TRAINING: AMTT SOFTWARE

Specifically designed for training aircraft maintenance technicians in basic team skills, AMTT uses a multimedia presentational approach with interaction opportunities between the user and the computer. The multimedia presentation includes full motion video which provide real-life examples of proper and improper team behavior, photographs and animation that illustrate difficult concepts, and voice recordings coupled with visual presentations
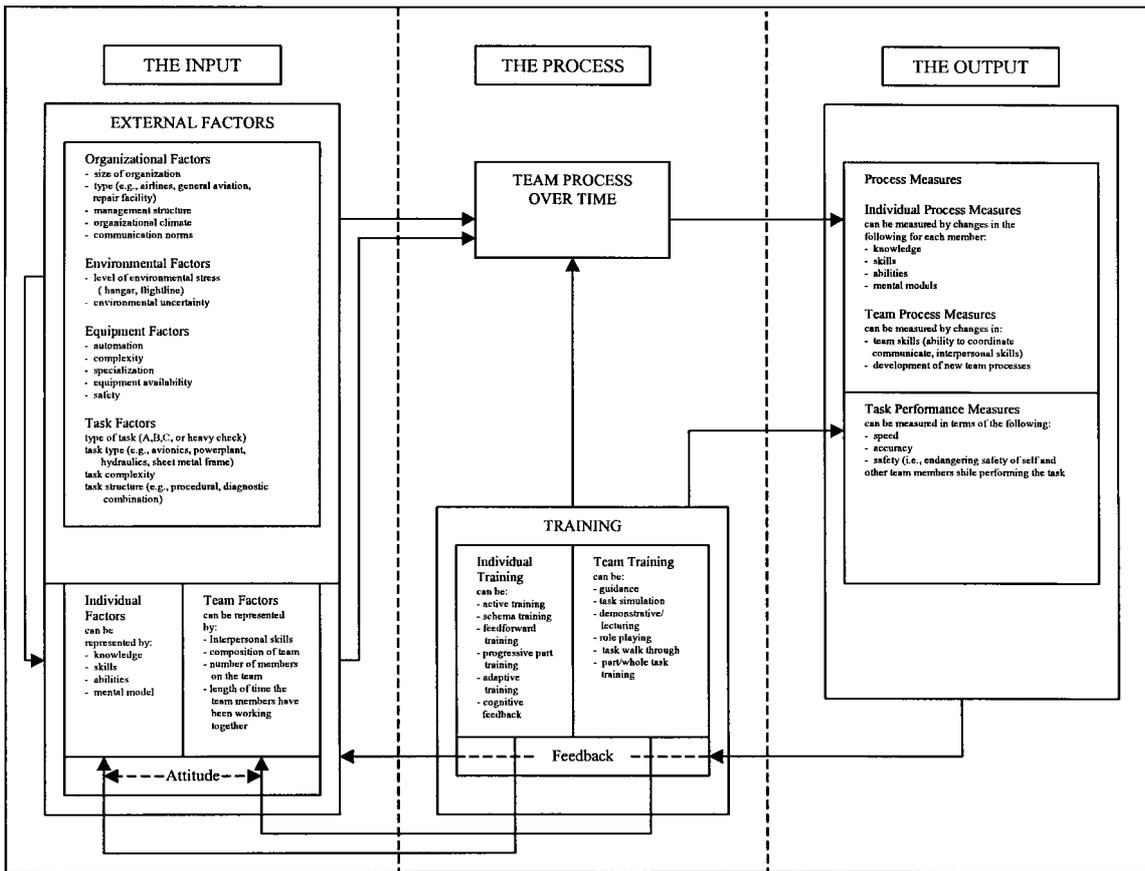
Figure 1. Team training model for the aircraft maintenance environment
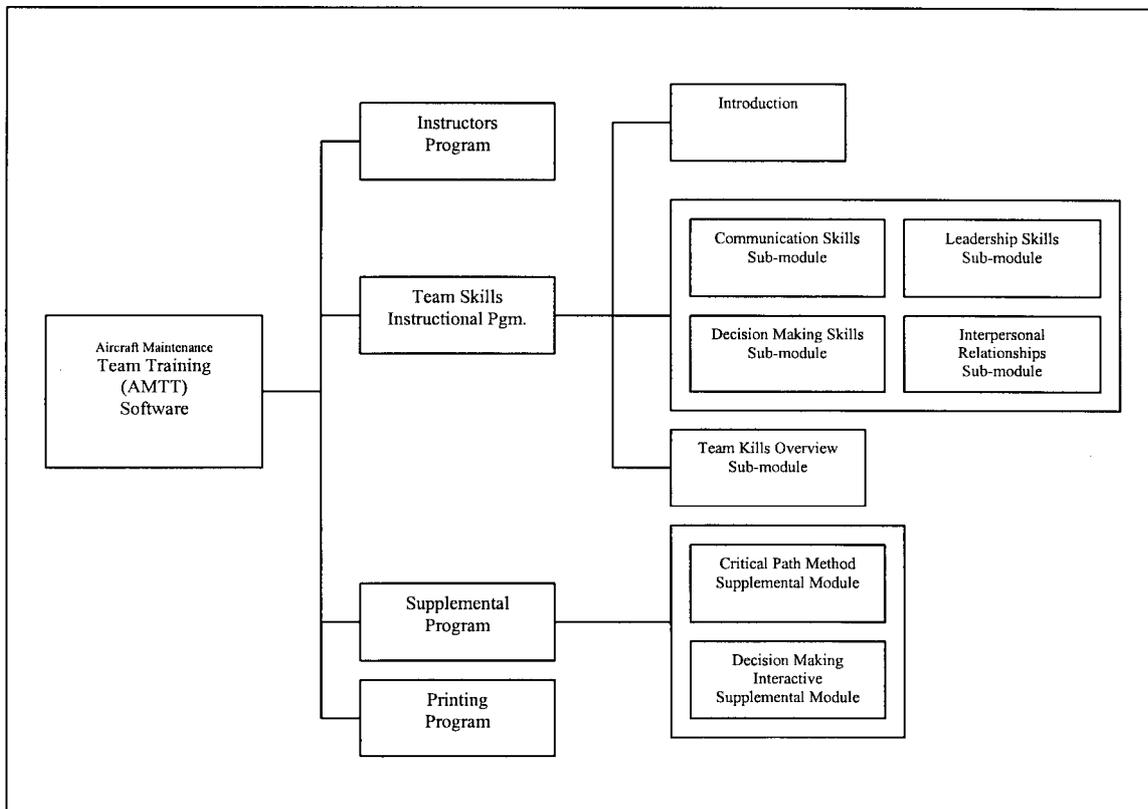
Figure 2. Layout of the AMTT software

of the main contextual material. Since the software was developed as a training and research tool, the software facilitates the collection of pretraining and posttraining performance data.

AMTT is divided into four major programs (Figure 2): (1) team skills instructional program, (2) instructor's program, (3) printing program and (4) the supplemental program. The team skills program and the supplemental program have been designed for use by the aircraft maintenance technician undergoing team training, but the remaining two programs are for use by the instructor or supervisor. An aircraft maintenance technician interacting with the AMTT software first uses the team skills instructional program, which initially provides an introduction to the software. Following this step, the AMT is provided with instruction on basic team skills through four separate team skill submodules: communication, leadership, decision making and interpersonal relationships. These submodules not only emphasize and cover generic material related to these skills but also relate the importance and use of the specific skills within the aircraft maintenance environment. These are the same four skills emphasized in an earlier FAA (1995) report which looked at the role of team training in the aircraft maintenance environment.

Figure 3 shows a prototypical layout of a team skills submodule. The right-hand side of the screen is dedicated to key points being discussed in the voice-over, whereas the left-hand side of the screen provides supporting material. This supporting material comes in a variety of formats which include, but are not limited to animation, video, photographs, diagrams and flowcharts. Buttons on the command line at the bottom of the screen can be clicked on to exit, advance, back up, stop and replay audio, replay video, and access the navigation map. Each of the team skills submodules has a similar structure.

3. METHODOLOGY

3.1. Test Site
The controlled study was conducted at the Aircraft Maintenance Technology Center of Greenville Technical College (GTC). The center houses classrooms for A&P training and a fully equipped hangar for conducting aircraft maintenance and repairs. The classrooms at the Aircraft Maintenance Technology Center provide seating for 20 students. Each classroom was equipped with a 25-inch color

Figure 3. Prototypical layout of a team skills submodule

television, video player, overhead projector, whiteboards and blackboards, and a lectern. In addition, the classrooms were equipped with four Pentium 75 MHz computers and 15-inch color monitors (1024 ¥ 768 resolution) installed with multimedia packages.

3.2. Test Subjects
The subjects for this study consisted of 12 students from a local aircraft maintenance technology program and 24 licensed A&P mechanics from a local aircraft maintenance facility. The subjects were compensated for their participation. The 36 subjects were assigned to two groups: IBT for instructor-based training, and CBT for computer-based training. Each group had equal number of subjects from the aircraft maintenance technology program and the maintenance facility.

3.3. Experimental Procedure
3.3.1 Instructional phase
Subjects in the IBT group were trained on team concepts using a traditional instructor-based training delivery system, whereas those in the CBT group received similar training on a computer using the AMTT software. Every effort was made to maintain a constant curriculum and presentation sequence for both groups. The only difference in the training between the two groups was the type of delivery system. The team skills training focused on four separate skills: communication, decision making, interpersonal relationship and leadership. More details on the structure and content of the team training program can be found in Gramopadhye *et al.* (1995). Note that in the instructional phase, team training was provided to individuals, and teams were not formed until the evaluation phase.

3.3.2 Evaluation phase
This phase examined the transfer effects of team training (IBT and CBT) on AMT performance. Upon completion of the instructional phase, subjects in each group were assigned to six three-member teams. Following the assignments, the teams were given two tasks representative of normal aircraft maintenance: a routine maintenance (RM) task to determine the center of gravity of an aircraft, and a nonroutine maintenance (NM) task to troubleshoot an electrical problem on an aircraft. The order in which the tasks were performed was balanced within each group.

4. DATA COLLECTION
4.1. Instructional Phase
Before the training was initiated, all subjects completed a team skills perception questionnaire for each team skill being taught (communication, decision making, interpersonal relationships and leadership); figure 4 shows an example. The questionnaire used elements from the crew resource management/technical operations questionnaire (CRM/TOQ), a modified version of Taggart's questionnaire (1990), Taylor's questionnaire (1990) and the critical team behavior form (CTBF) (Glickman *et al.* 1987). Each questionnaire consisted of 10 questions on a seven-point Likert scale, and each was designed to measure the subject's perception of a particular team skill. In addition to the team skills perception questionnaire, each subject completed a 20-question multiple-choice knowledge test before training (figure 5). The objective of the knowledge test was to measure each subject's knowledge on the different team skills before training.
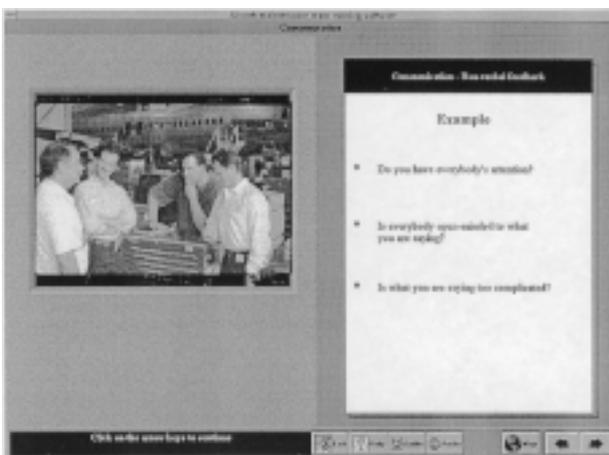
1. Good communication and team coordination are as important as technical proficiency for aircraft safety and operational effectiveness

| Very strongly disagree | 1 | 2 | 3 | 4 | 5 | 6 | 7 | Very strongly agree |

2. Crew leaders and supervisors should encourage questions during work and in special situations

| Very strongly disagree | 1 | 2 | 3 | 4 | 5 | 6 | 7 | Very strongly agree |

3. The start of shift team meeting is important for safety and effective team management

| Very strongly disagree | 1 | 2 | 3 | 4 | 5 | 6 | 7 | Very strongly agree |

Figure 4. Sample questions from team skills perception questionnaire: communication

After completing the team skills perception questionnaire and the knowledge test, the subjects received team skills training using the appropriate delivery system. Upon completion of the training, the same team skills perception questionnaires and the knowledge tests were readministered to the subjects. The purpose in repeating the same perception and knowledge tests was to measure the changes in perception and knowledge that had occurred during training. Each subject also completed two sets of usability questionnaires. The questionnaires collected subjective satisfaction ratings on the training delivery system using a seven-point Likert scale, where 7 indicated strong agreement and 1 indicated strong disagreement. The first questionnaire, called the general questionnaire (figure 6), contained questions relevant to both the training delivery systems; it was completed by subjects in both the groups. The general questionnaire addressed usability issues related to content, mechanics of presentation, format and usefulness. The second part of the usability questionnaire was specific to the training delivery system; it addressed usability related to presentation and format.

4.2. Evaluation Phase

As the teams performed the routine and nonroutine tasks, their performance on the tasks was evaluated by three independent evaluators on measures of speed, accuracy and safety. In addition, at the conclusion of the routine and nonroutine maintenance tasks, the evaluators (instructor's evaluation) and each individual subject (self-evaluation) completed a questionnaire evaluating their team on the application of various team skills

(communication, decision making, interpersonal relationships and leadership). The evaluation consisted of a questionnaire with seven questions on each of the team skills, wherein the evaluator rated the team on their application of the team skill on a seven-point Likert scale.

5. RESULTS

5.1. Analysis of Instructional Phase

5.1.1 Knowledge test.

The knowledge test was analyzed to understand the effect of team training on the subject's knowledge of team skills. The knowledge test analysis used a 2 ¥ 2 design – two delivery methods (CBT, IBT) by two trials (pretraining, posttraining) – with 18 subjects nested within each group. The method of training delivery was the between-subjects factor and the trial was the within-subject factor. A comparison of the pre- and posttraining knowledge test showed that in every skill category there was a significant increase in the test scores after training. The fact that both training programs showed comparable increases in test scores probably indicates the effectiveness of both methods of delivering team training. These results are encouraging about the potential of the team training program in imparting team skills knowledge. The fact that the CBT (specifically the AMTT software) was able to achieve the same scores as IBT (an equally well-developed classroom team training program) bodes well for the role of computers in imparting team skills knowledge. In other words, given equivalent content of the team training program, a well-

Your crew chief is talking to the team and when he says something significant and you nod your head in agreement…
    A. It displays a lack of desire to communicate.
    B. You are communicating with body language.
    C. You should keep very still so as to not confuse the speaker
    D. You should remain quiet since verbal and non-verbal communication do not mix.

When a supervisor summarizes what he said and then asks several questions of the team member, he is…
    A. Grading the team members.
    B. Using an investigative management technique.
    C. Using reverse psychology to get his team to work harder.
    D. Looking for feedback to see if he got his message across.

Figure 5. Sample questions from knowledge test: communication

Contents

The amount of information presented was adequate.

| 1 | 2 | 3 | 4 | 5 | 6 | 7 |

Very Strongly
Disagree   Neutral   Very Strongly
                                                                                                  Agree

Mechanics of Presentation

The videos were helpful in understanding the concepts presented.

| 1 | 2 | 3 | 4 | 5 | 6 | 7 |

Very Strongly
Disagree   Neutral   Very Strongly
                                                                                                  Agree

Format

The information presented flowed smoothly.

| 1 | 2 | 3 | 4 | 5 | 6 | 7 |

Very Strongly
Disagree   Neutral   Very Strongly
                                                                                                  Agree

Format

The knowledge gained from the communication module was useful.

| 1 | 2 | 3 | 4 | 5 | 6 | 7 |

Very Strongly
Disagree   Neutral   Very Strongly
                                                                                                  Agree

Figure 6. Sample questions from General usabiltiy Questionnaire

designed interactive computer-based team training program can be as effective as the traditional instructor-based team training program.

5.1.2 Perception

The team perception questionnaire was developed to quantitatively measure an individual's perception of team skills before and after training. The results of the pre- and postperception questionnaire on the four team skills showed that the training delivery system had comparable effects on the subject's perception of team skills. Figure 7 compares the perception scores for the two groups on each of the four skills.

5.2. Usability

The development of the AMTT software followed an iterative design process to ensure that problems with the software design
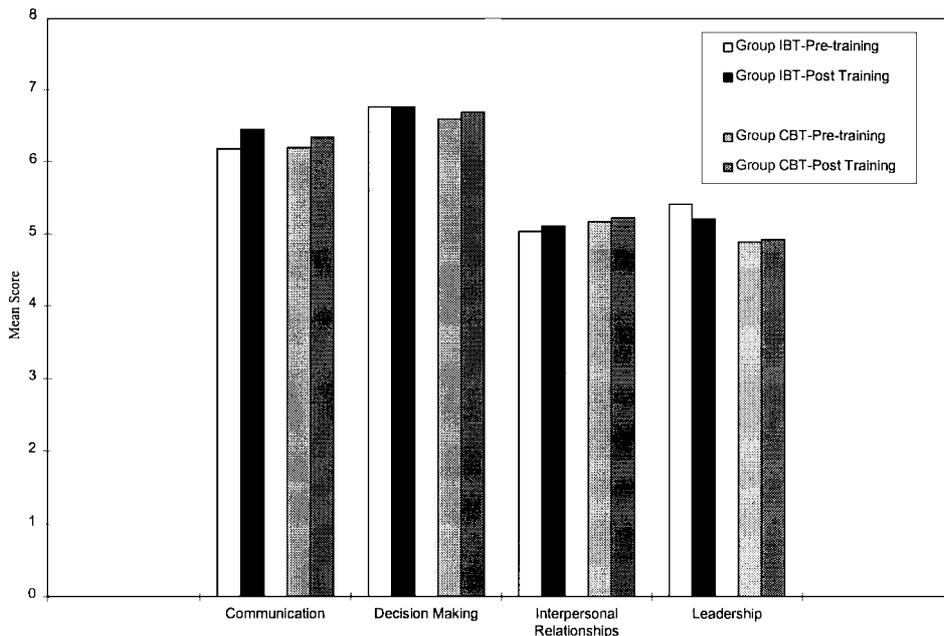
Figure 7. Team skills perception pre and post training for groups IBT and CBT

were identified and corrected before implementation. The cycle of design, test, measure and redesign was repeated a number of times in the development process (e.g., Gould and Lewis 1985). The user interface capitalized on the following: graphical user interface technologies, human factors research (i.e., color, formatting, layout, etc.), ease of use, and information utilization (Chabay and Sherwood 1992). Thus the AMTT software was developed after understanding the needs of the AMT, discussions with experts from Lockheed Martin and Greenville Technical College, following a process of iterative design and development, and eventually resorting to detailed user testing (with instructors, supervisors and AMTs).

Figure 8 shows the results of the general usability questionnaire with mean score for the four separate usability issues: content, mechanics, format and usefulness. For the various usability issues, there were no significant differences in the levels of satisfaction between the groups. Also, when the four usability issues were consolidated (combined average score), there was no significant difference between the groups. Both of the training delivery systems had high scores on system-specific usability issues. These results were encouraging since they indicated that the users were equally satisfied with either a highly interactive computer environment or a traditional classroom environment.

5.3. Analysis of Evaluation Phase

On completion of both the routine and nonroutine maintenance tasks, the subjects in groups IBT and CBT scored their team on the application of the four team skills. As each team completed the routine and nonroutine maintenance tasks, the evaluators also scored the teams on their use of the four team skills. The instructor's evaluation did not reveal any significant differences between the groups in each of the team skills categories (figure 9). Thus, according to the evaluators' scores, the training delivery systems had comparable effects on the subjects' use of team skills in the performance of maintenance tasks. While the teams conducted the two tasks (routine and nonroutine maintenance), the evaluators recorded the performance of the teams on accuracy, speed and safety violations. Using all three performance measures, there was no statistically significant difference between the two groups on either the routine or the nonroutine maintenance task. Thus, the training delivery system had no significant effect on task performance measures.
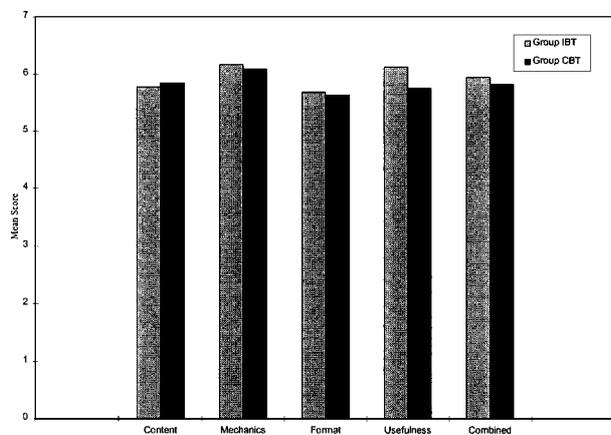
Figure 8. Usability scores for groups IBT and CBT on training delivery issues
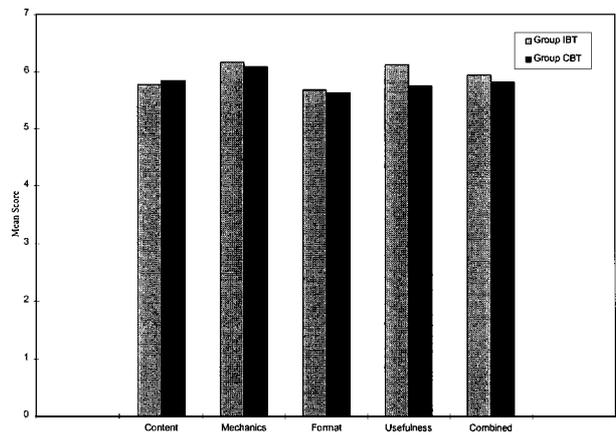
Figure 9. Evaluator's comparison of team skills while conducting tasks (RM and NM) for groups IBT and CBT

6. DISCUSSION AND CONCLUSIONS

The goal of this research was to understand the role of computer-based team training specifically within the aircraft maintenance environment. Part of the research involved development of the AMTT software, a computer-based team training package And its usefulness was tested against a traditional classroom method of instruction in terms of team knowledge acquisition, usability issues and knowledge transfer. Here are the important conclusions:

1. Team training enhanced the knowledge of individuals on team skills However, the type of training delivery system did not have a significant effect on the individual's ability to acquire team skills knowledge. Thus, a well-designed computer-based team training program was equally effective in imparting team skills.

2. Many times CBT fails because software designers do not design interfaces and systems that users can understand. A user-centered design approach is required with an iterative process of design, test, measure, modify and retest. This procedure was used in the development of the AMTT software, and a user-friendly product was produced. There were no significant differences between IBT and CBT in terms of user satisfaction. Both training delivery systems reported a high level of user satisfaction on the general and delivery-specific portions of the usability questionnaire.

3. The results are unequivocal. CBT (i.e., AMTT) was as effective in delivering team training instruction as IBT. Finally, the iterative design methodology employed in this study proved to be useful in designing effective computer-based team training software. These results have obvious ramifications for the use of AMTT in team training within the aircraft maintenance environment. Besides being as effective as existing instructor-based methodologies, AMTT team training has other obvious advantages:

- *Standardization*: The AMTT software provides a standardized and systematic team skills training program which aircraft maintenance instructors (at certified repair stations, airline companies, general aviation stations and A&P schools) can use to provide team skills training.
- *Adaptability*: Maintenance training has traditionally been accomplished via on-the-job training or classroom training,

both of which are labor-intensive. It requires careful scheduling of personnel or encumbers others in the training process. AMTT is adaptive, self-paced and can be done at convenient times when trainees are available, and it need only involve the person being trained.

- *Record keeping*: The record-keeping capabilities of AMTT tracks the student's progress. This information can be used by the instructor/supervisor to design remedial training.
- *Cost-effectiveness*: Team training using AMTT can be cost effective because: (1) It can be delivered on-site thus eliminating travel expenses for the trainer and the trainee. (2) It can minimize downtime by providing training at times that are convenient to the trainee and the company's work schedule. In larger organizations, AMTT can be delivered to many people at multiple sites thus proving to be cost effective.
- *Use of advanced technology*: Many facilities (e.g., A&P Schools and fixed based general aviation facilities) do not have access to larger aircraft. The AMTT software provides team skills training against the backdrop of maintaining a DC-9. Thus the trainees not only acquire knowledge and skills on teamwork, but also gain an understanding of the importance of teamwork in the maintenance of wide-bodied aircraft.

ACKNOWLEDGMENTS

REFERENCES

CHABAY, R.W. and SHERWOOD, B.A., 1992, A practical guide for the creation of educational software. In Larkin and Chabay (eds) *Computer-Assisted Instruction and Intelligent Tutoring Systems: Shared Goals and Complementary Approaches* (Hillsdale NJ: Lawrence Erlbaum).

DRURY, C.G., PRABHU, P., and GRAMOPADHYE, A.K., 1990, Training for visual inspection. *Proceedings of the Third Federal Aviation Administration Meeting on Human Factors in Aviation Maintenance and Inspection: Training Issue*, Atlantic City NJ (Washington DC: FAA).

FAA (FEDERAL AVIATION ADMINISTRATION) OFFICE OF AVIATION MEDICINE, 1991, *Human Factors in Aviation Maintenance – Phase One: Progress Report*, DOT/FAA/AM-91/16 (Washington DC: FAA).

FAA (FEDERAL AVIATION ADMINISTRATION) OFFICE OF AVIATION MEDICINE, 1993, *Human Factors in Aviation Maintenance æPhase Three, Volume 1 Progress Report*, DOT/FAA/AM-93/15 (Washington DC: FAA).

FAA OFFICE OF AVIATION MEDICINE, 1995, *Human Factors in Aviation Maintenance - Phase Four* Report, DOT/FAA/AM (Washington DC: FAA).

GLICKMAN, A.S., ZIMMER, S.R., MONTERO, C., GUERETTE, P.J., CAMPBELL, W.J., MORGAN, B.B. Jr, and SALAS, E., 1987, The evolution of teamwork skills: an empirical assessment with implications for training. *Technical Report 87-016* (Orlando FL: Office of Naval Research).

GOULD, J.D and LEWIS, C., 1985, Designing for usability: key principles and what designers think. *Communications of the ACM*, 28(3).

GRAMOPADHYE, A.K., KRAUS, D.C., PRADEEP, R., and JEBARAJ, D., 1995, Team training for the aircraft maintenance technician: the aircraft maintenance team training (AMTT) software. *Proceedings of the Tenth Federal Aviation Meeting on Human Factors Issues in Aircraft Maintenance and Inspection* (Washington DC: FAA).

HACKMAN, J.R., 1990, A normative model of work team effectiveness. *Technical Report 2* (New Haven CT: Yale University).

JOHNSON, W.B., 1990, Advanced technology training for aviation maintenance. *Final Report of the Third FAA Meeting on Human Factors Issues in Aircraft Maintenance and Inspection*, Atlantic City NJ (Washington DC: FAA).

KRAUS, D.C., GRAMOPADHYE, A.K., and BLACKMON, R.B., 1996, Teams in the aircraft maintenance environment. *Proceedings of the Fifth Industrial Engineering Research Conference*, pp. 441–46.

PHILLIPS, E.H., 1994, Focus on accident prevention: key to future airline safety. *Aviation Week and Space Technology*, 52–53.

SHEPHERD, W.T., 1991, *Human Factors in Aviation Maintenance, Phase 1: Progress Report*, DOT/FAA/AM-91/16 (Washington DC: FAA).

TAGGART, W.R., 1990, Introducing CRM into maintenance training. *Proceedings of the Third Federal Aviation Administration Meeting on Human Factors Issues in Aircraft Maintenance and Inspection* (Washington DC: FAA), pp. 93–110.

TAYLOR, J., 1990, Facilitation of information exchange among organizational units within industry. *Proceedings of the Second Federal Aviation Administration Meeting on Human Factors Issues in Aircraft Maintenance and Inspection* (Washington DC: FAA), pp. 69–78.

Teamwork

C.H. Antoni

ABO-Psychologie, Fachbereich I, Universität Trier, Universitätsring 15, D-54286 Trier, Germany

1. INTRODUCTION

Effective teamwork has been crucial for mankind all throughout history. It has proven helpful not only when hunting mammoths in prehistoric times, but also in contemporary research projects or in high-tech industrial and service companies. The need for teamwork even grew with increasing task complexity, international division of labor and interdependence. In companies all over the world this development is reflected in the multitude and variety of teams at all levels and in all functions. Teamwork is, however, no guarantee for success. On the contrary, social psychologists have found it much easier to prove process losses rather than synergy gains due to teamwork. Potential pitfalls such as group think due to selective information processing, social loafing due to reduced individual motivation in groups or coordination losses can easily reduce group efficiency. Nevertheless, growing task complexity and interdependence require more and more teamwork in different settings and different types of teams. A better understanding of the basic concepts of teamwork, different team models and critical factors of team effectiveness might help to implement teamwork successfully.

2. CONCEPTS AND CRITERIA

The widespread discussion of teamwork did not reduce but rather has led to increased confusion about the concept. The managerial demand "we've got to be a team" refers to companies with thousands of employees as well as to boards of management comprising just a few people. The term "team" is used to stress corporate or team spirit. In other cases "team" is used to label an organizational entity of people having equivalent functions, such as security, or having complementary tasks, such as product development. In the latter case, "group" is often preferred whereas "team" is used to stress the importance or existence of common spirit.

In the scientific discussion and in this text, "team" or "group" are, in most cases, used interchangeably because a clear distinction can not be made. Social psychologists define "teams" as two or more persons who are interacting and influencing each other. Small-group research customarily considers at least three people, because group processes such as coalition formation start at this size. On the other hand, they restrict teams to ~20 people to have significant direct interactions and mutual influence. In addition to size, interaction and influence there are other criteria such as group roles, norms, cohesion and goals. They are not always included in definitions of "team" or group" because they evolve automatically in all kinds of teams due to the specific interaction and influence processes. Hence, these aspects differ in type and strength between teams.

In contrast to naturally developing informal teams, formal teams in organizations are defined by a common task, which is usually assigned by a supervisor. A team might be formed according to task requirements or a suitable task is assigned to an already existing team. However, teamwork requires not only the assignment of a common task, but also a shared and accepted notion of this task by the team members. To develop this mental model of what has to be done a communication process between the supervisor or client but also within the team is required.

Teams differ in the type of tasks they are assigned and in the way they work on these tasks. These aspects are crucial for the kind of teamwork which will develop. It is particularly important whether and how the assigned task can be split reasonably into different subtasks and whether these subtasks are interdependent or not. For example, one person usually types a letter. Hence, typists in a typing pool usually divide their labor by quantity and according to a "first in, first out" principle. In this case individual performance is independent of other team members.

One might reasonably question whether this can be called "teamwork" at all, since at first glance no common goal exists. Nevertheless, this pool of, for example, say five typists might be referred to as the typing team of the company. They might feel as a team as well, because they share one room, they chat with each other and help each other out, for example, to spell correctly or to solve computer problems. These aspects suggest team goals, which might be implicit or explicit, given or self set, such as we want team spirit and satisfaction, and we help each other out, to achieve top quality and efficiency. It is easy to set quantified goals even for such teams, for example to have all letters completed the day they come in. The core task for such a group is to organize their work in such a way that this goal can be achieved, whereas no cooperation is needed during the regular work (letter typing) itself.

Obviously this kind of teamwork is quite different from tasks that all team members have to work on successively or simultaneously. For example, on assembly lines, team members work individually as well but the product is built step by step along the line, and the team members depend on each other successively. Task interdependence is even stronger if team members simultaneously work on the same product, for instance at manufacturing cells, or deliver the same service such as a surgery team. Particularly the latter teams share not only the secondary (or indirect) tasks such as planning, coordinating and problem-solving, but also the primary (or direct) task. For these reasons the task immanent cooperation requirements are much stronger and therefore the need for shared mental models, including goals, strategies, norms and role expectations. These consequences of task assignment must be kept in mind if one wants to implement teamwork in a company.

3. TEAMWORK IN INDUSTRIAL AND SERVICE ORGANIZATIONS

In industrial and service organizations one finds quite different models of teamwork (Antoni *et al.* 1994). Their functions differ according to their use as regular work units or as specific problem-solving instrument.

3.1. Team-based Work Organizations

Teams have been used for a long time as a means to get the daily work done more efficiently. Examples for such classic work teams are the above-mentioned typist teams, construction teams or teams in a research and development departments, and so on. In manufacturing departments, teams were traditionally organized

according to Tayloristic principles, that is division of labor and functional specialization. For example, team members might perform the same specialized logistic or pre-assembly tasks, such as filling up supply boxes or inserting screws, or working on specialized positions on an assembly line.

In the past decade, functionally integrated work teams with multi-skilled members replaced classic work teams more and more following the paradigm of lean production. This proved much more efficient because multi-skilled team members can replace one another if one is absent. Also indirect labor costs, maintenance and other unproductive time are saved because the team members can do quality control and minor repairs themselves.

A supervisor might regulate these teams or be given some autonomy to regulate themselves, as proposed in the concept of "self-regulating" or "semi-autonomous" *work teams*. This concept is based on *socio-technical system theory* (Emery 1972) and is also theoretically well supported by the *theory of action regulation* (Hacker 1998). Socio-technical system theory proposes that self-regulated groups can respond faster and better to changing internal or external requirements, given the team has the needed know-how and other resources. This conception holds particularly well under dynamic environmental conditions and technical interdependent tasks, which require employee information processing. Furthermore, giving discretion to the team fulfills members needs for control and promotes skill and personal development (Emery 1972).

Team autonomy or participation in decision making has long been discussed from a motivational perspective only. Even more important might be that participation in decision making has important cognitive effects: task-relevant information is exchanged and task adequate action strategies are commonly developed. As a consequence, decisions rely on a broader information base and shared mental models can evolve. Team members learn to develop action strategies and to adapt them to changing task requirements.

Similar effects occur if "complete tasks" are delegated. Complete tasks require goal setting, planning, organizing, performance and control. They also require regulation processes not only on a motoric or perceptual, but also on an intellectual control level. These sequentially and hierarchically complete tasks further the development and elaboration of task adequate and efficient action strategies (Hacker 1998). Thus, task immanent learning processes can develop as well as preventive action planning to avoid scrap and to solve problems in advance (Hacker 1998). This problem-solving capacity of teams has even led to special types of teamwork which focus on unique and complex problems that can not be dealt with efficiently during daily business.

3.2. Problem–solving Teams

Besides routine task accomplishment, teams have proven themselves particularly helpful in complex problem solving. *Project teams* or *task forces* are set up to solve new, unique and complex problems affecting several organizational functions. For example, in Germany 87% of the top 100 organizations (in terms of revenue) used this type of teams in 1994. Depending on the kind of problem, experts from several organizational functions are sent to a project team for a limited period to solve this problem.

Typical occasions might be new product development, software implementation or re-engineering projects which affect complete process chains.

Project teams can differ whether the team leader or team members do their job part or full time. If team members still have duties on their regular job this often leads to conflicts in time resources between task forces and line functions. Since team members mostly still report to their functional supervisor at least in Germany, project teams are likely to run into serious difficulties and power conflicts. To prevent or resolve these conflicts, powerful clients are chosen for important projects, clear client–contractor agreements are made and steering committees are implemented to define boundary conditions and to control project progress. The importance of project teamwork can be stressed by additional measures, such as team-based bonus systems dependent on goal achievement or individual performance ratings by the team leader complementary to ratings done by the line manager.

Another critical factor for project success is team member composition. Ideally members have task relevant but complementary skills and expertise and strive cooperatively for the common goal. Hence, team members should not behave primarily as representatives of their home departments and stress those particularistic interests. Furthermore, qualification and not status should determine membership. Consequently project teams should rather encompass different hierarchical levels and the relevant functions of the process chain than to be restricted to staff departments.

Similar considerations that expert knowledge can be found at all hierarchical levels has led to the concept of "quality circles". Quality circles are traditionally small groups of blue-collar workers who meet voluntarily to discuss and solve their quality and other work problems. The primary goal of quality circles is quickly to detect and solve problems that directly affect team members. Particularly, in Tayloristic work structures, quality circles may be the only means for blue-collar workers to discuss and solve their problems commonly.

Quality circles usually meet twice a month for one or two hours but might increase meeting frequency or length in accordance with problem and production requirements. The time spent in meetings is regarded as regular work and paid regardless whether circles meet before, during or after regular working hours. Particularly assembly line and other sequential work structures often do not allow meetings during regular work time, thus demanding additional motivation to work overtime.

Although participants are experts in their special field of work they are usually not familiar with collective problem solving. Hence, it is useful to qualify a member in group dynamics and problem-solving methods to be able to organize and moderate quality circle meetings. If necessary, quality circles might invite experts from other departments to get necessary information or discuss problem solutions. To solve more complex cross-functional problems, members from different affected departments might as well form a cross-functional quality circle. Quality circles are well advised to start with simple problems, to get used to teamwork and to experience quick success. Success will not only motivate circle members but also might convince critics outside. Long response times, rejected problem solutions and delayed solution implementation by line management are

key causes of quality circle failure. Similar to project teams, the effectiveness of quality circles heavily depends on organizational context factors.

3.3. Models of Team Effectiveness

In the meantime a number of models of team effectiveness have been proposed. Most follow an "input–process–output" scheme (Guzzo and Shea 1992). Supporting empirical evidence still remains a problem. This might be due to model complexity as well as to practical difficulties in realizing (quasi-) experiments in field settings. Besides socio-technical system theory outlined above, one of the most prominent models is still Hackman's (1987) normative model of group effectiveness.

Hackman (1987) defines group effectiveness in terms of three aspects: (1) to get acceptable results for those who receive or evaluate them; (2) to improve ability of group members to cooperate on future tasks; and (3) to satisfy member needs. Acceptable results and member satisfaction reflect a stakeholder view on group effectiveness whereas ability for cooperation takes group resources into account. These output variables depend primarily on three process criteria: the amount of effort expended by team members; their knowledge and abilities; the adequacy of the task strategies applied by the team. In addition to these process aspects, sufficient resources are needed as well for group effectiveness (such as sufficient material, technical, and personnel resources).

The three process criteria are influenced by organizational context variables, group design and group synergy. Important organizational context variables are the reward, training and information system. Group design influences task performance by task structure, group composition and group norms. Group synergies support group interaction by reducing process losses, such as group think or social loafing and by creating process gains, for example, by a strong team spirit. Leadership factors are not explicitly considered but are nevertheless important. Group tasks, goals, feedback and rewards are influenced directly by leadership behavior or by decisions concerning task structure, reward or information systems. No specific recommendations are given concerning leadership behavior. Management decisions about group autonomy should take into account organizational factors, such as organizational culture, as well as task characteristics and employee expectations and abilities.

In this respect socio-technical system theory is more specific. Contingency factors supporting team self-regulation are high employee social needs and needs for growth, high technical interdependence of tasks and high technical uncertainty requiring employee information processing and decisions, as well as high environmental dynamics demanding quick system adaptation. In contrast, low needs for growth, low technical uncertainty and low environmental dynamics all call for externally regulated groups.

In other respects the two models have similar or complementary propositions. For example, both stress the importance of task design for team member motivation. They are complementary because Hackman's focus is on team interaction, whereas socio-technical system theory focuses on the relation between the social and the technical system. For these reasons, both models can be easily combined to deduce recommendations for implementing and improving teamwork.

4. RECOMMENDATIONS

Socio-technical system theory has shown that effective teamwork can only be achieved if both social and technical factors are considered. To analyze them, proven methods for socio-technical system analysis exist that focus on employee needs, technical, social and environmental system variances. These analyses and consequent system design should be done participatively, joining managers, employees and consultants in a project to implement teamwork.

Participative implementation strategies have cognitive and motivational effects. They ensure that relevant information is gathered and shared. The development and implementation of specific team models can be based on this information. During the implementation process employees and managers learn about one another. They also start learning teamwork on the job as they participate in project teams doing the analyses, team model development and implementation. Furthermore, their motivation to engage in this process as well as their acceptance of process results is likely to rise, if they are informed and experience influence, and control.

Team effort, knowledge and task strategies as crucial variables for team effectiveness can be influenced by the way teams and their organizational context are designed. Teams need interdependent tasks. Otherwise common team goals cannot be set, and teamwork cannot develop. Task requirements have to be matched with employee skills and abilities. In a constant process of dynamic task design, task requirements can be increased with growing employee qualification. When teams are composed of members who have heterogeneous skills they can easily qualify each other on the job. Management has to ensure that sufficient resources in manpower are provided to handle daily business and on the job training.

In addition, management can support team effectiveness by designing the organizational context (Antoni *et al.* 1994). For example, pay systems considering employee qualification will reward team members efforts to qualify for new tasks. Team oriented pay systems can strengthen team resources as well as customer orientation. Customer focus can be supported by performance components in pay systems, for example, by rewarding improvements in timely delivery or reducing customer complaints. Pay systems, however, should avoid too much team pressure. This can be achieved if individual differences are considered and if individual salary is not too much dependent on team performance.

Performance based pay systems are only useful if they rely on a team based goal setting and feedback system. Otherwise teams have no chance to plan and to control their actions. Specific, challenging, group goals direct team member actions in the same direction, increase their effort, and stimulate the development of adequate task strategies. The pursuit of common goals requires shared mental models of what has to be achieved and how it can be done cooperatively. Team-building efforts can speed up this process, especially if teams get a kind of "kick-off" training.

Team development training can be used to clarify role expectations, to shape group norms, and to increase team spirit. These measures promote group synergy and reduce process losses. Similar effects can be achieved by assigning subgoals to team members and providing individual as well as team feedback. Furthermore, process gains can be attained if group working

techniques such as moderating and visualizing team meetings are used.

These recommendations will not guarantee effective teamwork but if they are kept in mind, teamwork might be implemented more effectively. Although teams are not the only building blocks of organizations, teamwork might become even more crucial for future company success.

REFERENCES

ANTONI, C.H, EYER, E. and KUTSCHER, J., 1994, *Das flexible Unternehmen* (Wiesbaden: Gabler). [http.//www.flexible-unternehmen.com.]

EMERY, F.E., 1972. Characteristics of socio-technical systems. In L.E. Davis and J.C. Taylor (eds) *Design of Jobs* (Harmondsworth: Penguin), 177–98.

GUZZO, R.A. and SHEA, G.P., 1992, Group performance and intergroup relations in organizations. In M.D. Dunnette and L.M. Hough (eds) *Handbook of Industrial and Organizational Psychology* (Palo Alto: Consulting Psychologists Press), 609–45.

HACKER, W., 1998, *Allgemeine Arbeitspsychologie. Psychische Regulation von Arbeitstätigkeiten* (Bern: Huber).

HACKMAN, J.R., 1987. The design of work teams. In J.W. Lorsch (ed.) *Handbook of Organizational Behavior* (Englewood Cliffs: Prentice-Hall), 315–42.

Technology Transfer

H. Shahnavaz

Division of Industrial Ergonomics, Department of Human Work Sciences, Luleå University of Technology, 971 87 Luleå, Sweden

1. INTRODUCTION

Technological development over the past century and especially since the Second World War has contributed enormously to economic growth and social progress in the industrialized world. It is beyond doubt that technological advancement has elevated the standard of living of mankind and has contributed greatly to the reduction of many sources of occupational accidents, injuries and stresses. However, advanced technology has also brought new sources of work stress and injuries. The complexity of the modern technology and the changing nature of work, work organization and production systems are placing extra high demands on the workforce. As a result, the degree and pattern of problems at work are changing along with the technological changes.

Further, competition at the global market imposes higher challenges on productivity, and product quality, which in turn require a knowledgeable workforce with broader skills, flexibility, good motivation, and self-regulation. As a result, progressive firms are shifting from a technology- to a human-centered approach, regarding people as central to all development initiatives. This notion highlights the importance of ergonomics in the development and application of technology to ensure that the technology is accepted and fitted to its users and the operating environment for safe and effective use. Only by considering ergonomics in the design, introduction and utilization of technology can system performance be optimized, achieving higher productivity as well as improved worker health and satisfaction.

2. ERGONOMICS: AN INVALUABLE TOOL FOR TECHNOLOGY DEVELOPMENT IN INDUSTRIALIZED COUNTRIES

Among industrialized countries (IC), the present technological advancement has likewise contributed to the growing awareness and recognition of the importance of ergonomics. It has become common experience that the same engineering and technical breakthroughs that propelled the current surge of industrialization can be harnessed to bring about technical innovations, answers and solutions to numerous work/work environment problems accompanying such growth. Consequently, the sphere of ergonomics, occupational health and industrial safety has expanded to help apply these realizations. For example, legal demands for risk assessment and preventive programs, when planning and organizing work have been initiated. Joint efforts of employers and employee organizations, as well as the establishment of joint industrial safety councils, have contributed a great deal to the improvement in industrial safety, occupational health and ergonomics' services. Even the development of recommendations, directives and standards concerning these issues has become an important part of work environment activities. Issues dealing with people and how they perform are considered as important as the technology *per se* and they are now linked to the overall company policies. At the same time, cooperation between managers and employees in participative problem-solving process has become a successful practice in many workplaces.

Thus, it can be said that IC are trying to keep pace with the technological advancement through the development of know how and legal and administrative structures for protecting the work force by making optimum use of technology.

3. INDUSTRIALLY DEVELOPING COUNTRIES: THE PREVAILING SITUATION

Even today, industrially developing countries (IDC) share some common problems of poverty, low productivity and low product quality. In their strive for an overall improvement of the quality of life through economic growth, an *industrialization* based mostly on importation of advanced technology is often considered as the best solution. Moreover, most of the IDC a rapid pace of industrialization — a speeded-up version of the historical process of industrial development in IC — is assumed to be the best way to overcome the economic problems. This concept of industrialization has, in fact, become for many synonymous with national development and social progress. Unfortunately, due to several complex socio-cultural, economic, human and technological factors, such policy is not always successful in terms of leading to any significant improvement in the countries' economy or the people's quality of life. In so doing, inappropriate technology transfer (TT) is considered as one of the reasons for the current situation that has contributed to unemployment, chronic ill-health, high rate of accident, low motivation level and increased physical and mental stress.

How such problems occur in TT will be closely examined. It is generally accepted that technology's characteristics are mostly determined by the prevailing conditions of the technology-producing country. They usually reflect the specific requirements and availability of both human and material resources, income level, resource costs, organization and political system, infrastructure, human factors/ergonomics and socio-cultural conditions of that country for which technology is designed. Then, when the very same technology is transferred from an IC to an IDC characterized by a vastly different set of requirements and characteristics, some kind of *adaptation* is imperative to fit the transferred technology. However, due to reasons ranging from ignorance to lack of demand from the technology receivers, very seldom do companies take the initiative to adapt their technology to the conditions of the recipient countries.

The developing countries, possessing 80% of the global workforce, usually gain access to advanced technologies without having sufficiently developed their technology-absorptive capacities, or even their legal or administrative infrastructures to control the adverse consequences of imported technology. (Here, the term technology pertains to the technology in use by IDC depending on their communication ability and selection mechanism.)

For TT to be successful and industrial development to occur,

they require, first, a conscious and rational selection of technology from the available world of technology. Second, there is need to prepare the local conditions for optimum absorption of the technology and to adapt the society and its social services to the changes required by the new technology for optimum use. Yet, many social functions in IDC are incompatible with the demand of industrial progress as developed in the IC. Lack of a safety-conscious culture, low educational standards, inadequate training in handling the new technology, inappropriate policies and organizational structure are but a few of the contributing factors to the failure of TT. In most cases, the rapid and hasty rate of change that takes place in some IDC is often too quick to suit either the individual or the society, threatening a failure to establish the required infrastructure and social structure for making maximum use of technology.

4. TECHNOLOGY TRANSFER PROCESS: THE NEED FOR A MORE HOLISTIC APPROACH

In the TT process emphasis is usually laid on the engineering aspect of technology (technoware). As a result of this narrow approach many IDC are struggling with a low rate of technology utilization, loss of production and low product quality, as discussed above. There is also ill health of the work force due to poor working conditions and other social problems.

To avoid the past mistakes, a more holistic and systematic approach that considers all the interacting components of technology should be adapted with regard to TT. Technology comprises of four interrelated components (Figure 1).

1. Technoware: tools, capital, intermediary goods, products, machinery, physical processes, etc.
2. Human: human labor, capacity for systematic application of knowledge, know-how, skills, ideas, problem solving capacity, etc.
3. Information: scientific and other organized knowledge, technical and social information, specifications, standards, soft wears, etc.
4. Organization: organization of products, processes, tools, services for use by people, social arrangement, means for using and controlling factors of production, etc.
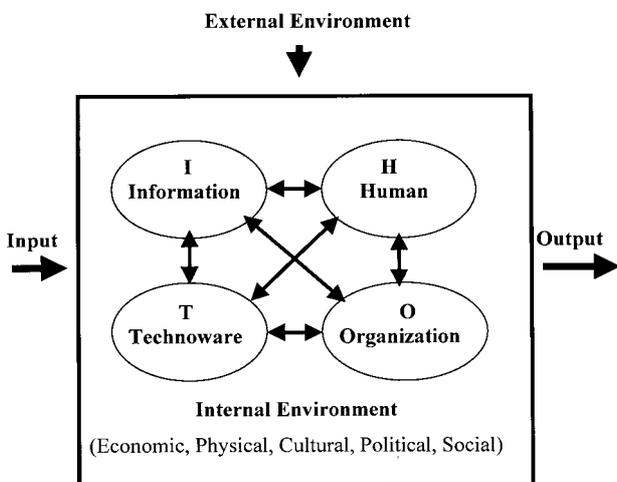
Figure 1. Integrating components of the technology transfer process.

The four components of technology are complementary to one another and are required for the production of goods and services (TAT 1987, Sharif 1988). Depending on the nature of activity, the relative importance of each of the four components may differ. Not only must the four components be fully considered, but also their interactions within the operating environment (technology climate). A change in one factor can affect the others and ultimately the system output. The technology climate includes both external and internal factors. The external environment pertains to factors such as physical infrastructures, support facilities, R&D institutions, political system and legal systems as well as country's culture and administrative institutions. The internal factors include the local firms' cultures, physical environment and economic condition as well as the social and political conditions of the firms. Thus, a system-based approach rather than a reductionist functional approach is required to optimize the system out put.

5. ROLE OF APPROPRIATE TECHNOLOGY

Technology, defined in terms of physical products, techniques, know-how, information, skill, labor and organization, is an integral part of a country's structure. Any changes in technology can have an impact on the social, political and economic system. The solution to many of the problems associated with industrialization in IDC lies in the analysis of their causes. By adapting better development policies, choosing appropriate technology and exerting a more efficient control, IDC can solve many of their problems and provide better living and working conditions for their population.

An appropriate technology is one consistent to the needs and resources of the country: it is suited to its user population, is sensitive to its operating environment. It takes into consideration the local infrastructures, as well as the educational, social, cultural, economic and political aspects. Technology should therefore not be regarded as the objective of development, but rather as the principal means for its attainment.

With regards to the large variety of circumstances that exist among countries, there is no uniform prescription or single system that can be regarded as the appropriate technology for all IDC at all times. It differs according to place and stage of development. The crucial issues that should be fully considered with regards to industrial development is to keep in balance these substantial laws:

1. Economy (basic needs and equality).
2. Ecology (environment and protection).
3. Energy (material and human resources).

It is also important for IDC to identify which technology to acquire, how to transfer it and how to implement it. They need to learn how to adapt it to its users' and its operating environment, how to maintain it and how to build upon it. In this process *ergonomics* is the missing link in the quest for appropriate technology.

6. ERGONOMIC CONSIDERATIONS

Technology can be inappropriately transferred in three ways:

1. Incomplete transfer: not considering all aspects of the technology in the transfer process, e.g. leaving out (neglecting) the maintenance system.
2. Imperfect: not considering the user's characteristics in the transfer process, e.g. not translating manuals and instruction

books into local language or not considering the anthropometric dimensions (body size) of the technology users.

3. Inadequate: not considering the environmental conditions of the recipient country such as the climate, infrastructure of society, finance, technology and culture, etc., e.g. transferring products such as protective wears made for cold countries to tropic ones.

Ergonomics considerations ensure appropriate transfer of technology. They result in a 'good fit' between technology, its users and the operating environment. Thus, ergonomics should be considered in the process of TT both at the micro and macro levels.

6.1. Microergonomic Level

Microergonomics is concerned with the design of product or system with the objective of creating a 'usable' user–machine–environment interface. Issues that need to be considered from an ergonomic point of view in the process of TT at a microergonomic level are given below.

6.1.1. Anthropometry

Differences in body size, reach and shape of different populations may mean that many products and machines imported from IC do not fit the anthropometric dimensions of the user population in the IDC.

Areas where these differences apply and need to be considered are in the design and use of workstations, machinery, equipment, tools, safety equipment and protective devises. Failure to do so may result in poor working postures — a major risk factor causing musculoskeletal disorders, accidents and low efficiency at workplaces. With regards to misfits of protective devices, they may expose operators to environmental hazards (Shahnavaz 1992).

6.1.2. Physical work capacity

This is the capacity of people to perform prolonged physical (manual) work and it is related to individual health, age, weight, sex, fitness, level of nutrition and the environmental conditions. Physical working capacity of the average worker in most IDC is much less than his counterpart in IC because of inadequate nutrition, poor physic, lower quality of life in general and harsher working conditions.

Considering the high workloads in IDC and the lower physical work capacity of the workers, it is not difficult to find that work demands in levels equal to that in IC are in many cases too high for IDC operators. These lead not only to higher rates of accidents and injuries but also to lower productivity. These and other factors that affect people's physical performance and well-being in IDC (e.g. longer working hours, improper shift-work schedules, hazardous work environment, nutrition inadequacies and the effects of tropical climate) are often not taken into account when technology is introduced. The total working time in IDC is usually much longer than 40 h a week. Taking into account the transport time to and from work (which can be very long and highly stressful) and often the extra work people do to earn their living, daily work can add up to 14–16 h, especially among females who also have domestic responsibilities at home.

The realities of the situation in IDC should therefore be considered and consequent changes made; for example, when introducing manual work schedules, rest pauses and welfare facilities that are originally designed for IC.

6.1.3. Functional ability

This pertains to differences in generation of force and muscular function. The average maximum force generated by the workers in IDC is often much lower than those from IC. This can be due to differences in weight, health and nutrition. The designs of equipment, tools, mechanical interfaces such as hand and foot controls, as well as manual handling procedures must consider these differences for a safe and efficient work.

6.1.4. Physical working environment

The typical physical environment of an IDC is generally that of a tropical or subtropical climate. Consequently problems of heat and ventilation will be encountered. The indoor climate at workplaces is also not usually regulated and is more severe than in IC. Furthermore, there is the additional heat produced by the hard physical work usually performed in IDC. Subsequently, if the working environment is such that this excess heat is not removed and the worker is unable to retain normal body temperature, then it can be very uncomfortable and dangerous (e.g. causing heat collapse).

Other physical environmental factors of similar importance that need to be taken into consideration are lighting, noise, vibration, air quality, chemical hazards, etc. Environmental problems are usually very common and severe in IDC. They need to be taken cared of when introducing technology and when designing work places, work organizing, jobs, working methods, work–rest schedules, protective wears and plant design. Failure to do so will result in reduced performances in both physical work and mental activity. In extreme cases the working environment can also cause serious health problems. There are also instances of transferring dangerous technology. For example, technologies and their products, such as asbestos, already banned in the IC are unfortunately still being transferred to IDC. For IDC purchasing technology, it is vitally important to lay special attention to their environment. It is necessary to make sure that industrial development is not harmful to their working and natural environment to achieve a sustainable development.

6.1.5. Cognitive differences

These refer to variations in skills and differences in perception and cognitive complexity. It has been observed that people from IDC have a different internal model (mental model). They have different operational images and population stereotypes and there is a difference in their pictorial perception and information-processing behavior. These differences will affect the operator's decision-making, action and performance. In areas such as communication design (verbal, written, pictorial), display design, training programs, organization of work and skill, system design (particularly complex systems) and system operation, these differences need to be considered.

6.1.6. Cultural differences

Proper application of technology depends to a large extent on the compatibility of the user's culture with the design specifications and operational procedures. Compatibility must exist at all levels of culture — at the social, organizational and

professional levels. In turn, cultural compatibility can have a strong influence on people's preferences and aptness to utilize a technology. It must be mentioned that there exists no uniform culture in IDC. Nevertheless, some very broad distinctions can be made regarding social, organizational, professional and individual cultural-based differences that can influence technology utilization in general.

At the microergonomic level, the usability of a product, which involves not only its functionality and reliability but also its product safety, comfort and acceptance, depends on the compatibility of the product specifications and the user's characteristics and preferences wherein the user's culture plays an important role. In areas such as products, icons, symbols and interface design (regarding shape, form, color), control/display compatibility design, workplace design (regarding sitting posture, working height and gender issues), as well as protective clothing (regarding religion and custom constrains), these differences must be accounted for.

6.2. Macroergonomic Level

Macroergonomics refers to the design of organization–machine interface technology (soft-system technology). It concerns the proper design of organization and management systems and optimization of the technical and personal subsystems. Evidence exists showing that because of differences in culture, socio-political conditions, educational and technological levels, several of the successful management and organizational methods developed in IC cannot be successfully applied to IDC. On the other hand TT is not just a transfer of technoware, but together with the hardware systems it also becomes the soft system technology. Organization and its structure, values and behavior are usually cultural products, and are more strongly emphasized in the traditional societies of IDC. In most IDC, an organizational hierarchy characterized by a down-flow of authority structure within the organization is the common practice. It has little concern for Western notions such as democracy or power-sharing in decision-making, which are regarded as key issues in modern management, being essential for proper utilization of human resources as regards intelligence, creativity, problem solving potential and ingenuity.

A major problem of many transferred technologies is the lack of macroergonomic consideration. Macroergonomic issues are more complex than microergonomics ones because they are widely influenced by the firm's culture, which in turn is much affected by the socio-cultural structure of the country.

A problem pertaining to the implementation of macroergonomics is the lack of awareness regarding its benefits. The concept is regarded as new and a Western product and therefore mostly resisted by IDC managers. For making macroergonomics acceptable, it is better to find justifications for the macroergonomics ideas and its principles within the national culture, such as in religion, folk tales, stories, phrases and expressions. According to McWhinney (1992), 'conflict is avoided by making changes in accordance with the path of the myth and stories that guide our lives. Changes are most effectively accomplished when we have uncovered the core stories of the relevant culture and use our skills and courage to advance along the paths that are natural to the person, the organization, and the culture'.

The macrolevel actors such as government, business organizations and organized labor are the major determinants of macroergonomics at work. The higher their knowledge and commitment to these issues the better will be the firm's organization from a macroergonomics viewpoint.

6.2.1. Organizational design and management

Organization is a social structure wherein employees play a decisive role in improving its performance. Higher demand on productivity and intensive production require a workforce with high and broad skills, flexible, well motivated and self-regulated. Decision-making and action-taking should be concentrated to the heart of operation. This is to reduce the risk and duration of system failure and better to utilize resources as well as to increase system reliability and availability. However, organizational changes are difficult, time-consuming and an expensive process. Cultural factors, including the way people interact with each other in an organization and commit themselves to organizational goals, are complex matters that have significant bearings on the success of an organizational change. It is therefore necessary to match the management methods and techniques to the local conditions. Combining bottom-up and top-down approaches, macroergonomics takes full advantage of utilizing broad participation within the organization. This will create a shared vision and a program for change acceptable to both employees and management. This is especially essential with the view towards development of more complex and sophisticated systems.

The organization should regard employees as its future problem-solving agent, and must thus allow for enhancement of their skills and preserving their innovative capacities. To facilitate employees' participation in solving technical and psychosocial problems at work and to create a safe and productive environment, management has to institute a vigorous program of education, training and self-improvement.

In IDC, narrow, tightly controlled, fragmented jobs and hierarchical organization are still common practice. They are usually the cause of low motivation among workers, having adverse effects on both the individual and organizational performances. On the other hand, advanced technology requires a human-centered organization for its effective operation. Introducing organizational changes and new management systems could be easily achieved in IDC through a process of training and education in macroergonomics for both management and employees. These will bring about awareness and action, aiding in the introduction of necessary changes. It has been shown that the synergistic effect of a macroergonomics approach is significant. It improves not only the employees' productivity but also their quality of working life.

6.2.2. Workers' participation

Studies have emphasized the significance of worker participation in the introduction of new technology since technology needs to be legitimized by both management and workers. Worker participation can help improve the quality of the decision and implement the technology to its fullest.

In this competitive world, a company cannot survive without the total involvement of its workforce. Although participation is a culture-free concept, political as well as socio-cultural factors

must be considered for its proper applications in IDC's workplaces. Participation is needed at all stages of TT. It should start at the design conceptualization stage and continue throughout the life cycle of TT to the actual utilization of technology.

6.2.3. Communication and action

Culture is the key to effective and clear communication and dialogue, which is the means of information flow between the technology supplier (S) and technology receiver (R). Since technology utilization is a continuous process, proper communication and interaction of the two partners (S&R) are essential. Bridging cultural differences and learning to understand each other is a key for successful TT and building a long-term relationship (Figure 2). Clear and effective communication plays a decisive role in all five phases of TT (i.e. initiation, analysis, selection, implementation, utilization). In each phase, the information necessary for making a mutual decision is exchanged between the two partners. Information is invaluable to assess the environmental characteristics of the technology receiver, e.g. the capacity to absorb a new technology. It is needed to predict the cultural differences as well as to provide knowledge about a firm's characteristics and the current conditions of macro- and microergonomics and the necessary actions for a successful transfer.

Meshkati (1986) described the systematic integration of economic and ergonomics analysis needed to be performed in each phase of TT through a joint action of technology supplier and receiver. According to him, each partner should perform the following ergonomic activities:

1. In Phase 1 (mutual conviction — TT may be mutually beneficial):
S: recognition of technology receiver's needs and abilities to absorb and use technology.
R: identification of relevant human factors considerations affecting intended technology.

2. In Phase II (preliminary formulation of TT project and identification of appropriate choice):
S: macro- and microergonomics study of man, machine and environmental factors of intended technology.
R: functional analysis and evaluation of human resources, environment and working condition.

3. In Phase III (final decision on choice of technology and determination of needed adjustments):
S: final identification and selection of major ergonomics requirements and decision on modification.
R: developing ergonomics checklists and guidelines, technical adjustments and ensuring its appropriateness.

4. In Phase IV (implementing technology and initiating start up process):
S: design and development of operational and maintenance manuals in local language; determine staff training requirements.
R: work place design, job design, personnel selection and training, organization and management system design.

5. In Phase V (TT is running. Compile relevant data regarding operation effectiveness and total system performance):
S: monitor operational system, oversee procedural issues and evaluate total system performance.
R: monitor safety, productivity and performance of the system, improve working conditions.

7. TOWARDS A SUCCESSFUL TRANSFER OF TECHNOLOGY

The transfer of a technology from a firm in IC to a firm in IDC depends on the decisions taken both by the technology supplier (IC) and the technology receiver (IDC), as well as the dynamic interaction and communication between the two partners. The environment surrounding the TT is likewise another major component in this process. The environment influences the development of a particular relationship between the partners. The success of the process is determined by the cooperation between partners, something that can only be developed during long-term contacts. Furthermore, it is generally agreed that three characteristics of the technology recipient's environment are decisive for successful TT: government policy, technical absorptive capacity and cultural distances (Robinson 1988). Figure 2 illustrates the model of TT between two partners.

Regarding the appropriateness of the transferred technology, the level of ergonomic awareness by the technology supplier and receiver firms as well as their commitment to ergonomics issues will greatly influence their decision regarding how such appropriateness will be put into reality. These can be analyzed by closely examining the firm's characteristics and attitudes since they generally reflect the firm's micro- and macroergonomic conditions. The better the ergonomics conditions of a company and the firmer its commitment to it (both at micro and macro levels), then the better the choice and utilization of technology, leading to a more appropriate TT. With regards to the environmental factors, government policies should consider several macroergonomic parameters. These involve proper assessment of the recipient countries' needs with regards to the nature and type of technology, further establishment of relevant
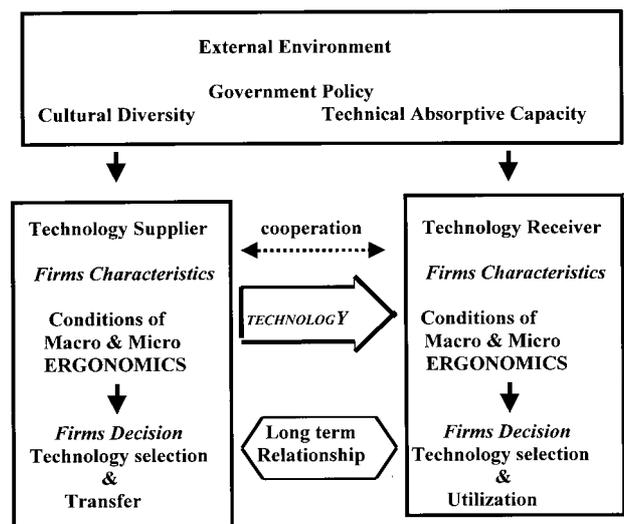
Figure 2. Technology supplier and receiver's decision-making and action.

research institutes and support organizations for the development of national statistics and data banks, as well as formulating recommendations, promotions, control and implementation of ergonomics at work. Since many factors influence the nature, extent and diversity of TT, problems are specific in each IDC and it is necessary for each country to make research and training an indispensable part of its industrialization/development policy. Likewise, proper legislation and inspection policies have significant roles in the encouragement of considering ergonomics issues and in creating a crucial impact on companies in the maintenance of ergonomics. Further, it is an important way of ensuring a healthy, safe and ergonomically sound workplace, contributing towards the national aims of prosperity and higher quality of working life.

In the context of technology absorption capacity, one of the main problems of IDC is the lack of scientific and technological infrastructures and training facilities for improving the level of education and skill of the workforce for safe and effective operation, maintenance and development of the imported technology. These are important since technology will be absorbed better at the local environment if it is in harmony with its users and its operating environment. Certain concrete steps can be taken to meet this problem. For example, in most cases the necessary ergonomics knowledge for making the right choice with regards to purchasing a new technology or its proper utilization does not exist among companies in IDC. In such cases assistance can be sought from an ergonomics consultancy or research organization. A well-established support system for providing ergonomics information at a country level would ensure proper choice and utilization of the imported technology.

Vis-à-vis cultural distances, the local culture of the technology-recipient firm and its similarities with the technology-producing firm greatly influence the success of the transferred technology. Cultural values and behavioral patterns have a direct bearing on the peoples' willingness and ability to adapt and absorb technology. A successful transfer requires, therefore, that the cultural barriers are overcome and that cultural issues are considered in the life cycle of the TT process. However, it should be remembered that the ultimate responsibility for successful TT rests with the technology receivers — the recipient country's policy-makers involved in development planning as well as the firms importing the technology.

At the national level, effective institutions are needed for the formulation and implementation of sound policies. It is vital that they support the local firms and conduct inspections and control to ensure appropriate TT and utilization. At the firm's level, serious consideration of macro- and microergonomics, on the proper interaction between technology suppler and technology receivers will greatly help not only in selecting the right technology, but also in utilizing it efficiently in the long-term.

8. CONCLUSION AND RECOMMENDATION

In conclusion, the criteria for selecting appropriate technology and utilizing it successfully are:

Selection criteria:
- Relevance to basic needs of local population.
- Availability of local resources.
- Protecting local environment (air, water and soil).
- Contribution to improving quality of life of the local population.
- Sound ergonomics design (both at macro and micro level), considering local conditions.
- Availability of skills, information, services and infrastructure for optimum utilization of the technology.
- Creation of job and income opportunity for local population.

Factors to be considered with technology utilization:
- Human factors and ergonomics.
- Civil work.
- Raw material and energy.
- Trial production.
- Technical exchange and net working.
- Quality efforts.
- Market support.

REFERENCES

FARHANG, M., 1994, Model for analysis IITT: decisions and interaction. In International transfer of technology. Doctoral thesis, Luleå University of Technology (138D), 71–96.

MESHKATI, N., 1986, Major human factors consideration in technology transfer to industrially developing countries: an analysis and proposed model. In *Human Factors in Organizational Design and Management Conference, ODAM–II* (North Holland), 351–63.

ROBINSON, R. D., 1988, *The International Transfer of Technology: theory, issues, and practice* (Cambridge, MA: Balliner).

SHAHNAVAZ, H., 1998a, Adapting macroergonomic approach for identification of workplace problems and development of low-cost/no-cost solutions in industrially developing countries. Case study. In *Glucosan-Iran, Human Factors in Organizational Design and Management Conference, ODAM-VI* (North Holland), 585–90.

SHAHNAVAZ, H., 1998b, Cultural differences. In *Encyclopedia of Occupational Health and Safety* (4th ed.) (Geneva: International Labor Office), 29.79.

SHAHNAVAZ, H., 1992, Ergonomics and industrial development. *Impact of Science on Society*, no. 165, 99–108.

SHARIF, N. M., 1988, Basis for techno-economic policy analysis. *Science and Public Policy*, 15, 217–29.

Technology Atlas Team (TAT), 1987, Components of technology for resource transformation. *Technological Forecasting and Social Change*, 32, 19–35.

Tools to Design and Evaluate New Forms of Work Organization

C. Axtell, K. Pepper, C. Clegg, T. Wall, and P. Gardner*

Institute of Work Psychology, University of Sheffield, Sheffield S10 2TN, UK and *Psychology Department, University of Leeds, Leeds, UK

1. INTRODUCTION

This article describes a set of tools that have been developed to help organizations design and evaluate new forms of work organization, especially when introducing new technology. This is important given the typical lack of attention to human and organizational issues when implementing new technology, and the paucity of practical guidance on how to implement new work designs. The existing literature describes various principles, but does not help practitioners decide which ones they should use (Section 1.1). We focus on a scenarios tool which enables planning of the work organization. The outcomes of using this method can be incorporated into questionnaire and interview tools, which can be used to evaluate the change. The tools are illustrated by a case study of group work design in a large organization which accompanied a major investment in new technology. The implications and conclusions from using the tools are also considered.

1.1. Work organization and new technology

The literature on work organization, sociotechnical systems, and teamworking offers an array of general principles for workgroup design which practitioners may find rather bewildering when considering how they apply to their own organization. For instance, a workgroup or team should have boundaries, shared goals, and interdependence among members and tasks, and although skills and tasks should be broadly related, there is likely to be some role differentiation (Alderfer 1977). A workgroup needs to be of manageable size, and although team size will vary, it is likely to be smaller than 20 members (West 1996) and perhaps more typically between 2 and 7 members (Bettenhausen 1995). A workgroup is a meaningful unit that is identified as a team from both inside and out (West *et al*. 1998). However, these units can be changed; for example, engineers may not have originally been considered as part of a team with machine operators, but such boundaries may move (Parker and Wall 1998). As can be seen, principles for team design are not necessarily hard and fast.

Sociotechnical theory (Cherns 1976, 1987) proposes the need for the joint optimization and parallel design of the social and technical subsystems. This theory advocates (among other things) that employees should deal with variances (breakdowns, changes in product requirements, etc.) at source. Work roles should be multifunctional and multiskilled; jobs should have variety; and methods of working should be minimally specified (Cherns 1976, Emery 1959). One way to optimize the social and technical systems is through autonomous workgroups or self-managing teams (Pasmore 1988, Goodman *et al*. 1988). Such teams are able to decide upon their own methods of working, and are responsible for handling any operational problems they encounter

(i.e., these teams are not under traditional supervision). In order to do this, barriers to effective interaction within the group need to be removed, and employees need to have access to relevant information and resources in order to carry out their work. Work may take on more of a process flow, with people moving from machine to machine to produce the whole product, rather than being based on one particular machine and completing only a small part of the product.

In a similar vein, the job characteristics model (Hackman and Oldham 1976, 1980) specifies that jobs should be designed so they have skill variety, task identity (whole and complete tasks), task significance (important tasks), autonomy, and feedback. With regard to workgroups, it is advocated that not only should group members have collective responsibility and accountability for completing group tasks, but also for monitoring and managing their own work and interpersonal processes (Hackman 1987). Clear goals should be set for all dimensions of performance that contribute to the overall effectiveness of the group, and feedback (from usable, measurable indexes) should be provided on the group's progress towards it's goals (Hackman 1989). In addition, individuals should feel that they are important to the fate of the group, to minimize social loafing (Guzzo and Shea 1992).

Although the supervisory level is often removed when creating autonomous workgroups, some organizations opt for having a team leader. This role is different to the traditional supervisory role, in that the team are coached and facilitated by the team leader, but the group as a whole still has responsibility for the work, and autonomy and involvement in decision making (Parker and Wall 1998).

Sociotechnical theory advocates that work redesign processes should be compatible with desired design outcomes (i.e., they should be highly participative), and that work redesign should be a continuous process, rather than a single event. Consequently, roles do not stand still, but evolve as employees become more skilled and learn more about how to improve performance. Indeed, jobs should offer opportunities for continuous learning (Thorsrud 1972).

There are different constraints and varying contexts within organizations which will affect the application and impact of work design principles. For instance, internal political considerations will need to be taken into account, as autonomous or empowered teamworking can change existing power distributions within an organization, such as including engineers in a team of operators, or removing the supervisory level (Clegg 1984). However, if organizations can be encouraged to consider some of the barriers and contingencies faced within their own organizations, they should be able to plan effective ways of dealing with them.

Despite a plethora of literature on the design of work, reports are less clear on how these principles are translated into practice, given the varying constraints and contexts. The two main approaches offer no easy methods and simple tools to help with design and evaluation (Clegg 1993, Mumford 1987). And the Ethics tool (Mumford 1986) or key variance analysis (Davis and Wacker 1987) may not be appropriate in all circumstances.

Clearly a gap in the literature exists because work design principles, although laudable, have in practice been difficult to implement in the absence of clear pragmatic guidance. Moreover, organizations rarely evaluate the outcomes of initiatives they

implement (Clegg *et al.* 1997), so they find it difficult to chart their own progress. Here are some pertinent questions which the following sections try to answer by describing a set of tools and their use in a particular case:

- How do organizations generate ideas for new practices in a participative manner?
- How does the theory translate into the division of responsibilities and individual tasks that team members undertake within a particular organization?
- How can organizations be helped to make decisions about integrating theory and practice?
- How can implementers deal with the constraints within their organizations?
- How should they measure and evaluate the progress of work design changes?

2. THE SCENARIOS TOOL

The scenarios tool has been developed by Clegg and colleagues since 1986 (Clegg *et al.* 1996). It is a structured and systematic technique for developing and evaluating a range of choices for how an organization (or part of it) works. The tool is used in a workshop setting involving all the relevant stakeholders, with the discussions, choices and outcomes recorded (on paper) for later use. Although it is preferable to involve frontline employees in these workshops it may not always be possible. Then the tool can be used without frontline staff, but it is recommended to use alternative mechanisms for employee participation (e.g., employee opinions sought in representative workshops before implementation; or local adjustment of the overall scenario).

First, the members of the workshop describe the current scenario (for benchmarking and comparisons) and then generate a range of explicit alternatives. The tool itself does not generate the best choice, but it does provide a systematic method through which users can determine a range of possibilities to consider. Seven standard headings are used to describe each scenario (Table 1) although additional or alternative headings can be used. The headings do not necessarily have to be considered in order, but the vision does help to guide the description of the rest of the scenario (i.e., the vision should be consistent with it). An example is a vision based on the idea of employee empowerment. Describing the logic clarifies the reason or theory behind a particular vision (e.g., empowered workers are better performers). The vision has implications for the organizational structure and employee roles, which can be considered in detail. The benefits of the scenario can then be explored. Consideration of the costs and implications highlight the disadvantages, allowing reflection on the constraints and the organizational context; this encourages preventive planning.

Table 1. Standard headings for the organizational scenarios tool.

1. Vision (of how the system works)
2. Logic (underlying this choice)
3. Organizational structure (that supports this way of working)
4. Roles (that exist within the structure)
5. Benefits (associated with the scenario)
6. Costs (associated witht he scenario)
7. Implications (of making that choice)

3. CASE STUDY

The case study took place within three sites of a large organization within the UK, which was implementing advanced processing machinery (APM) across numerous sites. The new technology was highly complex, integrating several existing processes and was computer controlled. Tailor-made for the organization, the machinery was strategically important for the business and represented a massive financial commitment (in the region of £200 million). The organization was a traditional one with specialist roles for employees, and a preference for a standardized work organization across the different sites. However, there was an interest in taking a different approach.

The organization was already taking steps to introduce various initiatives. They were planning to implement teamworking, total productive maintenance (TPM), and associated engineering initiatives (such as condition-based monitoring, zero defects, and engineers having more responsibility for strategic management of machinery). In addition there was a large focus on management behavior with initiatives planned for training managers in coaching, devolving responsibility, and involving employees in decision making. Another development in management thinking in the organization had been whether it would be appropriate for team structures to follow a cell-based process flow model rather than the current machine-based model (Figure 1).

The new technology had already been bought but not implemented, and the organization wanted help in organizing the work around it. Specifically they wanted help answering these questions:

- How should work be organized to get the most out of the machinery?
- What should be the job role boundaries?
- What tasks will employees need to conduct?
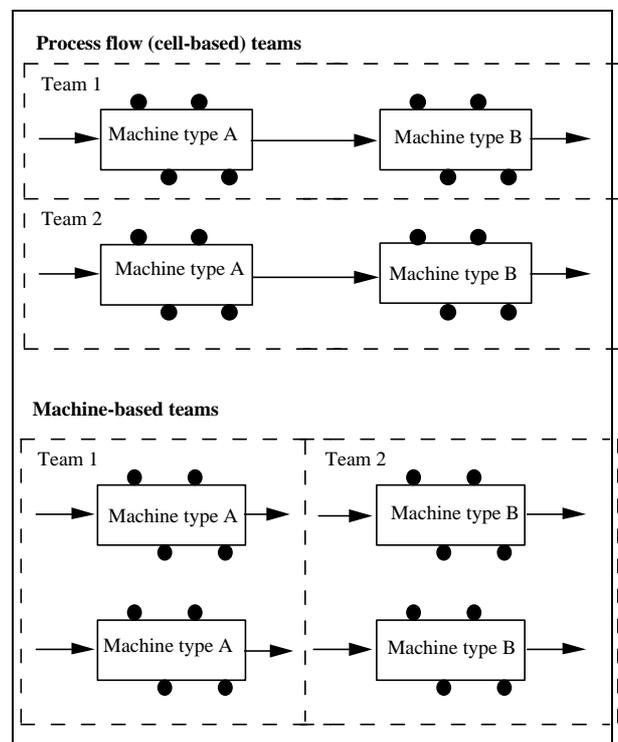
Figure 1. Process flow and machine based groupings.

- How many operators and managers will be required to run the APMs?

4. USING THE SCENARIOS TOOL

The scenarios tool was administered and facilitated by three of us over the course of two workshops. The workshops were attended by several people:

- The group of people responsible for implementation and policies surrounding the APM.
- A senior manager from the site where the new machinery would be piloted.
- Managers who had experience of previous technology implementations within the organization.

The focus of the first workshop was to consider the roles of employees working with the APM (e.g., how to organize work to optimize the use of the technology, job role boundaries, and the tasks employees undertake). The second workshop considered in more detail the management configuration (e.g., how many machines a supervisor should be in charge of, how to combine old and new machinery).

When considering employee roles, four possible organizational scenarios were developed (Table 2); scenarios 1, 3, and 4 were developed during the workshop and scenario 2 was developed later on. Although scenario 2 was developed later on, it is covenient to put it in the same table. The choices of scenario were strongly influenced by the existence of the initiatives described in Section 3. Consequently, the four scenarios were based on these visions:

- *Scenario 1*: the current situation with specialist operator, engineer, and management work boundaries based around single machines.
- *Scenario 2*: partial empowerment with only slight broadening of roles (assuming only partial implementation of the initiatives) and team structures based around machine grouping and work areas.
- *Scenario 3*: flexible working with empowered, multiskilled

operators (assuming full implementation of the initiatives) and team structures based around machine grouping and work areas.

- *Scenario 4*: flexible, empowered, multiskilled working (scenario 3) but with team structures designed around process flows rather than machine groupings.

Scenarios 3 and 4 are almost exactly the same except for the way in which the machines are grouped. Scenario 4 has a process (cell) flow, whereas scenario 3 is organized around individual machines or functions .

During the workshop the facilitators were able to input their knowledge of theory and their experiences in other organizations to help the group describe the scenarios and consider their benefits and costs. The benefits that were felt to be associated with the more empowered scenarios (3 and 4) were greater machine efficiency and utilization. They were also considered to lead to increased employee satisfaction and well-being due to employees having well-designed jobs.

The costs (or requirements) of scenarios 3 and 4 meant that considerable outlay and effort would be needed in order to convince members of the organization and unions to accept the new initiatives. Moreover, there would be the expense of ensuring people were trained in the appropriate skills (e.g., TPM, decision making, teamworking) and that appropriate supports were put in place (e.g., pay, appraisal, information).

The implications of moving towards scenarios 3 or 4 were that they would rest on the success of the new initiatives. It was recognized that both these initiatives would require a considerable shift in the organizational culture, as the organization was very large and traditional and there would a considerable amount of inertia to overcome. Although there was doubt about whether the organizational initiatives would be ready in time for the new technology's implementation, the group felt that their long-term aim should be a scenario based on these initiatives.

During the workshop, the group decided that, because scenarios 3 and 4 would require the success of the organizational

Table 2. The four scenarios (illustrative not comprehensive).

	Scenario 1	Scenario 2	Scenario 3	Scenario 4
Vision	specialist roles and machine groupings	partial empowerment and machine grouping	full empowerment and machine grouping	full empowerment and process flow
Logic	specialist roles are cheap and effective	partial multiskilling is better than specialist working	complex machines need skilled operator control	complex machine need skilled operator control and process management
Structure	hierarchical	flatter hierachy	flatter hierarchy	flatter hierarchy
Roles	narrow and simple jobs with close supervision	partial skilling with some empowerment, e.g., operators have more responsibility but no TPM	multiskilled, empowered operators have responsibility for orgainzing own work and conducting TPM	multiskilled, empowered operators have responsibility for organizing own workflows and conducting TPM
Benefits	cheap and fits with employee expectations	only requires small outlay to get slight increase in job satisfaction	more machine efficiency and increased job satisfaction	more machine efficiency, more process knowledge, and more job satisfaction
Costs	unlikely to give improvements in machine efficiency	won't get much improvement in machine efficiency	hard work to persuade others; needs training and supporting services (e.g., information appraisal)	even harder to persuade others; needs training and supporting services (e.g., inforamtions appraisal)
Implications	requires no change in culture	only requires partial change in culture	requires big change in culture	requires big change in culture; the process flow may be too complex

initiatives, they would have to plan the detailed breakdown of tasks related to the new machinery within the existing specialist scenario (scenario 1) as well as scenarios 3 and 4. This was because some of the tasks on the APM would be different to those conducted on the old machinery, due to the greater complexity of the new technology.

A detailed description of the roles was undertaken. For instance, within scenarios 3 and 4 the roles were described as being multiskilled with empowered operators who organize their own work, breaks, etc., and who maintain the machinery. Engineers would be highly skilled and able to concentrate on strategic management, machine development, etc. Managers and supervisors would be able to concentrate on managing boundaries and overall goals rather than closely supervising their staff. This stands in stark contrast to the specialist model (scenario 1) in which operators have very narrow roles where they just operate machinery and are closely supervised.

One outcome of the discussion about roles during the workshop was a matrix showing the breakdown of broad tasks for operators, managers, and engineers within each of the scenarios. The general description of roles was decomposed and the workshop participants listed and allocated the key tasks they had generated. This matrix was a useful mechanism for comparing the division of responsibilities across the different scenarios. For instance, "fix faults" was an engineering task in scenario 1 and an operator task in scenarios 3 and 4 (Table 3). Table 3 illustrates scenario 2, although this scenario was not devised during the workshop.

The group felt they needed input from others to decide which routine maintenance or fault-fixing tasks operators could perform on the APM. After the first workshop, the group of stakeholders used the scenarios tool themselves with some other managers (from engineering and operational areas) to refine the division of responsibilities matrix. This represents another strength of the tool, in that the stakeholders felt able to use it themselves. Within this refined matrix, tasks were more detailed (particularly those relating to TPM) so that "fix minor faults" became tasks like "replace faulty drive belts," "check print quality," and "change ink bottles."

At the same time the group used the tool to come to a decision about how many operators were needed to operate the APM, based on the description of the tasks and the vision of multiskilling. They decided on four operators, fewer than typical for scenario 1, but this allowed a broader range of responsibility for each operator. The group agreed that the number of operators would be reviewed after the pilot.

The scenarios can be elaborated by considering potential

Table 4. Criteria to assist in the configuration of teams.

CRITERIA FOR TEAMS
1. Size (perhaps the maximum number of people is about 12)
2. Proximity (consider trade-off between physical location and communication needs)
3. Interdependence (of inputs and outputs of each member)
4. Logical grouping (meaningful unit that matches mental model of members)
5. Complete task (conducted by team from begining to end)
6. Self-contained responsibility (so can stand alone and manage selves to meet targets and responsibilities)
7. Performance measurement (members have measurable contribution and usable performance indexes)
8. Mix of tasks (so good and bad tasks can be shared out)
9. Related skills (members have broadly related skills and competencies)
10. Internal politics (team should be consistent with important internal political considerations)
11. Natural grouping (for communication purposes, with shared information needs)

NOTES
There are no hard and fast rules
Several of the above items are interrelated
There may also be trade-offs (e.g., you may want to change the existing logical groupings and mental models)
The general trend is towards a product or process flow organization, rather than a machine-based organization
Teams in the form of autonomous workgroups provide a very effective and flexible way of working but can be difficult to manage
The role of team leaders needs very careful consideration (not equivalent to supervisor)

subscenarios. For instance, the group wanted to consider the management structures and team boundaries that would fit into the ideal scenarios 3 or 4, so we set up a second workshop to do this. A list of criteria to assist configuring teams was drawn from the literature and used to help the group make their decision (table 4). This list was not exhaustive but included some useful points for consideration by the stakeholders. In particular, the criteria could be used to assess the benefits and costs of the scenarios.

Seven possible management configurations were developed using the scenarios tool. They were based around different types of machine or process flow groupings which outlined particular combinations of machines or functions that could be managed under one supervisor at the pilot site. The number of managers required for particular machine or functional groupings was considered an important issue for the organization, who wanted to get the balance right between allowing operators autonomy and having enough supervisory support. Consistent with the themes of scenarios 3 and 4, we assumed a team leader model (rather than a traditional supervisory model). However, the group

Table 3. Illustrative excerpts from the division of responsibilities matrix.

Task	Scenario 1	Scenario 2	Scenario 3,4
Prepare work area	supervisor	operator	operator
Allocate tasks among selves	supervisor	sup/op	operator
Generate/write reports	supervisor	sup/op	operator
Brief the team	supervisor	supervisor	operator
Fix minor faults	engineer	engineer	operator
Monitor machine operation	engineer	eng/op	operator
Clean APM	engineer	engineer	operator
Call engineer if problem	supervisor	sup/op	operator
Fix major faults	engineer	engineer	engineer

did not want to consider a scenario where engineers were part of the team at this stage, partly due to political considerations.

A key initial decision was whether to adopt a cell-based process flow or a machine-grouped structure. The benefits and costs of each configuration were listed. It was decided that the idea of cell process groupings (which would involve team leaders being in charge of the APM and old machinery covering the whole process from beginning to end) should be abandoned as the full operation was too fragmented, with several different machines involved, and unpredictable levels of material collected and processed from hundreds of different customers. This ruled out scenario 4. Scenario 3 then became the overall aim. The chosen management configuration model that would fit into this (figure 2) depicted one team leader in charge of two of the APM machines (8 people or 2 teams). This represented a move away from the current scenario (scenario 1) where one supervisor was usually in charge of just one machine.

The logic behind the groups final choice was a management configuration considered the best compromise between the criteria for successful teams (i.e., manageable team size, clear boundaries, good groupings for communications) and consistency with the vision of reduced number of managers, multiskilled and empowered teams, etc. However, a cost of the chosen scenario was that the group felt it would not be consistent with employee mental models (or perspectives) of the work, particularly with regard to confusion over how to divide raw materials between the APMs and old machinery, the reduced number of managers, and a relatively small number of staff on the new technology. Therefore, the group recognized that the organization would need to carefully plan how to obtain employee acceptance and prevent this scenario from failing. However, this was not discussed in detail during the workshop.

The decision was made by the stakeholders to pilot the chosen organizational scenario (scenario 3) and associated management structure as part of the technical evaluation and pilot of the APM. It was recognized that careful negotiations with the unions would be required in order to achieve this. After the workshop, members of the group extended the new management structure themselves, to enable it to fit different sites with varying configurations of old versus new machines (i.e., some sites may have an odd number of APMs rather than multiples of two).

Scenario 2 was devised at a later date by members of the group. The realities of the organizational context meant that several months after the initial workshops, opposition to TPM
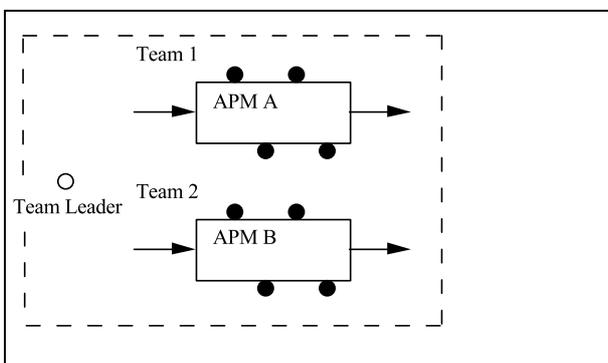
Figure 2. The chosen management and team structure.

and teamworking by staff and unions made the group rethink their options. By that stage, it became clear that only partial implementation of the organizational initiatives would be achieved. Thus, the scenarios tool helped the group develop a contingency plan, given the constraints within the organization. They used the division of responsibilities matrix to develop an interim goal (scenario 2). This scenario removed the TPM and teamwork activities that had been allocated to operators within scenario 3. So, although scenario 3 had originally been chosen due to the sense it made (for machine utilization, job satisfaction, etc.), it was felt that due to changing circumstances (i.e., politics within the organization) scenario 2 would be a good compromise and would now be used in the pilot of the APM.

But the long-term aim was still to reach scenario 3. Part of the power of the scenarios tool is that it enabled the group to think beyond the constraints of their organization (to some future state, i.e., scenario 3 or 4) and also allowed them to consider contingency plans given the reality. Moreover, it allowed them to contemplate how they would migrate from the current scenario to the desired one through the detailed discussions and outcomes of the workshops, and also via a substage (scenario 2).

5. OUTCOMES FROM THE SCENARIOS TOOL

Using the scenarios tool during a workshop produces a clear description (in the form of a table or similar document) of a range of possible organizational scenarios, the reasons behind them, the form they will take, and the pros and cons of each one. This helps the people using the tool to make decisions based on several different indexes. Another outcome of the scenarios tool is a detailed document or matrix which describes the roles, tasks, and division of responsibilities that different job grades will undertake in the new work organization.

5.1. Development and use of evaluation tools

Although the scenarios tool is primarily for planning, it is also valuable because its outcomes can be used to guide the development and structure of evaluation tools or instruments. Work redesign is a continuous process, so it requires evaluation and feedback on how work practices change in reality. The division of responsibilities matrix can be adapted for use within questionnaire and interview formats in order to measure and evaluate the extent to which the work organization has been met or modified in the light of changing circumstances. For instance, to help measure the distribution of tasks between different levels in the organization, employees can be asked to rate the extent to which they think different parties get involved in particular work-based activities. Besides that, perceptions and attitudes can be measured with regard to the properties of the existing work design using questionnaire scales. By the time evaluation took place it was obvious that the move towards scenario 3 would be limited and scenario 2 had become the interim goal. But it was still considered necessary to evaluate elements of scenario 3, just in case it was realized in the long term and people wanted to map out the route which had achieved it.

5.1.1. Questionnaire

A key element of scenario 3 was a broadening and empowering of operator roles. A questionnaire was developed to measure role breadth and empowerment. It was based on a measure of workgroup autonomy (Gulowsen 1972) and the division of

responsibilities matrix (operator and management tasks) generated from the first workshop. It began "To what extent do you and your immediate colleagues get involved in the following" and example items were "allocation of jobs amongst yourselves," "generating reports," and "applying the conduct code."

A measure of machine maintenance was also developed. It used the division of responsibilities matrix, including both operator and engineering tasks. This measure included key tasks such as Do you clean your own machinery? Do you fix minor faults? These measures used a five-point response scale: not at all, just a little, a moderate amount, quite a lot, a great deal. These tools may be unique to particular situations or scenarios. For more information on these measures, see Pepper *et al*. (1999).

The questionnaire was administered to all those involved in working with the APM at two different sites, around the time it was introduced (time 1) and again approximately one year later (time 2). Time 2 was after the pilot of the APM and shortly after some further successful negotiations with unions on progress towards scenario 3.

5.1.2. Interview

A structured interview was also developed to map the division of responsibilities between operators, managers, and engineers. This schedule asked the interviewee the extent to which each grade was involved in conducting certain tasks. These tasks included, machine maintenance, organization of work, and general use of the APM. The interview was derived from the division of responsibilities matrix as well as theory, so it could be unique to a particular scenario or situation. The interview enabled a more detailed and in-depth analysis of the existing work scenario and allowed us to investigate more of the specific tasks in the division of responsibilities matrix. It also allowed interviewees to comment on whether they felt that certain grades (such as operators) could potentially (or would like to) take over particular tasks (Figure 3). Therefore, the interview not only captured the status quo but also orientations towards change. A small selection from each grade were interviewed just before the APM became fully operational (to provide a benchmark at time 1) and again approximately one year later (time 2). This process was conducted over the two sites where the questionnaire was administered. The interview was also conducted at a third site several months after the new technology was implemented (time 2).

6. OUTCOMES FROM EVALUATION TOOLS

As expected, the findings from the interview and the questionnaire showed that movement towards scenario 3 was slow but scenario

2 looked promising. Previous union opposition meant that work roles were still fairly narrow and specialist but with some increases in variety and responsibility (consistent with scenario 2). Recent negotiations towards scenario 3 (e.g., regarding TPM) had not filtered through into organizational practice at that stage. However, where other initiatives were starting to be introduced (such as the management initiatives) there was change in line with scenario 2. Notably there were some improvements in general skill use, involvement in decision making, coaching, and management communication.

One problem with adopting scenario 2 was the many technical difficulties during implementation of the machinery and for some time afterwards. At the pilot site this meant there was not enough time to test out the work organization element before other sites started to use it. This prevented full evaluation or experimentation with scenario 2 and meant that frontline employee input was only on fairly minor work organization or design issues (e.g., delineating walkways around the machinery, design and arrangement of auxiliary equipment and communication devices).

From the information gathered using the evaluation tools, it also became apparent that the three sites worked with the APM in different ways and there was resistance to sticking to the one best way that the planning group had chosen. For instance, at one site the agreed number of team members was boosted by having an extra member of staff who did not work on the APM itself, but took on the tasks of fetching and removing materials for operators. Moreover, at another site the management structure had not been adhered to, because of local operational needs and the requirement for some additional machinery to process local material. This meant the evaluation tools gave an insight into change processes at the different sites.

Despite the slow start towards the ideal of scenario 3, there were employees who had attitudes consistent with it, as they showed a willingness to take on some of the duties such as machine maintenance. In general, managers were more positive towards the changes, followed by engineers and operators; this might be expected in a gradual shift in culture and attitudes. The interview evaluation tool was more sensitive at identifying these attitude shifts from scenario 1. The questionnaire was useful for gross measures across a large number of employees, but it was less detailed and less sensitive to smaller changes. We would expect the questionnaire to be especially useful when more substantial shifts in actual job content are made. So, although the scenarios tool was the driver, the evaluation tools were able to identify how well the ideal was being realized.

7. IMPLICATIONS AND CONCLUSIONS

This section outlines some theoretical and practical implications and conclusions from the use of the tools within this case study. There is a clear need for tools and methods to help organizations plan new work organizations. The scenarios tool was extremely useful for planning and enabling the clarification of key issues and tasks. It helped to identify a range of possibilities, a long-term objective (scenario 3), and the changes necessary to reach that objective; it also helped to structure some evaluation tools.

The scenarios tool provided the means to specify an interim solution (scenario 2) by enabling the group to consider the pragmatic "what are we able to do now" against the long-term

```
How often are the following people
involved in replacing
printer ink bottles?          Never Rarely Sometimes Usually Always NA/Don't Know

Operators
_____

Managers
_____

Engineers
_____

Someone else*
_____

* please indicate who

If you do not currently do this activity, do you think you/operators could do it?
```

Figure 3. Some interview questions.

objective. In other words, it helped to make explicit the link between designing shopfloor jobs and the political context, and enabled the planning group to eliminate aspects of the new work organization that were politically sensitive.

Although progress towards the full scenario (scenario 3) was slow (as shown by the evaluation tools), it would appear that the plan and aims were compelling enough to ensure the scenario was merely delayed and could remain on the agenda. The evaluation tools also prevented false claims of success as there was a systematic and measured assessment. Without these systematic planning and evaluation tools, the group may have had much more difficulty deciding what to do and evaluating their progress.

There are several implications for practical use of the scenarios tool. For instance, it was used in a nonsociotechnical way, as the technology had already been bought, and consideration was exclusively about the work organization that would fit around the given technology. Although the scenarios tool can provide several benefits even at this stage, it would be particularly useful for considering the work organization issues that arise during the design of new technology. Hence it would be beneficial to use the scenarios tool much earlier in the process (i.e., before the technology is purchased).

It may be fairly difficult for people in organizations to consider several different scenarios. But within this case study the participants were able to consider more alternatives in the second workshop. This may be due to the fact that the group was more open-minded about the possible machine groupings or the immediate management structures. One potentially problematic issue in the first workshop was that the existence of the long-term organizational initiatives (TPM, management training, etc.) prevented alternative (and perhaps better) scenarios from being considered for working with the APM. This occurred even though the group were encouraged to think more broadly, hence it highlights the possible effects of the organizational context on the ideal use of the tool.

The participants in the workshops were policy makers rather than frontline implementers, union members, or frontline employees. Again, this highlights the effect of the organizational context, for although the tool emphasizes the need to involve other stakeholders in the decision-making and implementation process, past involvement was low within this organization. However, a key point is that the scenarios tool can be used both online and off-line. The organization was not ready to use the tool online with frontline staff, but a great deal of progress could still be made by using it off-line with the planning group.

Moreover, the difficulty of involving a representative cross-section of frontline staff within a large multisite organization may mean that off-line use is sometimes more appropriate as long as employees are involved in other ways. Related to this are employee acceptance and the need for greater consideration (greater than occurred here) of preventive planning within or after the scenarios workshop. This planning might be for greater involvement of the unions in the initiative, or forums for educating and fine-tuning the responsibilities matrix involving employees and union staff at the relevant sites.

The findings from using the tools also reveal the slow and evolutionary nature of work organization implementation. The scenarios tool enabled the organization to devise small steps and

helped to provide structure and coherence to a cycle of planning, change, evaluation, and back to planning. The evaluation tools enabled the organization to observe the extent to which the goals were being met and identified small shifts in attitude towards the new scenario.

The scenarios tool raises awareness that there is not necessarily one best option. This is of theoretical and practical interest and was initially highlighted for the organization during the workshops when several possible scenarios were developed. And subsequently, at the individual sites, the preference for one best method has not yet worked in practic. Variations in local needs and requirements contribute to this, as well as employee adjustment towards a less specialized and leaner way of working. This suggests it may require some local control and flexibility in translating principle into practice at the different sites.

Indeed, the work organization literature advocates the empowering of sites or individuals to have some control over the way they work, within certain boundaries (e.g., broadly within a particular organizational scenario in which methods are minimally critically specified) (Cherns 1976, 1987, Hackman 1989). Moreover, local involvement and adjustments to the new ways of working would help to enhance ownership of them. These variations suggest that particular forms of work organization should not be considered as fully generalizable (i.e., slotted into place anywhere). We need to know more about the contingencies, circumstances, or boundaries under which these principles will be successful (Parker and Wall 1998).

Other issues relate to the local tailoring of the tool itself. The stakeholder group suggested some changes to the headings within the scenarios tool to make it more organization friendly. They suggested that the heading "logic" should be replaced by "reason for vision"; the heading "roles" should be expanded to "roles—who does what?" Alternative and additional headings have also been suggested elsewhere (Nadin 1996, Hesse 1998). We think the scenarios tool can easily be modified to particular organizations or problems in this way, as it is a guide to thinking through various issues, rather than a rigid, inflexible dogma. And we feel this is one of its strengths.

Ongoing research has updated the interview evaluation tool. Useful additions have come to light as TPM negotiations have progressed, as employees have gained more experience with the technology, and as ideas have become clearer about the detailed nature of the jobs and the division of responsibilities. This will allow further and more detailed mapping of the change in the organizational scenario (towards scenario 3).

In conclusion the scenarios tool seems to be robust within the realities of an organization undergoing a massive change, with all its cultural and political implications. The evaluation tools (questionnaire and interview) were useful for assessing the progress of the change and provided a necessary link between design and evaluation. Finally, the case study demonstrates that a multimethod range of tools can help with the planning and evaluation of change and therefore helps to fill the gap between principle and practice.

REFERENCES

ALDERFER, C.P., 1977, Group and inter-group relations. In Hackman and Suttle (eds) *Improving Life at Work* (Santa Monica CA: Goodyear), pp. 193–210.

BETTENHAUSESN, K., 1995, Group size. In Nicholson (ed.) *Encyclopaedic Dictionary of Organizational Behaviour* (Oxford: Blackwells), p. 206.

CHERNS, A.B., 1976, The principles of sociotechnical design. *Human Relations*, 29, 783–92.

CHERNS, A.B., 1987, The principles of socio-technical design revisited. *Human Relations*, 40, 153–62.

CLEGG, C.W., 1984, The derivation of job designs. *Journal of Occupational Behaviour*, 5, 131–46.

CLEGG, C.W., 1993, Social systems that marginalise the psychological and organizational aspects of information technology. *Behaviour and Information Technology*, 12, 261–66.

CLEGG, C., COLEMAN, P., HORNBY, P., MACLAREN, R., ROBSON, J., CAREY, N., and SYMON, G., 1996, Tools to incorporate some psychological and organizational issues during the development of computer-based systems. *Ergonomics*, 39(3), 482–511.

CLEGG, C., AXTELL, C.M., DAMODARAN, L., FARBEY, B., HULL, R., LLOYD-JONES, R., NICHOLLS, J., SELL, R., TOMLINSON, C., AINGER, A., and STEWART, T., 1997, Information technology: a study of performance and the role of human and organizational factors. *Ergonomics*, 40, 851–71.

CUMMINGS, T.G. and BLUMBERG, M., 1987, Advanced manufacturing technology and work design. In Wall, Clegg, and Kemp (eds) *The Human Side of Advanced Manufacturing Technology* (Chichester: Wiley), pp. 37–60.

DAVIS, L.E. and WACKER, G.J., 1987, Job design. In Salvendy (ed.) *Handbook of Human Factors* (Wiley: New York), pp. 431–45.

EMERY, F.E., 1959, Characteristics of socio-technical systems. *Document 527* (London: Tavistock Institute).

GOODMAN, P.S., DEVADAS, R., and GRIFFITH-HUGHSON, T.L., 1988, Groups and productivity: analysing the effectiveness of self-managing teams. In Campbell, Campbell *et al.* (eds) *Productivity in Organizations* (San Francisco CA: Jossey Bass), pp. 295–327.

GULOWSEN, J., 1972, A measure of work-group autonomy. In Davis and Taylor (eds) *Design of Jobs* (Harmondsworth: Penguin), pp. 374–90.

GUZZO, R.A., and SHEA, G.P., 1992, Group performance and inter-group relations in organizations. In Dunnette and Hough (eds.) *Handbook of Industrial and Organizational Psychology 3* (Palo Alto CA: Consulting Psychologists Press), pp. 267–313.

HACKMAN, J.R., 1987, The design of work teams. In Lorsch (ed.) *Handbook of Organizational Behaviour* (Englewood Cliffs NJ: Prentice Hall), pp.70–94.

HACKMAN, J.R., 1989, *Groups that work (and those that don't): creating conditions for effective teamwork* (San Francisco CA: Jossey Bass).

HACKMAN, J.R. and OLDHAM, G. R., 1976, Motivation through the design of work: test of a theory. *Organizational Behaviour and Human Performance*, 16, 250–79.

HACKMAN, J.R. and OLDHAM, G.R., 1980, *Work Redesign* (Reading MA: Addison-Wesley).

HESSE, D., 1998, An evaluation of the Sheffield method: a case study. MSc Thesis, Institute of Work Psychology, University of Sheffield.

MUMFORD, E., 1986, *Using Computers for Business Success* (Manchester: Manchester Business School).

MUMFORD, E., 1987, Sociotechnical systems design: evolving theory and practice. In Bjerknes, Ehn, and Kyng (eds) *Computers and Democracy: A Scandanavian Challenge* (Aldershot: Avebury), pp. 59–76.

NADIN, S., 1996, Participative tools to facilitate job design: further development of the scenarios tool and consideration of issues associated with participation. MSc Thesis, Institute of Work Psychology, University of Sheffield.

PARKER, S. and WALL, T., 1998, *Job and Work Design: Organizing Work to Promote Well-being and Effectiveness* (London: Sage).

PASMORE, W.A., 1988, *Designing Effective Organizations: The Sociotechnical Systems Perspective* (New York: Wiley).

PEPPER, K., AXTELL, C., WALL, T., CLEGG, C., and GARDNER, P., 1999, A measure of role breadth, forthcoming.

THORSRUD, E., 1972, Policy making as a learning process. In Cherns, Sinclair, and Jenkins (eds) *Social Science and Government: Policies and Problems* (London: Tavistock), pp. 39–67.

WEST, M. 1996, Working in groups. In Warr (ed.) *Psychology at Work* (4th ed.) (Harmondsworth: Penguin), pp. 359–82.

WEST, M.A., BORRILL, S., and UNSWORTH, K.L, 1998, Team effectiveness in organizations. In Cooper and Robertson (eds) *International Review of Industrial and Organizational Psychology 1998* (Wiley: Chichester), Vol. 13, pp. 1–48.

Training System Development in Ergonomics

M. M. Robertson

Liberty Mutual Research Center for Safety & Health, 71 Frankland Road, Hopkinton, MA 01748, USA

1. INTRODUCTION

Ergonomics training is recognized as a critical element in effective ergonomics, health and safety programs. When ergonomics training is part of a comprehensive, systematic approach to integrating ergonomics into an organization, it can play a key role in enabling an organization to link corporate goals with ergonomics practices, enhance organizational effectiveness and facilitate the change process. Training programs can modify how people work together, solve work-related problems and actively fulfill their role in ergonomics implementation. When a successful office ergonomics training program is implemented, the result is an increased ability for the worker to change his/her work environment, reduce the exposure to work-related risk factors and promote healthy work practices. High quality ergonomics training incorporates a "participatory" approach, in which end users, managers, engineers, designers, health professionals and others are involved. A participatory approach assures that each employee learns and develops the knowledge, skills and motivation to provide suggestions for improvement and change into the organization. It is this participatory aspect, along with the ergonomics training and work system design, that forms the basis for creating an improved work environment and continual change in the company.

Training alone does not constitute an ergonomics program, and it is not a panacea for reducing work-related musculoskeletal disorders (WMSD). Generally, work-related disorders include WMSD, such as low back pain, and other health disorders arising from exposures to hazardous conditions in the workplace. Typically, ergonomics training programs are designed to address these disorders as well as the contributing workplace risk factors. Further, training provides the mechanism by which workers' performance and well-being are enhanced to maximize an organization's investment in people and technology. Owing to the multifaceted nature of these workplace risk factors, a work system design or macro-ergonomics approach should be taken effectively to address the associated WMSD.

Training is an integral part of a larger work system with the individual or end user in the center. It is important that the entire work system be taken into account, including the balance of the interrelated system elements. These work system elements include: job design factors, work organization, environmental design, technology and organizational structure. All of these integrally related system elements should be considered effectively to minimize the negative health effects arising from poorly designed work systems. These training elements, processes and models are applicable to other work environments and conditions, such as aviation maintenance (Robertson 1998), and industrial ergonomics (McKenzie *et al.* 1985). Results of these training programs' effectiveness vary depending on several issues, like the social and organizational culture; however, significant and positive effects of training programs have been demonstrated on workers' ergonomics knowledge, reducing discomforts and increasing healthy work practices.

This chapter will focus specifically on the design, implementation and evaluation of office ergonomics training programs with the goal of reducing WMSD as related to office technology work environments. The importance of office ergonomics training programs, the elements necessary for successful programs, training models for designing and evaluating the effectiveness of training programs, and organizational examples of the application of these models are presented.

2. IMPORTANCE OF SYSTEMATIC OFFICE ERGONOMICS TRAINING PROGRAMS

2.1. Need for Training

Owing to the health effects, such as WMSD experienced in office environments, office ergonomics intervention strategies have been proposed and implemented throughout organizations (e.g. Sauter *et al.* 1990, Robertson and Robinson 1995). These strategies include: ergonomics training, re-design of workspaces, environmental re-design, biomechanics and physiological interventions (work–rest cycles); enhancement of user control, job and organizational re-design. All of these intervention strategies are critical elements that must be taken within a systems perspective to effectively minimize the negative health effects arising form the intensive use of office technologies (e.g. Robertson and Rahimi 1990, Robertson and O'Neill 1995, Hendrick 1997).

Several researchers and practitioners have established the need for and the importance of office ergonomics training (e.g. Verbeek 1991). A successful ergonomics training program should be incorporated into the overall organizational strategic plan for health and stress reduction programs. The linking of the ergonomics training program objectives to the organizational goals ensures the integration of ergonomics in the organizational culture. Based on organizational performance measures, it also is important to establish an evaluation plan to track the effectiveness of the ergonomics training program.

While engineering controls such as workstation re-design or the use of adjustable furniture are frequently suggested (Verbeek 1991), administrative controls, such as training must accompany them so that employees and management understand the need for change using video display terminals (VDT) and related new office technologies. In addition, training can assist in ensuring both managerial and employee support for the introduction and implementation of an ergonomics program. In addition to the basic principles of office ergonomics and approaches for incorporating them into work methods, training also is a necessity in order for employees to understand workstation set-up and the use of proper postures to avoid discomfort and WMSD (Gross and Fuchs 1990, Kukkonen *et al.* 1983, Robertson and Robinson 1995, Robertson and O'Neill 1995).

Without training employees on these ergonomics concepts, the presence of other administrative and engineering controls will result in only limited success. For instance, it was found the availability of adjustable furniture alone did not prevent the onset of overuse injury in some users. Verbeek (1991) did implement

a training program for use of adjustable furniture, and found that it significantly improved user posture when compared with an untrained group. Kukkonen *et al.* (1983) also were able to reduce neck and shoulder tension through education of workers who were found to be unfamiliar with the proper adjustment of their furniture and, as a consequence, were using poor working positions. Therefore, by providing employees with flexible, adjustable office workstations and environments, coupled with the knowledge of office ergonomics, the necessary elements can then be put into place to realize the benefits of a comprehensive, systematic office ergonomics program.

3. COMPONENTS OF A SUCCESSFUL OFFICE ERGONOMICS TRAINING PROGRAM

3.1. Using an Instructional Systems Design Approach

The design of an effective training program should include several processes and activities involving: (1) conducting a needs analysis, (2) designing the training materials, (3) developing the training materials, (4) implementing and delivering the training and (5) evaluating and measuring the effectiveness of the training. There are several instructional system design models, and each incorporates these basic processes and phases to create effective training programs (e.g. Gagne *et al.* 1988, Goldstein 1993). For each instructional system stage, different types of information are conveyed and different techniques are used to analyze the collected information. See Figure 1 for the instructional system design model and phase activities.

3.2. Needs Analysis

In the needs analysis phase, an organizational, task (job) and person analysis is conducted. Two questions need to be answered: what is the *current* performance of the organization and the workers? and what is the *desired* performance for the organization and the workers? The derivation of the training objectives and their linkage to the corporate goals are accomplished at this stage. If the performance problem is identified as a lack of knowledge, skill or ability of the workforce, then a training program should be designed as a necessary intervention strategy. However, if through the needs analysis it is discovered that the performance problem is due to a poorly design work system (e.g. workstation design, work demands), then training is not the *only* intervention solution. This phase also includes the development of the criteria for evaluating and measuring the training effectiveness. These criteria are linked to the training objectives and are established at the trainee, supervisor, departmental (strategic business unit) and organizational performance levels.
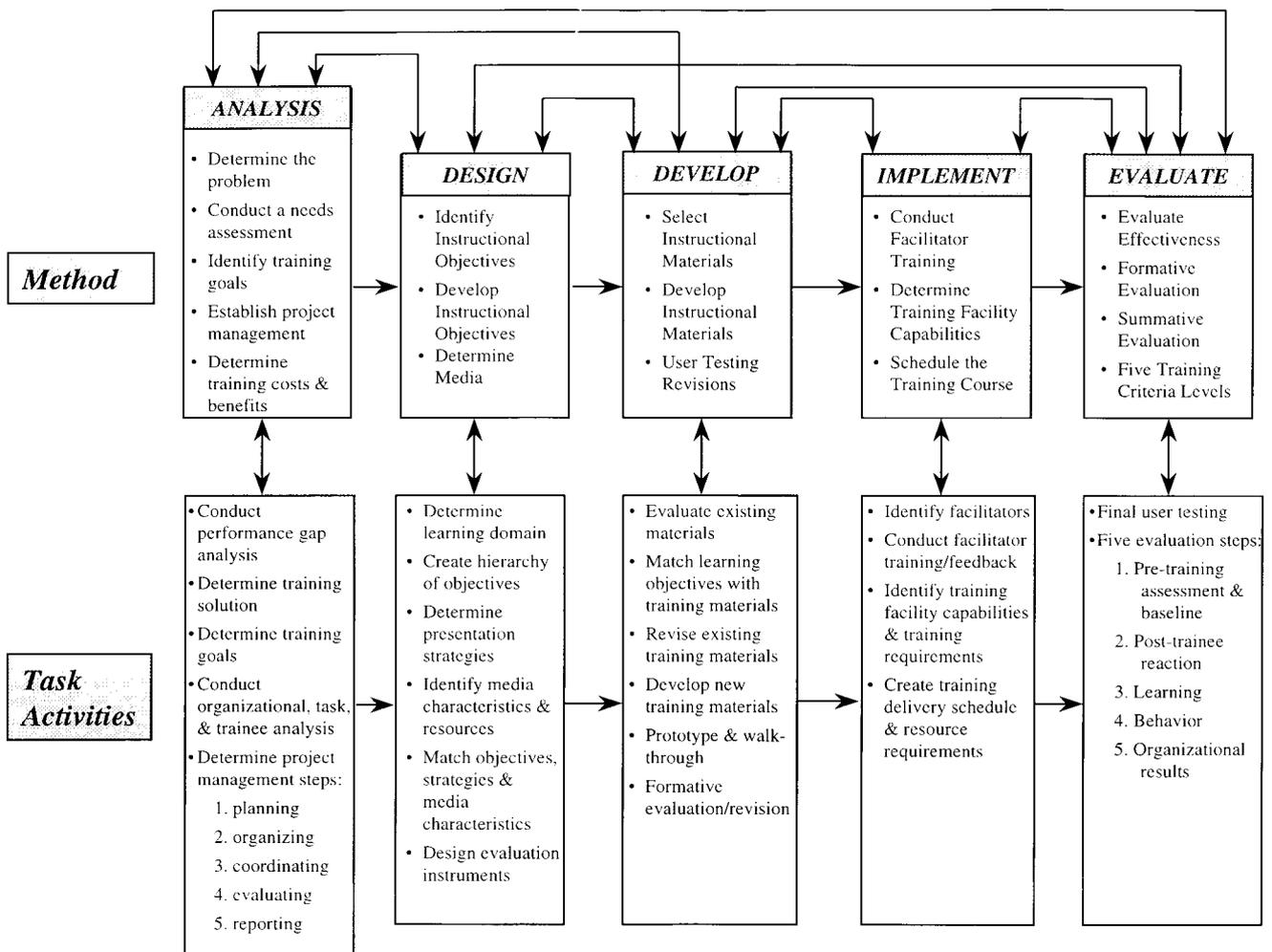
Figure 1. Instructional system design phase, activities and feedback.

3.3. Design

In the design phase, the needs analysis results are used to determine how the training objectives are going to be met. This involves establishing:

- training prerequisites;
- trainee population;
- desired learning outcomes (e.g. knowledge, skills, and abilities in cognitive; affective and psychomotor domains);
- training media and techniques;
- training environment;
- learning conditions (e.g. individual differences);
- instructional strategies; and
- learning principles.

Contextual factors involving organizational, environmental, and social issues, such as designing training for newly hired workers versus experienced workers are determined in this phase. Since training occurs within an organizational culture, how the organization values the training and its integration, including the link to corporate strategy, must be accounted for in this design phase. Information derived from the needs analysis phase is used to identify these factors.

Instructional strategies must be selected before developing the training materials themselves in order to outline how the instructional activities will relate to the accomplishment of the training objectives and goals. Gagne *et al.* (1988) proposed a series of nine instructional events that must occur for learning to take place which apply learning theories to the development of training. These nine events include:

- gaining attention of the trainee;
- informing the trainees of the training objectives;
- using recall or transferring from existing experience of the trainee;
- presenting training material to be learned;
- providing learning guidance or elaboration;
- eliciting desired performance;
- providing feedback;
- assessing performance; and
- enhancing retention and facilitating transfer of training to actual task performance.

Different instructional media have different capabilities and strengths for providing the various events (components) of instruction. Optimal selection of the training delivery system is achieved by matching the training media strengths with the training objectives as determined in each instructional event. In developing a strategy for instruction, the choice of the training delivery system can be assigned by event-by-event, objective-by-objective to accomplish the training goals.

3.4. Development

Development of training materials in whichever media are selected is the next phase in the instructional system design process. It is important to note that the instructional design process itself determines the effectiveness of a training program not the training media and technology. If sound instructional design principles are used, the instructional designer will choose instructional or other technologies that meet the functional and training requirements. Since various media interact with certain instructional methods, the development phase involves piloting and walkthroughs of all training modules. During this phase,

instructional strategies are applied sequentially to each training activity, then the most effective media delivery techniques are selected. Additionally, the training delivery format is determined. It may be instructor/facilitator controlled, performance-controlled or trainee self-paced style.

3.5. Implement

Implementing the training is the next phase consisting of scheduling how and when the training is to be delivered. Also, train-the-trainer or facilitation skills are developed and need to be taught and practiced before the trainers and/or facilitators present the training. It is important that the facilitator learns and practices how to lead and control active and interactive discussions among the trainees.

3.6. Evaluate

Evaluating the effectiveness of the training program and providing feedback to the organization and trainees is the last phase in the instructional system design. When the training results match the training goals and objectives, the training program can then be concluded to be effective. There are five major purposes for conducting training evaluation (Ford and Sego 1990). These purposes and definitions are:

- content validity: the relevancy between training content and the job;
- training efficiency: under or over training by the training program;
- training validity: the extent of the trainees' learning of the training material;
- transfer validity: trainees' job performance after training; and
- predictive validity: predicting job performance by training performance.

A systematic, multiple measures, four level training evaluation model can be used for evaluating training effectiveness (Kirkpatrick 1975). This four level evaluation model process includes:

- Level I: Reaction to the training program (How well did the trainees like the training program? How relevant, useful was the training?).
- Level II: Learning of principles, facts, techniques, and attitudes (What principles, facts, and techniques were learned? What attitudes were changed?).
- Level III: Behavior relevant to job performance (What changes in job behavior resulted from the training program?).
- Level IV: Organizational Results of the training program related to organizational objectives. (What were the tangible results of the training program in terms of reduced cost, improved quality, improved quantity, reduced injuries and lost work days, etc?).

Measurements that may be taken at each of these training evaluation levels are as follows: Level I: post-training questionnaire asking the trainee to evaluate the value and usefulness of the training; Level II: pre–post-questionnaires/tests assessing how well the trainee learned the information taught as well as observations/ interviews with the trainee; Level III: assessment of the trainees' behavior on the job — how well was the trainee able to transfer the knowledge and skills to the job — this may be completed by observations and/or interviews; and Level IV: results and impacts

of the training program on organizational performance measures, benchmarking and tracking of organizational performance measures.

4. OFFICE ERGONOMICS TRAINING PROGRAM EVALUATION MEASURES

For evaluating an ergonomics training program, the following measurements may be incorporated at each of the evaluation levels. Level I measurements may include questions on the post-training questionnaire regarding the usefulness and value of the training as well as the relevancy of the training to the workplace. For Level II, a paper and pencil *pre-knowledge* office ergonomics training test may be given as well as various measurements may be taken *before* the training such as an observational analysis of posture and work habits, and postural discomfort surveys. Also included in Level II, a *post-training* ergonomics knowledge test as well as a *post-training* discomfort survey identical to the pre-training survey measures may be administered. Level I and II also may serve as a type of formative evaluation of the instructional materials indicating if the training program was well received, the training materials were clear and understandable, and the training objectives were met. For Level III: various measurements may be taken *after* the ergonomics training to evaluate the transfer of the training by the trainee to the workplace. These may include an observational ergonomics analysis and interviews of what the trainee changed or did not change in their workplace as a result of the training. Additionally, on the post-training ergonomics knowledge test, open-ended questions may be asked of the trainee regarding how they are going to *use* the training when they return to work. For Level IV, various organizational performance measurements may be taken related to the results of office technology work environments, such as reported WMSD, time off work, worker's compensation rates, and health and stress-related costs. Pre- and post-training organizational health and safety performance measurements should be benchmarked and tracked over time. Tracking these pre- and post-health and safety performance measures, as well as other training costs, determines the basic variables for calculating a return on investment (ROI) of the training program.

4.1. Having Senior Management Support and Commitment

The foundation of any successful organizational program is senior management support. Senior managers must have the vision and commitment to reduce adverse health effects and increase quality of life and performance through the use of office ergonomics. When top decision-makers clearly support the mission and purpose of an office ergonomics program, an organizational culture and safety climate change can occur. Without such a commitment, a pervasive organizational change is unlikely.

4.2. Training for Supervisors and Middle Managers

Linked to senior management support, is training for supervisors and middle managers. These individuals interact daily with the workers who are ultimately responsible for implementing the new strategies. Mid-level mangers also need the support of upper-level management in implementing the new office ergonomics skills, knowledge and practices in the field. This support can

take many forms, but certainly includes the time to attend appropriate office ergonomics training courses. With this commitment, supervisors will have the opportunity to use their own ergonomics skills and knowledge in addition to managing a cultural safety and organizational change.

4.3. Creating a Responsive Environment

Training of managers and supervisors also is necessary in order to provide a responsive environment in which employees are encouraged through reinforcement and reward to utilize their training skills and knowledge. Since supervisors have more influence on the daily performance of individual employees, their participation in the training process is essential for the success of the ergonomics training program. It was found that supervisors responded best to training that emphasized situations over which they had some measure of control, and that such training made them more cooperative and supportive of change. This type of training assures that supervisors learn to respond effectively to suggestions regarding office ergonomics given by employees.

4.4. Supporting Active Participation

Participating in the creation, development and implementation of an office ergonomics training program stimulates a feeling individual ownership. Active involvement creates a sense of commitment to supporting the training program goals and a willingness to engage in the required cultural change process. Being a member of a team — that is designing and implementing an office ergonomics training program — is motivating, rewarding, and beneficial to the individual and the organization. Working together on a cross-functional, interdisciplinary team provides a unique strength in designing and developing a training program. If there is a lack of active worker participation in the training program, the worker's motivation for and understanding the material presented is low and their resistance to change is high (Luopajarvi 1987).

4.5. Developing Active Learning Experiences

More effective instructional methods, sometimes called "inquiry" or "discovery" learning, emphasize the involvement of learners. These methods involve students by having them participate in problem-solving activities and group discussions. The strength of this approach is that the use of group exercises and office ergonomics related case studies promotes an active and motivating learning environment. It creates an interactive, highly motivating approach, since students are doing more than just passively receiving the information – they are actively applying and using the concepts and skills. To strengthen further this approach, training courses can be co-facilitated by knowledgeable trained workers in office ergonomics and specific work processes. These facilitators can encourage students to participate, by bringing "real world" experiences into the classroom, and are viewed as relevant and valid experts. Also, if the class make-up consists of individuals from various job positions in the organization, active and interactive discussions can occur providing an ideal opportunity to discuss "real" problems together.

4.6. Continuous Learning and Improvement

Every work system changes over time. In the systems approach, office ergonomics training and practice must be viewed as part of the overall health, safety and ergonomics program. As such,

the training program must be adapted to changes that occur elsewhere in the system, as well as to changes in health and safety practices and technology. The idea of continuous change and adaptation is fundamental to making any system responsive to the needs of the workers. Financial and organizational resources must be committed actively to support the safety culture change process. This includes the commitment of human resources within the company, such as administrators, trainers and curriculum developers, and media and computer application developers. Continuous improvement is not a short-term activity. Rather it requires long-term management commitment to continuously adapt and improve the program.

4.7. Providing Continuous Feedback

Performance improves more quickly when people are given feedback concerning their successes (or lack of them). It is vital that timely feedback is given to all workers and managers about the results and effects of the office ergonomics training program. Feedback provides information to accomplish two performance improvement goals: (1) improve the program and identify necessary corrective actions and (2) reinforce the positive outcomes and benefits of using office ergonomics skills in the workplace.

5. OFFICE ERGONOMICS TRAINING CASE STUDIES

5.1. Background of the Office Ergonomics Training Programs

Two telecommunications companies designed, developed, implemented and evaluated an office ergonomics training program to (1) reduce adverse health effects from computer intensive work and (2) impart knowledge about how to effectively use their new ergonomically designed VDT workstations. Company 1 involved a large telecommunications company where 3500 telephone information operators, customer service representatives and supervisors were trained (Robertson and Robinson 1995). Company 2 involved 20 engineer designers, system engineers and manufacturing engineers who used office technologies extensively for their work at a large telecommunications company (Robertson and O'Neill 1995). In both companies, ergonomically designed workstations were provided with a moderate to high level of environmental control allowing for a high level of user control.

5.2. Office Ergonomics Course Content

The content of these office ergonomics training programs included: definition of ergonomics, basic physiology of the upper extremities, causes of discomforts and injuries, ergonomics principles regarding workstation layout, techniques on how to adjust and use their workstations properly, recommendations for analyzing the employees workstation, procedures to follow when they feel uncomfortable (management's policies on who to contact), and relaxation and exercise techniques to relieve office technology stress. For company 1, a three-tiered training program was designed for senior managers, supervisors and employees. Each training program included specific content areas related to their job responsibilities including how they were expected to respond when employees approached them with ergonomics problems. This allows the opportunity to support interaction and

participation between employees and managers to support the changes in the work environment. For company 2, a similar approach was taken; however, there was more focus on how to allow for a high level of user control as the ergonomics workstations were designed with the intent to allow for high flexibility and mobility. Managers involved in this training program were faced with different change issues as employees were applying their ergonomics knowledge in a more systematic manner by re-arranging their workstations and work environment components.

5.3. Office Ergonomics Training Evaluation Results

Overall, significant and positive effects of the ergonomics VDT training for both companies were demonstrated. For company 1, all four levels of the evaluation model were applied and clearly documented. Significant results of the training program was accomplished for all 3500 individuals included in the training program. Company 2 demonstrated positive results for three of the evaluation levels and results of the fourth level showed positive results in that there were significant decreases in self-reported musculoskeletal discomforts for the 20 individuals involved in the training.

5.3.1. Level I: Reaction

For Level I evaluation, the trainees from both companies rated the training highly favorable, useful, and informative. Over 90% of the trainees from both companies reported that the course material was relevant to their current office workspace design.

5.3.2. Level II: Learning

For Level II evaluation, there were significant changes in the amount of knowledge gained by the trainee's concerning office ergonomics as measured by pre- and post-training knowledge tests for company 1 and company 2.

5.3.3. Level III: Behavior

For Level III evaluation, significant positive behavioral changes of the trainees as measured by self-reported behavioral changes and observed changes were found and reported for both companies. For company 1, > 80% of the trainees reported that they had applied the ergonomics knowledge to their jobs, which included the correct placement of the VDT screen, the position of their wrists at the workstation, and their sitting posture. Follow-up observations and interviews by the corporate ergonomist and staff confirmed these self-reported behavioral changes. For company 2, all participants exhibited a high level of user control as they continually adjusted their workstations to meet various job demands. This involved arranging the workstation for sitting and standing postures. In addition, a significant decrease in overall discomfort was found in this group. Follow-up interviews and observations by the ergonomist supported these results, as > 80% of the trainees said that they were able to apply many of the principles taught in class to their workplace. Of the changes to the workplace reported, most were adjustments to the chair, placement of the monitor, workstations configuration and layout, and height adjustments of the keyboard and working surfaces. Many of the participants reported that the awareness developed from training led to changes in posture and an increase in the number of breaks for exercises or movement. In company 1, there

was some ability to change the workstation configuration within some defined constraints that provided some means of user control at the individual level. For company 2, the workstations were highly mobile and groups of workers were able to change not only the configuration of their individual workstation, but the overall configuration of the group's work area. One manager was very supportive and encouraged his workgroup to change their workstation configurations within certain boundaries to support their varying job processes.

5.3.4. Level IV: Results

For Level IV training evaluation, a ROI analysis for the office training program for company 1 revealed that the program resulted in a positive payback for the individuals trained and their company. The ROI analysis involved calculating the direct and indirect costs of company-defined WMSD, the training development and implementation costs, the workstation redesign costs, the ergonomics workstation evaluation costs as well as the number of reported WMSD that occurred after the implementation of the office ergonomics training program. Company 1 experienced a significant decrease of 16% in reported WMSD as well as a significant decrease of 21% in lost work days in the year following the office ergonomics training program. Additionally, in years 2–4 after the implementation of the office ergonomics training and workstation changes, this significant decrease in the number of WMSD continued.

For company 2, as the knowledge of control of their workstation increased, overall self-reported stress decreased. A significant reduction of 31% in self-reported upper/lower back discomfort was reported by the engineers who received the office ergonomics training and the highly adjustable workstation. For this same group of engineers, a 56% reduction in self-reported upper limb discomfort was found.

5.3.5. Successful components

Other successful components of the training were the commitment by top management to the ergonomics program itself, active involvement of the employees, positive response by management to employee's requests regarding office workstation redesign or reconfiguration, and continuous support of management in applying the office ergonomics principles to the work environment. Furthermore, an essential part of sustaining the results of a successful ergonomics training program is awareness by the supervisors in knowing what employee behaviors to reinforce. This was effectively completed in company 1 as part of the overall strategic plan to address health issues associated with office technology work. For company 2, a comprehensive ergonomics program is being designed and is planned to be implemented in the future.

6. SUMMARY AND CONCLUSIONS

In these two studies, overall significant and positive effects of the office ergonomics training programs were demonstrated for each of the training evaluation levels. These studies also support the contention that the combination of user control and training are important — since they can provide the worker with a high degree of environmental control through an increased knowledge of office

ergonomics and the ability to effectively apply ergonomics principles to their office work environment.

With an increase in office ergonomics knowledge as well as the implementation of an evaluation process the value of office ergonomics training can be demonstrated. A well-designed ergonomics program, coupled with an ergonomically designed work environment provides the foundation for creating a responsive environment for the employee and manager. As individuals interact with one another, applying and seeking ergonomics solutions, a sense of participation is created forming the basis of a positive change management process. Research and case studies show that a systems approach can decrease health and stress effects as well as reduce office ergonomics problems associated with WMSD. Coupling an effective training program with other engineering, work organizational factors provides a systematic approach to alleviating the multifaceted negative health effects experienced in office technology environments. Ergonomics training programs and environmental control can be successful intervention strategies at both the individual and organizational level in preventing WMSD, related office environment job stress and enhancing organizational effectiveness.

REFERENCES

FORD, J.K. and SEGO, D., 1990, *Linking Training Evaluation to Training Needs Assessment: A Conceptual Model.* Technical Paper, 90-69 (Texas: Air Force Human Resources Laboratory).

GAGNE, R.M., BRIGGS, L.J. and WAGER, W.W., 1988, *Principles of Instructional Design*, 3rd edn (New York: Holt, Rinehart & Winston).

GOLDSTEIN, I.L., 1993, *Training in Organizations*, 3rd edn (Pacific Grove: Brooks/Cole).

GROSS, C.M. and FUCHS, A., 1990, Reduce musculoskeletal injuries with corporate ergonomics program. Occupational Health and Safety, 1, 29–33.

HENDRICK, H., 1997, Organizational design. In G. Salvendy (ed.), *Handbook of Human Factors* (New York: John Wiley), 470–94.

KIRKPATRICK, D., 1975, Techniques for evaluating training programs. In D.L. Kirkpatrick (ed.), *Evaluating Training Programs* (Madison: American Society for Training and Development), 1–17.

KUKKONEN, R., LUOPAJARVI, T. and RIIHIMAKI, V., 1983, Prevention of fatigue amongst data entry operators. In T.O. Kvalseth (ed.), *Ergonomics of Workstation Design* (Guildford: Butterworths).

LUOPAJARVI, T., 1987, Workers' education. Ergonomics, 30, 305–11.

McKENZIE, F., STORMENT, J., VAN HOOK, P. and ARMSTRONG, T.J., 1985, A program for control of repetitive trauma disorders associated with hand tool operations in a telecommunications manufacturing facility, Journal of the American Industrial Hygiene Association, 46, 674–8.

ROBERTSON, M.M., 1998, Maintenance resource management. In M. Maddox (ed.), *Human Factors Guide for Aviation Maintenance* (Washington, DC: FAA Office of Aviation Medicine).

ROBERTSON, M.M. and O'NEILL, M.J., 1995, Effects of user and environmental control on stress, performance and group effectiveness. In *Proceedings of the 43rd Annual Meeting of the Human Factors and Ergonomics Society* (Santa Monica: HFES).

ROBERTSON, M.M. and RAHIMI, M., 1990, A systems analysis for implementing video display terminals. *IEEE Transactions Engineering Management*, 37, 55–61.

ROBERTSON, M.M. and ROBINSON, M., 1995, Enhancing user control

of VDT work environments: training as the vehicle. In *Proceedings of the 39th Annual Meeting of the Human Factors and Ergonomics Society* (Santa Monica: HFES), 417–21.

SAUTER, S.L., DAINOFF, M.J. and SMITH, M.J. 1990, *Promoting Health and Productivity in the Computerized Office* (London: Taylor & Francis).

VERBEEK, J., 1991, The use of adjustable furniture: evaluation of an instruction program for office workers. *Applied Ergonomics*, 22, 179–84.

Usability and Product Design

P. W. Jordan

Human Factors, Philips Design, Building W, Damsterdiep 267, 9700 AE Groningen, The Netherlands

1. INTRODUCTION

Perhaps the most important reflection on how seriously usability issues are now being taken is the sharp increase in the number of professionals employed by industry who are charged with ensuring that products are easy to use. These include human factors specialists and interaction designers.

Usability may be one of the few areas left to manufacturers where it is still possible to gain a significant commercial advantage over the competition. Manufacturing processes have now reached a stage of sophistication whereby any possible advantages in terms of manufacture quality or cost savings are likely to be marginal. Offering customers "user-friendly" products can be seen as something new in markets where the technical and functional specifications will vary little between brands.

Aside from the commercial implications, lack of usability can have effects that range from annoying the users to putting the users' lives at risk. While lack of usability in a video cassette recorder (VCR) may result in the user recording the wrong television program, lack of usability in a car stereo may put lives at risk by distracting drivers attention from the road.

2. USABILITY

Informally, usability issues can be thought of as pertaining to how easy a product is to use. They are, then, to do with the "user-friendliness" of products. More formally, the International Standards Organisation (ISO) defines usability as "the effectiveness, efficiency and satisfaction with which specified users can achieve specified goals in particular environments."

Effectiveness refers to the extent to which a goal, or task, is achieved. In some cases, the distinction between a task being achieved successfully or not may simply be success or failure on that task. However, there are also situations where effectiveness could be measured in terms of the extent to which a goal was achieved. Consider, for example, an industrial production machine. If the machine operator's goal were to produce 100 components per day, then if s/he were able to produce 80 per day, then this might be seen as an effectiveness level of 80%.

Efficiency, meanwhile refers to the amount of effort required to accomplish a goal. The less the effort required, the higher the efficiency. Effort might be measured, for example, in terms of the time taken to complete a task or the numbers of errors that the user makes before a task is completed. Consider the example of creating a new file with a word-processing package. If the user is able to create the new file instantly at the first attempt, then the package would be considered more efficient than if the user had to spend a couple of minutes thinking what to do, or than if s/he had activated several inappropriate commands before eventually finding the right one.

Satisfaction refers to the level of comfort that users feel when using a product and how acceptable the product is to users as a means of achieving their goals. In general, satisfaction might be seen as the most important aspect of usability for products whose use is voluntary. For example, if users of consumer products do not find them satisfying to use, they do not have to use them. Perhaps more importantly from a commercial point of view, they do not have to buy any more products by the same manufacturer. Conversely, in situations where people are "forced" to use products, such as manufacturing equipment or products used in professional environments (e.g. software packages), it might be argued that effectiveness and efficiency are equally or more important — at least from the point of view of the employers!

An important thing to note about the ISO definition of usability is that it makes clear that usability is not simply a property of product in isolation, but rather that it will also be dependent on who is using the product, the goal that they are trying to achieve and the environment in which the product is being used. Usability is, then, a property of the interaction between a product, a user and a task, or set of tasks, that s/he is trying to complete.

Examples of people characteristics that may have an effect on usability include:
- Experience (previous experience of the particular product).
- Domain knowledge (previous experience of similar products).
- Cognitive abilities.
- Physical abilities.
- Age and gender.

3. COMPONENTS OF USABILITY

Most of the user characteristics mentioned above may be thought of as comparatively stable. They are characteristics that, if they change at all, will probably change over a comparatively long period. The exception to this is "experience" — users' performance with a product is likely to improve significantly the more that they repeat particular tasks with a product. Thus, the usability of a product for a particular person completing a particular task may change very quickly as the task is repeated. To reflect this five separate components of usability have been identified.

3.1. Guessability

Guessability is a measure of the cost to the user in using a product to perform a new task for the first time — the lower the cost (e.g. in terms of time on task or errors made) the higher the guessability.
- *Guessability*: the effectiveness, efficiency and satisfaction with which specified users can complete specified tasks with a particular product for the first time.

3.2. Learnability

Learnability is concerned with the cost to the user in reaching some competent level of performance with a task, but excluding the special difficulties associated with completing the task for the first time.
- *Learnability:* the effectiveness, efficiency and satisfaction with which specified users can achieve a competent level of performance on specified tasks with a product, having already completed those tasks once previously.

3.3. Experienced User Performance (EUP)

This refers to the relatively unchanging performance of someone who has used a product many times before to perform particular tasks. Although performance may not always level off at an asymptotic level, there will probably come a stage with most products where significant changes will only occur over comparatively long time scales.

- *EUP*: the effectiveness, efficiency and satisfaction with which specified experienced users can achieve specified tasks with a particular product.

3.4. System Potential

System potential represents the maximum level of performance that would be theoretically possible with a product. It is, then, an upper limit on EUP.

- *System potential:* the optimum level of effectiveness, efficiency and satisfaction with which it would be possible to complete specified tasks with a product.

3.5. Re-usability

This component of usability refers to the possible decrement in performance after the user has not used the product for a comparatively long period of time.

- *Re-usability*: the effectiveness, efficiency and satisfaction with which specified users can achieve specified tasks with a particular product after a comparatively long period away from these tasks.

The comparative importance of the various components of usability will depend on the context of use of the product. For example, guessability and learnability will be important where there are a high proportion of one-off users or where users must familiarize themselves with a product within a comparatively short time. EUP will be important where a high level of expert performance is required and system potential is important when it is a limiting factor on performance. Re-usability is important when a product is used in intermittent "bursts."

Each of the components of usability is loosely associated with a different section of a notional learning curve. An example of such a curve is illustrated in Figure 1.

4. DESIGNING USABLE PRODUCTS

Ten principles of usable design have been identified and are summarized below. Products which display these characteristics will support superior levels of usability. Detailed descriptions of each of these characteristics are given in *An Introduction to Usability* (Jordan 1998).

- *Consistency*: designing a product so that similar tasks are done in similar ways.
- *Compatibility*: designing a product so that its method of operation is compatible with users' expectations based on their knowledge of other types of products and the "outside world."
- *Consideration of user resources*: designing a product so that its method of operation takes into account the demands placed on the users' resources during interaction.
- *Feedback*: designing a product so that actions taken by the user are acknowledged and a meaningful indication is given about the results of these actions.
- *Error prevention and recovery*: designing a product so that the likelihood of user error is minimized and so that if errors do occur they can be recovered from quickly and easily.
- *User control*: designing a product so that the extent to which the user has control over the actions taken by the product and the state that the product is in is maximized.
- *Visual clarity*: designing a product so that information displayed can be read quickly and easily without causing confusion.
- *Prioritization of functionality and information*: designing a product so that the most important functionality and information is easily accessible to the user.
- *Appropriate transfer of technology*: making appropriate use of technology developed in other contexts to enhance the usability of a product.
- *Explicitness*: designing a product so that cues are given as to its functionality and method of operation.

5. EVALUATING PRODUCTS FOR USABILITY

There is an ever increasing selection of methods available for evaluating product designs for usability. Seventeen of the most commonly used ones are described *in An Introduction to Usability* (Jordan 1998).

Broadly, methods can be classed as either empirical or non-empirical (Table 1). Empirical methods involve observing users interacting with products or asking users to comment on their perceptions of a product's usability. Non-empirical methods involve an investigator making an expert judgement about a product's usability or a structured check of a product's design qualities. As a rule, it may be preferable to take an empirical approach — there is no substitute involving users when it comes to learning about how to optimize user-product interaction. Nevertheless, there are circumstances where it may be preferable to take a non-empirical approach — for example, where it is
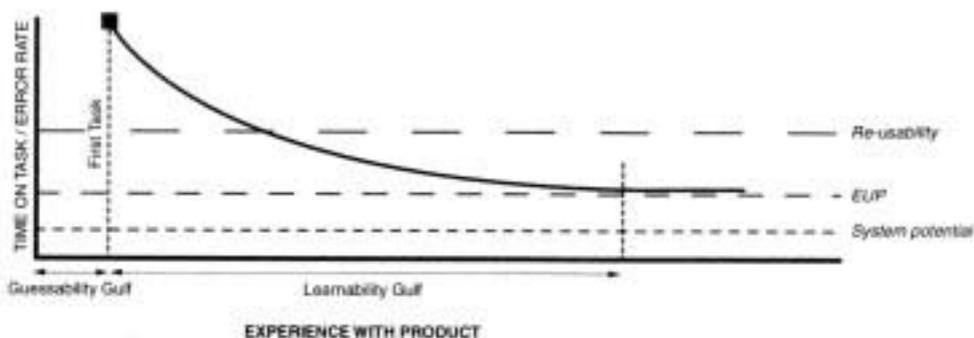
Figure 1. Idealized learning curve illustrating the five components of usability.

Table 1. Usability evaluation methods.

Empirical methods	Non–empirical methods
• Private camera conversations	• Task analyses
• Co-discovery	• Property checklists
• Focus group	• Expert appraisals
• User workshops	• Cognitive walkthroughs
• Think aloud protocols	
• Incident diaries	
• Feature checklists	
• Logging use	
• Field observation	
• Questionnaires	
• Interviews	
• Valuation method	
• Controlled experiments	

difficult to get hold of users for a test, where confidentiality is an issue, or where there are severe financial or time constraints on an evaluation.

Another option, where the circumstances of the evaluation present tight constraints, is to apply an empirical method, but in a "stripped down" form. This means setting aside some of the checks, balances and sampling techniques that would be included when a method was applied in its most rigorous form. The penalty of such informal (or "quick and dirty") approaches may be a reduction in the veracity of the data generated. Nevertheless, where the option is between a stripped down empirical evaluation and no empirical evaluation at all, the quick and dirty test can be a sensible approach. Indeed, compared with practice in research and academic environments, much of the empirical testing carried out by industry-based practitioners is informal in nature. The art of conducting an effective informal evaluation lies in knowing which elements of the test design can be relaxed without rendering the evaluation invalid. This is a complex and context dependent issue. For a comprehensive discussion of industry-based usability evaluation, including examples of best practice with respect to conducting evaluations within tight constraints, see (Jordan *et al.* 1996).

6. USABILITY EVALUATION IN THE PRODUCT CREATION PROCESS

6.1. Iterative Design and Prototyping

The principle of "iterative design" is that product concepts evolve through a design/evaluation cycle. This means that the first design idea will be evaluated in terms of how well it meets the users' needs. The strengths and weaknesses will be assessed and a further design developed on the basis of this. This in turn is evaluated and so on until a product is created that conforms to pre-specified levels of usability.

There are different prototyping options of differing degrees of realism and sophistication that can be used in the design/evaluation cycle. Some examples are given below.

6.1.1. *Product specification*

Perhaps the most basic form of prototyping is simply to give a verbal or written description of the form and functionality of the proposed product.

6.1.2. *Visual prototypes*

These are simply visual representations of a product. They could be paper-based sketches or drawings or on-screen representations created using drawing or computer-aided design packages. They may be supplemented with written or verbal descriptions of a product's functionality or procedures for operation.

6.1.3. *Models*

Sometimes physical representations or models of the product are created. Typically, these will be made of wood or of polystyrene foam. Sometimes they will include additional materials to weight the model so that its weight is representative of the proposed product.

6.1.4. *Screen-based interactive prototypes*

These are software-based representations of products. They offer simulated interactions — for example, by clicking the mouse on a representation of one of the product's controls. The screen-based representation will then change state to represent how the product would react to a particular interaction.

6.1.5. *Fully working prototypes*

These are prototypes that — at least from the point of view of the user — are barely distinguishable from a fully manufactured product. In the case of software products these will be the same as screen-based interactive prototypes — software that supports interaction and responds in the same way that the final product would do. However, where the product has physical dimensions then the prototype will also have these.

7. CONCLUSION

Usability is a "hot" issue at the moment. Users are demanding usable products and manufacturers are beginning to see designing for usability as bringing a competitive advantage. Certainly, usability is a commercial issue. However, it is more than just that. Providing adequate standards of usability is part of the responsibility that those producing products owe to those who use them. We live in a high-technology era and live with technology that has the power to enrich our lives — in our homes, workplaces and communities. Yet, all too often this technology proves difficult for us to control. The consequences of this can range from mild annoyance to serious injury or death. Usability is about giving control back to the user — producing useful, pleasurable, safe and helpful products that will enhance the quality of our lives.

REFERENCES

JORDAN, P.W., 1998, *An Introduction to Usability* (London: Taylor & Francis).

JORDAN, P.W., THOMAS, B., WEERDMEESTER, B.A. and McCLELLAND, I.L., 1996, *Usability Evaluation in Industry* (London: Taylor & Francis).

User-centered Systems Analysis in Aircraft Maintenance

D. A. Mitta

Georgia Tech Research Institute, GTRI/SEV, Georgia Institute of Technology, Atlanta, GA 30332-0840, USA

1. INTRODUCTION

This chapter documents the final results of a systems analysis of the task environment under which depot maintenance technicians perform their jobs — specifically, the programmed depot maintenance (PDM) environment for F-15 aircraft. This effort yielded a detailed functional breakdown of PDM activities from which an understanding of the PDM process, specific major jobs within the process and the types of support necessary for successful job completion were identified. The US Air Force, with all of Department of Defense (DoD), is under continuing pressure to reduce operational costs. One area offering the potential for significant cost reduction pay-offs is the preparation and presentation of technical and management data for aircraft maintenance. Previous assessments have demonstrated that the integrated presentation of technical and management data can reduce operational costs in flightline environments.

Recently, the Deployment and Sustainment Division of the Air Force Research Laboratory (AFRL/HES) directed a program of research focused on enhancing the task environment for aircraft maintenance, in particular PDM, technicians. One aspect of the program, referred to as Integrated Technical Information for the Air Logistics Centers (ITI-ALC), was to insert advanced technologies into the PDM task environment. Examples of such technologies include wearable computers, speech-to-text and text-to-speech interfaces, electronic technical orders, intelligent agents, web browser presentation systems, audio and video consultation capabilities, and a bar code capability to facilitate parts tracking. A second aspect of ITI-ALC was to demonstrate the insertion of advanced technologies into the PDM process.

One challenge for AFRL/HES was to identify those PDM functions for which the insertion of ITI-ALC technologies would be most beneficial, where in this context benefits have been defined in terms of a set of Business Process Improvements (BPI). The most pressing need was to introduce advanced technologies into those aspects of PDM functionality for which the greatest advantages would be realized. Identifying PDM functions in this manner would also allow AFRL/HES to identify the maintenance task scenarios most appropriate for ITI-ALC demonstrations.

The objective of this research was to demonstrate how a traditional human factors-based systems analysis enabled the identification of PDM functions and at the same time considered the end user, i.e. the maintenance technician. Research results reinforced the benefits of a traditional system analysis approach — one of the core analytical methodologies available to human factors specialists. Proponents of such analyses (e.g. Meister 1985, Chapanis 1996) have come from among leading researchers and practitioners within the field of human factors. This chapter shows how the results of a detailed systems analysis, when applied to the PDM task environment for F-15 aircraft, allowed the identification of the PDM functions most appropriate for the insertion of technologies anticipated within ITI-ALC. The systems analysis included a functional specification, detailed task breakdown and specification of the types of support required by the technician for successful task completion.

2. SCOPE

In this effort, we focused on the specific needs of maintenance technicians across several skill types and a range of PDM jobs. In focusing on technicians' needs, we identified areas of the PDM task environment that would most benefit from the insertion of ITI-ALC technologies. To identify the PDM functions most appropriate for ITI-ALC technologies, the following sources of information were used:

- Results of engineering analyses — in particular, Integrated DEFinition (IDEF) functional models developed as a means of documenting depot operations across all Air Force Air Logistics Centers (SRAC 1995a, b).
- A set of BPI derived from simulations of IDEF models.
- Analyses, insights and feedback provided by maintenance personnel from the Warner–Robins Air Logistics Center (WR-ALC) during on-site consultation sessions.
- A set of support requirements derived directly from the systems analysis effort, where the systems analysis included a specification of PDM functionality and a detailed breakdown of PDM jobs.

3. APPROACH

In analyzing data specific to the PDM task environment, the following four-step process was applied.

3.1. Specify PDM Functional Areas

- Assign each major PDM job to an appropriate functional area.
- Derive detailed breakdowns of critical path jobs, i.e. conduct a task analysis.
- Specify support requirements for critical path jobs, i.e. determine the types of support maintenance technicians will require to complete major jobs successfully (e.g. equipment, specialized knowledge, procedures, documentation, data).

Results obtained during steps 3 and 4 were validated against data obtained during on-site consultation sessions held at WR-ALC. These sessions provided an opportunity for questions and answers on the analysis results and allowed the elimination of any remaining analysis "holes."

The identification of critical path jobs was the result of an engineering effort completed at WR-ALC under the Multi-stage Improvement Program (MSIP). MSIP identified 52 major jobs within the PDM process. Of these, 15 jobs were defined as critical path jobs, regarded as so because they were known to create process bottlenecks if not completed according to schedule requirements. Ultimately, noncompliance with the job completion schedule will delay an aircraft's delivery to functional testing. The decision to focus on critical path jobs was based on our interest in analyzing the most important PDM activities.

Once the support requirements for each critical path job were specified in step 4 (where all major jobs were assigned to a

functional area), they were consolidated across functional area. In this manner, each function, and the jobs assigned to that function, were identified with a set of support requirements. Once consolidated, the support requirements were considered with respect to 10 BPI. By considering the relationships between support requirements and the BPI set, we assessed the extent to which advanced technologies would enhance the successful execution of PDM functions and, thus, determine where in the PDM process the greatest benefit would be realized.

4. RESULTS: ANALYSIS OF THE PDM PROCESS

To gain a more detailed understanding of the PDM process, as well as the types of jobs that contribute to process delays, we conducted a detailed analysis of PDM jobs — specifically, the 15 critical path jobs.

4.1. Functions and Critical Path Jobs

Seven areas of PDM functionality were specified: REMOVALS, MODS, BUILDUPS, INSTALLATIONS, CHECKS, INSPECTIONS and FACILITATIONS. Within each function, major jobs (including critical path jobs) were performed by five technician skill types: AC (aircraft), AE (electrician), AH (hydraulics), AN (weapons), AS (sheet metal). These skill types, specified in written documentation of the PDM process (WR-ALC 1997), were also identified during on-site data collection sessions held with WR-ALC personnel.

The process flow chart of figure 1 reflects the general sequence in which PDM functions are executed. The feedback loop has been included to indicate that various checks and inspections are included as tasks during BUILDUPS and INSTALLATIONS jobs. They might be considered as "mini" checks or inspections. In other words, they are distinct from the major PDM jobs assigned to the CHECKS and INSPECTIONS functions.

4.2. Support Requirements

Once each critical path job was decomposed into specific tasks and subtasks, a set of support requirements was specified. That is, each job was defined in terms of the types of support that would promote successful performance of job tasks by respective maintenance technicians. These support requirements facilitated job performance and were consolidated within each PDM function (i.e. across the major jobs assigned to each function). In defining support types, *generic* and *specialized* types of support were specified. Generic support referred to information requirements applicable to all jobs within a given functional area, while specialized support referred to requirements specific to a particular critical path job.

4.3. Targeting PDM Functions for ITI–ALC Technologies

Among the sources of information considered in identifying those functions to receive the greatest benefit from ITI-ALC technologies were each function's support requirements and the BPI that could be addressed by these support requirements. The 10 BPI identified below were considered. Specifically, each support requirement was evaluated in terms of the BPI to which it could contribute, and the support requirements for each functional area were mapped to BPI.

- B1. *Planning Process Enhancement*: application of previously developed plans, lessons learned, and fully defined work operations.
- B2. *Acquire Parts*: providing technicians with a kit containing all parts required for a specific maintenance repair.
- B3. *Electronic Signatures*: use of electronic signatures for "signing-off" that a task or inspection has been completed.
- B4. *User Technical Information Presentation System*: providing access to accurate and current technical information.
- B5. *Integrated Technical and Diagnostics Information*: facilitating automatic receipt of information required for the performance of maintenance activities — at required levels of detail.
- B6. *Visibility into Part Availability*: providing ready and reliable access to data on part availability.
- B7. *Multi-skilled Technicians*: certification in a broader range of skills.
- B8. *Data Sharing Among all Levels of Maintenance*: ready availability of accurate organizational level information to depot personnel.
- B9. *Performance Metrics Based on Actual Data*: maintaining data representative of actual task and process performance.
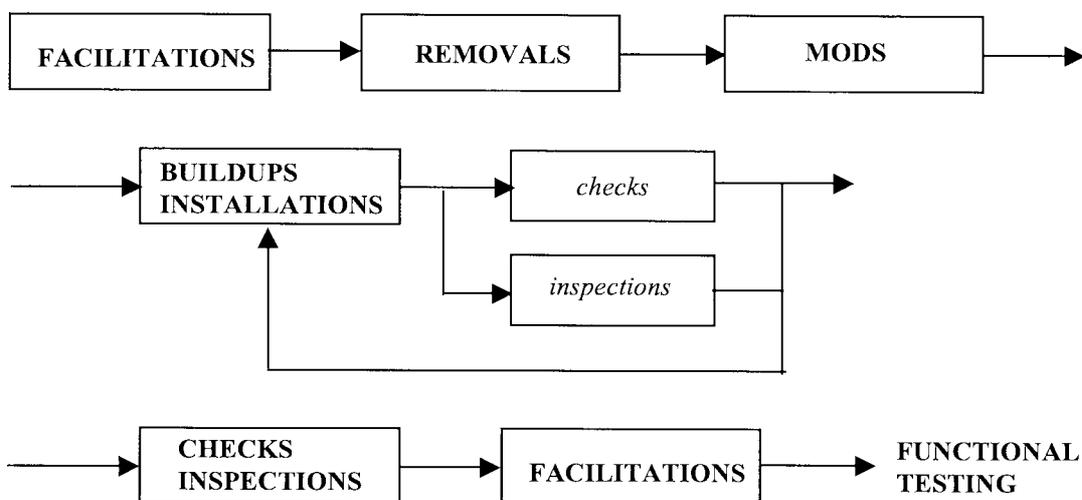
Figure 1. PDM process.

- B10. *Preplanned Over and Above/Unpredictables*: electronic cataloguing of planning packages for unpredictables and over and above discrepancies.

Table 1 provides an example of the support requirement-to-BPI mapping for the function REMOVALS. As one means of identifying the PDM functions for which advanced technology insertion would be most beneficial, one might examine the frequency with which BPI are addressed within a given function. That is, by analyzing the type of information contained in table 1, one can measure the extent to which the BPI set is addressed by the support requirements defined for each function. In effect, table 1 represents a *support requirement _ BPI* matrix M with elements m_{ij}. Furthermore, each *X* entry indicates that support requirement *i* addresses BPI *j*. By considering the data in table 1 (i.e. the number and location of matrix entries), one can determine — for each function — the range of BPI coverage by that function's respective support requirements. In other words, one can assess the frequencies with which BPI are addressed by the set of support requirements, as well as the proportion of matrix cells containing an entry. This proportion is referred to as the extent to which support requirements contribute to the BPI set. Note that the extent to which support requirements contribute to BPI is provided in the summary information of table 1.

5. IMPLICATIONS

The primary objective was to identify PDM functions that would receive the greatest benefit from ITI-ALC technologies. This type of identification required a ranking of the PDM functions according to a given set of criteria. In this effort, a single criterion was proposed: the extent to which each function's respective set of support requirements addresses the BPI set. One means of measuring this extent of coverage is to calculate the proportion of matrix cells that reflect a support requirement-to-BPI mapping.

As this proportion increases (i.e. as the density of the matrix representing a given function increases), greater benefit will be obtained through the insertion of ITI-ALC technologies into that particular function. In other words, satisfying the function's support requirements through the application of ITI-ALC technologies will effect greater improvement to the PDM process. Table 2 ranks PDM functions according to the extent-of-coverage measure. According to this measure, the REMOVALS function receives the greatest benefit, followed by BUILDUPS. The proportions offered in table 2 indicate that CHECKS, INSTALLATIONS, FACILITATIONS and INSPECTIONS benefit to a lesser extent.

One means of readily identifying the BPI addressed most (and least) often by the support requirements defined for a given PDM function is to examine the frequency with which each BPI is addressed by the support requirements defined for that function. By examining such a frequency distribution, one can identify the BPI most likely to be realized if the corresponding support requirements for the PDM function of interest are satisfied. Frequency based data are readily available from the support requirement-to-BPI mappings completed with respect to a given PDM function (table 1). Table 1, for example, indicates that for REMOVALS, the BPI addressed most frequently by the function's respective support types are *Visibility into Part Availability* (B6) and *Data Sharing Among all Levels of Maintenance* (B8).

The results of this work reinforce the value of human factors-based systems analyses — demonstrating how a detailed systems analysis enhanced our ability to identify those PDM functions for which the insertion of ITI-ALC technologies would be most beneficial. Our results also provided a means of targeting maintenance task scenarios most appropriate for ITI-ALC demonstrations. Specifically, the ranking of PDM functions (table 2) was driven directly by the mapping of support requirements (an outcome of the systems analysis) to a set of process improvements. Based on results of this work, AFRL/HES personnel were better informed about PDM functionality, the benefits of applying advanced technologies to a given PDM function, and the types of task scenarios in which to demonstrate the use of such technologies.

Table 1. Support Requirement-to-BPI Mapping: REMOVALS.

REMOVALS: AE Removals, AN Removals, AC Remove Fuel Tanks

	B1	B2	B3	B4	B5	B6	B7	B8	B9	B10
Generic										
ready access to proper tools/equipment for performing removals and disconnects		X				X		X		X
proper storage arrangements for removed parts	X	X				X		X		X
consistent inventory procedures for removed parts	X	X				X		X	X	X
knowledge of (or access to) inventory procedures	X			X	X	X		X	X	X
knowledge of (or access to) removal and disconnect procedures				X	X					
availability of required or specialized cleaning equipment		X				X		X		
Specialized: AC Remove Fuel Tanks										
proper storage arrangements for# 1 tee assembly parts	X	X				X		X		X
knowledge of/access to rules/procedures for performing manual corrosion inspection				X	X				X	
knowledge of/access to process to improving bottom backing board				X	X				X	

SUMMARY:
number of support types: 9
matrix dimensions: 9 x 10
number of matrix entries: 38
proportion of matrix covered: 38/90 = 0.42

Table 2. Ranking of PDM Functions.

	Function	Extent of BPI Coverage
1.	REMOVALS	0.42
2.	BUILDUPS	0.33
3.	CHECKS	0.31
4.	INSTALLATIONS	0.30
	FACILITATIONS	0.30
	INSPECTIONS	0.30

ACKNOWLEDGEMENTS

REFERENCES

CHAPANIS, A., 1996, *Human Factors in Systems Engineering* (New York: Wiley).

MEISTER, D., 1985, *Behavioral Analysis and Measurement Methods* (New York: Wiley).

SYSTEMS RESEARCH AND APPLICATIONS CORPORATION, 1995a, *Integrated Technical Information for the Air Logistics Centers (ITI-ALC): Architecture Report.* Contract no. F41624-94-C-5021 technical report (Systems Research and Applications Corp.).

SYSTEMS RESEARCH AND APPLICATIONS CORPORATION, 1995b, *Integrated Technical Information for the Air Logistics Centers (ITI-ALC): System/Segment Specification for the ITI-ALC System.* Contract no. F41624-94-C-5021 technical report (Systems Research and Applications Corp.).

WARNER–ROBINS AIR LOGISTICS CENTER, 1997, *Planned Operations for T/N 860185 — Descriptions of Major PDM Jobs: Skill Type, Ops Number, Standard Hours, Ops Description* (Warner–Robins Air Logistics Center).

User-centered Design: Needs Analysis

N. Marmaras

National Technical University of Athens, Department of Mechanical Engineering, GR-15780 Zografou, Greece

1. INTRODUCTION

A user-centered approach for the design of information technology (IT) systems puts the future users of the system at the center of the design process. The aim is to create computer systems that support the user within a working environment (WE) rather than the user supporting the system. The typical phases of the user-centred design are (1) user needs analysis or user requirements analysis, (2) system specification, (3) system design and (4) system evaluation. In what follows, the first phase of the design process, the user needs analysis, is dealt with.

User needs analysis can be carried out in the context of two different design paradigms: the technology-driven design paradigm and the problem-driven design paradigm.

The design of systems under the first paradigm is guided more by already existing IT solutions, as well as theoretical models and techniques of the domains of application, than by the problems and difficulties met by human problem solvers or operators. The main scope of these systems is either the automation of certain subtasks, or to offer some form of problem solution by stand-alone machine experts. In this case, the designers carry out a rather quick and superficial analysis of the domain of application and the future users' tasks (what they formally have to do), identify the tasks or subtasks that could be supported or automated by already existing IT solutions, and then start the development of the system. For example, in the domain of telematics, the designer, after identifying the formal data to be used by the future users of the system and the flow of these data, may develop a system to capture, transmit and store these data. Questions such as which is the most appropriate form or sequence of presentation of the data, what is the relative importance of these data for the users and for the different situations they cope with, what data treatments are required, etc. are more often addressed by the designers, based on their knowledge of the domain, their experience from previous applications and considering the constraints of the IT chosen to be used.

The technology-driven design paradigm has been proved efficient for the design of IT systems addressing simple, well-defined and repetitive tasks. In contrast, many systems developed under this paradigm and addressing complex cognitive tasks such as decision making, design and control, have had to be abandoned or used in a suboptimal way, not because they did not work, but because the users could or would not use them. More specifically, these systems:

- often do not provide the information that is really needed, or produce information in a form that is undesirable or incomprehensible to the user. On the other hand, these systems often provide information that is not required;
- can appear confusing to the users;
- sometimes do not provide the functions the user requires, and more often provide functions that the user does not need;
- force the users to perform tasks in undesirable ways; and

- can cause unacceptable changes in the structure and established practices of organizations, creating dissatisfaction and conflict.

Furthermore, the development of IT systems following the technology-driven paradigm increases the required time, effort and iterations at the systems' evaluation phase.

In the problem-driven design paradigm, the task requirements and constraints, as well as the way domain experts cope with them, would drive the choice of IT to be used and the design of the systems. The main scope of IT systems developed under this design paradigm is:

- to address the cognitive difficulties met by potential users in performing cognitive tasks;
- to improve their potential users' performance; and
- to achieve compatibility with potential users' competencies and work environment. This compatibility permits the user to remain in full control of the support provided by the system and ensure high usability, i.e. effectiveness, ease of learning and ease of use.

The design of systems under the problem-driven paradigm requires a much deeper and more extensive user needs analysis. The aim of the analysis being to identify (1) the cognitive difficulties potential users meet in performing their tasks, (2) the factors determining their performance and (3) possible ways to alleviate task difficulties and improve users' performance. This design paradigm has been proved particularly appropriate for the design of IT systems supporting complex cognitive tasks such as process and traffic control, medical diagnosis, design, managerial and production planning, computer programming, etc. (Pavard *et al.* 1990, Laios *et al.* 1992, Marmaras *et al.* 1992, 1997, Pougès *et al.* 1994, Nathanael *et al.* 1997).

Below is presented a methodological framework for user needs analysis in the context of user-centred design and under the problem-driven design paradigm.

2. CONCEPTUAL MODEL OF COMPLEX COGNITIVE TASKS

Complex cognitive tasks share some common characteristics both at the cognitive level and the environments within which they are performed. At the cognitive level, these tasks require different types of problem solving such as decision making, diagnosis, planning, and complex cognitive activities such as anticipation, monitoring, mental calculations, etc.

As far as the environments within which and for which these tasks are performed, they may have one or more of the following characteristics:

- they consist of many interrelated and interacting dynamic components and factors assuming different values each time;
- events can occur at undetermined times and the nature of the problem to be solved can change;
- there is uncertainty regarding the time at which one or more events occur, and the severity of changes they cause to the work system;
- there are multiple quantitative and qualitative objectives to be achieved, often conflicting, and with no predetermined hierarchy;
- they impose severe time constraints to the human operators;
- they are risky; and
- human errors may have serious consequences.

It could be concluded, therefore, that the difficulty of complex cognitive tasks is due both to the complexity of cognitive processes required for their completion, and to the complexity of the environments within which and for which they are performed.

To cope with the cognitive constraints and demands of their WE, domain experts develop and use specific cognitive skills or *competencies*. Elements of these competencies are:

- selective use of formal and informal sources of information;
- development and use of specific mental representations of the WE, or operative images;
- use of problem-solving strategies and heuristics such as problem decomposition, bounded rationality, satisfying and search costs, opportunistic thinking;
- ability to change appropriately between skill-, rule- and knowledge-based performance level, depending on the characteristics of the situation at hand; and
- use of components of the WE as memory supports.

3. FRAMEWORK FOR USER NEEDS ANALYSIS

Considering the above conceptual model, user needs analysis has to identify the main cognitive constraints and demands posed to the domain experts, as well as the main elements of the competencies they develop and use to cope with them. Given that the constraints and demands are determined by the WE's situations presented to the human agents, typical situations also need to be analyzed. Finally, the analysis of undesirable work outputs (e.g. systematic errors, suboptimal or erroneous decisions) will point out the situations where human experts' competencies fail. Figure 1 presents a framework for user needs analysis.

At a first phase the analyst, with the aid of domain experts, identifies the *typical work situations*. By this is meant both normal and characteristic exceptional situations. The way these typical situations are presented to the human agents as well as the cognitive demands, constraints and difficulties associated with them, are also investigated.
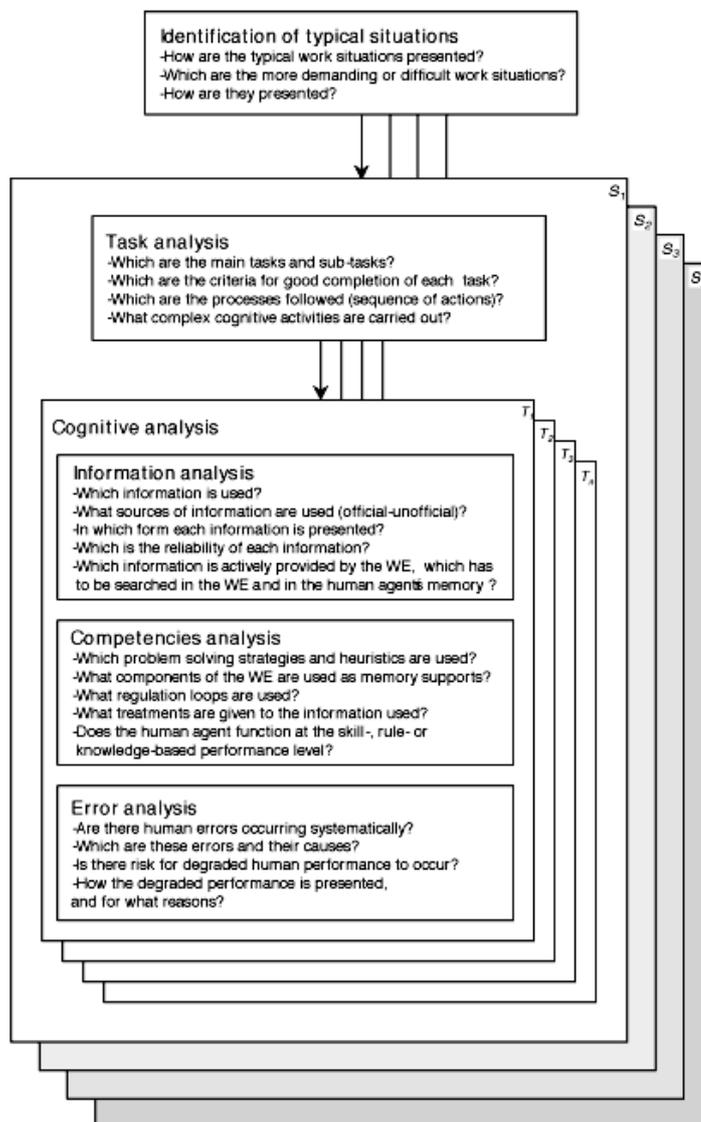
Figure 1. Framework for user needs analysis.

Task analysis is carried out at a second phase, for every work situation (S_1, S_2, S_3, … S_n) identified at the first phase. The main tasks and subtasks, the criteria for good completion of each task, the processes followed as well as the activities (both cognitive and physical) carried out for the accomplishment of each task, are identified and described.

The third phase deals with *cognitive analysis* of the tasks (T_1, T_2, T_3 … T_n) identified at the previous phase. Cognitive analysis comprises analyses at the levels of information, competencies and errors. *Information analysis* looks for the information used for the completion of each task, the sources providing the information (both formal and informal), the form each information is presented, its reliability and classifies them according to whether they are actively provided by the WE or whether they have to be found in it or the human agent's memory. *Competencies analysis* deals with the problem-solving strategies and heuristics used by the human agents, the components of the WE used as memory supports, the regulation loops eventually used, the mental treatments given to the information, as well as the performance level at which the human agents function, i.e. skill-, rule- or knowledge-based. Finally, *error analysis* looks for human errors occurring systematically and their causes, as well as for possible degraded human performance leading to suboptimal actions or decisions, and their causes.

The type of support that could be provided by an IT system can then be investigated, based on the results of the user needs analysis. This can be done by answering questions such as:

- What other information would be useful to the operators?
- Is there a more appropriate form to present the information already used as well as the additional new information?
- Is it possible to increase the reliability of information?
- Could the search for information be facilitated, and how?
- Could the treatment of information be facilitated, and how?
- Could we provide memory supports, and how?
- Could we facilitate the complex cognitive activities carried out, and how?
- Could we promote and facilitate the use of the most effective problem solving and decision making strategies, and how?
- Could we provide supports which would decrease mental workload and mitigate degraded performance, and how?
- Could we provide supports which would decrease human errors occurrence, and how?

Support of users' memory, computational aid, communication support, de-biasing, support to visualize or to make concrete the abstract and uninspectable, aid to the formation of a better global representation of the situation at hand, support to actions' planning, and support to evaluate alternative decisions are examples of possible functions of a system supporting complex cognitive tasks.

4. ACTIVITIES' ANALYSIS FOR USER NEEDS ANALYSIS

To carry out the user needs analysis described above, analysis of future users work activities is required. *Activities' analysis* deals with the responses human agents give to the tasks imposed on them or they wish to carry out, the tasks being goals to be attained within a set of conditions.

Activities' analysis is an iterative process, comprising the following stages:

1. Systematic observation and registration of observable activities, in relation to the implicated elements of the work environment. The observed activities may be body movements and postures, eye movements, verbal and gesture communications, etc.
2. Inference of cognitive activities, processes and strategies.
3. Formulation of hypotheses about competencies, by interpreting the cognitive activities and referring to the work environment's demands, constraints and supports, as well as to the limits of cognition.
4. Validation of hypotheses by repeating stages 1 and 2.

The iterations end-up when the developed model of human agents' competencies explains sufficiently and satisfactorily their behavior and performance.

Video- and tape-recording, paper and pencil or other specially designed displays (e.g. eye movement recorder) can be used for data gathering at stage 1. Consecutive or retrospective verbalizations supported by the data gathered at stage 1 (autoconfrontation) can be used at stage 2. Protocol analysis is the main method for stages 2 and 3.

Besides direct on-site observations, alternative scenarios and simulation methods may be used to analyze activities in rare or risky work situations. However, before using these methods, sufficient analysis of the work environment as well as selection and examination of historical data concerning the studied work are required. Comparative activities' analysis of equally experienced operators, who nevertheless demonstrate different performance (experts and super experts), is also useful in many cases.

Theories and models for human cognition and behavior, offered by cognitive psychology, artificial intelligence, ethnology, psycholinguistics, neuroscience, organizational psychology or philosophy, can be used for the inference of operators' cognitive activities and the formulation of hypotheses concerning their competencies. Examples of such theories and models are Newell and Simon's (1972) human problem-solving theory, Theureau's (1992) theory of action's course, Hutchins' (1990) and Hutchins and Klausen's (1992) theory of distributed cognition, Rasmussen's (1986) 'ladder' model of decision-making and Reason's (1990) model for human errors. However, these theories and models must have a rather hypothetical than a normative value for the analyst. They constitute his/her background knowledge and support the interpretation of observable activities and the inference of cognitive activities. Activities' analysis may confirm these models totally or partially, or it may indicate their limits or reject them.

5. EPILOGUE

The proposed user needs analysis may be difficult and time-consuming to carry out, requiring the use of specific work analysis methods and techniques, as well good cooperation between the analyst and the IT specialists who develop the system. These weaknesses are, however, counterbalanced by the fact that it ensures the design of systems which (1) improve their future users performance by decreasing the cognitive constraints and alleviating the cognitive demands imposed to them, (2) eliminate or mitigate error-prone and degraded performance situations, (3) facilitate the use of effective human experts competencies and eventually enrich them and (4) achieve compatibility with these competencies and the work environment. This compatibility will ensure the usability of the future system.

Furthermore, this way of analyzing user needs minimizes the risk to ignore the more demanding and therefore difficult and error-prone aspects of the target work, as well as providing sufficient data permitting to decrease the required time and iterations at the system's evaluation phase.

Finally, the outcomes of the proposed user needs analysis provide also sufficient data to specify the main features of the human–computer interface, ensuring compatibility with users' competencies. The way task's objects have to be represented by the system, the type of man–machine dialogues to be used, the procedures to be proposed and generic or customizable elements of the system are examples of human-computer interface features that can be specified, using the acquired data.

REFERENCES

HUTCHINS, E., 1990, The technology of team navigation. In Galegher, J., Kraut, R. E. and Edigo, C. (Eds) *Intellectual Teamwork* (Hillsdale: Lawrence Erlbaum Associates).

HUTCHINS, E. and KLAUSEN, T., 1992, Distributed cognition in an airline cockpit. In Middleton, D. and Engeström, Y. (Eds) *Communication and Cognition at Work* (Cambridge: Cambridge University Press).

LAIOS, L., MARMARAS, N. and GIANNAKOUROU, M., 1992, Developing intelligent decision support systems through user-centered design. In Galer, M., Harker, S. and Ziengler, J. (Eds) *Methods and Tools in User-centered Design for Information Technology* (Amsterdam: Elsevier), pp. 373–412.

MARMARAS, N., LIOUKAS, S. and LAIOS, L., 1992, Identifying compentences for the design of systems supporting complex decision making tasks: a managerial planning application. *Ergonomics*, 35, 1221–41.

MARMARAS, N., VASSILAKOPOULOU, P. and SALVENDY, G., 1997, Developing a cognitive aid for CNC-lathe programming through problem-driven approach. *International Journal of Cognitive Ergonomics*, 1, 267-89.

NATHANAEL, D., MARMARAS, N. and MATSAKIS, I., 1997, Ergonomic work analysis for the development of a telematics application for medical assistance. In Seppala, P. *et al.* (Eds) *From Experience to Innovation — IEA '97, Proceedings of the 13th Triennial Congress of IEA* (Helsinki: Finnish Institute of Occupational Health), 7, 489-91.

NEWELL, A. and SIMON, H., 1972, *Human Problem Solving* (Englewood Cliffs: Prentice Hall).

PAVARD, B., BENCHEKROUN, H. and SALAMBIER, P., 1990, La régulation collective des communications dans un center d'appel d'urgence: analyze et modélisation. In *Actes du Congrès Ergo IA '90*, Biarritz.

POUGES, C., JACQUIAU, G., PAVARD, B. GOURBAULT, F. and CHAMPION, M., 1994, Conception de collecticiels pour l'aid à la prise de décision en situation d'urgence: la nécessité d'une approche pluridisciplinaire et intégrée. In Pavard, B. (Ed.) *Systèmes coopératifs: de la modélisation à la conception* (Toulouse: Octarès), pp. 21–50.

RASMUSSEN, J., 1986, *Information Processing and Human-Machine Interaction: an approach to Cognitive Engineering* (Amsterdam: North Holland).

REASON, J., 1990, *Human Error* (Cambridge: Cambridge University Press).

THEUREAU, J., 1992, *Le cours d' action: analyze sémio-logique* (Bern: Peter Lang).

Work Design: Barriers Facing the Integration of Ergonomics into System Design

L.J. Steiner and C. Vaught

National Institute for Occupational Safety and Health (NIOSH), Pittsburgh
Research Laboratory, Mining Injury Prevention, PO Box 18070, 626
Cochrans Mill Road, Pittsburgh, PA 15236-0070, USA

1. INTRODUCTION

Many industrial injuries and illnesses can be attributed to poor design features. Mishaps are often the unintended consequences of either using or misusing designed products. At times these mishaps result in fatalities that are attributed to "human error." The question is error on whose part, the victim's or the designer's? The designer's job is to forecast potential uses and misuses of products through design. The integration of ergonomics principles at the product or process development phase helps to minimize unintended consequences. This attention at the design phase reduces and even prevents injuries and illness among end users. If ergonomics is to be integrated into the design process, however, organizational principles must be integrated into the design factors. This process is difficult and the barriers are multiple.

2. RELEVANCE

The main goals of any design effort are that the product works as it is intended to, trains easily, and performs safely when used (Broberg 1997). A product needs not be only beneficial but must possess no harmful side effects. Sometimes even these goals are not met. The automobile air bag is a good example. The goal was to protect automobile occupants during sudden impact. *Product requirements* documents spell out how the product will function. *Product specifications* stipulate the product's construction and performance. From an engineering standpoint, when the product meets these two sets of criteria it is considered a success. The air bag does function exactly as the design parameters said it would except that it sometimes kills small women and children. Any repercussion will not come back on the designer, but rather stops at his/her project manager. The design engineer is isolated from feedback until information for new specifications is received. The organization writes off its liability until the public or government becomes involved. This process, including the assignment of work roles and administrative procedures that shape the organization's social culture and conduct of work activities, is part the overall organizational structure.

Organizational structure and the thinking that reflects it play a major role in shaping the product development process. Broberg (1997) described the multistage and interconnected manner in which a concept becomes a product. Product requirements and specifications guide the development mechanism. These two basic design inputs were described earlier. What is missing from the process is a third document that sets forth *use parameters*. Dekker and van den Bergen (1996: 301) suggested that "integrity, reliability, efficiency, and usability" must all be characteristic of an interface that protects humans and optimizes their performance. Yet, the product–human interface does not show up in company standards governing the design process. Therefore, designers may continue working within their normal organizational framework and face little pressure to develop products with ergonomic input. What they do experience is social encouragement to repeat design approaches that have worked in the past (Broberg 1997).

The purpose of this chapter is to provide insight into barriers facing the integration of ergonomics into design. It is also intended to create an awareness of consequences that can result from ignoring ergonomics during the design process. Mistakes made by users are often invited by the way the process or product was designed (Chapanis 1996). Working through these implications and understanding how design engineers function may help determine how to expand the design approach to include ergonomics criteria.

3. ENGINEERING APPROACH TO THE PROBLEM

Typically, a design engineer is presented with a deadline to build a product that will do something. S/he conceptualizes it, and gathers whatever resources are necessary to achieve this goal. The designer will focus mostly on whether the product does what it is supposed to do. The concern is to make a product that is used according to the ideal manner in which it will be demonstrated. A designer's success or failure is measured by whether the product does what it was supposed to do and whether it was delivered in a timely manner. Such design approaches overlook the fact that ultimately, people use products. A typical engineering process does not consider this until the end of development, during usability testing. The product's team develops this testing with limited knowledge of the human element and in a controlled environment. They possess the same knowledge they had when designing the product, and, because of their familiarity with it, may be shortsighted as to how the product should be tested. Under such conditions, human interface problems will be hidden rather than highlighted.

4. BARRIERS TO INTEGRATION

The organizational structure is not supportive of ergonomics in comparison to other design parameters, and the lost benefits of such support are not recognized. Often designers are working under severe time constraints, and project managers assume that if the product functions as it was planned to function, ergonomics will work itself out. At best, the design can only be "tweaked" after completion. Design engineers may be lacking ergonomics knowledge or training, and simply not know the right questions to ask. Unless the organization's expectations reflect a necessity for ergonomics, their thinking will remain the same. Designers will continue to use the same tried and true methods for development. When there are product failures or customer complaints, design engineers are the last to know. When a developed product is determined to work but there are still usability problems, these issues are often blamed on the end user's inability to use the product properly. Products are almost never evaluated for their intended and unintended consequences. Sound ergonomics methods can be used to predict uses and misuses.

Unintended consequences are not brought back to the designer unless a re-design is demanded. The problem created is that the designer never gets a chance to build a database of

ergonomic knowledge or cause for failure. Therefore, s/he believes the design process works well and will use this belief in future designs. This practice encourages a "tunnel vision effect" because the designer is using a one-dimensional approach based on his/her previously tried and true algorithms. This thinking supports "blind spots" by not allowing peripheral information to enter. Blind spots keep the engineer from asking the right questions and allows him/her to ignore "extraneous" information. Most organizational structures tend to encourage this type of thinking. In the long run, though, the earlier human engineering is introduced into the design process, the better the product will be. Redesign for usability is more costly after production.

5. CASE ANALYSIS

Some of the ideas mentioned above can be illustrated by a program, begun in the early 1970s, to develop an oxygen breathing apparatus for US coal miners. There were some scientific and engineering principles that could not be violated during development and the process seemed to rely on past successes. Because there was little or no ergonomics involvement, it cannot be known how much impact ergonomics might have had on up front engineering design. However, it could reasonably be deduced that some user problems could have been avoided.

> Barrier: Because people are considered to be adaptable, there is no demand for use parameters. Users are therefore forced to adapt to designs that "make sense." A failure to understand that ergonomics is not just "commonsense" leaves the interface problem unrecognized.

Since its creation in 1910, the US Bureau of Mines was active in setting requirements and performing certification tests on oxygen breathing apparatus. These were intended to help workers survive irrespirable mine atmospheres. The weight and size of the devices proved to be too much for the typical mine worker and the daily apparatus came to be used only during rescue attempts. A succession of deaths among would-be rescuers underscored the fact that miners had little or no training in apparatus usability. The loss of lives led to the adoption of mine rescue standards in 1921. These standards provided a baseline for instruction. The schedule for initial training, which required 20 h over 5 days, included 9 h of practice wearing the device. The Bureau recommended additional instruction and practice every 6 months (Parker *et al.* 1934). Over time, mine rescue apparatus became more refined and rescue team members better trained in its use.

> Barrier: Even when a problem with the machine/human interface is recognized, the tendency is to believe it can be "trained out." Training then becomes the stopgap for product inadequacies.

In 1924, to encourage miners to wear the breathing device continually, the Bureau approved a different type of apparatus. This one measured 3 x 4 x 6 inches and weighed just 2.5 lb. Since then, the basic design has remained unchanged. Although the "filter self-rescuer" or "FSR" can easily be carried on the miner's belt, it does not provide oxygen. It converts carbon monoxide (CO) to carbon dioxide (CO_2) and provides some protection from hydrogen chloride gas. The FSR is of no use in oxygen deficient or highly toxic mine atmospheres.

> Barrier: The perceived solution to a complicated user problem, from an organizational standpoint, may be to make tradeoffs.

> Many times, the tradeoff criterion is arrived at non-inclusive of all critical parameters.

With passage of the 1969 Coal Mine Health and Safety Act, money became available to develop a self-rescue apparatus that supplies oxygen. During the requirements phase, officials established that the unit must be small, durable and weigh < 4.5 lb. Finally, the device had to provide complete respiratory protection. Performance specifications for the proposed apparatus were contained in Schedule E of the Federal Register. Schedule E set stringent metabolic requirements for oxygen delivery. In sum, the Bureau was aiming at an acceptable tradeoff that would supply oxygen like the mine rescue apparatus (for at least 1 h) but be worn like the filter self-rescuer. Lockheed Independent Development built a KO_2 (potassium super oxide) prototype under contract. According to Lockheed's Senior Research Engineer, human factors considerations heavily influenced the development of this device. "When an emergency situation arises, the unit can be activated quickly, even by an untrained and/or excited person. Once deployed, the breather is comfortable to wear, does not interfere with the user's activity, and gives the user confidence that it is working properly" (Perry and Wagner 1973). The Lockheed rescuer caught fire during a demonstration, leading to further research and development, as well as a need to defend similar apparatus. Another KO_2 unit widely used by Russians supposedly was used underground without any known problems. Also, the Chemox unit had been used for 30 years without technical problems in the unit itself. However, it was mentioned that there were problems with training and usage of it (Stein 1978).

> Barrier: When there are no carefully controlled studies, one is likely to see what one wants to see. Outside studies and suggestions helpful to building more sound designs are not welcomed by the organization research team leading to repeated designs that are ineffectual.

Despite the Lockheed engineer's endorsement of human factors, there is little evidence that the human/product interface was explored much beyond metabolic concerns. Lockheed never went to production with their rescuer but sold the technology to another manufacturer. In 1981, a federal law mandated that mines would provide miners with a one-hour oxygen delivery device. By the time of the Wilberg mine disaster in 1984, there were two types of oxygen delivery apparatus (chemical and compressed oxygen) and five different brands in US coalmines. All came to be commonly called "self-contained self-rescuers" or "SCSR." SCSR were readily available to the 27 miners who died at Wilberg, but investigations suggested they had not been able to use them. Since the suspect apparatus functioned properly when tested after the fire, inadequate training emerged as a determinant in the workers' deaths. Only later did the general question of usage emerge.

> Barrier: Time pressure and regulatory or other constraints might lead to solutions that seem neat and expedient, but are inappropriate and ineffective.

In 1986, researchers at the University of Kentucky (UoK) began a project in cooperation with the Bureau, several other agencies, and original equipment manufacturers. This study started with a task analysis and soon turned into a full-scale scientific investigation of worker proficiency with SCSR. In the study it

became clear that none of the apparatus was user-friendly. These devices were different in design, but all shared some traits with the earliest mine rescue apparatus that the Bureau had felt compelled to set stringent training standards for. Oxygen was activated manually, by either pulling a cord or turning a knob, each device had a mouthpiece with bite lugs and a flange that fit between lips and teeth, a nose clamp was affixed to the mouthpiece by means of a cord, and goggles were stowed inside the case. The notion that a unit "can be activated quickly, even by an untrained and/or excited person" was put to the test at a cooperating coal company. A group of surface workers (who are not required to carry SCSR) was brought into a training room where the device's function was explained in detail. Each person was then asked to don the apparatus while being videotaped. Only the company pilot was able to figure out the SCSR well enough to isolate his lungs, though he did not get the wearing straps tied nor goggles in place. No one else in the group of 14 was able to activate the device. Researchers concluded there are few built-in cues to help people activate SCSR, nor did it appear that the interface was characterized particularly well by the "integrity, reliability, efficiency, and usability" Dekker and van den Bergen (1996: 301) considered essential for an ergonomically correct product.

> Barrier: Claims are often not put to real-world tests. Without end user testing in the design phase, one's knowledge base limits his/her perception. The tendency is to see only confirmatory evidence.

The UoK study team noted that SCSR were the result of engineering research and development that had taken years. Nowhere, however, could they find evidence that ergonomic or cognitive studies had been conducted to determine what human design features the devices should have (Vaught *et al.* 1996). After the apparatus had tested out as reliable in field trials, been shown to function well mechanically, and to have a fairly long shelf life, they were placed in the mines. From that point workers must accommodate the SCSR in order to use them effectively. Training would have to be a large part of that accommodation, but the training research focused narrowly upon donning proficiency rather than on overall usability.

> Barrier: If a product is working according to specifications, any subsequent failures are likely to be blamed on the human user.

Lack of capability is clear in an account given to the UoK researchers by a safety instructor who had 96 trained workers don their SCSR and travel through theatrical smoke at a mine. His observations: (1) workers had to be helped to don their devices; (2) about one-quarter of the miners forgot to put on their nose clamps; (3) roughly a dozen individuals omitted their goggles; (4) nine people did not get their oxygen turned on; (5) several workers became entangled in the wearing straps; and (6) at least two miners failed to insert their mouthpieces correctly. Focusing upon this problem the study team, over the next 5 years, developed a rationalized donning sequence that could be used with all SCSR then on the market. Next, the researchers designed an evaluation method that would allow them to observe donning attempts and chart each miner's performance. Then, they constructed low cost training apparatus and tested them against the more expensive models offered by equipment manufacturers.

Finally, the study team employed their innovations in the field, training miners to proficiency and plotting their forgetting curves. One group of coal miners was trained to proficiency (able to perform five perfect sequences in a row) so that samples could be evaluated through the following year. Of those evaluated after 90 days, only 30% had retained enough knowledge to get their lungs isolated and the SCSR secured. It was apparent that there was little or nothing intuitive about the design of SCSR that will prompt a trained person's memory. Current guidelines are such that these miners had 9 months to go before they would be retrained.

> Barrier: A patchwork pattern of training and usability instructions developed when problems are encountered is not a substitute for anticipation of design problems prior to implementation.

Although they were committed to finding training solutions for the use of an apparatus that requires practice to master, under a standard that mandates little practice, researchers did encounter other usability issues. In interviews with 46 workers who used SCSR while escaping underground mine fires, the study team found 29 who reported having had breathing difficulties. Twenty-seven miners, though in smoke, removed their mouthpieces to breathe. As one individual stated, "I know there's a lot of CO … but I can't breathe. If I can't breathe in this thing [SCSR], I'm just going to collapse anyway" (Brnich *et al.* 1992). Potential breathing problems became greater when a second generation SCSR was introduced in the early 1990s. During the requirements phase it was decided that if these devices were made only twice the size and weight of a filter self-rescuer, they could be worn rather than carried or stored as the larger units had been. These size and weight specifications were met partly by crowding the metabolic limits for oxygen delivery. Where the first SCSR provided excess oxygen under certain conditions, the person-wearable SCSR (PWSCSR) measured it out more carefully. It can be imagined that if a miner decided s/he was not getting enough oxygen from a first generation apparatus, the problem would be magnified by a device designed to deliver a lesser quantity per minute.

> Barrier: A product may meet physiological requirements, yet fail to meet psychological ones. This must be recognized and attended to, or the device may not be successful.

A second usability problem related to the size and weight issue was wearability. It had been assumed by those establishing requirements that a device only twice the size and weight of a filter self-rescuer could replace the FSR on a miner's belt without undue complications. This reasoning might seem valid unless one applies the concept to more familiar objects replacing one's glasses with a pair twice the size and weight; or, exchanging the car in one's driveway for one twice the size and weight. In both instances, it might be expected there would be problems with mobility and physical impairment, among others. Such seems to have been the case at a mine where accident data collected over 10 years were reviewed by company personnel. They reported a dramatic increase in accidents due to fatigue and balance after introduction of the person-wearable units. In an investigation conducted by researchers at this mine, a sample of 51 workers was asked their opinions about wearing the PWSCSR. The miners' answers to a set of objective statements about the PWSCSR and FSR suggest the following: (1) they felt pain and soreness after

wearing the PWSCSR all shift; (2) they did not feel pain and soreness after wearing the FSR all shift; (3) they did not think wearing either device posed a danger of injury; (4) the weight of their belts, however, has at some time caused a fall due to loss of balance; and (5) given a choice, they would switch back to wearing the FSR even though it offers less protection.

> Barrier: Sometimes design tradeoffs are made arbitrarily without controlled studies based on proven principles. This may lead to unintended consequences, such as having workers who are adequately protected but uncomfortable or visa versa.

6. CONCLUSIONS

In general, the ignoring of human factors in the design phase constitutes an organizational blind spot. This blind spot has usually been recognized after the fact in partially failed attempts to "reinvent" organizations. A recent example is the result of re-engineering efforts here and abroad: "the basic assumption is that if you get the engineering right the human factor will fall into place. … As a result, the re-engineering movement has encountered exactly the same problems and failures experienced by older style classical management principles" (Morgan 1997: 22). The integration of ergonomics into system and product design is not accomplished by just putting an ergonomist on the project but requires a change in the cultural thinking of the designers and the organization. The *use parameters* document needs to be considered as important as the *product specification* and *product requirements* documents. An organization's social framework must be structured to accept ergonomics as part of the design criteria. That would increase the amount of interest design engineers need to have in ergonomics. Further, the design team should be educated about how to recognize and apply ergonomic principles during all development stages. This would help reduce the barriers to ergonomics integration. Methods used to assess ergonomic design criteria can be found in Chapanis (1996) and Ainsworth (1992). Chapanis (1996: 294) addressed

the difficulties and rewards of ergonomics integration, but clarified that "even the best efforts of human factors professionals and system designers cannot guarantee that a system will be successful because of the many factors over which they have no control." The way to move toward control over the system is to recognize that barriers do exist and the organizational structure may be part of the problem. Then, begin a culture change by using sound ergonomics and design principles to predict product usability and safety.

REFERENCES

BRNICH, M.J., VAUGHT, C. and MALLETT, L., 1992, SCSR proficiency requires hands-on practice. *Coal*, July, 52–4.

BROBERG, O., 1997, Integrating ergonomics into the product development process. *International Journal of Industrial Ergonomics*, 19, 317–27.

CHAPANIS, A., 1996, *Human Factors in Systems Engineering* (New York: Wiley).

DEKKER, E.F. and VAN DEN BERGEN, E.A., 1996, Human Factors in E&P facility design, a participatory approach. In *Proceedings of the International Conference on Health, Safety & Environment*, Society of Petroleum Engineers (SPE), #35792, New Orleans, LA, 297–302.

KIRWAN, B, and AINSWORTH, L.K., 1992, *A Guide to Task Analysis* (London: Taylor & Francis).

MORGAN, G., 1997, *Images of Organization* (Thousand Oaks: Sage).

PARKER, D.J., McCAA, G.S. and DENNY, E.H., 1934, *Self-contained Oxygen Breathing Apparatus: A Handbook for Miners* (Washington, DC: US Department of the Interior, Bureau of Mines).

PERRY, E.N. and WAGNER, P.A., 1973, One-hour self-rescue breathing apparatus. *American Institute of Aeronautics and Astronautics*, paper no. 73-1358.

STEIN, R.L., 1978, New respiratory equipment for the mining industry. *Industrial Hygiene for Mining and Tunneling*, November, 59–71.

VAUGHT, C., COLE, H.P. and WIEHAGEN, W.J., 1996, A program to enhance the self-contained self-rescuer donning proficiency of underground coal miners. OMB-No. 0990-0015.

Work Organization Interventions

M. C. Haims[†] and P. Carayon[‡]

†Department of Industrial Engineering, University of Wisconsin-Madison, Madison, WI 53706, USA

‡Ecole des Mines de Nancy, Parc de Saurupt, F-54042 Nancy Cedex, France

1. INTRODUCTION

The use of new technology in occupational settings has created changes in work tools, work methods and working conditions. These changes are likely affecting employee stress, health and quality-of-working-life (QWL) (Carayon and Lim 1994). Upper extremity work-related musculoskeletal disorders (WRMD) are increasingly becoming more prevalent among the US workforce (BLS 1997) and this trend seems to be related to the introduction of computer-based technology in the service sector. The introductions of new technology, and the accompanying changes in job design and work organization, have often failed to incorporate human factors principles of organizational design and management (ODAM) and socio-technical principles of job design and stress reduction. Accumulated research evidence has shown a relationship between the nature of work organization and various stress-related diseases (Cooper and Marshall 1976, House 1981), and there has recently been an interest in the potential relationship between work organization, psychological stress and WRMD (Smith and Carayon 1996).

There are theoretical reasons to believe that work organization factors can play a role in the report and development of WRMD (Sauter and Swanson 1996, Smith and Carayon 1996). Work organization is defined as "the way in which work is structured, supervised and processed" (Hagberg *et al.* 1995: 393–4). The model presented in Figure 1 (Carayon *et al.* 1997) suggests the potential relationship between work organization and WRMD. Mediating this relationship are (1) ergonomic risk factors (e.g. postures, force), (2) psychosocial work factors, that is "the individual subjective perceptions of the work organization factors" (Hagberg *et al.* 1995: 395) and (3) stress load, as well as the interactions among these factors. Organizational, ergonomic and psychosocial factors can produce a "stress load" on the individual: this stress load can have both physiological and psychological consequences (i.e. strain), such as biomechanical loading of muscles or joints, increased levels of catecholamine release, or adverse psychological mood states (Smith and Carayon-Sainfort 1989, Hagberg *et al.* 1995). These various concepts of work organization, ergonomic and psychosocial work factors, stress load, strain, musculoskeletal discomfort and WRMD are hypothesized to describe various stages of the chain of events that link work organization to WRMD.

Traditionally, physical ergonomic factors have been considered the primary risk factors for WRMD (e.g. Putz-Anderson 1988). Recent studies, however, have found associations between psychosocial factors and WRMD even after adjusting for physical demands, suggesting that the effects of these factors may be, in part or entirely, independent of physical factors (Kerr *et al.* 1997, NIOSH 1997, Wahlstedt *et al.* 1997). This research justifies the
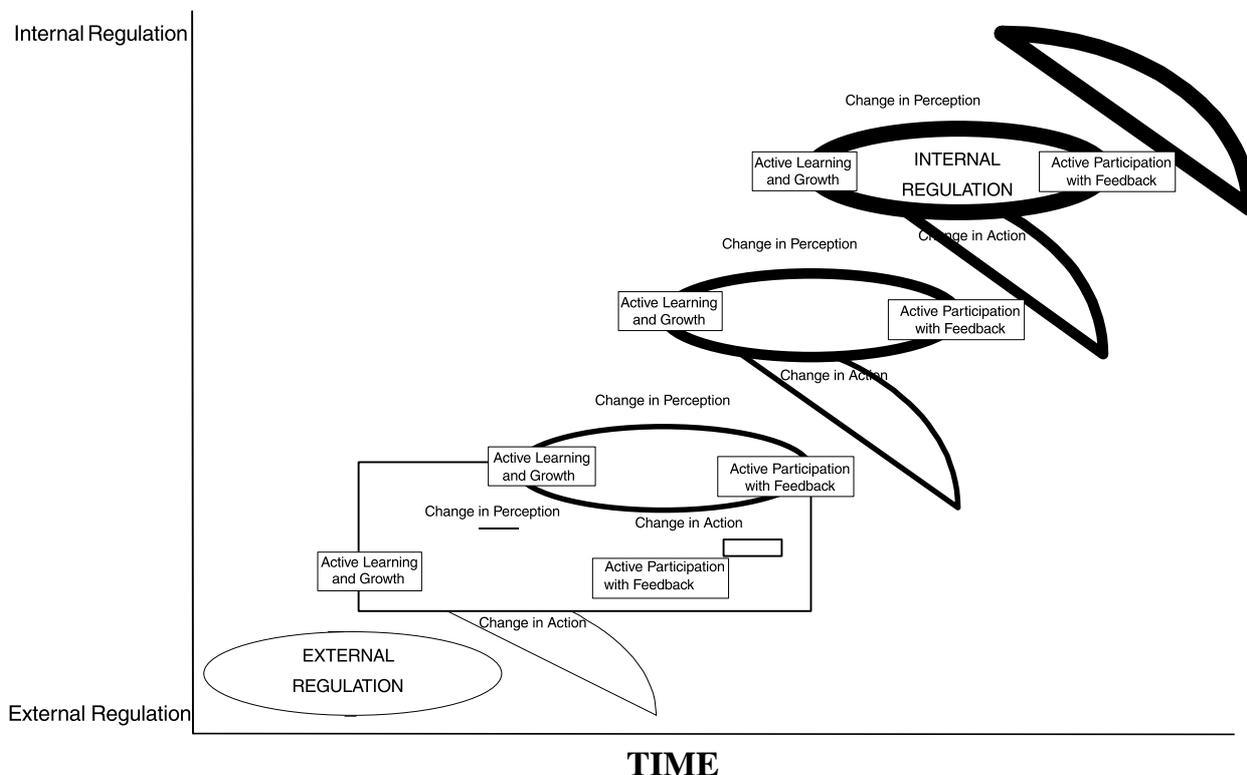
Figure 1. Model of the relationship between work organization and WRMD

exploration of the effects of interventions that go beyond changes to the physical aspects of work environment and job design. Work organization interventions encompass a systemic approach for improving QWL and health by taking into account macro-level organizational variables and both ergonomic and psychosocial work factors.

2. WORK ORGANIZATION INTERVENTIONS

Several work organization strategies have been proposed to improve QWL and health, and to prevent or reduce WRMD (Carayon 1996, Smith and Carayon 1996). Different types and examples of work organization interventions are listed in Table 1.

With respect to time management, various types of rest break schedule have been proposed to reduce exposure to ergonomic risk factors. Also, in order to reduce the QWL and health problems associated with shiftwork, different forms of work scheduling have been proposed: flextime, compressed work weeks and other alternative work schedules. More recently, with the use of new information and communication technology, new forms of work schedule have appeared, such as telework and job sharing. Increased job variety has long been recognized as a means for reducing exposure to ergonomic risk factors: workers are rotated and/or exposed to tasks with different physical requirements. From a psychosocial point of view, however, job rotation and job enlargement are of limited value because the basic tasks are not enriched and often do not add dimension, interest or challenge to work. Work organization interventions aimed at increasing

job content and responsibility or at changing the basic structure of work (e.g. team work and organizational design) are potentially more effective from a psychosocial point of view. Whereas job rotation and job enlargement are horizontal strategies (i.e. increasing the variety of jobs), job enrichment strategies, aimed at increasing job control and team work, are vertical strategies: employees are given more responsibility within their jobs. Team-work has different facets, such as temporary versus permanent teams and level of participation or involvement. Different forms of organizational design can contribute to improved QWL and health. Flat organizations and decentralization can help promote increased job content and responsibility for workers,, therefore, increasing satisfaction, motivation and QWL. Total quality management and other quality improvement strategies can be aimed at "internal customers", i.e. the employees, and can, therefore, foster improvement in working conditions. Finally, work organization interventions can target management, and indirectly contribute to better psychosocial and ergonomic work conditions.

It is important to recognize that these different forms of work organization have the potential to improve QWL and health, but may also have some negative effects on employees (e.g. increased workload and work pressure). Recognizing these negative effects and preventing or reducing them should be part of any work organization intervention.

3. PROCESS FOR WORK ORGANIZATION INTERVENTIONS

If not well managed, changes in work organization can become a source of employee stress. Change can lead to feelings of uncertainty, increased workload and feelings of decreased job control, all well-known psychosocial factors related to stress (Hagberg *et al.* 1995). It is crucial to develop appropriate processes for planning, implementing and monitoring intervention changes. One of the greatest challenges for successful interventions is overcoming resistance to change among various individuals and groups within an organization. Effective approaches for overcoming resistance include participation/involvement and knowledge/communication (Hagberg *et al.* 1995).

Participation has been used as a key method for implementing various types of organizational changes, such as ergonomic programs (Wilson and Haines 1998), continuous improvement programs (Zink 1996) and technological change (Eason 1988). Haims and Carayon (1998) have developed a model for the implementation of participatory work organization interventions (Figure 2) and have conducted several research efforts toward better understanding how to implement effective participatory work organization interventions and for defining best practices and design principles. The model is based upon the behavioral cybernetic theory of learning (Smith and Smith 1966) and asserts that effective participatory programs are developed through continuously providing opportunities for employee involvement linked with feedback and control mechanisms, which leads to continuous learning and the ability of employees to take increasingly higher levels of control over the intervention efforts over time. The model suggests that gradual increases in participation, learning, feedback and control that occur over time allow for the gradual transfer of the participatory program

Table 1. Types and examples of work organization interventions.

Types of work organization interventions	Examples	
1. Time management	Alternative rest break schedules	Compressed work week
	Flextime	Job sharing
	Self-scheduling	Telework
	Alternative work schedules	
2. Increased job variety	Job rotation	
	Job enlargement	
	Cross-functional training	
3. Increased job content and responsibility	Job enrichment	
	Vertical job loading	
	Control over work place	
4. Team work	Semi-autonomous work groups	Autonomous work groups
	Project teams	Self-managing teams
	Participatory ergonomics	Suggestion programs
	Quality circles	
5. Organizational design	Flat organization	
	Decentralization	
	Design of electronic performance monitoring systems	
	Total quality management	
	Healthy organizations	
6. Management	Supervisory training	
	Increased communication, feedback, information exchange	
	Just-in-time scheduling	
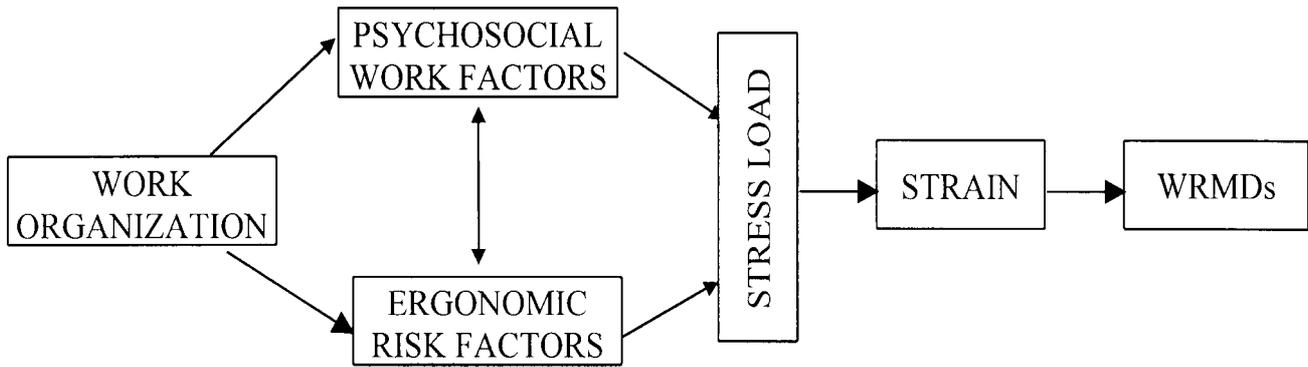
Figure 2. Research model for the implementation of participatory work organization interventions (from Haims and Carayon 1998).

from external regulation by outside experts to internal regulation by organizational members. We argue that this model can be used for the effective implementation of work organization interventions.

Several authors have proposed design principles and best practices for the process of implementing participatory work organization interventions (Spencer 1989, McLagan and Nel 1995, Mohrman and Mohrman 1997, Haims and Carayon 1998), and together they represent a "toolbox" for human factors and ergonomics professionals to use in the field. The subsequent list by no means is all-inclusive, but rather represents what these authors have expressed as key components for effective, continuous improvement work organization interventions.

- Collect data for identification of problems and identification of criteria and goals. Intervention design and assessment should "match up with" the data and established criteria and goals.
- Include all key stakeholders in the process, along with those who bring the necessary knowledge, skills and perspectives "to the table."
- Develop a staged process plan, including design/development, implementation, evaluation and continuous improvement phases.
- Strike a balance between structure (e.g. project management and adhering to the process plan) and flexibility (e.g. making necessary changes along the way).
- Ensure the necessary resources are available to carry out the process plan.
- Secure "buy-in" and commitment (including time) from top management.
- Share information by communicating all process plans to all organizational members.
- Use a systems approach to bring about change. That is, change should not be isolated, but rather widespread and embedded within the organization.
- Incorporate individual and organizational learning into the process. That is, treat change as a learning process and design for self-design.
- Adhere to established training principles, such as experiential learning through action, feedback and feedback control and staged levels of learning.
- Be a flexible, dynamic outside expert willing to wear many different hats throughout the process of intervention development and implementation.

4. EFFECTIVENESS OF WORK ORGANIZATION INTERVENTIONS

When examining the effectiveness of work organization interventions, it is important to consider both the process and content of the intervention (Haims and Carayon 1998). An effective process is one that reduces resistance to change, reduces stress related to change and leads to long-term, continuous improvements. Effective content (including the content of the process) leads to improved QWL and employee health. In this section, several examples of work organization interventions, using different processes and/or contents, are provided, along with discussion of their effectiveness and/or limitations.

Westlander *et al.* (1995) conducted a field study to examine the effects of an ergonomics intervention program in two different workplaces. The aim of the intervention was to create an ongoing intervention program, involving participatory problem solving, to deal with current and future ergonomics problems. Implementation of the interventions was not successful, and Westlander *et al.* (1995) discuss the environmental and organizational barriers they encountered in implementing such programs. Specifically, they found that turbulent economic conditions and concurrent organizational changes contribute to the difficulty of implementing interventions initiated by outside researchers. Åborg *et al.* (1998) also report organizational factors as limitations to the effectiveness of their reorganization of data entry work, such as lack of communication of changes, insufficient employee involvement in decision-making related to work organization changes and employee expectations of positive change not being achieved.

Smith (1994) used worker participation in a meat processing plant as a means for defining ergonomic problems and implementing improved work methods. Ergonomic training was provided to plant supervisors and employees, and an employee was named as a full time ergonomics coordinator to implement solutions, facilitate focus groups, and keep awareness levels of ergonomics high. Another component of the program was to involve all employees in ergonomic focus groups for recommending engineering controls, better work methods, and job enrichment strategies. Preliminary results showed a reduction in ergonomic risks, reduction in psychosocial stress and the "institutionalization" of ergonomics as part of daily life within the plant (Smith 1994).

The National Institute of Occupational Safety and Health reports on an additional three case studies of participatory ergonomic interventions in meatpacking plants (NIOSH 1994). Each

project used a participatory approach involving front-line workers, supervisory personnel and others organized into teams for problem-solving purposes. Findings from these case studies included: (1) successful participatory ergonomics programs require strong in-house direction, support and ergonomic expertise, (2) training programs must develop both team work and ergonomic skills among participants, (3) teams should include supervisors, maintenance and/or engineering staff (who actually implement the recommended changes), as well as production workers engaged in the jobs being studied, (4) realistic and measurable goals need to be set and communicated and (5) evaluation criteria must be planned.

Haims and Carayon (1998) implemented a participatory ergonomics program in a public service agency in the Midwest US with the aim of ergonomic and job redesign for reducing stress and improving health. A group of 13 employees, representing a total of 450 employees throughout the organization, directly participated in the program. Participants were provided with "hands-on" training and feedback and were encouraged to actively participate throughout the process. Findings indicated that participants in the program experienced a great deal of learning over time, increased levels of communication and information and gains in control over the program. Preliminary data showed a general trend of upper body discomfort, hand/arm discomfort and postural stressor complaints decreasing over the study period in one of the represented work groups within the organization. Results were positive but suggested the importance of increased consideration of macro-level organizational variables. For example, management support was crucial for providing participants with time, access and opportunities to exercise their gained knowledge and expertise for improving ergonomic work conditions and continuing program initiatives over time.

Haims (1998) reports on a longitudinal study of the implementation of a participatory work organization intervention among a group of office workers to examine the effects of the process and content of the intervention on both direct participants (project team (PT) members) and indirect participants (non-PT members) over time. The intervention process combined the approaches of participatory ergonomics, behavioral cybernetics and action research. The intervention content consisted of adherence to a document that specifies how work is to be organized and continuously improved. Data were collected using questionnaires (PT and non-PT members) and interviews (PT members only). Interview results suggested positive changes in participation in decision-making, communication, feedback and learning for both the PT and non-PT groups. However, work demands generally increased. Although non-PT were not directly involved in the participatory intervention process, they seemed to experience "spill over" effects, through interaction with PT and the process itself. It appears that learning among the PT was transferred or diffused to the non-PT over time. Owing to limitations, the questionnaire was generally unable to detect changes in either psychosocial factors or outcome stress and health measures. A multitude of outside environmental factors "overpowered" any effects of the intervention at the global level of questionnaire measurement. Results suggested the need for intervention research to focus at the global, work system level and for the development of methodologies and measures to examine the participatory learning process.

5. CONCLUSION

An inherent problem in intervention research is the plethora of uncontrollable organizational and environmental factors that always exist in concert with the intervention being put into place (Åborg et al. 1998). This, along with time issues (Carayon 1997), makes the measurement and assessment of interventions difficult. Solutions are not easy, but several are suggested. First, work organization interventions must adhere to a systems approach to change and focus at the global work system level (i.e. upper-level management) for making changes to the overall organizational structure. A work organization intervention will only be successful if a structure and culture are in place to support it. Second, work organization interventions must be designed for continuous self-design in order that the organization be capable of responding effectively to the reality of today's turbulent environmental conditions. This can be achieved by incorporating widespread involvement, participation and (individual and organizational) learning into both the process and content of work organization changes. Lastly, in terms of methodology, it is suggested that organizational variables (e.g. organizational climate and culture) be measured so that they can be controlled for, that quantitative and qualitative data be collected to provide both "hard results" and "contextualization" of the change process and that long-term longitudinal study designs be adopted to examine a variety of short- and longer-term effects over time.

REFERENCES

ÅBORG, C., FERNSTRÖM, E. and ERICSON, M.O., 1998, Work content and satisfaction before and after a reorganisation of data entry work. *Applied Ergonomics*, 29, 473–80.

Bureau of Labor Statistics, 1997, Safety and Health Statistics: Workplace Injuries and Illnesses in 1995 (Washington, DC: BLS, US Department of Labor); http://stats.bls.gov/news.release/osh.nws.htm

CARAYON, P., 1996, Longitudinal studies of job stress and upper extremity musculoskeletal discomfort among computer users. Presented at the International Conference on Occupational Disorders of the Upper Extremities, Ann Arbor, Michigan.

CARAYON, P., 1997, Temporal issues of quality of working life and stress in human–computer interaction. *International Journal of Human–Computer Interaction*, 9, 325–42.

CARAYON, P., HAIMS, M.C. and SUH, H-S., 1997, *Work Organization Interventions to Prevent Work-related Musculoskeletal Disorders in Office and Video Terminal Work*. Final Report of CDC/NIOSH Cooperative Agreements U60/CCU512018-01 and U60/CCU512983-01 (Madison: University of Wisconsin-Madison, Department of Industrial Engineering).

CARAYON, P. and LIM, S.-Y., 1994, Stress in automated offices. In A. Kent (ed) *The Encyclopedia of Library and Information Science*, vol. 53, suppl. 16 (New York: Marcel Dekker), pp. 314–54.

COOPER, C.L. and MARSHALL, J., 1976, Occupational sources of stress: a review of the literature relating to coronary heart disease and mental ill health. *Journal of Occupational Psychology*, 49, 11–25.

COTTON, J.L., 1993, *Employee Involvement: Methods for Improving Performance and Work Attitudes* (Newbury Park: Sage).

EASON, K., 1988, *Information Technology and Organizational Change* (London: Taylor & Francis).

HAGBERG, M., SILVERSTEIN, B., WELLS, R., SMITH, M.J., HENDRICK, H.W., CARAYON, P. and PÉRUSSE, M., 1995, *Work-related Musculoskeletal Disorders (WMSDs): A Reference Book for Prevention* (London: Taylor & Francis).

HAIMS, M.C., 1998, A longitudinal study of the process and content of a participatory work organization intervention. Unpublished doctoral dissertation, University of Wisconsin-Madison.

HAIMS, M.C. and CARAYON, P., 1998, Theory and practice for the implementation of "in-house", continuous improvement participatory ergonomic programs. *Applied Ergonomics*, 29, 461–72.

HOUSE, J., 1981, *Work, Stress and Social Support* (Reading, MA: Addison-Wesley).

KERR, M.S., SHANNON, H.S., FRANK, J.W., NORMAN, R.W.K., WELLS, R.P. and NEUMANN, P., 1997, The relative importance of bio-mechanical and psychosocial risk factors in a case-control study of occupational low-back pain. In P. Seppälä, T. Luopajärvi, C.-H. Nygård and M. Mattila (eds) *Proceedings of the 13th Triennial Congress of the International Ergonomics Association*, vol. 4 (Tampere: Finnish Institute of Occupational Health), pp. 64–5.

MCLAGAN, P. and NEL, C., 1995, *The Age of Participation* (San Francisco: Berrett-Koehler).

MOHRMAN, S.A. and MOHRMAN, A.M. JR, 1997, Fundamental organizational change as organizational learning: creating team-based organizations. In W. Pasmore and J. Woodman (eds) *Research in Organizational Change and Development*, vol. 10 (Greenwich: JAI Press), pp. 197–228.

NIOSH, 1994, *Participatory Ergonomic Interventions in Meatpacking Plants*. Publication 94-124 (Washington, DC: US Department of Health and Human Services).

NIOSH, 1997, *Musculoskeletal Disorders and Workplace Factors*. Publication 97-141 (Washington, DC: US Department of Health and Human Services).

PUTZ-ANDERSON, V., 1988, *Cumulative Trauma Disorders: A Manual for Musculoskeletal Diseases of the Upper Limbs* (London: Taylor & Francis).

SAUTER, S.L. and SWANSON, N.G. (eds), 1996, An ecological model of musculoskeletal disorders in office work. In *Beyond Biomechanics: Psychosocial Aspects of Musculoskeletal Disorders in Office Work* (London: Taylor & Francis), pp. 3–21.

SMITH, K.U. and SMITH, M.F., 1966, *Cybernetic Principles of Learning and Educational Design* (New York: Holt, Rhinehart, and Winston).

SMITH, M.J., 1994, A case study of participatory ergonomics and safety program in a meat processing plant. Presented at the 12th Triennial Congress of the International Ergonomics Association, Toronto.

SMITH, M.J. and CARAYON, P., 1996, Work organization, stress, and cumulative trauma disorders. In Moon and Sauter, S.L. (eds) *Beyond Biomechanics: Psychosocial Aspects of Musculoskeletal Disorders in Office Work* (London: Taylor & Francis), pp. 23–41.

SMITH, M.J. and CARAYON-SAINFORT, P., 1989, A balance theory of job design for stress reduction. *International Journal of Industrial Ergonomics*, 4, 67–79.

SPENCER, L., 1989, *Winning Though Participation* (Dubuque: Kendall/Hunt).

WAHLSTEDT, K., BJORKSTEN, M.G. and EDLING, C., 1997, Psychosocial work situation and musculoskeletal symptoms among Postal Workers. In P. Seppälä, T. Luopajärvi, C.-H. Nygård and M. Mattila (eds) *Proceedings of the 13th Triennial Congress of the International Ergonomics Association*, vol. 4 (Tampere: Finnish Institute of Occupational Health), pp. 413–15.

WESTLANDER, G., VIITASARA, E. JOHANSSON, A. and SHAHNAVAZ, H., 1995, Evaluation of an ergonomics intervention programme in VDT workplaces. *Applied Ergonomics*, 26, 83–92.

WILSON, J.R. and HAINES, H.M., 1998, Participatory design in the organisational context. In Scott, Bridger and Charteris (eds) *Global Ergonomics* (Amsterdam: Elsevier), pp. 11–17.

ZINK, K.J., 1996, Continuous improvement through employee participation: some experiences from a long-term study. In O. Brown JR. and H. W. Hendrick (Eds.) *Human Factors in Organizational Design and Management – V* (Amsterdam: Elsevier), pp. 155–60.

Work Organization, Job Stress and Occupational Health

M. J. Smith†, R. Kalimo‡ and K. Lindstrom‡

†Department of Industrial Engineering, University of Wisconsin-Madison, Madison, WI, USA

‡Department of Psychology, Finnish Institute of Occupational Health, Helsinki, Finland

1. INTRODUCTION

The purpose of this chapter is to provide an understanding of how work organization features can lead to stress and adverse health effects. It will discuss the working conditions that produce stress, and organizational interventions to reduce adverse health effects due to stress. The physiological basis of stress theory lies in the work of Selye (1956) who described a pattern of physiological reactions to environmental stressors which he called the "general adaptation syndrome" (GAS). This GAS was activated in a non-specific, stereotyped form by any environmental demand (Selye 1956). One of the characteristic features of this theory was that stress was understood as a stimulus–response phenomenon. The organism's reaction was activated in the same manner by any environmental demand which was called a stressor. Later theorists believed that this response-based definition was limited because the same reaction pattern could be evoked by a wide array of stimulus conditions.

Lazarus (1977) developed a cognitive theory of stress in which individual differences in cognition intervene between the stressor (stimulus) and the reaction (response) to define the nature of the reaction. These cognitive processes determine the quality and the intensity of the reaction to the stressor. A stressor will not have an effect unless it is recognized and assessed by the person in an appraisal phase. Three different kinds of stress situations are acknowledged, which are harm, threat and challenge. Harm refers to psychological damage that has already taken place, such as an irrevocable loss like the death of a loved one. Threat is the anticipation of harm that has not yet taken place but may happen. Challenge results from difficult demands that one feels confident about overcoming by the effective mobilization of resources.

Current stress theories usually involve the concept of a transaction between the environment and the individual. Some theories characterize the development of stress as an outcome of an imbalanced interaction of the environment and the person (poor "fit"). If the environmental demands are greater than the person's capacities and/or if the person's expectations are greater than the environmental supplies, then stress responses will occur (Smith and Sainfort 1989, Kalimo *et al.* 1997). These theories do not see the environment, such as work, only as a source of stressors. They see work as a source of a full range of demands and supplies that may match the individual. Thus, work can be a source of life satisfaction as well as a source of distress.

Work makes demands on the person regarding for instance workload and participation. The person responds to these demands according to her/his skills, motivations and other characteristics. The person has expectations regarding work, such as decision-making, opportunity for development and support. Work may present the opportunity for the achievement of such goals to a varying degree. In the case of a positive "fit" the working conditions put manageable challenge to the person, which results in positive satisfaction, feelings of competence, well-being and positive health. A poor "fit" causes stress responses and a series of negative reactions begin to develop. Negative emotions including frustration, dissatisfaction, fear and anger appear at first. These lead into the development of physiological changes involving the autonomic nervous system and the hormonal and immunological systems. If coping processes fail then the stress responses may change into a chronic state and eventual mental and/or physical disorders develop.

2. WORK ORGANIZATION AND JOB DESIGN

It is important to recognize those psychosocial characteristics in work, work organization and the environment which are potential contributors to stress and which adversely influence the health and well-being of individual workers, groups and the whole organization. The characteristics of the work and of the organization are overlapping and not independent. Their stress effects depend on their mutual interaction and on the specific characteristics of the individual (personality, capacity). We are interested mainly in the health relatedness of the psychosocial characteristics of work and work organization, but they can be also be important contributors to job satisfaction, work motivation and work performance.

2.1. Characteristics of Work Organization

Cooper (1986) has categorized organizational stress factors as those that are intrinsic to the job, role in the organization, career development, the relationships at work, as well as the organizational structure and organizational climate. In this section factors related particularly to the physical work environment and task demands are described. Factors intrinsic to the job are: (1) physical working conditions, (2) workload, both quantitative and qualitative, and time pressure, (3) responsibilities (for lives, economic values, safety of other persons), (4) job content, (5) decision-making and (6) perceived control.

Many blue-collar jobs entail a high physical energy expenditure or difficult work movements and postures. Ergonomic factors, such as repetition, posture, and force can contribute to the development of musculoskeletal problems (e.g. musculoskeletal discomfort or fatigue), and the features of a psychosocial work environment can also contribute to their development. The ergonomic aspects of the work place interact with the psychological factors to influences stress responses.

The most obvious job characteristic is workload. This describes both quantitative and qualitative aspects of the task. Stress can be due to overload (over work) or underload (boredom). Quantitative overload means that there is too much to do or the person has to work under time pressure or meet strict deadlines. Also, requirements to work excessive hours increase quantitative workload. Too much workload has been associated with stress in many occupations from assembly-line workers to nurses to computerized clerical staff.

Control at work can be defined as the ability of a person to exert some influence on one's environment and/or some

characteristic of the task or the work organization. In a work situation the person can have control over various job demands, like the task itself, pacing of the work, work scheduling, the physical environment, decision-making, other people or mobility. As such, and in combination with work demands, a lack of job control or decision latitude has been found to be an important contributor to stress and ill-health. When job control is high, the other job demands such as workload tend to have less potential for adverse stress effects.

Uncertainty or lack of clarity is closely related to a lack of control because it implies a lack of needed information for control. Uncertainty exists when that knowledge about events requiring action is experienced as being inadequate. Uncertainty has been emphasized as one of the most predominant single sources of occupational stress. It arises from job insecurity, task ambiguity, lack of knowledge of results of performance and unanticipated slowdowns in work production. With increasing use of computer technology, this kind of unpredictability has become more frequent and quite stressful.

Responsibility at work can be related to the success and safety of other persons or to material resources. These issues are prominent in customer or client-related operations. Responsibility is also associated with occupational roles and organizational hierarchy, such as being a supervisor or manager. High responsibility can lead to high stress levels.

2.2. Organizational Structure and Culture

Organization that fail to have a commitment to employee welfare, career stability and personal development have been shown to influence employee stress. Job insecurity is a primary issue for stress in many workers, particularly when work life is undergoing major structural changes associated with new technology, downsizing, mergers and business process re-engineering.

Supervisory style is important from the occupational health perspective. A hierarchical, authoritative management style in general has been found to have negative effects on workers' well-being and stress, whereas a participatory leadership style has been found to be advantageous.

In a job design context, the following characteristics of job content are of main interest: the variety of tasks and task identity, feedback about job performance, perceived personal contribution to the product or service, challenges and opportunities to use one's skills, and individual autonomy. If these characteristics are perceived negatively by the employee then the level of stress increases.

Three types of problems have been observed in the organizational role setting: role ambiguity, role conflict, and role overload. Role conflict is defined as a condition of incompatible role expectations placed on a person. Role ambiguity means a lack of clarity regarding the role expectations, and an uncertainty concerning the outcomes of one's role performance. Role overload means a scope of tasks that is too wide with too many different kinds of role expectations. Recently, role underload has been recognized as a problem. It is caused by too low and too few expectations placed on a person who would be capable of a greater contributions.

Interpersonal conflicts are a source of stress which can lead to violence at work. Recent research about interpersonal relations has been focused on aggravated conflict situations and bullying.

Bullying, mobbing or psychological harassment, are a form of expression of the problems in the interaction between people at the workplace. A person is being bullied when s/he is exposed repeatedly and over time to negative actions on the part of one or more persons at the workplace. It means teasing, pressure and unfair treatment of a person in situations where the target cannot defend him/herself.

2.3. Extra-organizational Influences

Work and private life were recognized as two interdependent spheres of life in work-related stress and health. This relationship can be characterized by: (1) a spillover effect of stress from one sphere to the other, (2) role conflict between work and family responsibilities, (3) a cumulative effect of work and family stressors and (4) the buffering effect of one sphere against stress in the other.

3. SOCIAL SUPPORT

Social support is a resource which has been shown to influence the effects of stress. Social support can include: (1) social integration, which is the size and structure of one's social network, (2) the availability of satisfying relationships characterized by love, trust and esteem, and (3) the actual receiving of supportive acts from others when in need. House (1981) emphasized the importance of the actual receiving of social support as the key factor for the beneficial impact. He defined four forms of such support: (1) emotional support such as care giving and affectionate concern, (2) appraisal support through evaluative feedback and affirmation, (3) informational support by giving suggestions or guiding, and (4) instrumental support through organizing opportunities.

Social support is a beneficial factor for well-being and job satisfaction and lowered stress. There are direct effects in which social support is positively related to person's health independent of environmental circumstances or stress level. In addition, social support provides a protective function during conditions of stress, that is to "buffer" a person against the harmful effects of the social environment. As a whole, evidence for the beneficial effect of social support is conclusive enough to merit the promotion of efforts aimed at the improvement of interpersonal interaction at the work place. The role of the supervisors as sources of support seems to be stronger than that of colleagues and work mates.

4. IMPROVING WORK AND ORGANIZATIONS

Improving work and work organizations is usually a practically oriented activity supported by consultants and other specialists from inside and/or outside the organization. The intervention traditions and models are also strongly dependent on current social and cultural values of the company and the employees. The goal of a specific intervention or program at work is to improve the functioning of an organization, its productivity, profit, quality of production and services, and to promote workers' health, aspiration and competence, by making work and its characteristics favorable.

4.1. Healthy and Productive Work Organization

The organizational health approach reflects the capability of an organization to function effectively in relation to various economic and market factors within the context of a supportive work

environment for employees. This provides the ability respond effectively to environmental changes. A value-based organization using the healthy company model involves strategies for profit-making and valuation of people. The management plays a key role in the implementation of these strategies, which are open communication and employee involvement, learning and renewal, valued diversity, institutional fairness, and equitable rewards and recognition, and general economic security. People-centered technology as well as a health-enhancing environment and meaningful work are important. The balance between work and family life is also emphasized.

4.2. Improving Environmental and Ergonomic Conditions

Ergonomics is the science of fitting the environment and activities to the capabilities, dimensions and needs of people. Good ergonomic practice suggests that all aspects of the work system be included in job redesign improvements. From an ergonomic perspective the design of the environment, workstations, tasks, work organization and the tools/technology should reasonably accommodate the employees' capacities, dimensions, strengths and skills. Critical aspects of the success of ergonomic interventions for providing psychosocial benefits are a strong management support for the ergonomics program; the involvement of managers and supervisors in the ergonomics program; the involvement of employees in the program and their willingness to participate in defining problems, proposing solutions and improving their work practices; and the application of engineering improvements to reduce biomechanical and physiological loads.

4.3. Strategy for Job Redesign

The following section will define a strategy for redesigning jobs to reduce job stress, but will not elaborate the specific characteristics of job elements needed to achieve proper job design. Smith and Sainfort (1989) have proposed a balance among various elements of the work "system." The essence of this approach is to reduce the negative health consequences caused by stress by "balancing" the various elements of the work system to reduce unwanted loads. It goes without saying that proper work organization and job design can best be achieved by providing those characteristics of each individual work system element that meet recognized criteria for proper physical loads, work cycles and job content, and that provide for individual physiological and psychological needs. The best designs will eliminate all sources of stress, job dissatisfaction and discomfort.

For instance, proper workload can be established by use of appropriate methods of work analysis, while aspects of job control can be developed through the use of worker participation. The negative influences of inadequate skill to use new technology can be offset by increased training of employees to acquire the needed skills. Or the adverse influences of low job content that creates repetition and boredom can be balanced by an organizational supervisory structure that promotes employee involvement and control over tasks, and job enlargement that introduces task variety if not complexity.

A major advantage of this model is that it does not highlight any one-job factor such as job content, or a small set of factors such as autonomy and social relations. Rather it examines the design of jobs from a holistic perspective to emphasize the potential positive elements in a job that can be used to overcome the adverse elements when such adverse elements cannot be eliminated or modified.

4.4. Organizational Change Management

All planned changes and development interventions should follow the developmental cycle. They start with problem or goal definition, and the committing and informing of all interest groups at the workplace. They then proceed to analysis of the present situation, based on survey results, and to joint planning of the intervention. The intervention itself needs clear goals, subgoals, steps, a time schedule, duties for people, and ways to overcome obstacles and problems during the process. The process seldom proceeds as planned, it needs reevaluation and checking of the methods and even the goals during the process. The intervention process and the outcome should be evaluated. This helps the organization to learn about its experiences for the next project. This part is usually left out because very often changes and projects are already arising before the end of the former changes and projects.

The successful implementation of changes in work processes, and subsequent enhancement of worker health, performance, and satisfaction requires organizations to have a transition policy that includes worker participation in all stages of the change process. That is, workers should participate in the planning, then in the selection of equipment, and finally in the daily operation of the work system.

5. SUMMARY

Psychosocial aspects of the workplace can have profound influences on the individual's health and well-being, satisfaction with work and productivity. There are various characteristics of working that have generally been shown to have negative physical and/or psychological consequences, for example, machine paced work, a lack of task control, high job demands, shiftwork, poor supervisory relations, and organizational downsizing. The extent of the negative consequences varies from individual to individual depending on the perceived threat, individual constitution and coping mechanisms. When the person is vulnerable, substantial harm can occur including disability and death.

Several organizational behavior (OB) approaches that can be applied which will diminish or eliminate the harmful effects of poor workplace design that leads to psychological distress. Examples of these approaches include removal of the sources of stress (stressors) through job redesign, balancing the positive and negative aspects of work to provide an acceptable level of stress, developing a corporate culture that promotes self worth and productive employment (the "healthy organization"), and enhancing employee skills and career development.

If the psychosocial aspects of work design are properly developed and implemented, there can be benefits for employee's health and performance, and a "healthy corporation" will emerge. If job stress is not properly controlled, then the consequences can be reduced employee job satisfaction, health, and morale, and reduced employee and corporate performance.

REFERENCES

COOPER, C.L., 1986, Job distress: recent research and the emerging

role of the clinical occupational psychologist. *Bulletin of the British Psychological Society*, 39, 325–31.

HOUSE, J.S., 1981, *Work Stress and Social Support* (Reading, MA: Addison-Wesley).

INTERNATIONAL LABOUR OFFICE, 1986, *Psychosocial Factors at Work: Recognition and Control*. Occupational Safety and Health Series no. 56 (Geneva: ILO).

KALIMO, R., EL-BATAWI, M.A. and COOPER, C.L., 1987, *Psychosocial Factors at Work and their Relation to Health* (Geneva: World Health Organization), 23–36.

KALIMO, R., LINDSTROM, K. and SMITH, M.J., 1997, Psychosocial approach in occupational health. In G. Salvendy (ed.) *Handbook of Human Factors and Ergonomics* (2nd ed.) (New York: Wiley-Interscience), 1059–84.

LAZARUS, R.S., 1977, Cognitive and coping processes in emotion. In A. Monat and R.S. Lazarus (eds) *Stress and Coping* (New York: Columbia University Press), 145–58.

LINDSTRÖM, K., 1991, Well-being and computer-mediated work of various occupational groups in banking and insurance. *International Journal of Human–Computer Interaction*, 3, 339–61.

LINDSTRÖM, K., 1995, Searching for the model of a healthy organization. In *Work, Stress and Health '95: Creating Healthier Workplaces*, American Psychological Association and National Institute for Occupational Safety and Health, 14–16 September 1995, Washington, DC, 37.

SELYE, H., 1956, *The Stress of Life* (New York: McGraw-Hill).

SMITH, M.J. and SAINFORT, P.C., 1989, A balance theory of job design for stress reduction. *International Journal of Industrial Ergonomics*, 4, 67–79.

Part 10

Health and Safety

Agriculture

D. H. O'Neill

Silsoe Research Institute, Silsoe MK45 4HS, UK

1. INTRODUCTION

More people gain their livelihoods through agriculture (including horticulture) than any other industry. About 50% of the world's economically active population produce food, most working at the subsistence or small-scale level as a family enterprise. This makes agriculture the world's largest industry, with > 3.5 billion people depending on it. Agriculture is also one of the most labor-intensive and energy-demanding occupations, alongside mining and construction work. This is true for all farming systems, whether small or large in scale, extensive or intensive. In terms of human effort agriculture is the most important industry on the planet and the ergonomics issues are as wide and varied as the industry itself. Different aspects of ergonomics will be prominent in different farming systems, but the design and use of equipment and associated working practices are the underpinning issues.

Agriculture is also one of the most hazardous occupations at whichever level it is practiced, but the risks and hazards depend on the farming system and these change with the degree of mechanization. Environmental and other stresses, such as caused by weather conditions, long working hours and isolation, are also likely to be more extreme in agriculture than in other industries.

Table 1. Three types of agriculture

	Small-scale	Intermediate	High input
Main locations	rainfed areas, hinterlands, most of SSA	irrigated and stable rainfall, high potential areas in the Third World	industrialised countries and specialised enclaves in the Third World
Main climatic zone	tropical	tropical	temperate
Major type of farmer	small and poor farm households	large and small farmers	highly capitalized family farms and plantations
Use of purchased inputs	low	high	very high
Nature of farming system	complex	simple	simple
Production stability	high risk	moderate risk	moderate risk
Priority for production	raise	maintain	lower
Production level relative to sustainability	low	near limit	too high
Environmental diversity	varied	uniform	uniform

Adapted from Chambers *et al.* 1989

Table 2. Contrasting characteristics of small-scale and high input farming

	Small-scale	High input
1	underproduction	overproduction
2	low yields	high yields
3	starvation	overeating
4	low body weight	obesity
5	poverty	affluence
6	crop product consumption	meat consumption
7	muscle powered	oil powered
8	high percentage of young people	high percentage of old people
9	79% of population rural	22% of population rural
10	poor hygiene	good hygiene
11	contaminated water	clean water
12	poor control (over conditions)	good control
13	food processed domestically	food processed industrially
14	poor infrastructure	sophisticated infrastructure

All these factors demand major ergonomics inputs to improve the productivity of the agricultural industry and the welfare of its workers in all countries of the world. Some examples of the problems and possible solutions are presented below for the main types of farming systems.

2. FARMING SYSTEMS

Farming systems may be categorized in a variety of ways but, from the ergonomics perspective, it is convenient to consider the level of mechanization. Three levels can be identified, described as (1) subsistence or small-scale farming, (2) intermediate mechanization and (3) high-input and high-output farming. The lowest level predominates in the poorest countries of the world, mainly in sub-Saharan Africa (SSA), while high-input–high-output farming is found in the richest countries of the world (typically Europe, North America, Japan and Australia), although not all the farmers in these countries operate at this level. The intermediate level of mechanization is found among the more successful farmers in countries with emerging industrial economies, typical of South-East Asia and South America. On average, farmers' potential to invest in mechanization reflects the local industrial situation. The main characteristics of these three significant types of farming system, described by Chambers *et al.* (1989), who referred to these systems as "Industrial," "Green Revolution" and "Third/CDR" (complex, diverse, risk-prone), are summarized in Table 1. Some contrasts between small-scale and high-input farming systems are given in Table 2.

Several of these factors have ergonomics implications, particularly those numbered 4, 7, 12, 13 and 14. Nutrition and body size have implications for work capacity, endurance and equipment design. Anthropometric differences may exist over and above those observed in national or ethnic populations. Oil- rather than muscle-powered equipment confers a role of control

rather than energy (mechanical) provision by the farmer. The final three all implicate contrasting levels of equipment design and use, including power tools, production lines, information processing, environmental hazards and transportation. The contrast, relating to age, indicated by the eighth factor may be diminishing in the late 1990s because of migration of younger people to cities to avoid the drudgery of agriculture and to find more rewarding work. Recently, issues relating to the first factor for high-input farming have become more prominent, with the sociological problems associated with overproduction achieving similar significance to those of poverty alleviation and sustainability of small-scale and subsistence farming.

3. WORKLOAD

Women do the majority of the work in the small-scale farming sector (Figure 1). This work is added to their domestic tasks, to which the male contribution is ~5%. Thus, for women problems of workload, in terms of both energy (and therefore nutrition), and time are amplified in the small-scale sector and must be considered alongside the other gender issues (e.g. income generation, decision-making).

For both men and women in small-scale agriculture the greatest constraints on crop production and productivity are land preparation and weeding; the former because of the energy required and the latter because of the time required. Workloads in small-scale agriculture in industrially developing countries may exceed energy intake, particularly during land preparation when food reserves are often depleted.

Typical levels of work intensity range from ~12 kJmin⁻¹ (at ~30% VO_{2max}) for simple weeding tasks to ~20 kJmin⁻¹ (at ~50% VO_{2max}) for digging in hard soil. Work intensity may also be expressed as PAL (physical activity level), a multiple of BMR (basal metabolic rate). When PAL > 2, as may be demanded for sustained periods in agriculture, considerable motivation is required to maintain work output. A PAL > 2 should be avoided by pregnant and lactating women.

4. ERGONOMICS FACTORS

In all farming systems, the principal ergonomics factors are the (1) design and use of equipment, particularly ensuring physical, economic and cultural compatibility, (2) information processing and consequent decision-making and (3) health and safety. The levels of engineering and technology increases from small-scale to high-input farming, but that does not necessarily mean that the complexity of the ergonomics issues increases, merely that it changes. The design criteria can be just as stringent for a low-cost, multi-purpose and durable handtool as for an engine-powered piece of equipment, such as a hedge-cutter or a power tiller. However, because of the value of the market (rather than the size) for the more sophisticated equipment, more research and development resources have been applied in that direction.

In the progression from small-scale to high-input farming, the most obvious changes in the characteristics of the equipment are the reduction in human energy expenditure and the increase in information processing required for operation. The
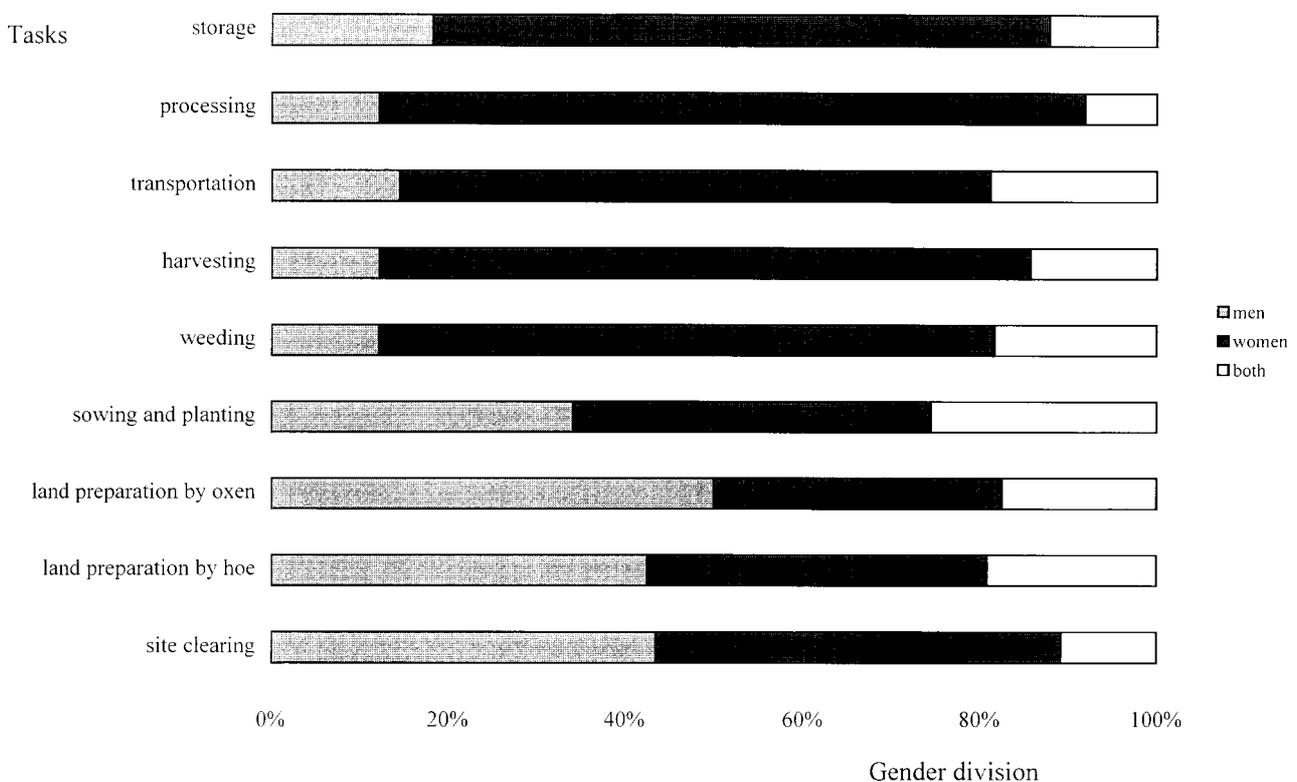
Figure 1. Typical gender division of labour in Africa

information-processing demand of the latest, most sophisticated equipment aimed at the high-input market has now been acknowledged and recent developments incorporate features to reduce the mental workload or to facilitate decision-making. Thus, modern high-specification tractors have automatic transmission and draught devices, controlled by signals from sensors to relieve the driver/operator from constantly having to make adjustments to achieve optimum performance. The use of computers, for both vehicles and fixed plant, from sprayers and combine harvesters to grain dryers and glasshouses, is rapidly increasing to leave the operator free to make and implement decisions at the highest (and most important) level. These applications have generated ergonomics effort aimed at making information technology and software packages more "user-friendly" for farmers.

The intensive raising of livestock and the production of milk are both prominent in commercial farming enterprises. The efficient raising of healthy stock depends on appropriate feeding, environmental control and waste management. Farmers need well designed interfaces to monitor herd (etc.) performance, initiate regular routines as effectively and safely as possible and to maintain an optimum working and living environment for both human and animal occupants of livestock buildings.

4.1. Ergonomics Case Study — Tractor Design

The ubiquitous tractor has been the target of most ergonomics research and development with efforts to alleviate the effects of noise, vibration, overturning, air quality and heat stress (Chisholm *et al.* 1992). Tractor ergonomics presents a complex array of issues: the adopted method of protecting the driver from some of the stressors can introduce additional ones, so careful design is required to achieve the best (and affordable) compromise. For example, enclosing a driver in a strong cabin with a suspended seat offers protection against overturning, noise, whole-body vibration and respiratory hazards but reduces the field of view, attenuates feedback from the machine, inhibits communication and can increase heat stress. Similar considerations apply to more specialist vehicles such as combine harvesters and sprayers.

As a case study, consider the reasons for and consequences of fitting a "Q-cab" (i.e. quiet cab) to an agricultural tractor. To protect the driver's hearing, an approved Q-cab must now be fitted to tractors that, in any operating condition, can generate a noise level of > 90 dBA at the driver's ear. Although there are a few special exemptions, this has led to almost every tractor with a power output > 25 kW (i.e. most tractors) being fitted with a cab. Historically, the cab was an adaptation of the mandatory rollover protection — a frame fitted over the seat to protect the driver in the (not uncommon) event of the tractor overturning. With the frame in place, it became popular to attach panels to protect the driver against bad weather (particularly in Northern Europe). These panels, usually of sheet metal and Perspex, were simply fixed to the frame and had the effect of amplifying the tractor noise to the extent of causing the drivers to suffer hearing loss (initially temporary, but permanent after sustained periods). When this problem was identified and it became clear that tractor drivers did not wish to lose their weather protection, cabs were developed, based on the ROPS (rollover protective structure), but these cabs had to meet the noise exposure at work restrictions. The resulting engineering design had the effects of reducing the

Table 3. Advantages and disadvantages of enclosed cabs

Advantages	Reasons
Less risk of ear damage Less stressful working environment	Reduced exposure to noise
Improved air quality	Protection from dusts and chemicals
Less protective clothing needed Opportunity to fit heater	Protection from cold and wet
Fewer postural problems	Potential to improve work-place layout
Greater safety of personal effects	Secure enclosure
Disadvantages	
Access to workplace impeded	Cab structural members and panels
Access to external controls impeded	Cab structure, panels and windows
Excessive temperatures in sunshine	Glazed enclosure acting as 'greenhouse'
Reduced field vision	Cab structure and panels
Isolation from useful auditory information	Necessary cab attenuation properties

driver's field of view, isolating the driver from other auditory (and sometimes important) auditory signals and, during periods of sunshine, making the driver uncomfortably hot (through the "greenhouse" effect). In other words, the solution to one ergonomic problem created new ergonomics problems. The main advantages and disadvantages are shown in Table 3.

As time has progressed, Q-cab design has improved to eliminate as far as possible the undesirable effects without compromising the benefits and, furthermore, to extend the range of potential benefits afforded by the presence of a cab. These include better vibration isolation than can be provided by a simple suspended seat, selective attenuation to improve the transmission of important auditory information, the fitting of climate control systems, greater proportion of glazed area to improve visibility and better design and location of hand and foot controls. Without such improvements, such as to the thermal environment, drivers would choose to avoid the immediate discomfort of heat stress and expose themselves to the potentially more damaging and insidious effects of noise and respiratory hazards (e.g. dust and chemicals). Vibration on tractors is a specific problem because the natural frequency of the large tires typically fitted to the rear of tractors is ~2–4 Hz. This form of mechanical energy, transmitted through the rigid tractor frame and seat, subjects the driver to relatively high amplitude low frequency vibrations which are particularly damaging to the human body. The latest designs of tractor display many engineering refinements, achieved through considerable research and development effort, such as suspended axles, air conditioning, electronic controls to improve man–machine performance over the long periods that tractors are driven day after day.

Enclosing the driver in a cab and reducing the information available, notably auditory and tactile (less so for visual information on modern tractors), concerning behavior of the machine and its attachments can reduce the perception of danger or hazards, so, despite the improvements described above, the incidence of overturning accidents has remained almost constant. This is likely to be reduced only when the driver can be provided with better, readily assimilated information about the stability of

the tractor, especially on slopes. While this remains an ongoing ergonomics challenge, the fitting of ROPS continues to save lives.

4.2. Livestock

The need to work with animals, often with body weights and strengths exceeding those of a typical human, sets agriculture apart from most other industries. The difficulties of handling and generally managing farm animals, from the shearing of sheep to the milking of cows, create ergonomics problems specific to agriculture. Irrespective of the degree of mechanization or automation, equipment designers must consider not only the human–machine interface but also the human–animal interface and, preferably, the animal–machine interface in pursuing complete system compatibility. Milk production is the principal enterprise for many farmers and the activities involved pose wide-ranging ergonomics problems. These include milking the cows (particularly equipment and workplace design), monitoring their health and productivity (identifying and responding to relevant indicators) and designing livestock housing to minimize human health risks. The design of milking parlors must minimize postural problems, promote safe and efficient work routines and provide an acceptable thermal environment. Satisfying all these demands within the prevailing economic constraints generally involves ergonomics and engineering compromises. Monitoring health and productivity becomes more difficult as herd size increases and can be facilitated by information technology, but care has to be taken that such technology is "user-friendly." Current research and development is moving towards robotic milking, which will reduce problems associated with the workplace but, with greater isolation of the herdsman, will place greater reliance on effective information systems to facilitate the husbandry of individual animals.

The growing concern for animal welfare in some developed countries, especially concerning the transport of animals (often to abattoirs), has often been based on ergonomics principles. The techniques for monitoring and evaluating whole-body vibration and thermal stress have been particularly useful. Techniques for assessing work performance and efficiency (e.g. the use of heart rate, oxygen uptake) have also been applied to draught animals (e.g. oxen, horses, camels) to improve productivity.

5. HEALTH AND SAFETY

There are many health and safety concerns in agriculture, apart from those associated with tractors (and other vehicles) and animals referred to above. These include:

- hand–arm vibration from power tools;
- entering hazardous situations (e.g. oxygen depleted atmospheres in store-places);
- slip and fall injuries associated with working on poor surfaces, ladders, roofs or in trees;
- diseases communicated from animals (zoonoses);
- respiratory diseases from exposure to organic or inorganic dusts, or chemical sprays;
- injuries from inadequate machine guarding; and
- environmental noise (annoyance) from fixed agricultural plant (e.g. grain dryers).

Although ergonomics alone can not deal with all these problems, acceptable solutions will not be realized without significant ergonomics and systems contributions.

POSTCRIPT

In so diverse an industry as agriculture, particularly with such a large informal sector worldwide, it is impossible to cover all the ergonomics issues in a few pages. The purpose of this chapter is to give an indication of the major issues and to emphasize the importance of the systems approach, fundamental to sound ergonomics practice. There is probably no other industry in which interactions with economic constraints, cultural factors and engineering design are so critical to the development of "ergonomics solutions," and in which participative approaches are so crucial to their implementation.

REFERENCES

CHAMBERS, R., PACEY, A. and THRUPP, L.A. (eds), 1989, *Farmer First: Innovation and Agricultural Research* (London: Intermediate Technology).

CHISHOLM, C.J., BOTTOMS, D.J., DWYER, M.J., LINES, J.A. and WHYTE, R.T., 1992, Safety, health and hygiene in agriculture. *Safety Science*, 15, 225–48.

Anthropometry for the Needs of Rehabilitation

E. Nowak

Institute of Industrial Design, świĺtojerska 5/7, 00-236 Warsaw, Poland

1. INTRODUCTION

At present there appears a new direction of ergonomics aimed at the needs of disabled people. This is *rehabilitation ergonomics* (Kumar 1992, Nowak 1992, 1996), roughly defined as an interdisciplinary field of science that aims at adjusting tools, machines, equipment and technologies as well as material work and life environments including objects of daily use and rehabilitation equipment to the psychophysical needs of the disabled. Rehabilitation ergonomics takes part both in the rehabilitation process and in equalizing chances.

Anthropometry plays an important role in rehabilitation ergonomics. It is defined as a set of measuring methods and techniques that allow to investigate exactly the differentiation of the measurements of the man and their changeability during ontogenesis and phylogenesis.

As ergonomics develops and in compliance with its needs anthropometry develops new methods that can be called ergonomic anthropometry. Many of these methods, concerning both classical and ergonomic anthropometries, can be applied to the rehabilitation ergonomics that can be divided into two parts:

- The first part is strictly connected with ergonomics, where anthropometry provides data for designing and shaping work and life environment of the disabled.
- The second part embraces all methods and measuring techniques that assist the rehabilitation process.

2. ANTHROPOMETRIC RESEARCH OF THE DISABLED — DATA FOR DESIGNING

2.1. Somatic Characteristics

Defining the possibilities and necessities of the disabled is a necessary condition for designing the material work and life environment for this population. Data that characterize the somatic structure constitute basic information.

The influence of disabilities on shaping the body structure was studied by many anthropologists (Pheasant 1986, Molenbroek 1987, Nowak 1989, 1996, Das and Kozey 1994, Jarosz 1996). The largest disproportions between the healthy and the disabled population can be found in a group of people with motor dysfunction. This is understandable since dysfunction results from the past or currently developing diseases that lead to joint disturbances of the osseous, ligament and joint, muscular and nervous systems. These disturbances lead to deformities and somatic changes of particular parts of the body. This then affects the final shape and dimensions of the body and its motorics.

Other factors that restrain the development and growth of the body include restriction of motion activities, neglected nursing, unproper or lack of rehabilitation, as well as stresses connected with pains, frequent stays in hospitals and rehabilitation centers and which accompany the pathological process. Descriptions of investigations of people with the lower

extremities dysfunction are usually found in the literature. They are usually wheelchair users. It should be realized, however, that this group embraces people with various degrees of motor-efficiency limitations. This depends not only on a type and stage of a disease, but also on time of its appearance. Therefore, researchers dealing with this problem face great difficulties in selecting subjects, and in describing results scientifically. This may be the cause of the small number of studies undertaken in this field. This particularly concerns studies where the results are to provide data for designing.

Not only anthropometric measurements of the disabled are important for the needs of design, but also differences between the disabled and healthy population. The majority of authors indicate that the body structure of disabled men and women differs significantly from the able-bodied population (Nowak 1989, Das and Kozey 1994, Jarosz 1996). This problem was illustrated in the example of data on the Polish population, girls and boys (Nowak 1989, 1996). It appears that these measurements show smaller values for the disabled than for healthy people (Figure 1).

2.2. Elements of Anthropometry in Clothing Design

Somatic characteristics are important for clothing design and the fact that its values are bigger in case of the disabled is of importance for designers. A person suffering from disability or permanent illness should get such clothing that makes his/her life easier, is functional, meets his/her emotional needs and, at the same time, is adjusted to the limitations caused by a disability. People with motion dysfunction have the most problems with clothes. Clothing designed for this group of people should assure appropriate physiohygienic properties, heat comfort, as well as ease of manipulation. Concepts of model-constructions solutions should take into account not only the usefulness of products, but also the economic effect. Designs must be simple in
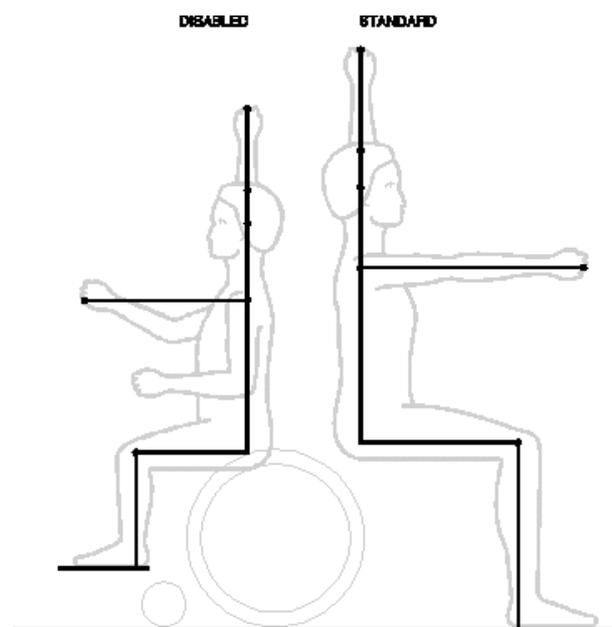
Figure 1. Comparison of body dimensions of the disabled and the able-bodied.

construction and easy to manufacture. The form of clothing is to integrate the group of the disabled with society, to conceal a disability, and to give the feeling of satisfaction at possessing clothes appropriate to needs and expectations. In addition to the body dimensions, determination of motor abilities is of vital importance. Basing on ergonomic investigations, assumptions for clothing design are determined. For example, clothing for the wheelchair user must fulfill (among others) the following requirements: (1) appropriate clearance allowing the hands to move easily up and forward and appropriately larger shoulder breadth. This is ensured by the special construction of sleeve and armpit cut. (2) Increased transverse dimensions related to the enlarged, muscular chest. (3) Widened sleeve finish related to the considerably enlarged biceps and triceps. (4) Adjustment of the bottom part of clothing length and cut (in case of a blouse, shirt, and vest) to the sitting position of its user. This is attained through the removal of excess fabric in front.

People with Down's syndrome (DS) require clothes of different construction (Harwood 1997). This group of people is always offered oversized garments to fit their large hips and thighs; as a result all of the features and proportions do not lie in their correct locations. Some of the common problems can be summarized: (1) when the waist of a garment fits, the bust/chest does not; (2) trousers that have the correct waist measurement are too tight on the hips and thighs; (3) dresses and coats that fit the shoulders are too long, i.e. sleeves, trouser legs and coats are too long; as a result many garment features are at the wrong level of the body; (4) bust contour features are below the bust line, trouser pockets are too low for inserting hands; (5) the full crotch measurement is never long enough to accommodate the over-developed abdomen of the wearers; (6) the overall effect of using standard size garments and adapting them to a reasonable fit is obviously unsatisfactory in terms of both their aesthetic and functional characteristics; maternity garments offer a closer approximation to the size requirements and are often used by DS females. The lack of provision is disappointing particularly when block patterns, based on DS body dimensions, are available from which garments could be designed for industrial production.

Many investigators and designers of clothing for the elderly and the disabled underline the fact that in the process of clothing design they are governed by the idea that clothing, in addition to its functional characteristics, should give its user physical comfort. The disabled want to live and work among healthy people. Therefore, they have to accept the clothes they wear. Clothing should not deform the shape of the body, but just the opposite, it should cover anatomical defects.

2.3. Workspace Measurements

Essential characteristics exerting an influence on workspace shaping are functional characteristics of the upper extremities, i.e. reaches. Values observed in these characteristics are significantly lower in persons with the lower extremities dysfunction, although their upper extremities are qualified as 'efficient'. Lower values of reaches result not only from lower values of the arm and forearm length, but also from limitations in shoulder and elbow joints. In connection with the above, the disabled have difficulty in performing the movements of abduction and extension (Nowak 1989). Many authors found that in case of persons suffering from rheumatoid arthritis

workspace of the upper extremity is 7–10% lower if a shoulder joint is constrained and 25–33% lower when there is the constraint of the elbow joint movements. It is obvious that disorders of these two joints increase the limitation of the upper extremity movements, and thus the efficient workspace is significantly reduced. This is confirmed by investigations conducted by Nowak (1989) and Jarosz (1996). It should be pointed out that particular values of reaches of the disabled refer to the straight position of the body. Thus, they can be increased through the trunk movements forward and lateral.

Nowak (1989), basing on the Das and Grady's method (1983) developed a simple method of defining workspace for arms. This space was determined for disabled young people with the dysfunction of lower extremities, using the wheelchair.

Figures 2a and b show the graphic way to determine this space in the sagittal plane (maximum sagittal reach, MSR) and in the transverse plane (maximum transverse reach, MTR). They show the difference for both reaches between disabled and healthy young people. Differences in maximum reach measurements were
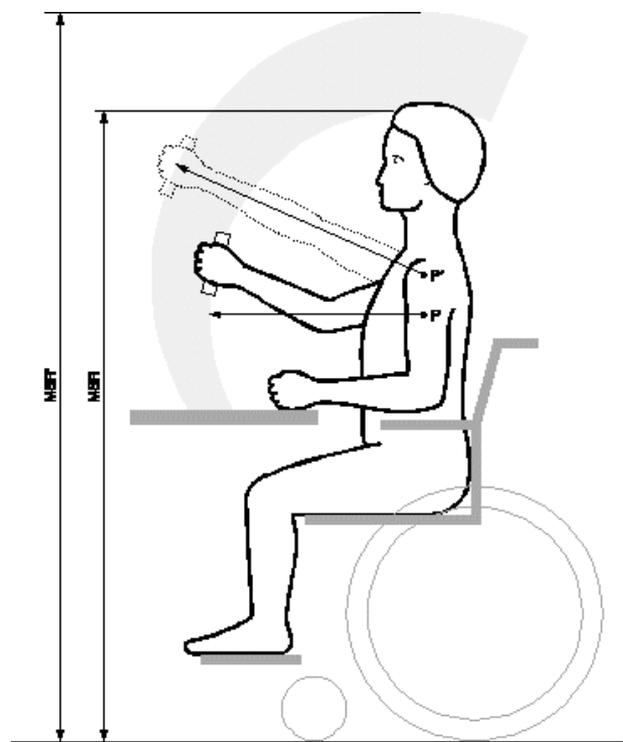
Figure 2a. Maximum sagittal reach for the all-Polish population (MSR') and the disabled population (MSR).
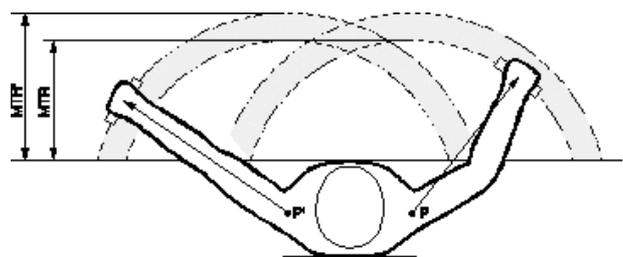
Figure 2b. Maximum transverse reach for the all-Polish population (MTR') and the disabled population (MTR).

significant and amounted for the 5th percentile to 300 mm. Results of the study were used to ergonomic analysis at school and to designing school workshops, laboratories and rehabilitation centers. This method can be recommended for workspace design for the disabled. It is simple and easy to use. Using only five anthropometric characteristics one can obtain the graphic representation of workspace of any population or an individual person.

According to Pheasant (1986), a wheelchair user (whose upper extremities are unimpaired) can reach a zone from ~600 to 1500 mm in a sideways approach, but considerably less 'head on'. It may well be that the location of fittings within this limited zone will prove entirely acceptable for the ambulant users of the building, but in case of working surface heights no such easy compromise is possible.

As the review of investigations has proved, the body structure of disabled men and women differs considerably from that of the healthy population. Anthropometry, providing data concerning the body structure of disabled people, makes it possible to adjust designs of products and spatial structures to the possibilities and predispositions of this group of users.

3. ANTHROPOMETRY FOR THE NEEDS OF REHABILITATION

Anthropometric data and methods can be used for the evaluation of the disabled children development and the rehabilitation process of the impaired organ or function.

To check and assess the process of children's physical development biological reference systems, the so-called standards, are developed. The values of standards are determined on the basis of the population investigations, embracing people qualified as 'normally' efficient. Standards are prepared with the use of mathematical and statistical methods.

For the needs of pediatrics two border regions of developmental standards are distinguished:

- The so-called narrow developmental standard, comprising values between the 25th and the 75th percentile and covering 50% of all observations.
- The so-called wide standard, comprising values between the 10th and the 90th percentile and embracing 80% of the observations (children included within these limits need observation and concern of a pediatrician).

Developmental standards can also be used to assess the physical development of disabled children. Rehabilitation activities include, among others, the monitoring of somatic development, aimed at the assessment of the child's biological potential. The development of a child depends not only on his genetic potential, but also on the environmental conditions.

As the result of disadvantageous events, which can take place during the ontogenesis, injuries inducing disturbances in the nervous system activity, growth or functioning of the whole body or its part can occur. Biological reactions of the body depend first on the releasing factors, that is, diseases and personal injuries. These factors have a direct effect on the development of the body, but reactions of the body are also indirectly influenced by other factors, such as the type of medicine and the schedule of its application, as well as undertaking of rehabilitation. Important factors, which influence the process of development, are the time of disorder occurring and the kind of factor that caused the disorder. Approaching the assessment, one should analyze and determine possible ways of the child's physical development. Developmental standards can be helpful in a single assessment of a child, as well as in continuous investigations; they can also be used in determining the differences between the development of the body structure characteristics of the children with deviation and this of the healthy children of the same age. The assessment is usually done with the use of the percentile nets representing the height and body mass characteristics. In a single assessment of the child's development one determines the position of the child on the percentile net for the above characteristics and the channel in which this point is located.

Based on the values of the two characteristics one can determine normality or disorders in the body proportions, for example, the child can have the big body mass (it is in the channel between the 97th and 90th percentile) and low height (between the 10th and the 25th percentile). If one tests the child regularly, the individual curve of their development can be obtained.

Children with motor dysfunction are shorter, they have very narrow hips and laterally flattened chest. They are similar to healthy children in that they have the same shoulder dimensions. They have a low body mass, and this deficiency increases with age.

Anthropometry can be particularly useful in diagnosing and assessing the motor efficiency of the human body. Rheumatic diseases and mechanical injuries result in pathological changes in joints, ligaments, tendons, muscles and lead to considerably restricted movement ranges. Thus, the assessment of motor efficiency of the affected joints is essential for monitoring the rehabilitation processes.

The simplest way of assessing motor efficiency is by comparing a restricted motion range of an affected joint with its initial range of motion, that is before injury or disease. Unfortunately, after injury or disease has already prevailed, it is impossible to determine what was the initial motion range in the healthy patient. It may hardly be expected that a rehabilitated patient had his initial motion ranges measured just before the injury.

Thus, motor efficiency of a rehabilitated patient can only be assessed based upon data of the healthy population. A method based upon this kind of data and allowing the quantitative assessment of rehabilitation progress was developed by Nowak (1992).

Based on the investigation of the ranges of motion of arm, leg, hand, foot and head standards for the healthy adult population were developed (Nowak 1992). These standards were developed in three motion classes: wide (W), average (A) and small (S). The maximum and the minimum movement ranges were developed for particular classes.

Three age categories are subordinated to the classes of movement shown. The youngest persons, aged 18–30, belong to class W, characterized by the maximum values. Class A includes adults aged 31–40, and class S includes the eldest subjects, aged 41–65, whose movement range in particular joints has the smallest values.

Knowing the standard value of particular age classes (f) and the values of movement ranges after the rehabilitation process (f') one can calculate the absolute decrease (Ar) and a percentage decrease (Pr) in the movement range. Using a four-grade scale one can calculate a degree of efficiency.
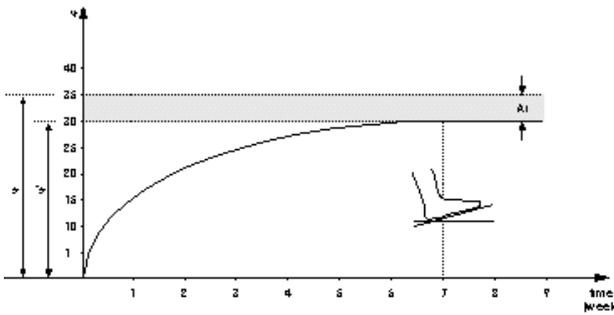
Figure 3. Example of the ankle joint rehabilitation.

- Very good: decrease up to 20% — efficiency from 80%.
- Good: decrease up to 40% — efficiency from 60%.
- Poor: decrease up to 60% — efficiency from 40%.
- Very poor: decrease > 60% — efficiency < 40%.

The application of the method above is presented on a graph (Figure 3).

REFERENCES

DAS, B. and GRADY, M., 1983, Industrial work place layout design. An application of engineering anthropometry. *Ergonomics*, 26, 433–47.

DAS, B. and KOZEY, J., 1994, Structural anthropometry for wheelchair mobile adults. In *Toronto: 12th Triennial Congress of IEA (Rehabilitation Ergonomics) (Human Factors Association of Canada)*, 3, 63–5

HARWOOD, R.J., 1997, The ergonomics of clothing for the disabled and elderly. In *IEA'97 Proceedings of the 13th Triennial Congress of the International Ergonomics Association (Tampere, Finland)*, 5, 430–2.

JAROSZ, E., 1996, Determination of the workspace of wheelchair users. *International Journal of Industrial Ergonomics*, 17, 123-33.

KUMAR, S., 1992, Rehabilitation: an ergonomic dimension. *International Journal of Industrial Ergonomics*, 9, 97-108.

MOLENBROEK, J.F.M., 1987, Anthropometry of elderly people in The Netherlands. *Research and Application*, 18, 187-99.

NOWAK, E., 1989, Workspace for disabled people. *Ergonomics*, 9, 1077-88.

NOWAK, E., 1992, Practical application of anthropometric research in rehabilitation. *International Journal of Industrial Ergonomics*, 9, 109-15.

NOWAK, E., 1996, The role of anthropometry in design of work and life environments of the disabled population. *International Journal of Industrial Ergonomics*, 17, 113-21.

PHEASANT, S., 1986, *Bodyspace: Anthropometry, Ergonomics and Design* (London: Taylor & Francis).

Assessing the Risk of Upper Limb Disorders

E. N. Corlett

Institute for Occupational Ergonomics, University of Nottingham, Nottingham, NG7 2RD, UK

1. INTRODUCTION

The preliminary assessment of a problem to understand the key areas where attention should be focussed is a frequent activity for the practicing ergonomist. It is not satisfactory, at least for the client, to state a view based solely on a judgement unless, of course, the problem is clear and can be adduced from the visible evidence when demonstrated to the client by the ergonomist. More usually both ergonomist and client are more confident if some measured data can be given that can identify the dimensions of the problem, and show how the attack on the problem will counteract the effects arising from the causes identified by that evidence.

Early attempts to achieve this situation without using a full-scale research program depended on checklists. In spite of considerable refinement these have major flaws. Among them are that they analyze the problem within the structure of the list, thus perhaps missing any new features present in the problem; they can be so exhaustive that they are counter-productive because of the time involved in their use and analysis, or they just serve as a checklist to assure the user that no item has been missed.

Undoubtedly valuable as this last point is, it is not enough. A skilled ergonomist's use of feature checklists (Sullivan and Corlett 1998), supported by other appropriate methods, can reduce these problems and serve to estimate the influence of the major factors affecting a situation. This technique, first illustrated by Shackel *et al.* (1969), sets out all the influencing factors in a given situation and requires the subject to assess each on the basis of its level of adequacy or unsatisfactoriness. This results in a profile of the contribution of each factor to the overall condition of the situation under study, from which a decision can be taken on which point of inadequacy should be attacked first.

2. UPPER LIMB DISORDERS

When it comes to assessing the situation surrounding the occurrence of work-related upper limb disorders (WRULD), the investigator is faced with the same problem, where to start. The understanding of the causation of WRULD has increased considerably over recent years and it is too simple to assume that it is purely a physical cause. Nonetheless, the physical situation is frequently the major causative factor and rarely would an attack on the presence of WRULD be successful without modifications to the workplace or work routine.

Faced with the situation where the causes of WRULD in a major industry had to be tackled rapidly, and in an economic manner it was necessary to develop a systematic procedure that would cover the major known areas contributing to such problems and identify the causes which, when attacked, would most effectively minimize or eliminate them. The context of the work was the requirement of the UK Health and Safety Executive, responding to a Directive of the European Union, that companies

should be proactive in identifying unsafe situations and show that they had undertaken safety assessments should injuries have taken place. Hence, the investigation had to include the development of a company system to monitor the ongoing situation, together with appropriate training courses for company personnel.

Within this whole investigation, which covered work organization, work attitudes and company recording systems (McAtamney and Corlett 1992) a prime need was to assess the work places and activities of operators, in factory and office, so that work and engineering changes could be carried out to reduce any existing hazards. Furthermore, this had to be done using existing personnel, primarily from the company's methods engineering group.

The approach to this was to devise a rapid survey method using the concepts underlying the OWAS procedure (Karhu *et al.* 1977). In this method, for the analysis of work postures, each major body part is assessed for its displacement and a number is allocated according to the increase in the departure of that body part from a defined neutral position. From combined values for the various body parts, achieved by the use of tables, ultimate values are found that indicate the severity of the posture and the level of urgency for change.

To utilize the idea for the assessment of hazards relevant to WRULD causation it was necessary to identify from the literature those postures of the upper limbs, and the contributory postures of the whole body, which had been identified as contributing to the disorder. Diagrams were then developed, on the lines of those used for OWAS, to provide the investigator with guidance when making an assessment (Figure 1). The assessment was done for the upper limb and for the whole body postures separately, groups A and B respectively in Figure 1. The results of the estimates of postural angles, after adjustment for force exertions and muscle activities, were combined to give a final score from which the likely level of hazard and the urgency of action could be obtained. The structure of the method also permitted the user to see what must be done to reduce the hazard.

3. RULA

RULA stands for rapid upper limb assessment and it is emphasized that it is an assessment method. There will be occasions when the results merely point the user in the direction of the required changes but where further analysis is needed before a full understanding of the problem is achieved.

The first part of the procedure is to decide where, in the work activities, are the tasks that give the most extreme angles of the upper limbs and where the forces occur which may aggravate the effect of the angles. These are the points where a RULA assessment should be done. Assistance in this task may come from the use of a body part discomfort diagram (Corlett and Bishop 1976) which will demonstrate the growth of discomfort over the working period in different parts of the body, and highlight possible sources of the most serious muscle loadings.

Having extracted those tasks in which the suspect postures and loads occur, the analyst then uses the ranges of the angles shown in Figure 1 to obtain the numbers to go into the tables of part A of Figure 2. Each of the boxes for upper arm, lower arm, wrist and wrist twist are completed in part A, for the upper limb. The muscle and force scores should then be estimated, using the

Diagram 1. RULA Group A body parts

UPPER ARMS

ADD 1. if shoulder is raised
ADD 1. if upper arm is abducted
SUBTRACT 1 if leaning or supporting the weight of the arm

LOWER ARMS

ADD 1. if working across the midline of the body or out to the side

WRIST

ADD 1 if wrist is bent away from the midline

WRIST TWIST

1. Mainly in mid-range of twist

2. At or near the end of twisting range
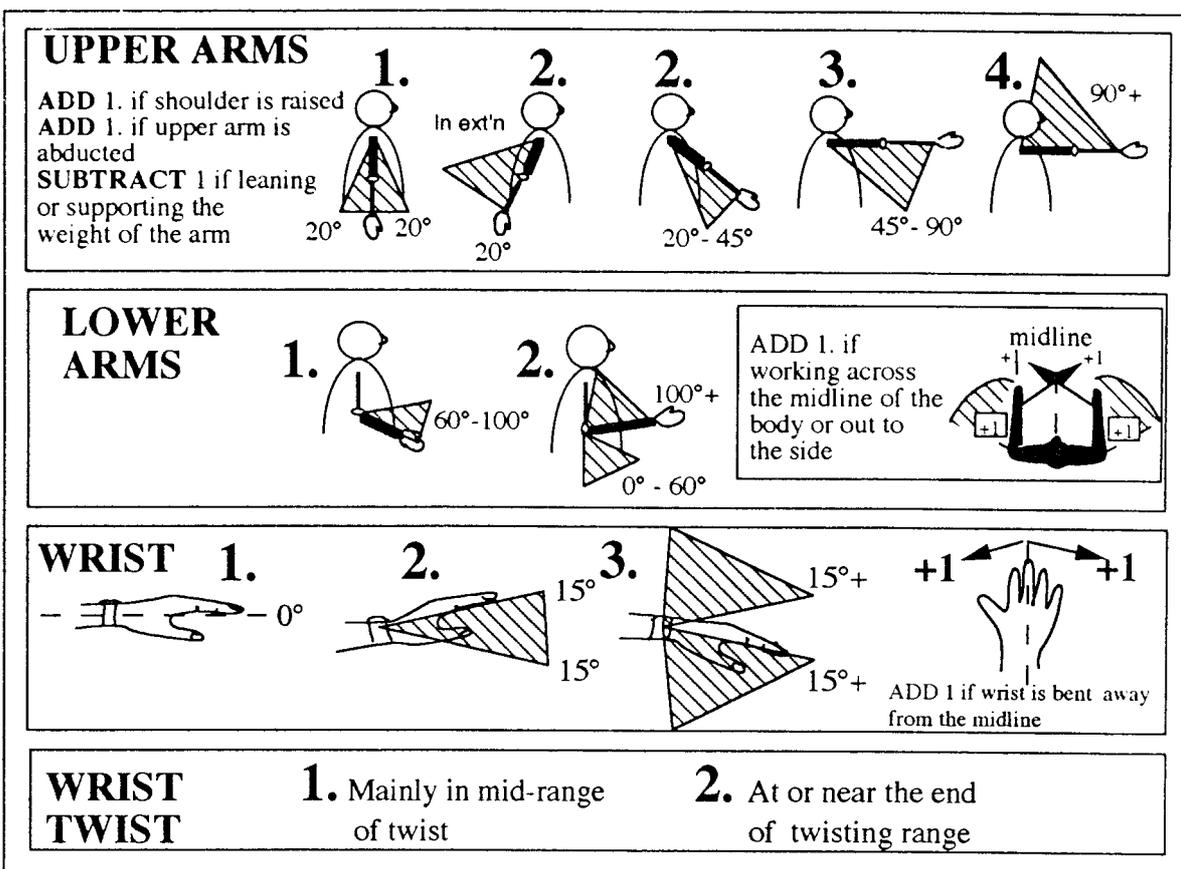
GROUP A

Diagram 2. RULA Group B bodyparts

NECK

ADD 1 if the neck is twisting

ADD 1 if neck is side-bending

TRUNK

ADD 1 if trunk is twisting

ADD 1 if trunk is side-bending

1. also if trunk is well supported while seated

4. also if unsupported in extension

LEGS **1.** if legs and feet are well supported and in an evenly balanced posture **2.** if not
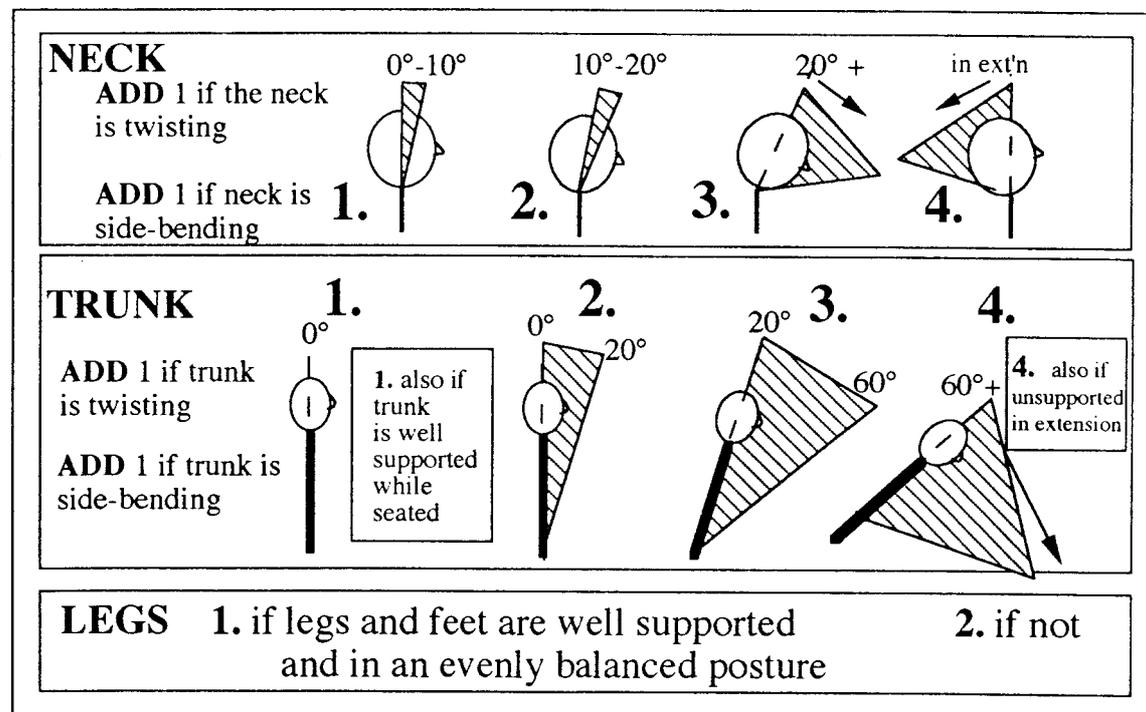
GROUP B

Figure 1. Diagrams to guide the assessment of joint angles and loadings, providing basic information for entering into the left-hand column of the score sheet of Figure 2.

values from the muscle use and force or load score tables of Figure 3, and entered in the boxes for the upper limb.

The same procedure should then be used for the whole body, but using the values from the Group B diagrams and entering them
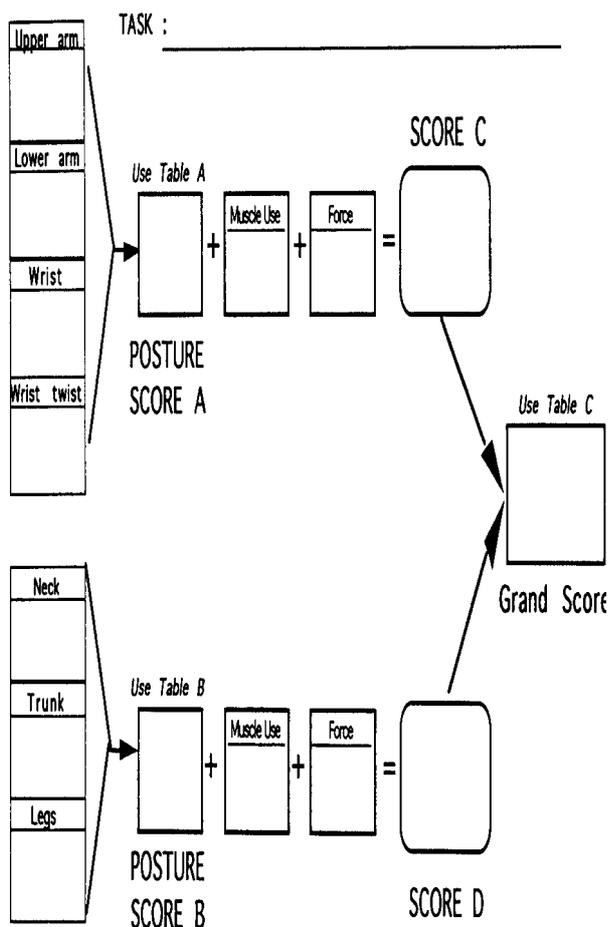
Figure 2. RULA score table in which the values obtained by study of the task are compiled for later evaluation concerning their potential level of risk.

MUSCLE USE SCORE

Give a score of 1 if the posture is;

mainly static, e.g. held for longer than 1 minute
repeated more than 4 times/minute

FORCES OR LOAD SCORE

0.	1.	2.	3.
●No resistance or less than 2kg intermittent load or force	● 2-10kg intermittent load or force	● 2-10kg static load ● 2-10kg repeated load or force ● 10kg or more intermittent load or force	●10kg or more static load ●10kg or more repeated loads or forces. ●Shock or forces with a rapid buildup.

Figure 3. Values of factors that modify the effects of the postures identified and entered into Figure 2. Values from these tables should be entered into the relevant spaces in Figure 2.

TABLE A   Upper Limb Posture Score

UPPER ARM	LOWER ARM	WRIST POSTURE SCORE 1		WRIST POSTURE SCORE 2		WRIST POSTURE SCORE 3		WRIST POSTURE SCORE 4	
		TWIST 1	TWIST 2	TWIST 1	TWIST 2	TWIST 1	TWIST 2	TWIST 1	TWIST 2
1	1	1	2	2	2	2	3	3	3
	2	2	2	2	2	3	3	3	3
	3	2	3	3	3	3	3	4	4
2	1	2	3	3	3	3	4	4	4
	2	3	3	3	3	3	4	4	4
	3	3	4	4	4	4	4	5	5
3	1	3	3	4	4	4	4	5	5
	2	3	4	4	4	4	4	5	5
	3	4	4	4	4	4	5	5	5
4	1	4	4	4	4	4	5	5	5
	2	4	4	4	4	4	5	5	5
	3	4	4	4	5	5	5	6	6
5	1	5	5	5	5	5	6	6	7
	2	5	6	6	6	6	6	7	7
	3	6	6	6	7	7	7	7	8
6	1	7	7	7	7	7	8	8	9
	2	8	8	8	8	8	9	9	9
	3	9	9	9	9	9	9	9	9

TABLE B   Neck, Trunk, Legs Posture Score

NECK POSTURE SCORE	TRUNK POSTURE SCORE 1		2		3		4		5		6	
	LEGS 1	2	LEGS 1	2	LEGS 1	2	LEGS 1	2	LEGS 1	2	LEGS 1	2
1	1	3	2	3	3	4	5	5	6	6	7	7
2	2	3	2	3	4	5	5	5	6	7	7	7
3	3	3	3	4	4	5	5	6	6	7	7	7
4	5	5	5	6	6	7	7	7	7	7	8	8
5	7	7	7	7	7	8	8	8	8	8	8	8
6	8	8	8	8	8	8	8	9	9	9	9	9

Figure 4. Tables that combine the values identified from the observations into a single score, for later combination into an estimate of the risk. Tables A and B give scores that should be entered into the 'score C' and 'score D' boxes respectively in Figure 2.

in part B of Figure 2. Again the force and load scores should be added, taking the values from the same tables on Figure 3 as before.

The necessary data has now been collected from the workplace. The analyst should now refer to table A of Figure 4. Taking the values for upper arm, lower arm and wrist posture the table should be entered, selecting one of the two columns given under each wrist score according to the value of the wrist twist score. The value thus identified in the body of the table should be entered in the space on the score sheet of Figure 2 marked posture score A. The two values for muscle use and force can now be added to posture score A, giving score C.

Repeat this procedure for the whole body score, but using table B of Figure 4, to give posture score B. Again add this to the muscle use and force scores to arrive at the number for score D, to be entered in the appropriate space in the score sheet of Figure 2.

To combine the two scores so as to arrive at a score representing the hazard value of the posture under study, combine scores C and D by using table C of Figure 5, giving the grand score. This value, from the body of the table, represents the probable hazard, and should be assessed by the recommended action levels given below table C at the bottom of Figure 5. Note

SCORE D (NECK, TRUNK, LEGS)

	1	2	3	4	5	6	7+
1	1	2	3	3	4	5	5
2	2	2	3	4	4	5	5
3	3	3	3	4	4	5	6
4	3	3	3	4	5	6	6
5	4	4	4	5	6	7	7
6	4	4	5	6	6	7	7
7	5	5	6	6	7	7	7
8	5	5	6	7	7	7	7

SCORE C (UPPER LIMB)

<u>TABLE C</u> Grand Score Table

ACTION LEVEL 1 A score of one or two indicates that posture is acceptable if it is not maintained or repeated for long periods.

ACTION LEVEL 2 A score of three or four indicates further investigation is needed and changes may be required.

ACTION LEVEL 3 A score of five or six indicates investigation and changes are required soon.

ACTION LEVEL 4 A score of seven or more indicates investigation and changes are required immediately.

Figure 5. Here, table C gives a final score to evaluate the potentially hazardous nature of the posture under study and its muscular activity. Scores C and D from the score sheet are combined using this table and the grand score found. Action Levels listed at the foot of the figure indicate the appropriate action to take as a result of the score found from this table.

that table C of Figure 5 is graded from light to dark to give some indication of increasing hazard with the darker area

4. USING THE RESULTS

Inspection of the procedure will show that the more serious is the posture, the larger is the number that is registered in the analysis. This gives a clue to what may be done to improve the situation, since any changes which will reduce the magnitude of the number will probably result in improvements in the postures and consequent reduction of the hazard. However, a change of one or two in one box is unlikely to change the hazard score significantly, major changes to body postures and forces are usually required.

If it is considered that both the upper limbs should be analyzed, the left and right upper limbs should be scored separately, and the diagram of Figure 6 shows how an economical score page may be laid out, with the left and right scores on the same diagram. When defining the grand score, there will be two, one for each of the two upper limbs. Decisions concerning what action to take will then depend on a study of the particular situation, bringing into
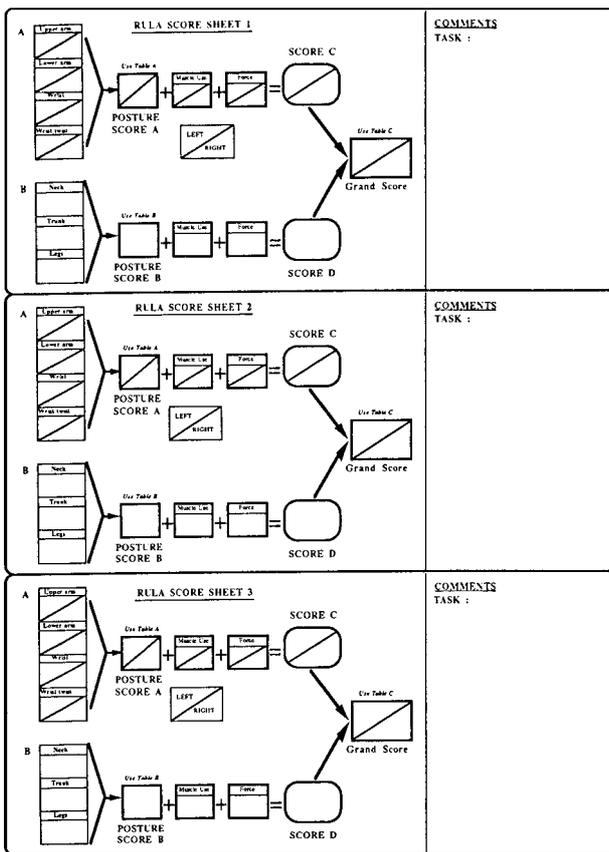
Figure 6. Score sheets may be combined as shown here. Diagonal lines within the boxes for the upper limb allow both left- and right-hand values to be entered on the same figure.

consideration the actual activities being undertaken, but the same point given above will apply, lower numbers arising from changes will most probably lead to lower risks to the operator.

After changes have been made RULA can again compare results with the state before the intervention. Confirmation of the effectiveness of the intervention, at a later stage, can be obtained from reports from workers or by using other measures such as the body part discomfort diagrams mentioned earlier. The use of such measures gives an independent validation of the decisions arising from the RULA analysis and confidence in the use of the method.

When new workplaces are being proposed a mock-up of the workplace can give opportunities to simulate the work activities and RULA can be used before any expensive decisions are made. Thus, the common situation, of finding that problems exist after equipment has been installed, can be avoided.

Of course, the use of RULA does not exclude the need to

consider other basic ergonomic points. What kind of grip is being used and what reaches are needed in the workplace? Are handles too big, hard-edged, too small or in difficult places? Are the hands or fingers doing pounding, long periods of gripping, doing stretching or twisting actions? Are the work postures aggravated because the worker cannot see easily, or reach or exert the necessary forces in advantageous postures? All such points should be noted and taken into account both during the study and when the remedial actions are decided upon.

5. CONCLUSIONS

Users of RULA will find that it can be easily learned, although some revision will be valuable at intervals to ensure that all users are consistent. It can be added to a normal methods engineering tool-kit and undertaken as a regular part of the methods engineer's activities. Experience shows that most findings of existing hazards can be reduced or eliminated immediately and at no expense. Some 20% may require some delay but very few problems require major costs for material or time.

The method has been taught to workers and they have subsequently made improvements to their own workplaces. The recognition by them that this can be done, and the increased understanding it generates about the possible causations of WRULD give an extra dimension to the safety program of the company.

Finally, it should be stressed that the implementation of a comprehensive program with recording of industrial injuries and assessments of management–worker relationships can add to the better reduction of WRULD. Although biomechanical factors may be the most readily identified of all the causative factors the psychosocial problems are real, and important. Hence, the need for the comprehensive program under which RULA was developed, described in McAtamney and Corlett (1992).

REFERENCES

CORLETT, E.N. and BISHOP, R.P., 1976, A technique for assessing postural discomfort. *Ergonomics*, 19, 175–82.

KARHU, O., KANSI, P. and KUORINKA, I., 1977, Correcting working postures in industry: a practical method for analysis. *Applied Ergonomics*, 8, 199–201.

MCATAMNEY, L. and CORLETT, E.N., 1992, *Reducing the Risks of Work Related Upper Limb Disorders: A Guide and Methods* (Nottingham: Institute for Occupational Ergonomics, University of Nottingham).

SHACKEL, B., CHIDSEY, K.D. and SHIPLEY, P., 1969, The assessment of chair comfort. *Ergonomics*, 12, 269–306.

SULLIVAN, A. and CORLETT, E.N., 1998, Transferring ergonomics: a methodology for self-diagnosis of workplace problems. *Occupational Ergonomics*, 1, 223–7.

Assessment of Combined Occupational Low Back Disorder Risk Factors

F. A. Fathallah

Liberty Mutual Research Center for Safety and Health, 71 Frankland Road, Hopkinton, MA 01748, USA

1. INTRODUCTION

The proper quantification of risk factors in occupational settings is an essential part of controlling occupational injuries and illnesses. Commonly, the first step in ergonomic interventions is to assess the problem at hand by quantifying the "known" potential risk factors identified through prior epidemiological evidence and/or through ergonomic guidelines and practices. For example, there is epidemiological evidence that an increase in the weight beyond a certain level in manual materials handling (MMH) jobs may lead to an increase in the likelihood of a low back-related disorder. An ergonomist evaluating an MMH job with historically high incidence of low back injuries would consider the weight of objects that workers handle as a primary factor for the observed high incidence of back injuries. In many instances this approach may prove successful. However, there are instances where simply reducing the weight of the object may not be sufficient substantially to reduce the injury experience of a given job, or may not be a feasible approach. This could be due to the fact that there are factors that may be interacting with the object weight. It is the simultaneous or combined effects of the object weight with other risk factors that may better explain the link between the job exposures and the injury outcome. When such an interactive relationship is ignored or poorly quantified, intervention measures that rely solely on reducing the weight may prove to be ineffective.

The purpose of this article is to highlight the importance of the simultaneous or combined quantification of risk factors in occupational injuries with special emphasis on occupational low back disorders (LBD) risk factors.

2. GENERAL GUIDELINES FOR QUANTIFYING COMBINED RISK FACTORS

The approach of quantifying combined risk factors attempts to answer the following two questions: Did a given risk factor simultaneously occur with other risk factors?; and What are the frequency and magnitude of these observed combined, or simultaneously occurring, risk factors? Challenges to answering these questions lie partially in the level of risk factors' measurement type. Within the context of this article, measurement type may be classified into two main classes: direct and indirect measurements. In general, direct (or semi-direct) measurement would indicate the measure of a risk factor via an instrument or an apparatus (e.g. an electrogoniometer to capture posture). On the other hand, indirect measurement usually entails observation workers and/or the use of questionnaires/self reports (e.g. asking a worker to estimate push force when pushing a cart). When assessing risk factors in occupational settings, it is desirable to obtain direct (continuous) measurement of these factors since the approach provides the most comprehensive level of information. However, several issues may limit the use of direct measurement, including cost, lack of technology and feasibility/practicality. Hence, indirect measurement has been alternatively used to overcome some of these limitations. In assessing combined risk factors, a hybrid of direct and direct measurements may be required, depending on the financial and technological limitations as well as practical issues. The following sections will discuss some examples of how hybrid (direct/indirect) measurements could be assessed within the context of combined risk factors related to occupational LBD.

3. COMBINED RISK FACTORS IN OCCUPATIONAL LOW BACK DISORDERS

LBD are considered the most prevalent musculoskeletal disorders in occupational settings. Lifting, in particular, has been implicated most often as an occupationally related low back disorder risk. A clear understanding about the underlying mechanism(s) that lead to back disorders would provide a means for minimizing LBD in industry. However, the current state of knowledge about such mechanisms is still incomplete. Thus, quantifying risk factors associated with occupationally related LBD is an important step towards a better understanding of the etiology or cause(s) of such disorders.

The spine (especially the lumbar region) undergoes complex motion patterns when performing dynamic lifting tasks in occupational settings. The sources of these complex motions can be attributed to both the inherit complexity of the spinal structure and its supporting elements (muscles, ligaments, etc.), and to task demands (weight of object, origin and destination, etc.). These complex spinal motions generate complex loading patterns experienced by the spinal elements (e.g. combined lateral shear and compressive loading). There is significant epidemiological and biomechanical evidence that implicates combined motions and loading as important risk factors for LBD. Unfortunately, few studies have quantified combined or simultaneously occurring motions and loading (in conjunction with other workplace factors) which could help in reducing the risks of occupational LBD.

Figure 1 operationally defines combined motions and demonstrates the potential significance of these motions in identifying high-risk situations. The captured continuous motion profiles were obtained through the use of a triaxial electrogoniometer, the Lumbar Motion Monitor (LMM), developed at the Ohio State University Biodynamics Laboratory (Marras *et al.* 1992). Figure 1 shows trunk motion components obtained from actual task cycles of two typical MMH jobs; one job had no history of LBD (low risk), and the other had several (> 12 per 200 000 man-hours) recorded back disorders (high risk). For a sample pair of jobs, the figure depicts the temporal occurrence of simultaneous magnitudes of trunk lateral and twisting velocities. If one considers the statistical profiles of each motion variable as temporally *independent*, the parameters for both jobs are rather similar. For example, the magnitude of maximum lateral velocity attained for the "low-risk" job was very close to that of the "high-risk" job. However, the *timing* of the maximum lateral velocity value for the high-risk job occurred simultaneously with the maximum twisting velocity. Whereas,
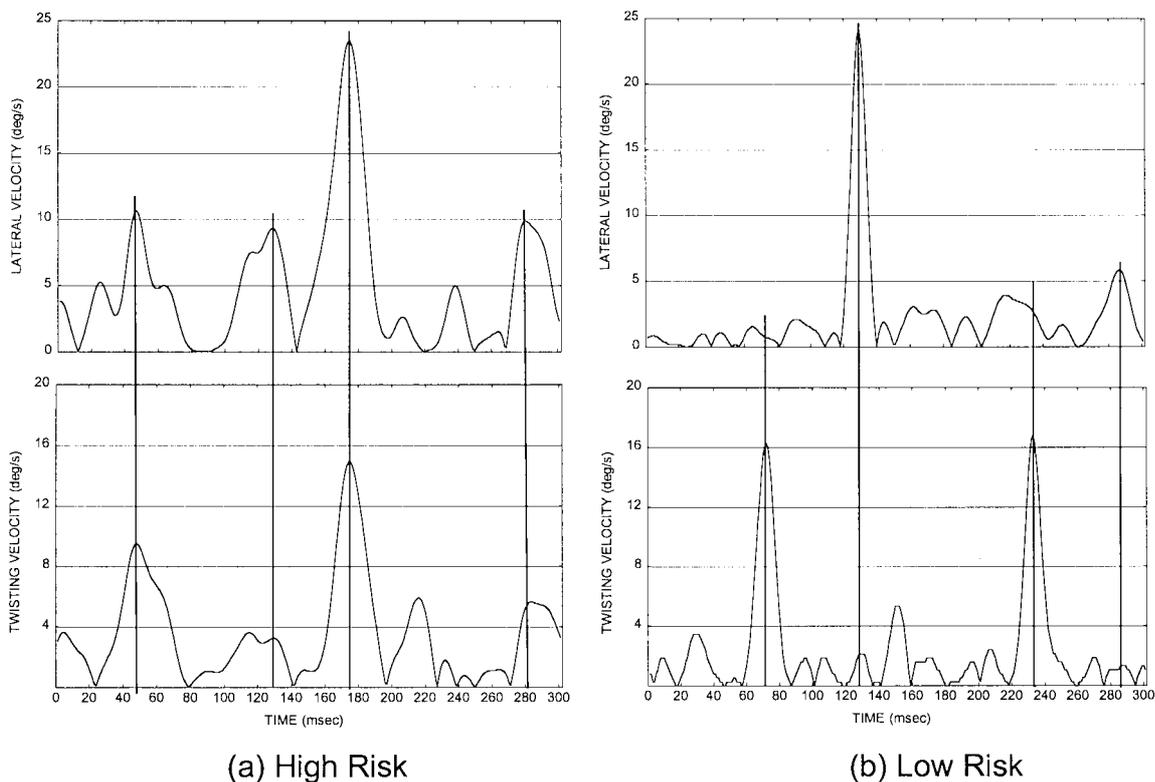
Figure 1. Simultaneous lateral and twisting velocities of actual MMH high-risk (a) and low-risk (b) jobs.

for the low-risk job, the maximum lateral velocity occurred at a distinctively different instant in time than the occurrence of the maximum twisting velocity. Similar observations can be made for submaximal values under which simultaneous high magnitudes of both velocities were occurring in the high-risk job.

3.1. Assessing Combined LBD Risk Factors

The preceding example emphasized the importance of quantifying the simultaneous occurrence of high velocity magnitudes in multiple planes of the trunk. In past studies involving the LMM in industrial settings (e.g. Marras *et al.* 1993), there were five risk factors identified as being related to the risk of occupational LBD. Two of these factors are task related, namely, the load moment (load weight times distance from the base of the spine) and lifting frequency (lifts per h). The remaining three factors were summary statistics obtained from the LMM. These included maximum sagittal (forward) flexion (deg), maximum lateral (side to side) velocity (deg/s), and maximum twisting velocity (deg/s). However, the combined effect or simultaneous occurrence of these motion risk factors was not documented. The following strategy provides one possible approach to quantify such a combined effect.

For each task within an MMH job, the continuous trunk motion of at least five task cycles should be captured. Data from these representative cycles are combined into *one* task motion profile. Finally, it is necessary to determine the proportion of the total time spent in situations where the sagittal flexion is > 20° combined with lateral and twisting velocities > 10 deg/s. These cutoff values represent the corresponding average values from the "high risk" group defined in Marras *et al.* (1993). The objective

is to minimize the percent of time spent in this region of combined sagittal flexion and lateral and twisting velocities. For the entire job, data from all the captured tasks could be combined into one representative data set and the proportion of time spent in the region defined above would be recalculated.

One potential way to incorporate the two task factors with the continuous combined motion factors is to determine the proportion spent in combined motion within various combinations of load moment (or weight) and frequency (Fathallah *et al.* 1998). The levels for each factor correspond to a low ($\geq$ 25th percentile), medium (> 25th percentile and $\leq$ 75th percentile), and high levels (> 75th percentile) of either the moment or frequency. Again, the objective is to minimize the percent spent in combined motions within a given moment/frequency combination.

Another way to quantify the combined motion is to take the average of the continuous product between sagittal position, lateral velocity and twisting velocity. This combined motion variable could be again linked with the task factors in a similar fashion as described above or through logistic regression approach, as described by Fathallah *et al.* (1998), where a probability of high risk group membership could be determined.

4. SUMMARY

When dealing with occupational injuries, especially LBD, risk factors are commonly assessed in an independent fashion where there is no indication of the temporal or combined relationship(s) among these factors. This article attempted to highlight the importance of addressing the issue of combined or simultaneously occurring risk factors since it could play an important role in explaining the occurrence of occupational injuries. The main

question to raise is: does a given risk factor simultaneously occur with other identified risk factors? The article discussed two potential ways of quantifying combined occupational LBD risk factors.

REFERENCES

FATHALLAH, F.A., MARRAS, W.S. and PARNIANPOUR, M., 1998, The role of complex simultaneous trunk motions in the risk of occupationally-related low back disorders. *Spine*, 23, 1035–42.

MARRAS, W.S., FATHALLAH, F.A., MILLER, R.J., DAVIS, S.W. and MIRKA, G.A., 1992, Accuracy of a three-dimensional lumbar motion monitor for recording dynamic trunk motion characteristics. *International Journal of Industrial Ergonomics*, 9, 75–87.

MARRAS, W.S. *et al.*, 1993, The role of dynamic three-dimensional trunk motion in occupationally-related low back disorders: the effect of workplace factors, trunk position, and trunk motion characteristics on risk of injury. *Spine*, 18, 617–28.

Back Belts

S. M. McGill

Occupational Biomechanics and Safety Laboratories, Faculty of Applied Health Sciences, Department of Kinesiology, University of Waterloo, Waterloo, Ontario, Canada N2L 3G1

1. INTRODUCTION

Debate over the use of abdominal belts (also known as back belts, back braces (*sic),* lumbar supports) in industrial settings continues. The premier question still remains "Should abdominal belts be prescribed to workers in industry to perform manual materials handling tasks?" Abdominal belts continue to be sold to industry in the absence of a regulatory requirement to conduct controlled clinical trials similar to that required of drugs and other medical devices. Many claims have been made as to how abdominal belts could reduce injury. For example, some have suggested that the belts remind people to lift properly. Some have suggested that belts may possibly support shear loading on the spine that results from the effect of gravity acting on the hand-held load and mass of the upper body when the trunk is flexed. Compressive loading of the lumbar spine has been suggested to be reduced through the hydraulic action of increased intra-abdominal pressure associated with belt wearing. Belts have been suspected of acting as a splint, reducing the range of motion, and thereby decreasing the risk of injury. Still other hypotheses as to how belts may affect workers include (1) providing warmth to the lumbar region, (2) enhancing proprioception via pressure to increase the perception of stability, and (3) reducing muscular fatigue.

A NIOSH (1994) publication entitled "Workplace Use of Back Belts", contained critical reviews of a substantial number of scientific reports evaluating back belts and concluded that back belts do not prevent injuries among uninjured workers nor do they consider back belts to be personal protective equipment. While this is generally consistent with our position stated in 1993, my personal position for belt prescription is somewhat more moderate.

The following sections have subdivided the scientific studies into clinical trials and those that examined biomechanical, psychophysical, and physiological changes from belt wearing. Finally, based on the evidence, guidelines are recommended for the prescription and usage of belts in industry.

2. CLINICAL/FIELD TRIALS

Many clinical/field trials that were reported in the literature were fraught with methodological problems and suffered from the absence of a matched control group, no post-trial follow-up, limited trial duration, and insufficient sample size. While the extreme difficulty in executing a clinical trial is acknowledged, only a few trials will be reviewed in this chapter.

One trial divided 81 male warehouse workers into three groups: a control group (n = 27); a group that received a half-hour training session on lifting mechanics (n = 27); and a group that received the one-hour training session and wore low-back orthoses while at work for the subsequent six months (n = 27). Instead of using more common types of abdominal belts, this research group used orthoses with hard plates that were heat molded to the low-back region of each individual. Given the concern that belt wearing was hypothesized to cause the abdominals to weaken, the abdominal flexion strength of the workers was measured both before and after the clinical trial. The control group and the training-only group showed no changes in abdominal flexor strength nor any change in lost time from work. The third group, which received both training and wore the belts, showed no changes in abdominal flexor strength or accident rate, but did show a decrease in lost time. However, it appears that the increased benefit was only to those workers who had a previous low-back injury.

In a larger field trial 642 baggage handlers who worked for a major airline were divided into 4 treatment groups; a control group (n = 248); a group that received only a belt (n = 57); a group that received only a one-hour back education session (n = 122); and a group that received both a belt and a one-hour education session (n = 57). The trial lasted eight months and the belt used was a fabric weight lifting belt, 15 cm wide posteriorly and approximately 10 cm wide anteriorly. There were no significant differences between treatment groups for total lumbar injury incident rate, lost workdays, or workers' compensation rates. While the lack of compliance by a significant number of subjects in the experimental group was cause for consideration, those who began wearing belts but discontinued their use had a higher lost-day case injury incident rate. In fact, 58% of workers belonging to the belt-wearing groups discontinued wearing belts before the end of the eight-month trial. Further, there was an increase in the number and severity of lumbar injuries following the trial of belt wearing. This general paradigm has recently been repeated in Europe where similar results were found in that no difference in injury rates were observed between the belt wearers and the non-belts. Interestingly, a small subgroup of pained workers reported less pain when wearing belts.

One large retrospective study administered questionnaires to 1,316 workers who performed lifting activities in the military. While this study relied on self-reported physical exposure and injury data over six years prior to the study, the authors did note that the costs of a back injury that occurred while wearing a belt were substantially higher than if injured otherwise.

A most recent study that has been widely reported, by Kraus and colleagues (1996), surveilled the low back injury rates of nearly 36,000 employees of the Home Depot Stores in California from 1989 to 1994. As reported, the company implemented a mandatory back belt use policy. Although the authors claim that belt wearing reduced the incidence of low back injury, analysis of the data and methodology suggests a much more cautious interpretation may be warranted. The concern is based on two issues: the lack of a robust effect and co-interventions (in addition to belts). The data suggest that the beneficial effect was limited to men with 1 to 3 years of employment (but not for longer or shorter lengths of employment) and limited to women employed 1 to 2 years - this is a very narrow band of effected employees. However, of greatest concern is the lack of scientific control over co-interventions to ferret out the true belt wearing effect - there was no comparable non-belt wearing group which is critical given that the belt wearing policy was not the sole intervention at Home Depot. For example, over the period of the study, the company increased the use of pallets and forklifts (altering physical demands), installed mats for cashiers, implemented post-accident

drug testing, and enhanced worker training. In fact, the company made a conscious attempt to enhance safety in the corporate culture. This was a large study and the authors deserve credit for the massive data reduction and logistics. However, despite the title and claims that back belts reduce low back injury, this uncontrolled study cannot answer the question about the effectiveness of belts.

In summary, difficulties in executing a clinical trial are acknowledged: the placebo effect is a concern, as it is difficult to present a true double-blind paradigm to workers since those who receive belts certainly know so; and there are logistical constraints on duration, diversity in occupations, and sample size. However, the data reported in the better-executed clinical trials cannot support the notion of universal prescription of belts to all workers involved in manual handling of materials to reduce the risk of low-back injury. There is weak evidence to suggest that those already injured may benefit from belts (or molded orthoses) with a reduced risk of injury reoccurrence. However, there does not appear to be support for uninjured workers wearing belts to reduce the risk of injury, and in fact, there appears to be an increased risk of injury during the period following a trial of belt wearing. Finally, there appears to be some evidence to suggest that cost per back injury may be higher if the worker was wearing a belt than if injured otherwise.

3. BIOMECHANICAL STUDIES

Biomechanical studies have examined changes in low back kinematics, posture and issues of specific tissue loading. Some studies have suggested that wearing an abdominal belt can increase the margin of safety during repetitive lifting but these studies generally did not measure this parameter directly. It was generally assumed that intra-abdominal pressure is a good indicator of spinal forces, which is highly contentious. Nonetheless, they assumed the higher recordings of intra-abdominal pressure (IAP) indicated an increase in low-back support which, in their view, justified the use of wearing belts. Spinal loads were not directly measured or calculated in these studies.

Several studies have questioned the hypothesized link between elevated intra-abdominal pressure and reduction in low-back load. There is no debate that IAP is increased but so too is concomitant abdominal wall activity (a build-up of intra-abdominal pressure requires additional activation of the musculature in the abdominal wall), resulting in a net increase in low-back compressive load and not a net reduction of load as had been previously thought. Therefore, it would seem erroneous to conclude that an increase in intra-abdominal pressure due to belt wearing reduces compressive load on the spine. In fact, it may have no effect or may even increase the load on the spine.

Several studies evaluating belt wearing and changes in muscle activity have noted no change in muscle activation amplitude of the low-back extensors nor in any of the abdominal muscles (rectus abdominis or obliques). Furthermore there has been no evidence for changes in muscle strength or endurance. Generally belt wearing appears to add stiffness to the torso particularly about the lateral bend and axial twist axes but not when subjects were rotated into full flexion. Thus, it would appear from these studies that abdominal belts assist to restrict the range of motion about the lateral bend and axial twist axes but do not have the same effect when the torso is forced in flexion, as in an industrial lifting situation.

Posture of the lumbar spine is an important issue in injury prevention for several reasons, but in particular the compressive strength of the lumbar spine decreases when the end range of motion in flexion is approached. Therefore, if belts restrict the end range of motion one would expect the risk of injury to be correspondingly decreased. While, the splinting and stiffening action of belts occurs about the lateral bend and axial twist axes, stiffening about the flexion-extension axes appears to be less. Some data have suggested that some belt styles are better in stiffening the torso in the manner described above, namely the taller elastic belts which span the pelvis to the rib cage. However, even in well controlled studies, it appears that belts can modulate lifting mechanics in some positive ways in some people and in negative ways in others. It remains to be seen if belts perturb the motor control system including the kinesthetic apparatus in the future.

4. PHYSIOLOGICAL STUDIES

Blood pressure and heart rate have been monitored in a few studies. Generally, both systolic and diastolic pressures can rise when wearing belts, over a wide variety of tasks from quiet sitting to heavy lifting. Given the relationship between elevated systolic blood pressure and an increased risk of stroke, some authors have concluded that individuals who may have cardiovascular system compromise are probably at greater risk when undertaking exercise while wearing back supports.

Over the past three or four years I have been asked to deliver lectures, and participate in academic debate on the back belt issue. On several occasions, occupational medicine personnel have approached me after hearing the effects of belts on blood pressure and intra-abdominal pressure, and have expressed suspicions that long term belt wearing at their particular workplace may possibly be linked with higher incidents of varicose veins in the testicles, haemorrhoid and hernia. At this point in time, there has been no scientific and systematic investigation of the validity of these claims and concerns. Rather than wait for strong scientific data to either lend support to these conditions, or dismiss them, it may be prudent to simply state concern. This will motivate studies in the future to track the incidents and prevalence of these pressure-related concerns to assess whether they are indeed linked to belt wearing.

5. PSYCHOPHYSICAL STUDIES

Studies based on the psychophysical paradigm allow workers to select weights that they can lift repeatedly using their own subjective perceptions of physical exertion. Generally, it has been observed that wearing belts increased the load that subjects were willing to lift by up to 20% in some cases. There has been some concern that wearing belts fosters an increased sense of security, which may or may not be warranted. This evidence may lend some support to this criticism of wearing abdominal belts.

6. BACK BELT PRESCRIPTION

My earlier report (McGill, 1993) presented data and evidence that neither completely supported, nor condemned, the wearing of abdominal belts for industrial workers.

Definitive laboratory studies that describe how belts affect tissue loading and physiological and biomechanical function have yet to be performed. In addition, clinical trials of sufficient scientific rigor to comprehensively evaluate the epidemiological risks and benefits from exposure to belts must be done. The challenge remains to arrive at the best strategy for wearing belts. Therefore, the available literature will be interpreted and given placement, and also combined with "common sense" to derive the most sensible position on prescription.

Given the available literature, it would appear the universal prescription of belts (i.e., providing belts to all workers in a given industrial operation) is not in the best interest of globally reducing both the risk of injury and compensation costs. Uninjured workers do not appear to enjoy any additional benefit from belt wearing and in fact may be exposing themselves to the risk of a more severe injury if they were to become injured and may have to confront the problem of weaning themselves from the belt. However, if some *individual* workers perceive a benefit from belt wearing then they may be allowed to conditionally wear a belt, but only on trial. The mandatory conditions for prescription (*for which there should be no exception*) are as follows:

(1) *Screen for cardiovascular risk*. Given the concerns regarding increased blood pressure and heart rate, and issues of liability, all those who are candidates for belt wearing should be screened for cardiovascular risk by medical personnel.

(2) *Enlist candidates in education programs*. Given the concern that belt wearing may provide a false sense of security, belt wearers must receive education on lifting mechanics (back school). All too often belts are being promoted to industry as a quick fix to the injury problem. Promotion of belts, conducted in this way, is detrimental to the goal of reducing injury as it redirects the focus from the cause of the injury. Education programs should include information on how tissues become injured, techniques to minimize musculoskeletal loading, and what to do about feelings of discomfort to avoid disabling injury.

(3) *No belts will be prescribed until a full ergonomic assessment has been conducted of the individual's job*. The ergonomic approach will examine, and attempt to correct, the cause of the musculoskeletal overload and will provide solutions to reduce the excessive loads. In this way, belts should only be used as a supplement for a few individuals while a greater plant-wide emphasis is placed on the development of a comprehensive ergonomics program.

(4) *Enlist candidates in a mandatory exercise program that must continue well after belt wearing has ceased*. Belts should not be considered for long-term use. The objective of any small-scale belt program should be to wean workers from the belts by insisting on mandatory participation in comprehensive fitness programs and education on lifting mechanics, combined with ergonomic assessment. Furthermore, it would appear wise to continue vigilance in monitoring former belt wearers (together with structured exercise) for a period of time following belt wearing, given that this period appears to be characterized by elevated risk of injury. Advice for exercise program development together with evidence to justify specific components has been provided in McGill, 1998).

REFERENCES

KRAUS, J.F., BROWN, K.A., McARTHUR, D.L., PEEK-ASA, C., SAMANIEGO, L. and KRAUS, C., 1996, Reduction of acute low back injuries by use of back supports. *Int. J. Occup. Environ. Health* 2:264–73.

McGILL, S.M., 1993, Abdominal belts in industry: A position paper on their assets, liabilities and use. *Am.Ind.Hyg.Assoc.J.* 54(12):752–54.

McGILL, S.M., 1998, Low Back Exercises: Evidence for improving exercise regimens, *Physical Therapy*, 78(7):754–65.

National Institute for Occupational Safety and Health (NIOSH), 1994, Workplace use of back belts, U.S. Department of Health and Human Services, Centres for Disease Control and Prevention. National Institute for Occupational Safety and Health, July.

REDDELL, C.R., CONGLETON, J.J., HUCHINSON R.D., and MONGOMERY J.F., 1992, An Evaluation of a weightlifting belt and back injury prevention training class for airline baggage handlers. *Appl. Ergonomics* 23(5):319–29 .

Building and the Construction Industry

H.F. van der Molen
Arbouw Foundation, Amsterdam, The Netherlands
N.J. Delleman
TNO Human Factors Research Institute, Soesterberg, The Netherlands
P.L.T. Hoonakker
P2 Research and Consultancy, Utrecht, The Netherlands

1. GENERAL

Building and construction are among the oldest activities of mankind. Important improvements in history were the invention of nails and screws, the introduction of ceramic building materials and the invention of prestressed concrete. The first important step in the field of work organization was the introduction of professional tradesmen. Specialization of trades was introduced in the early middle ages. In the second half of this century many new specialists arose, mainly as jobs were split up into several new trades, which often only consist of one or just a few tasks. In most cases the only change for the construction process was the appearance of several tradesmen in a sequential order. Teamwork still does not exist in road construction and other civil work. As we look at the pace of innovation in other sectors of industry, the construction industry can be characterized as conservative. Work is still physical straining and work organization and working methods are traditional. Koningsveld and van der Molen (1997) described the history and future of ergonomics in building and construction at the First International Symposium of Ergonomics in Building and Construction, which was part of the 13th triennial congress of the IEA in Tampere in 1997. There it was concluded that radical changes in working methods, work organization, and working conditions are necessary for a secure future for companies and a safe and healthy future for the workers. Ergonomists can help in the improvement process on evidenced-based ergonomic knowledge.

In the Netherlands, the last decade has shown growing attention to working conditions. In the construction industry *physical workload* and *work stress* are the two most important causes of absenteeism and disablement from work. Two-thirds of all health complaints are caused by these two working factors.

Section 1 describes the guidelines for assessing all aspects of physical workload of construction tasks in the construction industry. The assessment of physical workload on the basis of an adequate task analysis is, especially for ergonomists, a basic step in reducing the physical workload. The selection of technical solutions is easier than the assessment of physical workload. Section 2 describes the causes and relevant measures of work stress in the construction industry. Since, by comparison with physical workload, there are no validated and reliable guidelines available for work stress in the construction industry, this section focuses on causes on the basis of qualitative assessments. The different directions for solutions are also described. These more complex organizational solutions are, in most cases, primary preventive measures for eliminating physical overload and work stress.

2. GUIDELINES ON PHYSICAL WORKLOAD

In general, more than 50% of sick leave among Dutch construction site workers is a result of musculo-skeletal complaints and disorders, mostly related to the lower back region (Arbouw 1990–5). According to Hoonakker *et al.* (1992), in a group of about one thousand Dutch construction workers, the following hazards were mentioned: repetitive work (61%), heavy materials handling (59%), poor working postures (52%) and high force exertion (49%). In addition, occupational profiles based on the results of periodic health examinations show a high percentage of complaints concerning the musculo-skeletal system (Broersen *et al.* 1992). To reduce or eliminate health risks concerning the back, neck, and limbs, one needs to know which tasks in particular overload construction workers, and so guidelines and/or standards on physical workload are essential.

2.1. Background

Most guidelines on physical workload focus on the lifting of materials. The collective labor agreements for the construction industry in the Netherlands generally specify a maximum weight of 25 kg. In countries such as Australia, the United Kingdom and the USA it is common practice not to consider the object weight as the only risk factor in quantitative guidelines. The so-called Arbouw guidelines therefore follow this multi-factor approach by also taking into consideration, for instance, handling frequencies, task duration, or whatever is known to most affect physical load in particular work activities. Quantitative guidelines concerning lifting and carrying, pushing and pulling, static postures and repetitive work have been developed at the request of the parties involved in negotiating the labor agreements. The first edition of the Arbouw guidelines was based on a literature review and discussions among experts on physical workload in 1992 and 1993. Subsequently, 10 health and safety professionals evaluated the guidelines by actually working with them. In 1996 the guidelines were slightly revised in order to include the latest information from standards, reviews, etc. The health limits in the revised version of the Arbouw guidelines are based on Waters *et al.* (1993), Mital *et al.* (1993), Kilbom (1994a, 1994b), ISO/DIS 11226 (1999), NF X 35-106 (1985), and prEN 1005-4 (1998).

2.1.1 Evaluation scheme

All construction guidelines are based upon the following evaluation scheme:

green zone: basic, i.e. non-increased health risk for ≥90% of males (P90);

yellow zone: increased health risk; action may be planned in stages; immediate action is to be preferred. For the guidelines on external force exertion, such as for lifting and carrying, pushing/pulling and repetitive arm work, this zone denotes that between 25% and 90% of males (P25–P90) are able to exert a certain force. Static postures and repetitive movements associated with an increased health risk are included in this zone in the case of a task duration of between 1 and 4 hours;

red zone: strongly increased health risk; immediate action is necessary. For the guidelines on external force exertion (refer to above) the zone denotes that ≤25% of men (P25) are able to exert a certain force. Static postures and repetitive movements associated with an increased health risk are included in this zone in the case of a task duration of >4 hours.

The boundary between the green zone and the yellow zone is called the Action Limit (AL), whereas the boundary between the yellow zone and the red zone is called the Maximum Arbouw Limit (MAL). The AL-MAL concept has been used earlier for manual lifting (i.e. in terms of AL-MPL by NIOSH 1981), while the green-yellow-red concept has also been described before, e.g. for repetitive work (e.g. Hedén *et al*. 1993). The Arbouw guidelines use both concepts as one, with their own definitions, and in a consistent way for all kinds of physical workload.

2.1.2 Criteria for AL and MAL

AL represents a health limit, i.e. it should be the ultimate goal for action programs to get working conditions into the green zone. MAL was introduced for setting priorities within these programs. Strongly increased health risks (red) are to be tackled first, followed by those conditions associated with an increased health risk (yellow). Nonetheless, it should realized that an essential element of priority setting is also to check whether possible actions are reasonably practicable in economic, organizational and technical terms.

The MALs for external force exertion (lifting and carrying, whole body pushing/pulling and repetitive arm work) are set to a level at which a majority of the workers (75%) are not even able to exert that particular force. Employers do not desire such a situation in particular as it excludes far too many workers from doing the jobs. Thus there is a strong motivation for setting in motion a process of change, in addition to the obvious reasons for reducing health risks. Currently, most Dutch construction workers are male, and so the guidelines on maximum forces are based on male population data. It would be possible to formulate guidelines for females as well.

For static postures and repetitive movements, task duration was used to set AL and MAL. Primarily, tasks lasting more than 1 hour a day are included in the risk assessment procedure, based on the results of a review by Kilbom (1994a, 1994b). For repetitive work primary attention is on movement frequencies ≥ 2 per minute. Positions of body segments and joints are evaluated based on CEN and ISO standards. Those postures and movements observed that are associated with an increased health risk are classified in the yellow zone if the task lasts between 1 and 4 hours, and classified in the red zone if it lasts more than 4 hours. As a matter of course, durations of tasks loading the same body region are to be added together. It is recommended that each evaluation should be interpreted in relation to the level of complaints for the associated body region for the group of workers involved. Although the 4-hour limit, and to a lesser extent also the 1-hour limit, are arbitrary, the authors are of the opinion that the current procedure does at least, to some extent, take the risk factor task duration into consideration.

If application of the guidelines does not disclose an increased health risk, while an increased level of physical complaints for the group of workers involved exists, further analysis is considered necessary.

2.2 Application of the Guidelines

The Arbouw guidelines are meant for health and safety professionals. The guidelines show the user the most effective way to get to the green zone, or, for the time being, to the yellow zone. That is, they provide guidance as to whether actions are best directed towards the workplace (fixtures, transport, machines, tools/objects), the work organization and/or the workers. The results of analyses regarding a certain profession or job are discussed with employers and employees in order that a decision may be taken concerning actions that are reasonably practicable. Then a straightforward document is written for employers, employees, commissioners of work, architects or manufacturers of equipment and tools. Various so-called A(rbouw) documents (state-of-the-art documents) are available, e.g. for paving, scaffolding, and installing windowpanes. More than 20 such documents are available; in addition, a solutions database contains hundreds of technical solutions for reducing physical workload.

2.3 Guidelines for Five Areas of Physical Workload, i.e. Lifting, Pushing/Pulling, Carrying, Static Load, and Repetitive Work

2.3.1 Guideline on lifting

(mainly based on NIOSH, 1991, Mital *et al*. 1993, and NF X 35-106, 1985)

The following conditions are *red*:

- weight more than 25 kg when in a standing position;
- weight more than 10 kg when in a sitting, squatting, or kneeling position;
- weight more than 17 kg when lifting one-handed;
- horizontal location (H) more than 63 cm;
- asymmetry angle (A) more than 135°;
- vertical location (V) more than 175 cm or less than 0 cm;
- frequency (F) more than 15 lifts/minute.

Tables 1 and 2 below distinguish between symmetric and

Table 1: Lifting index for *symmetric* lifting (asymmetry angle (A) 0°), vertical location (V) 175 cm, vertical travel distance (D) 175 cm, work/lifting duration 8 hours and no recovery time. The figures in brackets are for lifting under optimal symmetric conditions, i.e. vertical location (V) 75 cm, vertical travel distance (D) 0 cm, work/lifting duration 1 hour and recovery time 7 hours. A lifting index ≤1 is *green*, a lifting index 1-3 is *yellow*, and a lifting index ≥3 is *red*.

| Weight | horizontal location (H) | Frequency | | |
		5–9 lifts/ minute	1–5 lifts minute	<1 lift minute
0-5 Kg	≤25 cm	2.8 (0.5)	1.2 (0.3)	0.5 (0.3) green
	25-40 cm	4.2 (0.7)	1.9 (0.5)	0.9 (0.4)
	40-63 cm	7.1 (1.7)	2.9 (0.8)	1.4 (0.6)
5-10 Kg	≤25 cm	5.9 (0.9)	2.3 (0.6)	yellow 1.1 (0.5)
	25-40 cm	8.3 (1.5)	3.7 (1.0)	1.7 (0.8)
	40-63 cm	14.2 (2.3)	5.9 (1.5)	2.7 (1.3)
10-15 Kg	≤25 cm	8.3 (1.4)	3.5 (0.9)	1.6 (0.8)
	25-40 cm	12.5 (2.2)	5.6 (1.4)	2.6 (1.2)
	40-63 cm	21.4 (1.9)	8.8 (2.3)	4.1 (1.9) Red
15-20 Kg	≤25 cm	11.1 (1.9)	4.7 (1.2)	2.2 (1.0)
	25-40 cm	16.6 (2.9)	7.4 (1.9)	3.5 (1.6)
	40-63 cm	28.5 (4.7)	11.7 (3.0)	5.4 (2.6)
20-25 Kg	≤25 cm	13.8 (2.3)	5.8 (1.5)	2.7 (1.3)
	25-40 cm	20.8 (3.7)	9.3 (2.4)	4.3 (2.0)
	40-63 cm	35.7 (5.8)	14.7 (3.8)	6.8 (3.2)

Table 2: Lifting index for *asymmetric* lifting (asymmetric angle (A) 135°), vertical location (V) 175 cm, vertical travel distance (D) 175 cm, work/lifting duration 8 hours and no recovery time. The figures in brackets are for lifting under optimal asymmetric conditions, i.e. vertical location (V) 75 cm, vertical travel distance (D) 0 cm, work/lifting duration 1 hour and recovery time 7 hours. A lifting index ≤1 is *green*, a lifting index 1-3 is *yellow*, and a lifting index ≥3 is *red*.

Weight	horizontal location (H)	Frequency (F) 5-9 lifts/ minute	1-5 lifts minute	<1 lift minute
	≤25 cm	4.6 (0.8)	2.0 (0.5) yellow	0.9 (0.5)
				green
0-5 Kg	25-40 cm	7.1 (1.3)	3.3 (0.8)	1.5 (1.7)
	40-63 cm	12.5 (2.0)	5.0 (1.3)	2.4 (1.1)
	≤25 cm	9.1 (1.6)	4.0 (1.1)	yellow
				1.9 (0.9)
5-10 Kg	25-40 cm	14.2 (2.6)	6.7 (1.7)	3.0 (1.4)
	40-63 cm	25.0 (4.0)	10.0 (2.6)	4.8 (2.3)
	≤25 cm	13.6 (2.5)	6.0 (1.6)	2.8 (1.4)
			yellow	
10-15 Kg	25-40 cm	21.4 (3.9)	10.0 (2.5)	4.6 (2.1)
	40-63 cm	37.5 (6.0)	15.0 (4.0)	7.1 (3.4)
	≤25 cm	18.1 (3.3)	8.0 (2.1)	3.8 (1.8)
15-20 Kg	25-40 cm	29.1 (5.1)	13.3 (3.4)	6.1 (2.9)
	40-63 cm	50.0 (8.0)	20.0 (5.3)	9.5 (4.5)
	≤25 cm	22.7 (4.1)	10.0 (2.7)	4.7 (2.3)
20-25 Kg	25-40 cm	35.7 (6.4)	16.6 (4.2)	7.6 (3.6)
	40-63 cm	62.5 (10.0)	25.0 (6.6)	11.9 (5.7)

asymmetric lifting, where a lifting index ≤1 is *green*, a lifting index 1–3 is *yellow*, and a lifting index ≥3 is *red*.

In the case of limited headroom the lifting index found in one of these tables has to be multiplied by a multiplier taken from Table 3 below.

Table 3: Limited headroom multiplier

	multiplier
fully upright	1.00
95% upright	1.67
90% upright	2.50
85% upright	2.60
80% upright	2.78

An index greater than 1 means an increased health risk, concerning musculo-skeletal complaints/disorders in particular. A greater index means a higher risk. Table 4 presents the guidelines for lifting in a sitting, squatting, or kneeling position or for lifting one-handed.

Table 4: Guidelines for lifting in a sitting, squatting, or kneeling position or for lifting one-handed infrequently

	Weight green/yellow limit	yellow/red limit
sitting, squatting or kneeling	4.5 Kg	10.0 Kg
one-handed	7.5 Kg	17.0 Kg

For frequent lifting (≥2/minute) and longer durations (>1 hour/day) the guidelines on repetitive work apply.

2.3.2 Guidelines on pushing/pulling
(mainly based on Mital *et al.* 1993, and NF X 35-106, 1985)

2.3.2.1 Pushing/pulling with the whole body while walking

The tables below distinguish between pushing to set an object in motion (initial force exertion), pulling to set an object in motion (initial force exertion), and pushing or pulling to keep an object in motion (sustained force exertion). The guidelines are valid for two-handed pushing/pulling for a whole working day (8 hours), and an optimal height of the hands during force exertion (95–130 cm). A detailed analysis is necessary if slipping is likely to occur (or actually occurs) or in the case of high movement speed, awkward posture, asymmetric force exertion (one-handed, course changes), or a bad view of the surroundings (surface, obstacles, etceteras). All cells of the tables contain a pushing/pulling-index, besides the green, yellow, or red evaluation result. An index greater than 1 means an increased health risk, concerning musculoskeletal complaints/disorders in particular. A greater index means a higher risk.

2.3.2.2 Pushing/pulling with the whole body while staying on the spot

The guidelines for setting an object in motion (refer to tables 1 and 2 in section 2) also apply to pushing/pulling with the whole body while staying on the spot.

2.3.2.3 Pushing/pulling with the upper limbs

The guidelines below are valid if the postures of the trunk and the upper limbs are evaluated as green (refer to the guideline on repetitive work), if the hands do not reach further forwards than three-quarters of the maximum reach distance (trunk upright), and if the hands are between pelvis height and shoulder height.

Table 5: Pushing to set an object in motion (initial force exertion)

force	frequency 2.5x/minute– 1x/minute	1x/minute– minutes	1x/5 minutes– 1x/8 hours	≤1x/8 hours
0-25 kgf	1.0 green	1.0 green	0.9 green	0.8 green
25-30 kgf*	1.2 yellow	1.2 yellow	1.1 yellow	1.0 green
30-45 kgf*	1.8 yellow	1.7 yellow	1.6 yellow	1.5 yellow
45-50 kgf*	2.0 red	1.9 yellow	1.8 yellow	1.7 yellow
50-65 kgf*	2.6 red	2.5 red	2.3 red	2.2 yellow
>65 kgf*	>2.6 red	>2.5 red	>2.3 red	>2.2 red

Note: * = usually not possible because slipping is likely to occur

Table 6: Pulling to set an object in motion (initial force exertion)

force	frequency 2.5x/minute– 1x/minute	1x/minute– 1x/5 minutes	1x/5 minutes– 1x/8 hours	≤1x/8 hours
0-25 kgf	1.0 green	1.0 green	1.0 green	1.0 green
20-40 kgf*	2.0 yellow	2.0 yellow	2.0 yellow	2.0 yellow
40-45 kgf*	2.3 red	2.3 red	2.3 yellow	2.3 yellow
45-50 kgf*	2.5 red	2.5 red	2.5 red	2.5 yellow
>50 kgf*	>2.5 red	>2.5 red	>2.5 red	>2.5 red

Note: * = usually not possible because slipping is likely to occur.

Table 7: Pushing or pulling to keep an object in motion (sustained force exertion)

force	distance	frequency 2.5/minute–1/minute	1/minute–minutes	1/5minute–hours	≤1/8 hours
0-5 kgf	2–15 m	0.8 green	0.4 green	0.4 green	0.3 green
	15–30 m	-	0.8 green	0.4 green	0.3 green
	30–60 m	-	-	0.6 green	0.4 green
5–10 kgf	2–15 m	1.5 yello	0.9 green	0.7 green	0.6 green
	15–30 m	-	1.5 yellow	0.8 green	0.6 green
	30–60 m	-	-	1.2 yellow	0.9 green
10–15 kgf	2–15 m	2.3 red	1.3 yellow	1.1 yellow	0.9 green
	15–30 m	-	2.3 red	1.3 yellow	0.9 green
	30–60 m	-	-	1.8 yellow	1.3 yellow
15–20 kgf	2–15 m	3.1 red	1.7 red	1.5 yellow	1.2 yellow
	15–30 m	-	3.1 red	1.7 yellow	1.2 yellow
	30–60 m	-	-	2.3 red	1.7 yellow
20–25 kgf*	2–15 m	3.8 red	2.2 red	1.9 yellow	1.5 yellow
	15–30 m	-	3.8 red	2.1 yellow	1.5 yellow
	30–60 m	-	-	2.9 red	2.2 red
25–30 kgf*	2–15 m	4.6 red	2.6 red	2.2 red	1.8 yellow
	15–30 m	-	4.6 red	2.5 red	1.8 yellow
	30–60 m	-	-	3.5 red	2.6 red
>30 kgf*	2–15 m	>4.6 red	>2.6 red	>2.2 red	>1.8 red
	15–30 m	-	>4.6 red	>2.5 red	>1.8 red
	30–60 m	-	-	>3.5 red	>2.6 red

Notes: * = usually not possible because slipping is likely to occur.

- = combination of frequency and distance is not realistic.

2.3.3 Guidelines on carrying

The following condition is *red*:

- weight more than 25 kg;
- weight more than 10.5 kg for infrequent one-handed carrying.

The guidelines below are valid for two-handed carrying for a whole working day (8 hours), and the hands at an optimal height (knuckle height, arms hanging down). All cells of the tables contain a carrying-index, besides the evaluation result green, yellow, or red. An index greater than 1 means an increased health risk, concerning musculoskeletal complaints/disorders in particular. A greater index means a higher risk

2.3.4 Guidelines on static postures

It is recommended to use the guidelines below to begin with on tasks lasting longer than 1 hour (continuous or a total of distinct periods) per working day (note: shorter task durations cannot be considered safe in all cases). In the case that two or

Table 8: Guidelines for infrequent pushing/pulling with the upper limbs

	force green/yellow limit	yellow/red limit
A	7.5 kgf	17.0 kgf
B	19.0 kgf	43.0 kgf

Notes: A and B refer to the various types of force exertion distinguished in the guideline on repetitive work.

For frequent pushing/pulling (≥2/minute) and longer durations (>1 hour/day) the guidelines on repetitive work apply.

Table 9: Guidelines on carrying (mainly based on Mital *et al.* 1993)

weight	distance	frequency 3/minute–1/minute	1/minute–1/5 minutes	1/5 minutes–1/8 hour	≤1/8 hour
0–10 kg	≤2.0 m	0.5 green	0.5 green	0.4 green	0.4 green
	2–8.5 m	0.8 green	0.6 green	0.5 green	0.4 green
10–15 kg	≤2.0 m	0.8 green	0.7 green	0.7 green	0.6 green
	2–8.5 m	1.2 yellow	0.9 green	0.8 green	0.6 green
15–20 kg	≤2.0 m	1.1 yellow	1.0 green	0.9 green	0.8 green
	2–8.5 m	1.5 yellow	1.2 yellow	1.0 green	0.8 green
20–25 kg	≤2.0 m	1.3 yellow	1.2 yellow	1.1 yellow	1.0 green
	2–8.5 m	1.9 red	1.5 yellow	1.3 yellow	1.0 green
>25 kg	≤2.0 m	>1.3 red	>1.2 red	>1.1 red	>1.0 red
	2–8.5 m	>1.9 red	>1.5 red	>1.3 red	>1.0 red

Table 10: Guideline for infrequent one-handed carrying (standing or walking up to 90 meters)

	weight green/yellow limit	yellow/red limit
infrequent one-handed carrying (standing or walking)	6 kg	10.5 kg

more tasks load the same body region (through static postures and/or repetitive work), the durations of these tasks are to be taken together. It is recommended to interpret the results of guideline application in relation to the complaints and disorders of the particular body region found for the group of employees involved. An evaluation result yellow/red becomes yellow in the case of a task duration of between 1 and 4 hours, whereas it becomes red in the case of a task duration >4 hours.

The guidelines on low back, shoulder and shoulder girdle as well as on neck and upper back are based on a description of the actual posture (observed/measured) with respect to a reference posture, i.e. a sitting or standing posture with a non-rotated upright trunk, a non-kyphotic lumbar spine posture, and the arms hanging freely, while looking straight ahead along the horizontal. Guidelines on sitting also include raised sitting.

- Asymmetric trunk posture (axial rotation and/or lateral flexion): *yellow/red*
- For sitting: kyphotic lumbar spine posture (valid for: trunk inclination between 0° and 20°): *yellow/red*

Table 11: Guidelines on static postures (mainly based on ISO/DIS 11226, 1999): *low back*

trunk inclination <0°	0–20°	20–60°	>60°
yellow/red**	green	green/yellow/red*	yellow/red

Table 12: Guidelines on static postures (mainly based on ISO/DIS 11226, 1999): *shoulder and shoulder girdle*

upper arm elevation		
0–20°	**20–60°**	**>60°**
green	green/yellow/red*	yellow/red

- • * = with full trunk support: green; without full trunk support: consult an expert for evaluating holding time — recovery time regimes (see ISO/DIS 11226); if that is possible, the evaluation result is *yellow* in the case of a task duration of between 1 and 4 hours, whereas it is red in the case of a task duration > 4 hours (expert guess of the authors); if two or more tasks load the same body region, the durations of these tasks are to be taken.
- • ** = with full trunk support: *green*
- • Upper arm retroflexion (i.e. elbow behind the trunk when viewed from the side of the trunk), upper arm adduction (i.e. elbow not visible when viewed from behind the trunk), or extreme upper arm external rotation: *yellow/red*
- • Raised shoulder: *yellow/red*
- • = with full arm support: green; without full arm support: consult an expert for evaluating holding time – recovery time regimes (see ISO/DIS 11226); if that is not possible, the evaluation result is *yellow* in the case of a task duration of between 1 and 4 hours, whereas it is red in the case of a task duration >4 hours (expert guess of the authors); if two or more tasks load the same body region, the durations of these tasks are to be taken together.

2.3.4.1 Neck and upper back

Green:
- head inclination 0°–25° and neck flexion (i.e. head inclination minus trunk inclination) 0°–25°.

Green/yellow/red:

- head inclination 25°–85°:
- with full trunk support, consult an expert for evaluating holding time — recovery time regimes (see ISO/DIS 11226); if that is not possible, the evaluation result is yellow in the case of a task duration of between 1 and 4 hours, whereas it is red in the case of a task duration >4 hours (expert guess of the authors); if two or more tasks load the same body region, the durations of these tasks are to be taken together;
- without full trunk support, the holding time for trunk inclination is critical, and should be evaluated.

Yellow/red:

- head inclination <0°:
- changing to *green* in the case of full head support;
- head inclination >85°;
- neck flexion (head inclination minus trunk inclination) <0° (i.e. neck extension) or >25°;
- asymmetric neck posture, i.e. axial rotation and/or lateral flexion of the head versus the upper part of the trunk.

2.3.4.2 Other joints

The following conditions are *yellow/red*:

- extreme joint positions;
- for a kneeling position: no adequate knee protection;
- for standing (except when using a buttock rest): flexed knee;
- for sitting: knee angle >135° (180° = the upper leg in line with the lower leg), unless the trunk is inclined backwards and fully supported (as in car driving);
- for sitting: knee angle <90° (1) (180° = the upper leg in line with the lower leg).

2.3.4.3 Pedal operation

The following conditions are *yellow/red*:
- standing;
- sitting, leg-actuated;
- sitting, ankle-actuated, force exertion >5.5 kgf;
- sitting, ankle-actuated, pedal operating already by merely supporting foot weight.

2.3.4.4 General

The following conditions are *yellow*:
- standing continuously ≥1 hour/day;
- standing for a total of distinct periods ≥4 hours/day.

2.3.5 Guidelines on repetitive work

It is recommended to use the guidelines below to be begin with on tasks lasting longer than 1 hour (continuous or a total of distinct periods) per working day (note: shorter task durations cannot be considered safe in all cases). If two or more tasks load the same body region (through static postures and/or repetitive work), the durations of these tasks are to be taken together. It is recommended to interpret the results of guideline application in relation to the complaints and disorders of the particular body region found for the group of employees involved. An evaluation result yellow/red becomes yellow in the case of a task duration of between 1 and 4 hours, whereas it becomes red in the case of a task duration >4 hours.

Table 13: Guidelines on repetitive work (mainly based on NF X 35-106, 1985, and prEN 1005-4, 1998): *low back*

trunk inclination		
<0°	**0–20°**	**>20°**
yellow/red[1]	green	yellow/red[1]

Asymmetric trunk posture (axial rotation and/or lateral flexion): *yellow/red*

Note: 1 An increased health risk is present if ≥2/minute an evaluation result yellow/red is found (N.B. lower frequencies cannot be considered safe in all cases).

Table 14: Guidelines on repetitive work (mainly based on NF X 35-106, 1985, and prEN 1005-4, 1998): *shoulder and shoulder girdle*

upper arm elevation		
0–20°	**20–60°**	**>60°**
green	green	yellow/red 1

Upper arm retroflexion (i.e. elbow behind the trunk when viewed from the side of the trunk), upper arm adduction (i.e. elbow not visible when viewed from behind the trunk), or extreme upper arm external rotation: *yellow/red* 1

Raised shoulder: *yellow/red* 1

neck flexion (i.e. head inclination minus trunk inclination) 0°–25°.

Note: 1 An increased health risk is present if ≥2/minute an evaluation result yellow/red is found (N.B. lower frequencies cannot be considered safe in all cases).

2.3.5.1 Neck and upper back

Green:
- neck flexion (i.e. head inclination minus trunk inclination) 0°–25°.

Yellow/red:
- neck flexion (head inclination minus trunk inclination) <0° (i.e. neck extension) or >25° (1);
- asymmetric neck posture, i.e. lateral flexion and/or extreme axial rotation of the head versus the upper part of the trunk (1).
1 An increased health risk is present if ≥2/minute an evaluation result yellow/red is found (N.B. lower frequencies cannot be considered safe in all cases).

2.3.5.2 Other joints

The following conditions are *yellow/red*:
- extreme joint positions (1);
- for a kneeling position: no adequate knee protection;
- for standing (except when using a buttock rest): flexed knee (1);
- for sitting: knee angle >135° (1) (180° = the upper leg in line with the lower leg), unless the trunk is inclined backwards and fully supported;
- for sitting: knee angle <90° (1) (180° = the upper leg in line with the lower leg).
1 An increased health risk is present if ≥2/minute an evaluation result yellow/red is found (note: lower frequencies cannot be considered safe in all cases).

2.3.5.3 Pedal operation

The following conditions are *yellow/red*:
- standing;
- sitting, leg-actuated.

2.3.5.4 Force exertion

1 General force limits, i.e. one-handed or two-handed, sitting or standing, almost all directions.

Table 15: General force limits

| A | force | |
frequency	green/yellow limit	yellow/red limit
≥2 and <3/minute	3.0 kgf	6.5 kgf
≥3 and <4/minute	2.0 kgf	4.0 kgf
≥4 and <5/minute	1.5 kgf	3.0 kgf
≥5/minute	1.0 kgf	2.5 kgf

1 Specific force limits: fore/aft direction; upward/downward direction; force limits are also valid for combined fore/aft and upward/downward directions. For one-handed pinching and pedal operation: see NF X 35-106.

Table 16: Specific force limits

| B | force | |
frequency	green/yellow limit	yellow/red limit
≥2 and <3/minute	8.0 kgf	18.0 kgf
≥3 and <4/minute	5.0 kgf	11.5 kgf
≥4 and <5/minute	3.0 kgf	7.0 kgf
≥5/minute	2.0 kgf	4.5 kgf

3. WORK STRESS IN THE CONSTRUCTION INDUSTRY

In the Dutch construction industry over half of the employees experience considerable pressure of work, and they think that measures should be taken either to prevent high work pressure or reduce it. Employers also admit that work pressure is sometimes high. One quarter of the employers taking part in a Dutch Monitoring Study on Stress and Physical Workload admitted that work pressure is indeed high. However, only 3% of the employers say that this causes serious problems.

3.1 Causes and Consequences of Work Pressure in the Construction Industry

The ultimate cause of high work pressure in the construction industry lies above all in the (economic) demands which are imposed on (the employees of) construction firms. The current system of public tenders (contracts) and the intense competition among construction firms compels such firms to try to achieve maximum production at minimum cost. This means being as efficient and flexible as possible and adapting quickly to the constantly changing demands.

This has several consequences for the organization (of the work). The activities are split up into parts which are small and standardized (Taylorism). This marked division of labor results in adjustment problems; creates organization of work which is more prone to disruption (interruptions) and increased control problems. It is the lack of decision latitude, in particular, that creates a high work pressure. The causes and consequences of work pressure are illustrated in Figure 1.

3.1.1 The division of labor

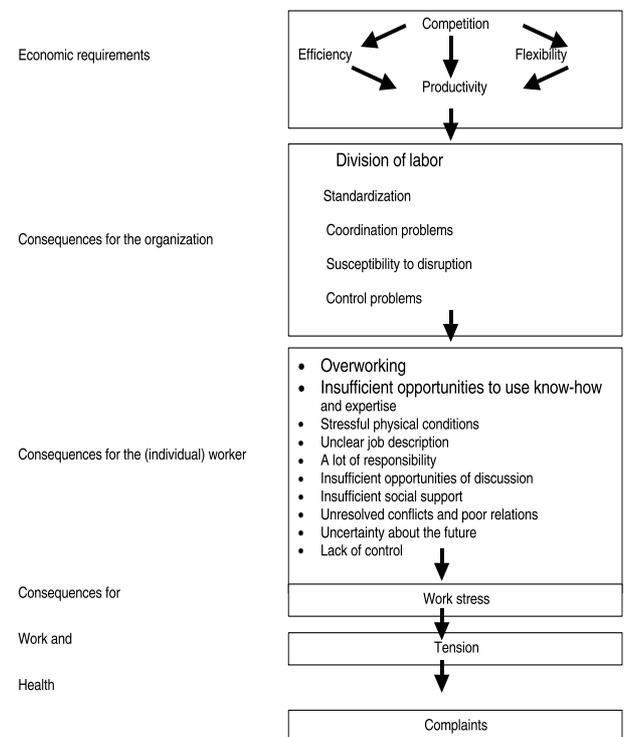The division of labor refers to the way work is organized.

Figure 1: The causes and consequences of work pressure

Minimum division of labor means that successive items of work are performed as part of one job or in one department. Maximum division of labor means that successive items of work are divided between different jobs or departments. Division of labor in the construction industry occurs at two levels:

3.1.1.1 The organization of the overall construction process

The activities in the building process are distributed between different parties. The building process starts with the customer (client). The client, nowadays, has a choice between various forms of real estate development. Traditionally, the design has been produced by the architect. Work preparation is done by the contractor. The suppliers take care of the materials and the subcontractors carry out the work.

3.1.1.2 The organization of construction companies

The organization of building companies is based upon division of labor. Different activities make up different functions. Before a construction worker starts his activities, the cost accountant, the planner, the project manager, the foreman, and the person responsible for the working conditions have already been involved in the process. Moreover, there are many construction companies which specialize in specific activities of the building process. Examples of these include companies specializing in foundation work, concrete formwork, bricklaying, carpentry, painting, roof laying, plastering, ceiling assembly, glazing etc.

A characteristic feature of the division of labor is the interdependence between the different parties involved. The later a party is involved in the process, the stronger its dependency on other parties and the smaller its influence on decision-making.

Maximum division of labor entails certain risks (see figure 1): standardization problems in the building process, adjustment problems, greater risk of disruption, low commitment of employees and a lack of decision latitude (autonomy). These effects reinforce one other and are important causes of work pressure.

The large number of parties involved in the building process demands standardization of the work process and rigid planning. Consequently, problems are often identified at the last moment and solved on an ad hoc basis. Little attention is paid to the solution of structural problems because the building process is characterized by unique collaborations: another characteristic of the construction industry is the project-based organization of work. The construction of each building is a project involving different parties. Because of this variety of buildings and parties involved, each project is unique and requires specific collaborations and organizations.

The complex organization of the building process demands good coordination between the parties involved. Coordination problems are very common. For example, in the process design phase decisions may be made which result in a working method that is difficult or presents unusual or even increased risks to the firm which has to carry it out.

Another consequence of the high division of labor is that the building process is very sensitive to disturbances. An example of such a disturbance is the failure to deliver materials on time. Employees have to wait unnecessarily in order to continue working. As a result, the sub-contractors at the end of the building process often have to work under high work pressure. Indeed, they are the most likely to have to cope with the negative consequences of high work pressure.

Because many parties are involved in the building process, it is difficult for people to feel responsible or be committed to the work as a whole. Construction companies and individual employees have a tendency to concentrate solely on their own interests. As a result there is little feedback.

The high division of labor means that employees have little decision latitude. Decision latitude is the opportunity for employees to solve problems themselves with the aid of their colleagues or in consultation with others. In the construction industry, employees are dependent on the decisions of their managers, who in turn are dependent on the choices made earlier in the building process.

The consequences of the high division of labor are interdependent and amplify one another. The splitting up of the activities makes the companies involved in the building process more and more dependent on each other. As a consequence, the building process is very prone to disruption. And these disruptions (for example; the late delivery of materials, the failure of subcontractors to show up) often interfere with the planning.

Apart from that, the division of labor considerably reduces the decision-making latitude. To increase the latter, excellent co-ordination is required between the parties involved in the building process. This is often difficult. The need to increase the decision latitude often requires standardization of the work processes. This higher standardization often means greater division of labor. And thus the wheel comes full circle.

3.1.2 New technologies

The process described above may be intensified by the development of new technologies. The development of such new technologies means that a great deal of work is being transferred from the construction site to factories and suppliers. Hard physical work is often being replaced by mechanization (for example, the more frequent use of heavy machines). The development of new technologies also affects the division of labor and increases work pressure. Because machines take over much of the work, only minor activities are left to be carried out by human beings. These minor activities are often short-term and monotonous. As a result of the introduction of new technologies, employees may be subjected to unilateral and sustained stresses both mental and physical.

3.1.3 Consequences for employees

For the (individual) employee this can upset the balance between the workload and his capacity to cope with it. Research has shown that several important factors play a role in this process (e.g. Warr, 1990). The figures reported here for the building industry are taken from research performed by the Economisch Instituut voor de Bouwnijverheid, the EIB (the Netherlands Economic Institute for the Construction industry) and from the Arbouw database which contains data for over 60 000 construction workers who took part in a routine company health survey (PBGO).

3.1.3.1 Overload

We speak of overload when the activities carried out by an employee overtax his capacities to cope with them. Overload

often occurs when there is not enough time to recover from a very demanding situation. There is a difference between overload due to *the amount of work* and overload caused by *the nature of the work*. In the first case, the work pressure is too high, simply because there is too much work to be done, or because the work tempo is too high (quantitative workload). In the second case the work is too complex or too difficult. This is known as the qualitative workload.

3.1.3.2 Underload

It is well known that overload can cause stress. But the reverse situation, underload, can also be a cause of stress. Stress can arise because an employee has too little to do, or because the work is too simple, too monotonous or is not demanding enough. This can be "soul-destroying".

Overload and underload can increase stress. In both cases the balance between workload and work capacity is disturbed.

3.1.3.3 Taxing physical working conditions

Overload can also occur as a result of very taxing physical working conditions. Examples include noise, vibrations, toxic substances, temperature and humidity. This not only has physical effects but also increases the mental stresses. A very specific demanding situation — especially in the construction industry — is working in hazardous situations.

3.1.3.4 Role ambiguity and role conflict

When it is not absolutely clear what one's responsibilities are, or what one is expected to do, we talk about role ambiguity. When there is a lack of specific procedures, it is impossible to know whether one has done a good or a bad job. This insecurity about the results of the work can cause a lot of stress; no one knows down to the very smallest details what is expected of them, and this insecurity sometimes becomes too great. It can result in job dissatisfaction; people feel redundant and experience a lack of self-confidence. Sometimes jobs are clearly described (on paper), but then bear little resemblance to the work that has actually to be carried out. Also, work in which there are a lot of conflicting demands can be harmful to well being. It creates a situation full of conflict and insecurity (what am I supposed to do?) and this in its turn can be a cause of stress.

3.1.3.5 Excessive responsibility

Excessive responsibility for other people's health, safety and job security can also be a cause of stress.

The same goes for excessive responsibility for (valuable) materials.

3.1.3.6 Insufficient opportunities for consultation and participation

Stress can also be caused when there are too few opportunities to participate or make decisions concerning the job that has to be done. The employee has little influence on the activities that have to be carried out and is at the mercy of others. A regular discussion of progress can be one way to exercise an influence on the decisions that are being made.

3.1.3.7 Insufficient social support

Lack of social support from colleagues can also be a cause of stress. Two sorts of social support are important: instrumental

and emotional support. Instrumental support means that colleagues or the management can actually give help to solve problems. This can be advice on how to solve the problem, but may also be an offer of practical assistance. Emotional support is also very important. This implies respect and consideration for one another, commitment, understanding and sympathy. It creates a situation of security and trust.

When people get enough social support, they are better equipped to deal with other stress factors. An important aspect of social support is the person who gives it. One's immediate boss and colleagues are important "supporters".

Table 17: Percentage prevalences of stress factors in construction employees

Source:: EIB: 1995, PBGO: 1993–1994, Arbouw

Overload	
– tasks are complex	18%
– much knowledge of materials required	67%
– work requires a lot of improvisation	54%
– much professional knowledge required	87%
– many professional skills required	90%
– work is mentally demanding	25%
– high work tempo	55%
Underload	
– work lacks interest most of the time	21%
– work too simple	10%
– work not challenging enough	6%
Taxing physical working conditions	
– toxic substances	17%
– climate	26%
– safety problems at work	16%
– physical demands	32%
– noise	37%
– vibrations	19%
Role ambiguity and role conflict	
– does other work than his 'own job'	40%
– unclear responsibilities	4%
– doesn't know what others expect	8%
Much responsibility	
– responsible work	78%
Lack of consultation and participation	
– insufficient consultation	10%
– insufficient participation	15%
Lack of social support	
– bad day-to-day management	19%
– bad social climate	9%
Unresolved conflicts and bad social climate	
– annoyed by others	20%
– annoying absence of others	11%
– tension on the work site	20%
Job insecurity	
– insufficient security	19%
– unfavourable changes	10%
– unfavourable prospects	13%
Lack of decision latitude	
– insufficient opportunities to take decisions	41%

3.1.3.8 Unresolved conflicts and bad work relations

If one lacks social support and is forced to work in a situation full of irritations and conflicts, this can also cause stress.

3.1.3.9 Job insecurity

Because of the project-based organization of work one of the factors which has always played an important role in the construction industry, and especially now in these days of mergers, reorganizations and cutbacks, is insecurity about the future. This relates mostly to the availability of work and career opportunities. The threat of losing your job or the insecurity of not knowing whether there will be work after the latest project creates a lack of job security. This can have serious social and financial consequences. Research has shown that job insecurity is as bad as actually being sacked (Dekker and Schaufeli 1995).

3.1.3.10 Lack of decision latitude

When someone finds it impossible to exert an influence on their work, for example affect the work tempo, take breaks or organize the work in the way he (or she) prefers, we talk of the lack of decision latitude or autonomy. Lack of control over the work situation can also be a serious cause of stress.

The reason is that the employee is not able to intervene to adapt the work to his own situation (e.g. to work more slowly when he is tired, work faster when he is fit).

The construction industry has conducted research into these aspects. For example, the research carried out annually by the EIB and by Arbouw is looking at these aspects. Table 17 below shows some of the results.

Research has shown (e.g. Karasek 1979; Karasek and Theorell 1990) that the combination of (high) demands and (low) decision latitude, in particular, is important in preventing work pressure and the strains resulting from it. Figure 2 below illustrates the model:

Explanation of the model:

1. There are low (psychological) job demands, the job is not too difficult and employees have little decision latitude. It is a passive job.
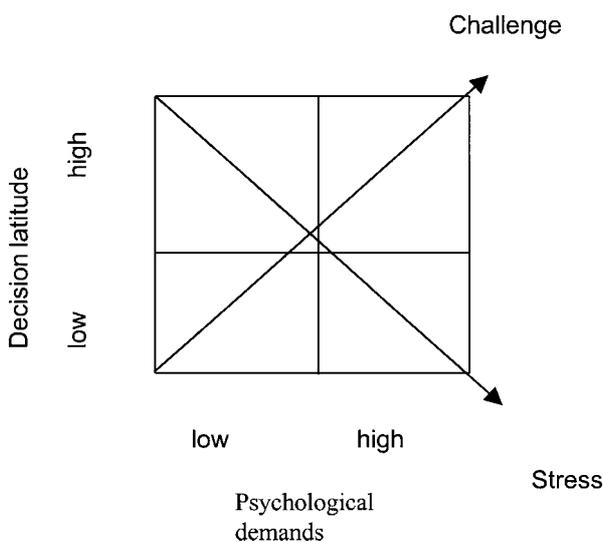
2. Despite the low job demands, employees have a lot of decision latitude. These are jobs with a risk of underload.

3. Job demands are high, but the job contains a lot of autonomy (decision latitude). This is an active job, with a lot of challenges, in which the employee can grow and develop his skills.

4. Much is asked of the employees: job demands are high. However, employees have little decision latitude. They are bound by procedures and obligations. In this situation there is a high risk of stress.

The above factors, which create excessively high work pressure and resultant stresses, are caused by the organization of work. The stresses can, in some cases, create health risks.

3.1.4 Consequences of high work pressure in the construction industry

As Figure 1 shows, high work pressure can result in stress. This stress can lead to various complaints: complaints of a behavioral, psychological and physical nature. Examples of behavioral complaints include: sleeping problems, reduced appetite, greater use of medication, increased smoking, drinking more alcohol, increased restlessness and avoidance behavior. This can, in certain cases, result in (higher) absenteeism and (higher) turnover. At the psychological level, high work pressure can result in the employees feeling stressed, tense, anxious and insecure, having concentration and decision-making problems, and creating depression. Physically, high work pressure can result in complaints of fatigue, nausea and dizziness, chest and stomach pains, headaches, painful muscles, especially in the neck and shoulders.

Everyone can occasionally suffer from the above complaints. And they don't have to be a problem. They start becoming a problem when they occur too often, persist or get worse and worse. They then can become a serious health risk. It is important, of course, to prevent this from happening.

Complaints like these are also occurring more and more frequently in the construction industry. They are one of the biggest reasons for sickness and absenteeism of foremen in the construction industry. But they are also becoming more and more of a problem for the other employees on the construction site. Analyses of periodic health examinations show that psychological problems have occurred more often in the last few years, and that they are the fastest growing reason for absenteeism and incapacity for work.

When stress complaints occur too often, it is important to ascertain whether employees are suffering from high work pressure.

Figure 2: Model illustrating some pyschological demands of work

Table 18: Measures to reduce high work pressure

Measures	Affect created	Prevention	Implementation
Adjustment measures	Little	Low	Easy, quick
Improvement measures	Medium	Medium	Relatively easy
Redesing measures	Large	High	Difficult, time-consuming

3.2. Measures to Reduce High Work Pressure

In this section we outline measures which can be taken to reduce high work pressure in the construction industry. There are three sorts of measures (see Table 18): adjustment measures, improvement measures and redesign measures. They differ from each other in a) the extent to which they affect the organization of work and the employees involved; b) the effectiveness (the degree of prevention) and c) the speed and complexity of implementation.

The different measures all have the same aim: to create an optimum match between the work demands and the ability to cope with these demands. The measures are characterized by several critical success factors. Practice in different branches of industry has shown that every measure should contain the following elements. The elements are also embodied in the Netherlands Working Conditions legislation.

- Clarity and identification of tasks
- An employer must be properly informed about the demands asked of him and about the purpose and content of the work.
- Cooperation and support
- Employees must be able to cooperate with and support one another.
- Autonomy
- Employees must have the opportunity to solve problems independently, without being dependent on others.
- Feedback
- It is very important to give employees feedback on the quality of work and the way the work is carried out. Feedback provides opportunities to learn. Learning has a positive influence on the motivation and involvement of employees. Apart from that, it makes optimum use of the potential of an organization.

The following sections explain the different measures in more detail. Examples from practice are used to illustrate how the various measures can be implemented in the building industry. The examples are taken from "Welfare measures in the building industry: analysis and discussion" (Manshanden 1996). The section concludes with the result of the scientific research into the implementation of the measures.

3.2.1 Adjustment measures

These measures are important, but usually have a low prevention value. The aim of the measures is to create "realistic" work demands. The organization of work is not changed, but employees are better equipped to cope with work demands. The probability of work pressure is reduced.

The measures can be taken relatively quickly and easily. Foremen can carry out these measures independently from the rest of the organization. It is important, in the preparation of the measures, to involve as many employees as possible. Some examples are given below.

Clarity about goals and results of work:
- clear two-way communication: employees can indicate what their limits are;
- realistic scheduling and reasonable standards relating to quality and quantity;
- good training for inexperienced employees;
- clear orders that can be understood by everyone involved.

Materials:
- delivery of materials on time
- clear agreements on when and what materials are needed;
- a reserve stock that can be used as an extra buffer to keep employees going;
- an incoming inspection; to check when the materials arrive on site and to prevent the quality of materials from constantly changing.

Equipment:
- presence of the right tools and equipment;
- the tools and equipment store to be as close to the workplace as possible;
- regular inspection of tools and equipment, timely replacement of worn tools and equipment.

Reliable feedback of results:
- feedback, aimed at learning: not only negative feedback but also positive feedback is desirable;
- the possibility for employees to create their own feedback.

Working conditions:
- optimal protection from bad working conditions, i.e. protection against noise, humidity, rain, cold and from heavy work.

Cooperation:
- good coordination between the different activities and tasks;
- creating opportunities to discuss and tackle problems together; one way of doing this is by introducing regular and structured work meetings.

3.2.1.1 Adjustment measures: an example from practice.

In the last few years, (structured) work meetings have been introduced at several construction companies. The experiences of one of these construction companies are described here.

One of the purposes of introducing work meetings was: "to get things running better and more smoothly. This means you need clearly defined topics, which you can discuss with the men and the foremen. Sometimes there are situations being complained about. This can signal problems."

The results of introducing work meetings are said to be as follows: "if the work meetings go well, it means that you as a foreman can organize things better. You make everything clear and create a situation where the men put forward their experience. We are doing this in such a way, but . . . can we not do it better? They then come up with suggestions as to how to improve the work process. Much has happened as a consequence of the work meetings. The foremen get more feedback — also about the work of the subcontractors. The work is better coordinated, and matched to suit the different teams."

3.2.2 Improvement measures

Improvement measures are more preventive in nature. The central feature is the content of a job. There are various sorts of improvement measures, such as job enlargement, job enrichment and task (or job) rotation. Job enlargement means combining several tasks into one job. For example, employees on the

construction site carry out tasks relating to both foundation work, erection and roofing.

Task rotation is a variant of this measure. A job constitutes only one task, but employees change jobs every now and then. Examples include: concrete shuttering workers who work occasionally with steel reinforcement, or bricklayers who now and then install windows and door frames. The construction companies believe that the benefits of task enlargement and job rotation are as follows: variety of tasks for the employees, wider experience and multi-functionality of personnel; fewer hiatuses in planning; greater responsibilities of employees; greater commitment to the quality of work (and the results) and reduced numbers of mistakes and problems in the building process.

For the individual employees, combining tasks means greater responsibilities and more decision latitude when it comes to solving problems. Employees also get more involved because they have a better view of the end result.

The other measures are geared towards ensuring that jobs fulfil specific requirements. One of the most important elements which a job must embody is containing adequate latitude of decision. Second, a job must involve as few repetitive tasks as possible and have the maximum amount of variation in the degree of difficulty. The last requirement which can be demanded of a job is the opportunity to find out as much as possible about the result of the work. This can be achieved by a proper information supply and targeted feedback.

3.2.2.1 Sufficient decision latitude
Decision latitude relates to the opportunities granted to the employee of being able to solve problems in the job. A job must incorporate the possibilities of being able to influence (1) the tempo of work, (2) the working method, and (3) the sequence of working operations. The solving of problems is made easier if there is an opportunity for mutual contact and sufficient cooperation. Having decision latitude enhances the feeling of responsibility for the product being provided and also increases the opportunity to learn from mistakes made earlier. Employees who, for example, are not able to complete a work schedule (i.e. an organized task) drawn up by themselves feel more responsible for such a failure than when this schedule has been drawn up by their boss. Afterwards, they will prepare a different work schedule.

3.2.2.2 No repetitive tasks
A job must embody the minimum number of repetitive (short and repetitious) tasks. Instead of being assigned tasks for short periods to be performed at a fixed tempo, a worker can be given tasks for longer periods with greater scope to devise the schedule and fix the tempo and method of work to suit him- or herself.

3.2.2.3 Varying the degree of difficulty
Simple and difficult tasks will be alternated as much as possible in a job. Whether a task is perceived as difficult or easy depends on the person carrying out the task. On the other hand, it is possible to make a "difficult" task easier by getting workers to learn about their job. This can be done by proper instruction, tutoring by qualified instructors, training, work meetings and sufficient decision latitude.

3.2.2.4 Information and feedback
A worker needs information on the purpose and results of the work. This is required to help him understand the demands which are made on him and enable him to learn from this. The information must be available in full and when it is needed. Information is also needed about the ins and outs of the site where he works and his own particular department.

As regards the different sorts of improvement measures, it may also be said that a balance also needs to be found here between the load imposed on the employees and their ability to cope with it. For example, tasks may not be added to a job as the work stress on the employee increases.

3.2.2.5 Improvement measures: a practical example:
A specific form of job enlargement (in the construction industry) relates to so-called multi-functional working. Experience with multi-functional working has been acquired at a construction company in the west of the Netherlands. One of the aims behind the changeover to multi-functional working is described as follows: "We noticed that there was a demand for people in maintenance who, in addition to bricklaying, could also do the finishing work. For example, concrete formwork carpenters who could also carry out tiling or carpenters who could also lay bricks or glaze a window. Then we deliberately introduced multi-functional working via the sort of maintenance, which was being done at that time. Although maintenance work was being carried out on large numbers of houses, not much needed to be done in a particular house — for example, the replacement of kitchen units, bathroom fittings or a sash window. Does a glazier have to do ten houses once a week and make the inhabitants wait? Cannot this man of ours fit the window straightaway?"

> For maintenance and renovation work it was important to have fewer people moving in and out of the client's house. It would be better to deal completely with a few houses than to be doing the same job successively in a large number of houses. Because different specialist disciplines were being employed, there were often gaps in the planning. It was then a matter of working more efficiently for which the social aim — the broader use of people — also came into the picture.

Multi-functional working can also be used in the construction of new buildings: "At the moment, in new building work, we are attempting to get the people doing the foundations to also do the complete job. They are both laying the foundations, doing the assembly work and roofing and thus advancing."

Multi-functional working can also be used when different companies are involved:

> In new building constructions we use our own steel benders. If a steel bender has to fit reinforcement grids on tunnel formwork or his walls, he can get help if he needs it. You can see here where the subcontractor sometimes helps our people. While our steel bender is fixing the grids, he can also, for example, be laying the electricity conduits. It becomes a joint operation. This is Cupertino and has developed over a period of time.

As for the results of multi-functional working it can be said that no computation has yet been made as to precisely what benefits it brings about. In various fields, however, the results of multi-functional working are clearly to be seen. Recently our company was awarded the "Consumer Building Prize", the prize

for the construction company capable of organizing the building process to the maximum benefit of the consumer, tenant or buyer. This Consumer Building Prize is an evaluation criterion; it shows that the occupants are happy with us. But that apart, you may be wondering whether things are better in the company. I believe they are because it is now possible to do things that would previously have been impossible or more difficult.

What is the actual difference?

Fewer people are involved and we work more quickly. I do not need to measure what the result is. You can see for yourself. We wish to optimize things still further and can hold our own in the market. Thanks to our added value we are frequently upfront in terms of negotiations and tenders. The results are geared to the market. In the company you will see motivated people created by job enlargement and greater latitude of decision-making.

And what's more:

the percentage of mistakes is small because the people work better. There is a greater readiness to think things over and to cooperate. Better quality and fewer mistakes help to save money. Also, people like working in the company; you can see this at all levels, and also on the construction site. We never need to advertise for recruitment.

3.2.3 Redesign measures

Another effective way of overcoming work pressure is to take so-called redesign measures. These measures, like the improvement measures, aim to increase the decision latitude. Working in teams is central to this. (The idea of autonomous teams arises from the social engineering field. It is a business trend, which creates links between the division of labor and the productivity of a company.)

A team consists of a fixed group of employees who are jointly responsible for (part of) the (total) process in which products or services are created. The team draws up the plans itself and monitors the progress of the process, jointly solves day-to-day problems and improves processes and working methods. When carrying out these tasks the team does not need to refer constantly to the management or supporting services. In summary, a team has executive, organizational, preparation and supporting functions.

Other names used for such teams are self-governing or self-reliant teams, autonomous task groups or independent construction teams.

The introduction of team-oriented working often means fundamental changes in the organization of the company. However, increasing numbers of companies and institutions in differing branches of industry are utilizing the principles of team working. A basic principle behind team-oriented working is that people are the most important resource in the company and that a team is a powerful means of nurturing and utilizing this resource.

In teamwork, control does not take place at individual level, but at team level. This means that each team has its own tasks, responsibilities and powers. Practice has shown that successful team working features various critical success factors. They are described as follows:

- The group task must be complete, have clearly defined limits and be linked to a measurable result. The group task is a complete whole. When carrying out the group task, the team is independent of its environment.
- The entire team must have sufficient decision latitude.

- Team members are interdependent.
- The team consists of a minimum of four and a maximum of twenty persons.
- The members of the team are available for several tasks within the team.
- The team has one or more fixed reference points for the outside world, but also for its own team.
- The team must have its own room, own production resources and own information.
- Planning and quality systems and the budgeting need to be adapted for working with teams.
- The wages system must be adapted for working with teams.

The decision latitudes of a team can relate to different tasks and powers. It is possible that a team will gradually take on more and more tasks while some tasks may continue to be performed to a limited extent. A team which, for instance, controls its own planning must take account of the existing executives drawing up the general plans. Listed below are examples of tasks and powers, which can be assigned to a team:

- Planning tasks, work preparation
- Training and induction of new employees
- Administration
- Timekeeping, price control and costing
- Quality assurance
- Looking after materials (ordering, maintenance)
- Looking after working conditions and environment (accident prevention, security, order and tidiness on the site, but also recording of absenteeism)
- Commercial responsibilities, budgeting
- Contact with occupants in maintenance and renovation projects
- Direct contact with (sub)contractors and suppliers
- Higher quality of labor with, as a result, lower work pressure, higher commitment and motivation among the workers, lower absence due to sickness and a lower turnover.
- Higher effectiveness (the degree to which an enterprise achieves its objectives): high quality, short delivery time, high technical flexibility, less time consumed by executives.
- Higher efficiency (the degree to which a company is capable of achieving optimum output at the minimum possible cost): higher productivity, reduced stock costs, lower material losses and, generally speaking, a shift in the types of costs (from fixed to variable and from indirect to direct).

When autonomous teams are introduced, however, allowance must be made for the fact that this will have an influence on *all facets of business management:* the organization structure, the organization culture, the production structure etc. It also concerns redesign measures. Some of these facets are set out below:

- Extra instruction and training courses are often part of the change process. Ultimately, employees are given new and sometimes more taxing tasks.
- Wages and job evaluation systems should be changed because the present system places little value on the ability to perform more than one task. For instance, individual wage-rate systems can be replaced by team wages systems.
- The present control systems (financial or logistics control, planning systems) usually inhibit the independence of teams. Frequently, plans are too detailed and leave no scope for independent action.

- The reduction of the indirect costs is an important effect of the introduction of autonomous teams. This enables the number of staff and support services and the number of executives to be reduced to a minimum. However, this advantage will be achieved only gradually.
- Working with autonomous teams requires another style of management. The team is central to this and the management supports the team.
- The period of change while autonomous teams are being introduced is frequently a long process during which simultaneous changes also occur to the division of labor, the overall control and contacts between people. There is no instant blueprint available for the process of change: every company has specific problems, which require a specific solution. The change process is a learning process for everyone in the company.

3.2.3.1 Redesign measures: a practical example

As part of its work on redesign measures a construction company in the south of the Netherlands has been using autonomous construction teams on a renovation project. This team was made up of carpenters, bricklayers, a painter, a tiler and a plasterer. The plumber and electrician of the specialized contractor were permanently assigned to the team. Without the workers ever really practicing a different trade, there were three things which differed from the "normal manner of working". First, a helping hand was given to the other specialist disciplines. This can be seen as job enlargement in several disciplines. Second, job enrichment was created by allowing the teams to also carry out preparatory and organizational tasks. The workers organized the work themselves. For example, arrangements were made with the occupants: "We'll come to you tomorrow to do this or that". They also were responsible for accepting and ordering materials and equipment. The organization, moreover, was under the control of the foreman. Third, there was the matter of learning. Workers in an autonomous construction team learn as a result of the control tasks and by organizing the work. The results show that working with autonomous construction teams has delivered both social and economic benefits.

The following are some figures based on (scientific) research in practice:

- A study at a Swedish construction company, which was organized on the principles of team-oriented working, has stated that the following benefits were obtained by working with teams. The productivity of the construction company rose by 15%, there were fewer disruptions, 20% fewer mistakes, 4% cost savings, the workers suffered less stress. Material losses were reduced from 10–15% to 3–4% and the team spirit increased markedly. Absence due to sickness fell, as also did the number of accidents.
- In 1993 the Arbouw and Bouw Research foundations undertook an experiment with the fabrication of plaster block partition walls in house construction. Despite the fact that this work is not complicated, there are a lot of operations and specialized activities involved. This experiment revealed that team working resulted in benefits both in terms of production efficiency and working conditions. The production time per m² wall surface and the total construction time were reduced and there were fewer

coordination problems. If, for example, the electrician was not present at the agreed time, the autonomous team could continue working at half strength. With a more conventional working method such a hold-up virtually puts a stop to the building process. The delay initiates a chain reaction causing things to get worse and worse. This can result in high costs. In the experiment, use was also made of new technologies, and improvements have been reported in the area of physical stress. During the experiment the team did not perform any organizational tasks in administration, budgeting, timekeeping, etc.

4. FUTURE

As we look at the pace of innovation in other sectors of industry, the construction industry should be characterized as conservative. Work is still physically straining and work organization and working methods are traditional. However, construction companies start more and more to experiment with different ways of work organization. And, as shown in this chapter, this can result in improving the working conditions, reducing physical workload and work stress.

REFERENCES

General

KONINGSVELD E.A.P and MOLEN, van der H.F., 1997, History and future of ergonomics in building construction. In Abstracts from The First International Symposium on Ergonomics in Building and Construction (part of the 14th IEA Congress), Tampere, Finland.

Physical workload

ARBOUW, 1990–5, Surveys of sick leave data in the Dutch construction industry.

BROERSEN, J.P.L., BLOEMHOFF, A., VAN DUIVENBOODEN, J.C., WEEL, A.N.H., and Van DIJK, F.J.H., 1992, *Figures of the Atlas of Health and Work Perception in the Construction Industry* (in Dutch), (Amsterdam: Arbouw).

HOONAKKER, P.L.T., SCHREURS, P.J.G., Van der MOLEN, H.F., and KUMMER, R., 1992, *Conclusions and Evaluation of the Research Projects Concerning the Psychosocial Workload of Six Professions in the Construction Industry* (in Dutch), (Amsterdam: Arbouw).

HEDÉN, K., ANDERSEN, V., KEMMLERT, K., SAMDAHL-HIDEN, L., SEPPÄNEN, H., and WICKSTRÖM, G., 1993, Model for assessment of repetitive, monotonous work: RMW. In Marras, W.S., Karwowski, W., Smith, J.L., and Pacholski, L. (eds) *The ergonomics of manual work* (London/Washington DC: Taylor & Francis), pp. 315–17.

ISO/DIS 11226, 1999, *Ergonomics: Evaluation of working postures* (ISO/TC159/SC3 - JISC, Osaka).

KILBOM, Å., 1994a, Repetitive work of the upper extremity: Part I - Guidelines for the practitioner, *International Journal of Industrial Ergonomics*, 14, 51–7.

KILBOM, Å., 1994b, Repetitive work of the upper extremity: Part II - The scientific basis(knowledge base) for the guide, *International Journal of Industrial Ergonomics*, 14, 59–86.

MITAL, A., NICHOLSON, A.S., and AYOUB, M.M., 1993, *A guide to manual materials handling* (London/Washington DC: Taylor & Francis).

NF X 35-106, 1985, *Ergonomie: Limites d'efforts recommandées pour le travail et la manutention au poste de travail* (AFNOR, Paris).

NIOSH, 1981, *Work practices guide for manual lifting*, NIOSH Technical Report No. 81-122, US Department of Health and Human Services, National Institute for Occupational Safety and Health, Cincinnati, OH.

prEN 1005-4, 1998, *Safety of machinery - Human physical performance - Part 4: Evaluation of working postures in relation to machinery* (CEN/TC122/WG4 - NNI, Delft).

WATERS, T.R., PUTZ-ANDERSON, V., GARG, A., and FINE, L.J., 1993, Revised NIOSH equation for the design and evaluation of manual lifting tasks, *Ergonomics*, 36, 749–76.

Work stress

ARBOUW, 1996 and 1997, Working pressure in the building industry, (Amsterdam: Arbouw, internal Arbouw publication, in Dutch).

DEKKER, S.W.A. and SCHAUFELI, W.B., 1995, The effects of job insecurity on psychological health and withdrawal: A longitudinal study. *Australian Psychologist*, 30(1), 57–63.

KARASEK, R., 1979, Job demands, job decision latitude and mental strain: Implications for job redesign. *Administrative Science Quarterly*, June 1979.

KARASEK, R. and THEORELL, T., 1990, Healthy work. *Stress, Productivity, and the Construction of Working Life* (New York: Basic Books).

MANSHANDEN, 1996, Welfare measures in the building industry: analysis and discussion (FNV, in Dutch).

WARR, P.B., 1990, Decision latitude, job demands and employee well-being. *Work and Stress*, 4, 285–94.

Construction

C. L. Zimmermann†, J. C. Rosecrance†, T. M. Cook† and S. P. Schneider‡

†Department of Occupational and Environmental Health, The University of Iowa, Ames, IA, USA

‡CIH, Director of Occupational Safety and Health, Laborers' Health & Safety Fund of North America

1. INTRODUCTION

Seven million workers are employed in the US construction industry. As identified in many other countries (Holmstrom *et al.* 1995), the construction industry in the USA accounts for some of the highest rates of workplace injuries and deaths. High rates of work-related injuries and illnesses have traditionally been accepted as being part of the construction worker's job. According to the Bureau of Labor Statistics (BLS) Annual Survey of Occupational Injuries and Illnesses (BLS 1998), construction workers experienced > 189 900 illnesses and injuries causing lost work days in 1997. The actual number occupational injuries and illnesses in construction is probably under reported. For instance, the annual survey does not include companies with < 10 employees or self-employed workers, which combined account for ~80% of the construction establishments. Additionally, injury or illness underreporting in construction may be associated with injured workers' concerns about losing work and contractors concerns about their impact on workers' compensation premiums.

Although work-related musculoskeletal disorders among construction workers are less dramatic than traumatic injuries, they can be as debilitating and very costly. Many of these conditions have a long latency between initial exposure and symptom manifestation, which may contribute to their under reporting as an occupational illness. Chronic low-back pain and other musculoskeletal disorders like carpal tunnel syndrome cost millions of dollars in medical expenses, lost wages and workers' compensation payments. The disability and costs of these conditions has prompted federal agencies, trade unions, trade associations, contractors and researchers to investigate methods to prevent these disorders. Primary prevention efforts in the past 10 years have focused on the multidisciplinary field of ergonomics to assist in the reduction of the risk factors associated with work-related musculoskeletal disorders.

Although ergonomics is commonly associated with manufacturing facilities (electronics, meat packing, automotive industry) and the use of video display terminals, there has been considerable interest and research in construction ergonomics since the mid-1990s. Construction is a very broad field and the ergonomic challenges and solutions are difficult to summarize in a short review. There are many different types of construction work (residential, commercial, highway, industrial, demolition, etc.) and many different types of construction workers (carpenters, laborers, electricians, bricklayers, etc.). Clearly, there are ergonomic challenges everywhere in the industry and injury rates are high. The rate of back injuries and sprain and strain

injuries in construction is ~50% higher than the rates for manufacturing plants in the USA. Construction materials are heavy and often require considerable materials handling, frequently by hand. Building ceilings requires extensive overhead work for installation. Floors and roofs require work to be done at floor level.

Because the worker's job-site is constantly changing in construction, it is not possible to design a workstation as you would in an industrial facility, which may not change for several years. Most US construction companies are small, with 82% employing < 10 workers. Most job sites are multi-employer sites, where one contractor may create ergonomic challenges for workers working for another contractor. Construction workers change jobs sites frequently (one job may last a few days or several months) and often change employers frequently as well. Despite these differences, many of the same principles of ergonomics that apply to industrial ergonomic problems can be used in construction. The main approach that has been taken is to divide the construction worker's job into "tasks" and to look at the ergonomic problems and solutions for each task individually (Schneider 1995). Though solving task-specific ergonomic problems may have beneficial effects for individual workers, a systems approach involving workers, contractors, and owners may have a greater impact.

The purpose of this review is to provide a brief description of some of the physical demands, ergonomic challenges and ergonomic strategies that have been identified in several construction trades.

2. CONSTRUCTION LABORERS

In construction laborers end up doing many of the materials handling tasks and are, consequently, at perhaps the greatest risk of acute and chronic low back injuries. Laborers act as "mason tenders" handling all the materials necessary for bricklayers to perform their jobs. They do demolition work, where heavy construction materials are demolished and must be manually lifted, carried and thrown into refuse containers. Laborers use manual tools to dig trenches in pipeline work and to move asphalt in paving. All these tasks place them at great risk of injury. For most materials handling jobs there are carts, dollies and cranes that can be used to reduce the amount of manual handling stress. Unfortunately, little research has been done thus far on the ergonomic problems of laborers, in part due to the diverse nature of their work.

3. BRICKLAYING

Bricklaying, by contrast, has little task variation and consists of handling mortar, bricks, and blocks and moving scaffolding. There has probably been more work done on bricklaying ergonomics than most other trades. Bricklayers place mortar on the wall, lift bricks or blocks, set them in place, and remove the excess mortar. The ergonomic hazards of the job revolve around the placement of the mortar mix, the placement of the brick/block supply, the weight and size of the bricks/blocks, the location of the work (e.g. height of the wall), the work rate, rest cycle, and duration of work. Luttmann *et al.* (1991) measured muscle activity, time in awkward postures, brick holding time and productivity relative to wall height. They found that muscle activity and awkward postures were highest when bent over.

Holding time increased when the wall height was higher. Productivity was highest when the wall was low. They concluded that a wall height of ~80–100 cm is optimal for all factors.

Thus, one approach to improving ergonomics for bricklayers would be to maintain wall height, as much as possible, at 80–100 cm. The increasing use of adjustable scaffolding, which can be raised in ~30-cm increments, can help keep the wall height close to 80–100 cm. Studies done at the University of Texas have shown a 20% increase in productivity by using adjustable scaffolding. Bilevel scaffolds and mortar pan stands are also available to keep the brick and mortar supplies just below waist height.

The size and shape of bricks and blocks can also be modified. In Germany, block manufacturers worked together with contractors and unions to design a block with handholds to make it easier to lift and carry. In the US, the Army Corps of Engineers worked with the University of Nebraska to design the Nebraska A block, which weighs half as much as a normal block but with the same strength characteristics. Mechanical hoists are also available in Europe for lifting and placing blocks. Their use was spurred on by a law requiring any blocks weighing > 20 kg to be lifted mechanically. The Army Corps of Engineers in the US is also working on a mechanical arm to lift masonry blocks in place.

The other main ergonomic problem for bricklaying is the manual handling of bricks and mortar. Brick tongs are available for individual masons to use to handle small numbers of blocks. For bulk handling of bricks, the Dutch have developed an integrated materials handling system which includes special carts and dollies for handling bricks and new methods for packaging and shipping bricks allowing much of the handling can be done mechanically. Equipment is also available to pump mortar up to point of operation and apply the mortar pneumatically.

4. DRYWALL WORK

Drywall or plasterboard installation has been identified as one of the most hazardous occupations, according to workers compensation records from Washington State in the USA. Drywall installers have very high rates of back injuries and most people cannot perform this work for many years without injuring themselves. The job includes three main tasks: manual handling of boards, cutting boards to size and screwing the boards into the studs. Drywall installation is either vertical (for walls) or horizontal overhead (for ceilings).

Most of the ergonomic research on drywall tasks has focused on the handling of drywall boards. Drywall boards are heavy and awkward to carry. They are a minimum of ~30 kg and ~120 _ 240 cm in size. Often they are larger and heavier. Handling them outdoors, they can act as a sail and catch wind gusts, making them even harder to handle. There are several handles available for workers to use for carrying drywall. The handle merely hooks under the board and allows the worker to carry it with a much more comfortable, less extended grip closer to their body. Alternatively, carrying handles, which attach to either end of the board, allow two people to carry boards easily, which is probably a better solution for manual handling given the weight and size of the boards. Drywall carts are also available for transporting boards, but most require a flat surface, such as a finished concrete floor. A third option was developed in Sweden, where 90-cm wide boards were introduced. Research in Sweden and Finland showed the boards are much easier to handle, including allowing for more visibility during carrying, and reduced ergonomic risk. Research has not yet been done to look at the ergonomic tradeoffs of using 90-cm boards, e.g. they require more screwing.

Other interventions have focused on cutting and fastening drywall. Ergonomically designed utility knives with a bent handle are available for cutting the boards as are cutters that can be placed over the edge of the board and pulled. Screwing the boards into the studs has become much easier with the widespread use of powered screw guns, either electric or battery operated. Little attention has been paid thus far to the design of these tools. Workers often operate then by pushing on the back of the gun and using their little finger to operate the trigger. Automatic feed screw guns are widely available and extension handles allow for screwing drywall into the ceiling without an overhead reach.

The organization of the work is also important. One Swedish study showed that taking short rest breaks, "micropauses," while installing drywall reduces fatigue and increases productivity.

5. REBAR TYING/STEEL FIXING

Modern concrete is generally reinforced by the placement of iron reinforcement bars in the floor, bridge or column before pouring. The bars are tied together to make sure they maintain their position during and after the pour. Traditionally tying has been done by hand. Workers, called rodmen or steel fixers, tie the bars together at junctions using tie wire from a roll on their belts. They maintain a bent over posture for most of the day resulting in back problems and often have hand and wrist problems from twisting and cutting the tie wire. Finnish researchers first studied rodmen in the 1970s. Several different rod-tying machines have been developed in Europe, Japan and the USA for tying from a standing height. They eliminate the bent over posture and hand and wrist movements. While these machines have resulted in enormous interest among contractors worldwide, their use has not yet become widespread. There has been significant resistance to their use from workers who see the machines as deskilling their work, transforming their skilled and highly paid jobs into simpler, low paid work. Unlike the other ergonomic interventions previously described, this machine essentially changes the job completely. This makes the success of such interventions considerably more difficult, no matter what their ergonomic benefits.

6. WORK ORGANIZATION

Nothing is more important to the ergonomics of a construction site than proper work organization. The job superintendent and foremen must ensure that materials are delivered to the job site in a manner than minimizes manual handling. They must make sure that the proper equipment (cranes, hoists, dollies) is available, in good working order and easily accessible to the workers. They must plan the supply of materials so there is not a storage problem, i.e. materials get in the way of the work. They must make sure they have the right number of workers, i.e. they are not understaffed, and they have to plan the work to ensure that the various trades do not create obstacles to each others' work. Proper work scheduling to avoid excessive production pressures is also essential to good ergonomics. Walk-around checklists have been developed to help workers and management evaluate their worksites and improve ergonomics by addressing these issues.

7. PARTICIPATORY ERGONOMICS

Participatory ergonomics has long been an important part of ergonomics in industrial workplaces. During the past few years, it has also begun to be practiced more in the construction setting. The approach has been to develop a small ergonomics steering committee composed of workers, a management representative and a researcher to facilitate the process. Through a series of short meetings the committee develops a short list of problem tasks to address, a list of potential solutions, ranks those solutions and tests them out on the job site. Solutions are evaluated and the most successful ones adopted. The best example of this in construction has been a Dutch project to develop ergonomic interventions for scaffold erectors.

Scaffold erection had been identified as one of the highest risk trades. One of the major problems the committee identified was that scaffold parts arriving from the rental company were thrown disorganized into the back of a pickup truck. By insisting that the parts be placed in the truck in the order they were to be used, an hour of back breaking work was avoided. They also recommended the use of smaller ladders and planking. They recommended using a small electric winch to raise parts to the top of the scaffold construction. These interventions were evaluated by measuring heart rate, percentage of time in awkward postures and frequency of heavy materials handling (> 20 kg). The results displayed improved ergonomics and increased productivity. This participatory approach shows enormous promise in construction. Construction workers are generally very good problem-solvers. Every day construction activities result in problems that must be solved on the spot. Ergonomic problem solving is no different and should be a natural extension for construction workers.

8. OTHER ISSUES

8.1. Ergonomics Training

Training is an essential part of any successful ergonomics program. Workers need initial awareness training on the principles of ergonomics and methods for recognizing hazardous tasks. They also need training on how to identify potential solutions and solution implementation strategies. Several training programs are now available for construction workers in the US. Most focus on training apprentices, those just beginning their careers in the trade. Training would also be helpful for foremen and superintendents. They run the job sites and if they understand and are supportive of ergonomics change comes much easier.

8.2. Stretching Programs

During the past few years pre-work stretching programs have become popular. Workers go through a series of routine stretches each morning before the day begins. Studies in the USA and Sweden have shown these programs to be beneficial, particularly if they are mandatory and on paid time.

8.3. Personal Protective Equipment

Personal protective equipment, such as kneepads, is essential for construction workers, primarily to reduce the contact stresses from kneeling, etc. Some work in kneeling postures is unavoidable in construction as floors and roofs must be installed. Good kneepads can help although they can create pinch points in the back of the legs and may be uncomfortable to wear. Newer models are available which show some promise in reducing discomfort. Studies are underway to evaluate kneepads and their use. Carrying materials on your shoulder makes some sense ergonomically, in that you keep the materials close to your center of gravity and use your legs to support the weight rather than your arms. However, there is little padding on the shoulders. Shoulder pads are available for reducing contact stresses associated with carrying materials on the shoulder. Standing on a concrete floor all day has been associated with back problems. Rubber matting and shoe insoles can help reduce the pressure from standing on concrete and may help reduce the risk of back pain. Lastly, back belts have become popular in construction. Unfortunately, there have been few studies of their efficacy in preventing injuries and no studies in the construction industry. Recommendations on their use must wait the outcome of research currently underway.

9. CONCLUSION

The construction industry presents a high risk of musculoskeletal injury and many ergonomic challenges. Despite these difficulties, many strategies can be implemented to reduce the risk of injury (Schneider 1998). New materials, new tools and ergonomically designed equipment all have a role to play. As do stretching programs and personal protective equipment, such as kneepads. Improved work organization and better job planning are probably the most important ergonomic interventions in construction while training of workers and supervisors is also critical. Participatory approaches to ergonomics in construction hold great promise for making effective changes in work methods, work organization and injury prevention.

REFERENCES

BUREAU OF LABOR STATISTICS, US DEPARTMENT OF LABOR, 1998, *Workplace Injuries and Illnesses 1997*. News Release, 17 December, Table 2. Number of nonfatal occupational injuries and illnesses by selected industries and case type, 1997.

HOLMSTRÖM, E., ULRICH, M. and ENGHOLM, E., 1995, Musculoskeletal disorders in construction workers. In Ringen, Englund, Welch, Weeks and Seegal (eds) *Occupational Medicine: State of the Art Reviews V.*, *Construction Safety and Health, 10*(#2) (Philadelphia: Hanley & Belfus), 295–312.

LUTTMANN, A., JAGER, M. and LAURIG, W., 1991, Task analysis and electromyography for bricklaying at different wall heights. *International Journal of Industrial Ergonomics*, 8, 247–60.

SCHNEIDER, S., 1995, Implement ergonomic interventions in construction. *Applied Occupational and Environmental Hygiene*, 10, 822–4.

SCHNEIDER, S., 1998, Ergonomics in the construction industry. In W. Karwowski and W. S. Marras (eds) *Occupational Ergonomics Handbook* (Boca Raton: CRC Press) (in press).

SCHNEIDER, S. and SUSI, P., 1994, Ergonomics and construction: a review of potential hazards in new construction. *American Industrial Hygiene Association Journal, V*, 55, 635–49.

Diagnosis of Work-related Musculoskeletal Disorders

C. Dillon† and M. Sanders

†School of Medicine, University of Connecticut, Farmington, CT 06030, USA

‡Department of Occupational Therapy, Quinnipiac College, Hamden, CT 06518, USA

1. INTRODUCTION

The diagnosis of a cumulative trauma disorder (CTD) poses a challenge to practitioners who treat workers with prolonged exposure to workplace ergonomic hazards. Although controversy surrounding the diagnosis of a CTD is not unlike that of other areas of medicine, a coherent understanding of the problem is necessary since CTD have such wide reaching impact on industry, insurance and medical systems. This chapter discusses CTD from the perspectives of these various sectors and reviews the nomenclature used to reference CTD. It explains the major diagnostic criteria that identify the presence of a CTD and addresses specific diagnoses that typify CTD in the workplace. The controversial issues relative to CTD diagnoses are examined.

2. DEFINITION AND NOMENCLATURE OF CTD

2.1. Definition of CTD

CTD is a broad term that refers to injuries of soft tissue structures of the body such as muscles, tendons, nerves, blood vessels and joints which have identified causal link to ergonomic risk factors. As the name implies, CTD is not the result of a fall or a sprain that may have occurred in a single incident. CTD develops insidiously from prolonged exposures to repetitive motions, awkward postures, static posture, vibration and forceful exertions that injure soft tissues and prevent them from healing (Buschbauer 1994). While researchers equate CTD primarily with workplace ergonomic factors, central nervous system and psycho-organizational factors also contribute to the problem (Gordon *et al*. 1995, Sanders 1997).

CTD encompass specific medical diagnoses including tendinitis, epicondylitis and entrapment neuropathies. While diagnostically similar disorders result from ergonomic exposures in sports or avocational activities, or from inflammatory arthritis, endocrine disorders and certain hereditary syndromes, these conditions should not be confused with CTD resulting from workplace ergonomic exposures.

2.2. Nomenclature Used to Reference CTD

The scientific and popular literature is inconsistent with regard to nomenclature for workplace CTD. The various terms for this phenomena include CTD, repetitive strain injuries (RSI), repetitive motion disorders (RMD), work related musculoskeletal disorders (WRMSD), occupational cervicobrachial disorder (OCD) and upper extremity repetitive strain disorders (UERSD). Each term conveys important meaning, yet is inaccurate in the entire portrayal of the condition. For example, the term "injury" is misleading since these disorders are characterized by a gradual onset and exacerbation; "musculoskeletal disorders" lacks diagnostic specificity for the area targeted. Since no truly satisfactory term for CTD exists, the term CTD is chosen here for the greatest currency and recognition.

3. DIAGNOSTIC CRITERIA FOR CTD

An interconnected variety of clinical syndromes exist under this "umbrella" of CTD. Confusion surrounds the issue of diagnostic criteria for CTD due to a lack of clinical case definitions for many CTD; and, equally as important, the differing clinical and scientific diagnostic perspectives held by the various contributors (Rempel *et al*. 1998). A brief review of these different perspectives and accompanying criteria is presented.

3.1. Medical CTD Case Definitions

A medical diagnosis relates to a clearly defined disease process involving pathology of an anatomically defined region or physiologically defined process. The medical diagnosis determines the nature of the disease and enables appropriate treatment selection. Medical diagnostic criteria facilitate making the diagnosis and consist of a structured set of symptoms and physical exam signs. Ideally, a confirmatory lab-based test or "gold standard" acts as the ultimate objective test to verify the existence of a disorder.

The medical diagnostic criteria are combined with a medical diagnosis in a medical case definition which describes an instance of a disease with its accompanying circumstances. A medical case definition should be both sensitive and specific. That is, the case definition should identify the overwhelming majority of clinical cases, yet not falsely include those cases superficially similar to but inherently different than the condition of interest.

For most CTD rigorous case definition criteria are not met because the ultimate pathophysiology and manifestation of disorders are not thoroughly understood. For conditions such as carpal tunnel syndrome (CTS), consensus case definitions have

Table 1. Diagnosis and Classification of Cumulative Trauma Disorders

Surveillance Case Definition Used to Determine the Presence of a Cumulative Trauma Disorder

Interview

- Symptoms of pain, numbness, or tingling
- Symptoms lasting more than 1 week and/or occurring more than 20 times in the last year
- No evidence of acute traumatic onset
- No evidence of systemic disease
- Onset of symptoms occurring with present job

Physical examination

- Characteristic signs of specific muscle, tendon, or nerve disorders
- Rule out other conditions with referred symptoms

Sources: Adapted from SILVERSTEIN, B.A., FINE, L.J. and ARMSTRONG, T.J. 1986, Hand wrist cumulative trauma disorders in industry. *British Journal of Industrial Medicine*, 43, 779; and SILVERSTEIN, B.A., FINE, L.J. and STETSON, D. 1987, Hand-wrist disorders among investment casting plant workers. *Journal of Hand Surgery* (Am), 12, 838.

been developed, which identify typical cases and allow for standardized patient enrolment into medical and scientific clinical trials (Rempel *et al.* 1998). Consensus case definitions are designed to recognize the typical case and usually do not encompass the early, mild or atypical patients with an existing disorder. Additionally, published medical case definitions in the CTD literature may not accurately reflect the circumstances of an individual patient examined in everyday clinical practice. Table 1 summarizes the classic physical signs and symptoms associated with common medically defined CTD diagnoses.

3.2. Surveillance Case Definitions

Surveillance case definitions are primarily used in regulatory monitoring to track the impact of the workplace on the incidence of CTD. Surveillance case definitions may incorporate all or part of existing medical case definitions. These definitions may consist of brief symptom surveys likely to be attributed to CTD or may employ a full range of symptom, physical exam and laboratory testing data. A key characteristic of surveillance case definitions is the frequent inclusion of ergonomic exposure variables. For example, work-related carpal tunnel syndrome might be defined as a case with characteristic symptoms, exam findings, lab tests *and workplace exposure history* for surveillance purposes.

The Occupational Safety and Health Administration (OSHA) (1990) in the USA utilizes broad surveillance criteria for reporting, recording and tracking CTD as part of compliance with regulatory standards. According to OSHA, if a worker's symptoms are *caused or aggravated by exposure to work* and meet recordability criteria then the case is considered to be a work-related CTD. The recordability criteria specifies the presence of subjective or objective findings with action taken as a consequence. A condition qualifies as a "recordable" CTD, if the following conditions 1 *or* 2, AND condition 3 are met: (1) the worker must possess one or more physical findings such as redness, loss of motion, deformity, swelling, a positive Phalen's, Tinel's test, or other positive provocative tests *OR* (2) the worker must possess one or more subjective findings such as pain, numbness, tingling, aching, stiffness, or burning AND (3) actions were taken as a result of this condition such as medical treatment (including self-administered), lost workdays, loss of full-duty status or transfer to another job. The presence of work-related exposures is inherent in the definition of a work-related CTD.

If a worker's condition fulfils these parameters, the condition is considered to be an occupational illness and must be recorded on the OSHA 200 Log under the 7f column, "disorders associated with repeated trauma" (OSHA 1990). Table 2 provides an example of a more strict surveillance case definition utilized in the studies by Silverstein *et al.* (1986).

3.3. Epidemiological Case Definitions

Epidemiologic case definitions are a third type of case definition found in the CTD literature (Rempel *et al.* 1998). These are stricter scientific case definitions with more rigorous, reproducible protocols for research data acquisition and inference. Epidemiologic CTD case definitions are typically employed for investigations of etiologic factors and natural history of CTD as well as for clinical intervention trials. While many epidemiologic CTD case definitions attempt to achieve fidelity with clinical case definitions, this is not always the case. Epidemiologic case definitions typi-

cally avoid the inclusion of ergonomic exposure data for studies examining the etiology of CTD.

4. PERSPECTIVES ON CTD DIAGNOSES

Parties that routinely deal with workplace CTD include med-changes of the tendon occurring in the presence or absence of inflammation (Gordon *et al.* 1995). The most common sites for tendon injury from repetition or overuse are the tendon attachment to bone, the myotendinous junction (point at which tendon attaches to muscle) and sections of the tendon that pass beneath a tunnel or pulley such as in the wrist and fingers.

5.3.1.1. Factors affecting the healing process.

Factors such as age, disease, genetics and medications significantly affect the healing process and integrity of tendons. The tendon strength of a 70-year-old person is ~21% less than that of a 25-year-old person because the collagen stiffens (loses elasticity) which reduces tensile strength (Leveau 1992). It is unclear the extent to which reduced strength is due to intrinsic qualities of the collagen or due to years of repetitive use and microtears.

5.3.2. Nerve-related disorders

Nerves relay messages of intent from the brain to muscle and receive sensory information from the periphery about the environment that is delivered to the brain. Hence, nerves have motor and sensory components. *Peripheral nerve compressions* occur as nerves are compressed by adjacent tissues or an abnormality in tissue structures which reduce the space through which the nerve travels. A *peripheral nerve entrapment* is caused by repetitive trauma to the nerve, which results in inflammation and later fibrotic changes. *Double crush syndrome* refers to the concept that nerve entrapment may occur at several points along the nerve. Compression and entrapment are likely to occur together as a nerve that is already traveling in limited space may be especially vulnerable to repetitive trauma by adjacent tissue.

In the early stages of nerve compression or entrapment, intermittent tingling and parasthesias may be provoked by certain positions but that later resolve with rest. Over time, the numbness and tingling become more pronounced, with numbness occurring first occasionally then constantly. Later, motor changes ensue; these changes may begin with muscle aches that may progresses to weakness and atrophy if nerve damage persists. Proximal sites of nerve compression in the neck and shoulder may increase the susceptibility of nerve compression at more distal sites such as in the wrist and hand (Novak and Mackinnon 1997).

Certain joint positions increase the shear forces between nerve and surrounding structures causing the nerves to be more vulnerable to entrapment if these positions are sustained. For example, extremes of wrist flexion and extension increase tension and shear forces on the median nerve within the carpal canal; elbow flexion increases tension on the ulnar nerve within the cubital canal. Irritation of the nerve at these sights will lead to edema fibrosis of the nerve, decreased excursion, and potentially impaired blood supply.

6. CONTROVERSIAL ISSUES RELATED TO CTD
6.1. Causal Factors in CTD Diagnoses

Researchers debate whether CTD should be attributed primarily to psychosocial, organizational or biomechanical factors within

Table 2. Common diagnoses related to CTDs. Medical and lay terms for each dondition are included as well as a description of the condition, the classic signs and symptoms, and causes of common CTDs.

Common Term	Medical Term	Description	Signs/Symptoms	Causes
Tendon-Related Disorders				
Tendinitis	Tendinosis Tenosynovitis	Inflammation of tendon sheath; becomes thickened, and frayed	activity-related sharp pain specific to an anatomical location	Repetitive use of tendons especially in awkward positions; assembly work, VDT work, wiring, packaging
Trigger finger or thumb	Stenosing tenosynovitis of the fingers or thumb	Thickened synovium accumulates as a "nodule" on the tendon. Nodules may catch and prevent movement of tendon under A1 pulleys in palm	Finger becomes "locked" when attempting to extend. Motion occurs in a snapping fashion	Repetitive pinch and grip; operating a trigger, use of hand tools with edges that compress tissue
deQuervain's tenosynovitis	Stenosing tenosynovitis of the tendons at the radial wrist	Inflammation and entrapment of the extensor pollicis brevis and abductor pollicis longus tendons of the thumb under the extensor retinaculum	Pain with thumb extension and abduction; positive Finklestein's	Typing with frequent use of space key, buffing grinding, sanding, use of small tools, forceful hand wringing
Tennis elbow	Lateral epicondylitis	Degeneration of the extensor tendons originating at the lateral epicondyle of the humerus	Pain at lateral elbow; pain with resisted wrist extension and supination	Repetititve wrist extension and repetititve grasp; VDT work, hammering, meat cutting, small assembly, musical instruments
Golfer's elbow	Medial epicondylitis	Degeneration of the flexor tendons originating at the medial epicondyle of the	Pain at the inside of elbow; pain with resisted wrist flexion and pronation	Repetitive wrist flexion; wringing or twisting hands, VDT work, golf
Shoulder tendinitis	Rotator cuff tendinitis, Supraspinatus tendinitis	Inflammation of the rotator cuff tendons as they pass beneath the acromion process and/or insert into humerus	Pain at mid range of shoulder motions (painful arc) and with resistance to abduction and external rotation	Overhead work; work with the arms away from the body; electrical work, grinding, sanding, soldering
Nerve-Related Disorders				
Carpal tunnel syndrome	Carpal tunnel syndrome	Compression of the median nerve in the carpal canal at the wrist	Numbness and tingling in the thumb, index, third and radial aspect of fourth finger; difficulty with dexterity, loss of prehension strength; positive Phalen's and Tinel's sign	Repetitive or forceful grip and prehensile tasks especially in extremes of flexion, extension and ulnar deviation; dental hygiene, surgery, VDT work, scrubbing, hammering, cutting
Cubital tunnel syndrome	Ulnar nerve entrapment	Compression of the ulnar nerve at the inside of olecranon groove of the elbow	Tingling, numbness or pain at elbow radiating to the ring or little finger; intrinsic hand weakness	Resting the elbow or forearm on a sharp edge or hard surface; repetitive or static elbow flexion
Guyon tunnel syndrome	Biker's fingers	Entrapment of the ulnar nerve at the wrist	Tingling, numbness or pain in the ring and little finger; intrinsic weakness	Repeated/prolonged pressure on the outside of the palm; prolonged flexion and extension of the wrist; biking, carpentry, use of pliers, hammering
Neurovascular-Related				
Thoracic Outlet Syndrome	Brachial Plexus compression injuries	Compression of the neuro-vascular bundles between the clavicle and first and second ribs at the brachial plexus; scalene triangle	Numbness/tingling in the ring and little fingers; pain at the brachial plexus; cold hands; intrinsic muscle weakness	Work with arms overhead or in front of the body; mechanical stresses applied to the brachial plexus; office workers, electricians, cashiers, postal letter carriers, musicians, waitresses

the workplace. The predominant evidence suggests that bio-mechanical factors inherent in the performance of work tasks initiate the development of tissue damage and resulting CTD diagnosis; personal, psychosocial and organizational factors such as work style, control over one's job, machine pacing, and stress in the workplace further contribute to or exacerbate the physical stressors in the disease process (NIOSH 1997, Sanders 1997, National Research Council 1998).

A group of prominent physicians view sociopolitical, le-gal and worker's compensation issues as the primary causal factors in the development of CTD. This conviction is partly driven by investigation of the 1980–85 Australian "RSI epi-demic" — a time period in Australia when reports of repeti-tive strain injuries rose > 300% over 3 years. Researchers criticized the medical system for inconsistencies in diagnosing the source of pain and observed that the magnitude of injury did not always seem to correlate with the work being performed.

While disputes continue as to whether political inter-ests, strong union support, litigation and substantial work-er's compensation awards are major factors driving CTD claims (Hadler 1992) research shows that the pool of unre-ported CTD cases is much larger than those ever reported to worker's compensation systems. The more salient question to address is the relative contribution of biomechanical and psycho-organizational factors to the prevalence of workplace CTD.

7. CONCLUSION

CTD diagnoses require expertise in evaluating workers in light of the biomechanical, personal and work organzational factors that cause such a condition. Utilitarian variations on CTD case definitions exist, including those for traditional medical diag-nosis and treatment planning, case definitions for safety and reg-ulatory monitoring and case definitions for epidemiological research on the natural history of CTD. Although each definition has a different practical purpose, these seemingly distinct demarcations often overlap in content and have complementary roles in the common ultimate goals of treating or preventing the development of CTD.

REFERENCES

BUSCHBAUER, R., 1994, Tissue injury and healing. In Buschbacher (ed.) *Musculoskeletal Disorders: A Practical Guide for Diagnosis and Rehabili-tation* (Boston: Andover Medical).

GORDON S.L., BLAIR, S.J. and FINE, L.J. (eds), 1995, *Repetitive Motion Disorders of the Upper Extremity* (Rosemont: American Academy of Orthopedic Surgeons).

HADLER, N.M., 1992, Arm pain in the workplace. *Journal of Occupational Medicine*, 34, 113–19.

KROEMER, K., 1992, Avoiding cumulative trauma disorders in shops and offices. *American Industrial Hygiene Association Journal*, 53, 599–600.

LEVEAU, B.F., 1992, *Williams and Lissner's Biomechanics of Human Motion* (3rd ed.) (Philadelphia: Saunders).,

NATIONAL INSTITUTE FOR OCCUPATIONAL SAFETY AND HEALTH, 1997, *Musculoskeletal Disorders and Workplace Factors: A Critical Review of Epidemiologic Evidence for Work-related Musculoskeletal Disorders of the Neck, Upper Extremity and Low Back.* DHHS (NIOSH) Publication No. 97-141 (Cincinnati: US Department of Health and Human Services).

NATIONAL RESEARCH COUNCIL, 1998, *Work-related Musculoskeletal Disorders: A Review of the Evidence* (Washington, DC: National Academy Press).

NOVAK, C.B. and MACKINNON, S.E., 1997, Repetitive use and static postures: a source of nerve compression and pain. *Journal of Hand Therapy*, 10, 151–9.

OCCUPATIONAL SAFETY AND HEALTH ADMINISTRATION, US DEPARTMENT OF LABOR, 1990, *Ergonomics Program Management Guidelines for Meatpacking Plants*. OSHA Publication 3123 (Wash-ington, DC: US Department of Labor).

REMPEL, D., EVANOFF, B., AMADIO, P.C., DEKROM, M., FRANKLIN, G., FRANZBLAU, A., GRAY, R., GERR, F., HAGBERG, M., HALES, T., KATZ, J.N. and PRANSKY, G., 1998, Consensus criteria for the classification of carpal tunnel syndrome in epidemiologic studies. *American Journal of Public Health*, 78, 1447–51.

SANDERS, M. (ed.), 1997, *Management of Cumulative Trauma Disorders* (Boston: Butterworth-Heineman).

SILVERSTEIN, B.A., FINE, L.J. and ARMSTRONG, T.J., 1986, Hand–wrist cumulative trauma disorders in industry. *British Journal of In-dustrial Medicine*, 43, 779–85.

SILVERSTEIN, B.A., FINE, L.J. and STETSON, D., 1987, Hand–wrist disorders among investment casting plant workers. *Journal of Hand Surgery (AM)*, 12, 838.

Epidemiology: Principles and Approaches to Prevention of Occupational Injury

G. S. Sorock† and T. K. Courtney‡

†Liberty Mutual Research Center for Safety and Health, Hopkinton, MA 01748, USA

‡Harvard School of Public Health, Boston, MA 02115, USA

1. INTRODUCTION

Epidemiology is a set of established and developing methodologies used to describe the distribution and determinants of adverse health outcomes in human populations. The subsequent evaluation of interventions to prevent injury or disease is the ultimate objective of epidemiology.

2. FUNDAMENTALS OF EPIDEMIOLOGY

2.1. Fundamental Aspects of Epidemiologic Methods

People who develop a health condition tend to be exposed to antecedent risk factor(s) more than persons who do not develop the condition. Epidemiologic methods have evolved to approximate the effects of potential risk factors (also known as exposures) on health outcomes (such as injuries, disorders, diseases) in groups of individuals. Associations between exposures and outcomes may be causal or due to alternative explanations such as study bias, confounding or chance.

In epidemiology, study bias refers to some systematic error in study methods that precludes a clear judgment about the nature of the association between the exposure and the disease outcome. For example, parents of children injured at home may recall certain exposures more thoroughly than parents of uninjured children. This *recall* bias may be estimated by comparing questionnaire responses to a "Gold Standard" source of information, perhaps a medical record or a direct observation of an exposure.

Confounding refers to the effect of an intermediary factor associated with both outcome and exposure. That is, the exposure–disease association may be artificially produced or obscured by this other factor. For example, cigarette smoking potentially confounds any observed association between alcohol use and lung cancer because many alcohol users also smoke.

Chance or sampling variability operates in most scientific research. It is a potential explanation for an observed association between an exposure and a disease outcome. It may be less likely, however, with larger sample sizes and as a study's findings are replicated by subsequent investigations.

2.2. Measures of Outcome Frequency

Important terms in interpreting and reporting outcomes in populations include incidence and prevalence. Incidence refers to the number of new instances of a health outcome in a population during a specified time period (Last 1988). For example, an *incidence rate* is defined as the number of new cases of a condition per persons at risk (usually 100 or 1000) per unit of time (usually one year). Prevalence is the count of all existing cases during a specified time period or point in time divided by total persons at risk. When not otherwise indicated, the prevalence usually reported is the *point prevalence* or the count of all cases in existence at one point in time.

2.3. Common Measures of Exposure–Outcome Association

Epidemiologic data are often summarized in tabular form in the initial stages of analysis as the presence or absence of exposure to some potential risk factor and the presence or absence of a health outcome in a group of individuals. Table 1 presents the results of a hypothetical study of the effects of a single exposure on a health outcome. Each cell (a, b, c, d) contains the number of persons with a particular combination of exposure and outcome status. The data in these four cells are used to derive various measures of effect. Two common measures of the effect of exposure on outcome are relative risk (RR) and odds ratio (OR). These are calculated and interpreted below.

If 1000 persons *with* and 1000 persons *without exposure* are followed-up over a specified period for outcome occurrence, RR is found by dividing the risk in the exposed group (75/1000) by the risk in the unexposed group (25/1000). The result is 0.075/0.025 or 3.0. This means that this sample of exposed persons had three times *the risk or probability of developing the outcome* than unexposed persons over the specified period.

If individuals are *selected on the basis of their outcome status*, the totals for the rows (a + b) and (c + d) are not predetermined. The relative risk can be estimated by calculating the ratio of the odds of exposure among cases (75/25) to that among controls (925/975) or ad/bc = 75 x 975/25 x 925. The resulting OR of 3.2 is close to the RR of 3.0 found above. OR is interpreted to mean that the cases are 3.2 times more likely to have been exposed than non-cases.

Any estimate of effect varies within a range of values that is termed a confidence interval (CI). CI depends on the magnitude of the estimate, the specified probability of including the true value of the estimate (usually 95%) and the sample size in each of the four cells of the table (the larger the sample, the smaller the range of the interval). For example, the 95% CI for the calculated OR from the study above ranges from 2.1 to 4.9 (for details of CI calculation and underlying assumptions, see Hennekens and Buring 1987: 255). This means that if the same study was repeated 100 times, 95 out of 100 times the true OR would lie between 2.1 and 4.9. If OR = 1.0, then there is no observed

effect of the exposure on outcome status. Because the interval calculated previously did not include the null value (1.0), CI indicates that the association is statistically significant. If the 95% CI covered 0.9–7.1, the result would not be statistically conclusive but strongly suggestive of an effect since most of the interval is > 1.0. If the point estimate and the interval are < 1.0 (e.g. 0.3, 0.1–0.5), then the exposure significantly protects against the outcome.

2.4. Epidemiologic Study Designs

The techniques illustrated above are applied to the study of exposure and outcome associations. As with most scientific inquiry, however, the techniques are merely assets supporting the execution of a study design. There are two main classes of study design in epidemiology: descriptive and analytic. Descriptive epidemiology describes *how* an outcome such as an injury or disorder is distributed in a population; analytic epidemiology tries to explain *why* (Kelsey *et al.* 1989). A brief description of each design, with examples, follows. More specific information about epidemiologic study methods can be found in standard introductory texts (Kelsey *et al.* 1986, Hennekens and Buring 1987).

2.4.1. Descriptive studies
2.4.1.1. *Case report and case series*
The case report is the complete description of the clinical characteristics of a single patient. The single patient may present some unusual finding that can suggest the need for a larger study of an exposure–outcome relation. Case series describe the clinical characteristics of two or more patients with a health outcome and perhaps an exposure in common. An advantage of the case series is that multiple observations allow compilation of potential trends in the causes, diagnosis and treatment of the disorder.

These types of studies are most useful for identifying and describing emerging outcome trends and for generating hypotheses concerning potential exposure–outcome relations, which may be tested in more analytic studies. However, the conclusions that can be reached about the existence and magnitude of such relations are limited with these designs due primarily to the lack of both exposure data and a comparison group.

2.4.1.2. *Cross-sectional*
Cross-sectional studies take a simultaneous look at the presence or absence of both exposure and disease in individuals. Such studies ask the question: what is the relation between exposure and disease at the same point in time? For example, a study of sedentary workers and low back pain examined the relationship between constrained, sedentary work and prevalence of low back pain (Burdorf *et al.* 1993). Occupations with the greatest proportion of time spent in non-neutral trunk postures were observed to have a significantly increased prevalence of low back pain. In a cross-sectional design, it is difficult to establish whether the potential exposure variable preceded the outcome (e.g. work posture differences contributed to the development of low back pain) or whether the potential exposure variable exists as a result of the outcome (e.g. workers differ in posture as an adaptation to low back pain). Hence, cross-sectional studies are useful for identifying potential exposure–outcome associations but not for establishing causality.

2.4.2. Analytic studies
Case-control, cohort and intervention studies are generally classified as analytic studies. These study designs make use of the temporal relation between exposure and outcome (i.e. exposure precedes the outcome) and identify potential causes of outcome more definitively than descriptive studies.

2.4.2.1. *Case-control*
In a case-control study, people with the outcome (cases) and people without the outcome (controls) are compared with respect to the proportion in each group with the historical exposure of interest. An advantage of this design is that both cases and controls may be matched on potential confounding variables, such as age. This design is particularly useful for the study of rare health outcomes. A limitation of this design is its susceptibility to recall and other forms of information bias since the exposure must typically be recalled by the cases and controls or be present in recorded data such as hospital records.

2.4.2.2. *Cohort*
A cohort study refers to a study design in which a group of persons who share a common exposure are studied over a period of time. Cohort studies are distinguished from case-control studies by two main features. First, classification into comparison groups is based on the exposure factor not the outcome. Second, the cohort study looks from the exposure forward rather than from the disease backward in time.

There are two main types of cohort studies: prospective and retrospective. The feature that distinguishes a prospective from a retrospective cohort study is whether the outcome of interest has occurred when the investigator initiates the study. In a prospective cohort study, the outcome occurs after the exposure is measured. In a retrospective cohort study, the investigation is initiated after both the exposure and the outcome have occurred.

2.4.2.3. *Intervention*
An intervention design is a prospective cohort study that investigates the impact of a treatment or control action on the exposure–outcome association. Study participants are assigned the treatment (intervention) at random and then followed to determine the proportion of persons exposed who develop the outcome compared with the proportion of persons unexposed (non-intervention). Assignment to an exposure category at random, also achieves, on average, equal distribution of other known and unknown confounding variables.

The study designs just discussed are presented in figure 1 to demonstrate how increasing knowledge of an exposure–disease relationship might develop for a hypothetical disorder (D) with a given exposure (E). Designs that accommodate the temporal relationship between an antecedent exposure and the outcome are considered most convincing of the strength or weakness of the E–D relationship. Where alternative explanations for this relation can be ruled out, for example by statistical adjustment of potential confounding variables, the results are more convincing.

3. APPLICATION OF EPIDEMIOLOGY TO WORKPLACE INJURY

Epidemiologic methods can be applied to the evaluation and prevention of injury in the workplace. Ultimately any attempt to
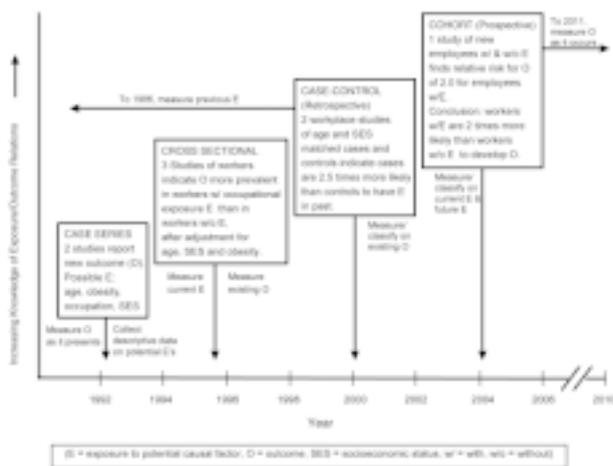
Figure 1.

prevent or reduce the impact of injury in a population, whether a model for risk evaluation or an engineering design intervention, can only be validated using epidemiologic methods. However, epidemiologic methods are primarily useful for existing or emerging injury problems that have sufficient numbers of cases to support analysis. Therefore epidemiology is inherently reactive, identifying its solutions *a posteriori*.

Injury has been defined as "any damage inflicted to the body by energy transfer during work with a short duration between exposure and the health event (usually less than 48 hours)" (Hagberg *et al*. 1997). Work-related musculoskeletal disorders with longer time periods between exposure and injury onset (i.e. gradual onset) may also be addressed epidemiologically.

3.1. Evaluation of Injury at Work by Surveillance (Passive or Active)

Injury prevention at the workplace requires surveillance for injuries among workers, both before and after any intervention effort to prevent injuries. Surveillance has been defined as "the ongoing and systematic analysis and interpretation of health data in the process of describing and monitoring a health event" (Centers for Disease Control 1988). Workplaces may then use this information to plan, implement and evaluate the health effects if any from interventions or changes in the workplace.

There are two essential elements of any surveillance system. The first is a clear definition of the injury cases to be counted also known as the case definition. Very often in practice, existing data sources such as dispensary records, required injury logs (e.g. OSHA log) or cases receiving wage replacement or sickness/absence benefit provide initial case definitions (e.g. counting all hand injuries treated in the plant clinic). This particular approach is referred to as passive surveillance since the data already exist. A limitation of this approach is that one has to accept the data elements and quality offered by the existing system.

An alternative that can improve the upon the performance of passive surveillance is active surveillance. In this instance new information is solicited from the workforce specifically for the purpose at hand. Instruments typically used include discomfort or symptom surveys and risk factor checklists or other exposure evaluation tools. This approach offers improved information and control at the cost of resources to actively collect the information.

Once the case definition of interest is established all reporting sites must adhere to it. This will necessarily entail the support of the system with administrative and management resources including training in the appropriate use of the system by those entering and interpreting the data.

The second essential element is a count of the number of work hours in the population exposed to relevant work conditions in a particular plant, department, assembly line or workstation. These two elements, cases divided by work hours constitute the rate of occurrence of the injury event, which is the primary measure of effect of any intervention. The rates may be subdivided for categories such as length of employment, workstation location, shift schedule, etc.

3.1. Surveillance System Applications at Work

Consider a company with multiple locations producing similar products. Substantial variation in injury rates between these plants would suggest that injury circumstances, prevention efforts or injury hazards were somehow distributed differently between plants. Only by establishing comparable and consistent surveillance systems can these "natural experiments" be exploited. Similarly, if an injury intervention being installed company-wide can only be put in a few plants at a time, injury rate comparisons between plants with the intervention compared with plants that are designated to receive it at a later date can contribute to evaluating the intervention's efficacy.

For example, a large manufacturing company developed a new employee conditioning program in one plant and wanted the first 6 months of experience with it evaluated before duplicating the program company-wide. Back and shoulder injury cases were counted 6 months before the program was implemented and then 6 months while the program was in operation. The rate for cases with days away from work increased from 14 to 22 per 200 000 h worked. The number of cases of back and shoulder musculoskeletal disorder increased from four before the program's start to nine during the first 6 months of its operation. This was thought to be a large enough increase in cases to warrant a major modification of the conditioning program (Sorock *et al*. 1997).

3.2. Use of Narrative Data from Injury Records to Improve Injury Surveillance

Many injury surveillance systems include a brief narrative section describing in a few sentences how the "accident" happened. In most databases, only a single code is assigned to each narrative that best describes the "causes" of the "accident." The narratives can be searched electronically and certain keywords selected that can be used to identify cases not found by searching for codes. Narrative analysis has been used, for example, to identify fractures from "powered industrial vehicles," finger lacerations while using "gloves," injuries from contact with industrial "robots," and compensatable claims for back injuries while using a "hand cart" or "dolly" (Sorock *et al*. 1997).

3.3. Recommendations for Workplace Injury Surveillance

(1) Generally the utility of a surveillance system will improve as case definitions are refined and as progressively less severe and perhaps even near-miss types of events are cataloged.

However, it is important to note that systems offering very specific levels of detail, hence a greater number of potential categories or descriptors, can be compromised by efficiencies in use. For example, if the system allows a coder to select from 500 outcome codes, the coder will typically habituate to assigning the vast majority of incoming cases into a significantly smaller subset of more frequently used codes (see the entries on Human Information Processing for more on this topic). Hence, a trade-off between specificity and efficiency exists in surveillance systems. The appropriate level of detail is best selected by examining the goals of the system, the skills and training of coders, the anticipated management of the coding system, and the level of performance in detail of any existing systems which the new system is to complement or improve.

(2) Management and supervision should be involved in all efforts with respect to surveillance and injury control. This may be optimized through awareness training for management and supervision.

(3) Records for surveillance should include a brief narrative of how the injury occurred with special note to contributing exposures (e.g. equipment involved).

(4) Surveillance systems should record demographic information such as location, shift, age of worker, gender, job class and occupation.

(5) The system should incorporate where possible the costs associated with injuries to permit cost analysis.

(6) Organizations should computerize their surveillance systems and ensure sufficient storage space for archived data (Sorock *et al.* 1997).

3.4. Occupational Injury Intervention Studies

Workplace injury prevention efforts can be categorized as either engineering interventions (targeting the design of the job, tools and workspaces); administrative interventions (focusing on work procedures and policies); personal interventions (addressing worker behavior, training and education); or a combination of strategies (multiple interventions) (Zwerling *et al.* 1997). In general, worker behavior, training and education interventions alone have not been demonstrated to be as successful in injury

prevention as have other approaches. Multiple intervention approaches that seek to optimize the management of an injury problem in each major aspect appear the most effective (McGorry and Courtney 1995).

REFERENCES

BURDORF, A., NAAKTGEBOREN, B. and DE GROOT, H.C.W.M., 1993, Occupational risk factors for low back pain among sedentary workers. *Journal of Occupational Medicine*, 35, 1213–20.

CENTERS FOR DISEASE CONTROL, 1988, Guidelines for evaluating surveillance systems. *Morbidity Mortality Weekly Report*, 37, 1–18.

HAGBERG, M., CHRISTIANI, D., COURTNEY, T.K., HALPERIN, W., LEAMON, T.B. and SMITH, T.J., 1997, Conceptual and definitional issues in occupational injury epidemiology. *American Journal of Industrial Medicine*, 32, 106–15.

HENNEKENS, C.H. and BURING, J.E., 1987, *Epidemiology in Medicine* (Boston: Little Brown).

KELSEY, J.L., HUEBNER, W.W. and BODIAN, C.A., 1989, Statistical and epidemiologic concepts for the reluctant reader. Part II: Basic epidemiologic concepts. *Surgical Pathology*, 2, 281–91.

KELSEY, J.L., THOMPSON, W.D. and EVANS, A.S., 1986, *Methods in Observational Epidemiology* (New York: Oxford University Press).

LAST, J.M. (ed), 1988, *A Dictionary of Epidemiology* (2nd ed.) (Oxford University Press, New York).

McGORRY, R.W. and COURTNEY, T.K., 1995, Exercise and cumulative trauma disorders – the jury is still out. *Professional Safety*, 40, 22–5.

MURPHY, P.L., SOROCK, G.S., COURTNEY, T.K., WEBSTER, B.A. and LEAMON, T.B., 1996, Injury and illness in the American workplace: a comparison of data sources. *American Journal of Industrial Medicine*, 30, 130–41.

SOROCK, G.S. and COURTNEY, T.K., 1996, Epidemiologic concerns for ergonomists: Illustrations from the musculoskeletal disorder literature. *Ergonomics*, 39, 562–78.

SOROCK, G.S., SMITH, G., REEVE, G., DEMENT, J., STOUT, N., LAYNE, L. and PASTULA, S., 1997, Three perspectives on work-related injury surveillance systems. *American Journal of Industrial Medicine*, 32, 116–28.

ZWERLING, C., DALTROY, L., FINE, L., JOHNSTON, J., MELIUS, J. and SILVERSTEIN, B., 1997, Design and conduct of occupational injury intervention studies: a review of evaluation strategies. *American Journal of Industrial Medicine*, 32, 164–79.

Ergonomics Considerations for Reducing Cumulative Trauma Exposure in Underground Mining

K.M. Cornelius and F.C. Turin

National Institute for Occupational Safety and Health (NIOSH), Pittsburgh Research Laboratory, Mining Injury Prevention, PO Box 18070, Pittsburgh, PA 15236-0070, USA

1. INTRODUCTION

Underground mining in the USA has undergone significant change in the past 20 years. Two key elements have been increased mechanization and a more educated work force. In spite of these changes, many jobs continue to be labor-intensive and repetitive in nature. They entail tasks that, performed over time, can take a toll on the soft tissues and joints. The problem may be compounded by an aging mining workforce. In 1996 the mean age of the coal mining work force was 45 years and the median total years of experience was 20 (NMA 1998). As a person ages, the body's resilience to chronic wear and tear is reduced, which may cause a worker to pay an increasingly higher health price for performing the same task. Mining companies, like many others, are becoming more aware of cumulative effects to the worker as reports of these types of injuries rise.

Conducting a job analysis is an important step when considering a job redesign or modification to reduce worker cumulative trauma exposure. A basic approach to job analysis is to examine the types of aches and pains reported, the tasks performed, and work site conditions. The US Bureau of Mines (USBM) conducted an evaluation of roof-bolting tasks performed at an underground coal mine concerned about early warning signs of cumulative trauma. This evaluation will comprise the primary focus of this chapter. The approach used for the roof-bolting case study may be applicable to other work environments.

2. FUNDAMENTALS OF CUMULATIVE TRAUMA EXPOSURE

Musculoskeletal injury is a term used to describe a wide range of soft tissue disorders that affect the nerves, tendons and muscles. Common examples include lower back pain, tendinitis and carpal tunnel syndrome. The majority of these injuries are not the result of sudden mishaps, but usually develop gradually from repeated wear and tear. Symptoms may not appear immediately, but can take weeks, months or even years. Symptoms may result from many types of activities, performed at work or at home, and it is often difficult to attribute them to a single event. In fact, it is more common to identify factors that may have contributed to the development of the condition. The terms repetitive strain injuries or cumulative trauma disorders (CTD) have been commonly used to refer to disorders that have occurred due to work-related activities.

Three main risk factors contribute to CTD: force, repetition and awkward postures. Any one or combination of these may contribute to the development of CTD. Therefore, the design of equipment in conjunction with the required tasks should be evaluated when attempting to reduce these risk factors. Examining the layout of the work area to help identify tasks which may contribute to cumulative trauma is necessary. The following list (Putz-Anderson 1988), describes ergonomics concerns that, overall, should be *minimized* at the work area:

- *Crowding or cramping the worker*: a work area layout may unnecessarily constrain movements of the worker.
- *Twisting or turning*: placement of tools and materials may require the worker to twist the spine to fulfill the requirements of the job.
- *Repeated reaching motions*: the layout of the work area may require the worker to lean to reach and grasp the necessary tools and controls.
- *Misalignment of body parts*: the arrangement of the work area may require the worker to frequently have one shoulder higher than the other or have the neck or spine bent to one side.

While many of these concerns are a function of equipment design and environmental conditions, making workers aware of these issues may help them to adapt their work habits to reduce risk of injury. Additionally, this information is useful when conducting an ergonomic evaluation of a work area and associated tasks.

3. UNDERGROUND MINING ENVIRONMENT

The underground mining environment is a unique challenge. It is more difficult to develop controls for an underground mine site as compared with a factory setting where equipment and facilities can be more easily designed to reduce worker force, posture and repetitive exposures. In an underground mine, workers are required to perform labor-intensive tasks that often cannot be avoided due to environment constraints. The dynamic nature of the environment does not allow easy implementation of mechanical assists to reduce force exposure. Many of the tasks performed by workers are repetitive. Restrictive work areas due to low ceiling height, low lighting levels and large pieces of equipment cause workers to perform these tasks in postures that are not desirable. Designers of underground mining equipment can control how a machine will function, but not the environment in which it will be used. Hazards in an underground mine cannot be completely removed by redesigning the system. There are many hazards and information sources that must be continually monitored by workers including their position in relation to large pieces of mobile equipment and unpredictable geological anomalies. Thus, immediate dangers may take priority over awareness of ergonomic considerations while performing a job.

4. CASE STUDY OF UNDERGROUND ROOF BOLTING

4.1. Roof Bolting and Cumulative Trauma Exposure

In an underground coalmine, after an area is mined it is necessary to support the roof to keep it from collapsing. Since 1950, the primary method for supporting the mine roof has been installation of roof bolts. Long bolts installed into the roof compress the layers of strata achieving a uniformly distributed support anchorage. Roof bolts, typically 6–8 feet long, are installed by workers using large roof-bolting machines. There are different types of machines

Figure 1. Roof bolting in a low seam mine often requires workers to maintain awkward postures.

used for high, medium and low coal seams. Typically in medium and low seams, the operator works in a small area between the machine and walls of coal called ribs. A roof bolter operator working in a low seam mine is illustrated in Figure 1. In high seams workers often work from a platform on the bolting machine. This case study evaluates workers roof bolting in a high seam mine. In general, bolting machine operators work in tight spaces.

Roof drilling and bolt installation in underground coal mines is labor intensive, repetitive and exposes operators to many hazards that can result in injury. In response to cumulative trauma exposure concerns at an underground coal mine, a case study was conducted to examine roof bolter tasks that performed over time could put workers at risk. For this study, the three data collection activities were analysis of lost time incident descriptions, interviews and observation of roof-bolting tasks.

4.2. Evaluation Approach

Researchers analyzed 43 lost time incident descriptions to identify roof-bolting activities and operator injuries having characteristics consistent with cumulative trauma exposure. Second, researchers conducted a series of interviews with the roof bolter operators, supervisors and the staff nurse. The objective of the interviews was to learn about bolting tasks and working conditions, to identify safety hazards and to discuss the details of injuries. The interview data were analyzed to identify similarities in injuries and pains; tasks that may contribute to cumulative trauma; and aspects of the working environment that may contribute to cumulative trauma. Finally, roof bolter operators were observed performing tasks, bolting activities were videotaped, still

photographs were taken of bolting equipment and mine conditions, and an experienced bolter operator discussed the layout and operation of a roof-bolting machine.

4.3. Results

From the injury analysis, 14 incident descriptions were identified as describing injuries that could have occurred from cumulative exposure and contained the following characteristics:

- Five of the 14 incidents involved pain in the back, neck, shoulder or elbow.
- Two incidents occurred while putting a roof bolt into a drilled hole.
- Two incidents occurred while lifting bolting supplies.
- One incident occurred while torquing a roof bolt.

Nine of the 14 incidents involved a strain or sprain injury to the ankle, knee or hip resulting from a slip, trip or misstep. Seven incidents involved stepping or kneeling on uneven floor, loose materials on the floor or equipment cable. Two incidents involved an operator stepping into or out of the bolting machine platform.

Interviews were conducted with 12 roof bolter operators. The most common injuries cited were:

- face and arm lacerations and cuts;
- shoulder, neck, and arm strains and pains;
- ankle sprains and twists, back pain and strains, and knee strains; and
- leg numbness.

Operators said that roof-bolting tasks require a lot of lifting, carrying, bending, reaching and stretching. Common activities cited as contributing to their pain and discomfort included: leg pains while leaning out to see the drill hole; hand and elbow

pain from using the controls; sore knees, back and shoulders from bending and twisting to install bolts or lift and position drill steels, wrenches and bolts; shoulder and elbow aches from picking up and holding drill steels; and knee and back aches at the end of the shift from standing all day.

The layout of the work area was examined and taken into consideration to minimize crowding the worker, twisting and turning, repeated reaching motions, and misalignment of body parts. After reviewing observation notes, videotape, and still photographs, key items were identified and are listed in Table 1.

4.4. Issues and Recommendations

The following cumulative trauma exposure issues were identified from the analysis: materials handling, operator orientation in workspace, vision obstruction, equipment design, and slipping and tripping hazards. Recommendations were categorized within each issue. Recommendations focused on reducing the three main risk factors that contribute to cumulative trauma disorders: force, repetition and awkward postures. They also addressed the three elements that define a system: human, equipment and environment. Recommendations directed at the human element are intended to increase worker awareness of risk factors. This knowledge can then be motivation for workers to modify their behavior to reduce exposure. Equipment recommendations address modifications to existing equipment, which can be performed at the mine site or retrofitted by the manufacturer and more significant modifications that should be addressed in

Table 1. Observations and issues concerning high seam roof-bolting machines.

Observation	Issue
Confined operator platform causes operators to twist and stretch to get drill steels, bolts, plates and wrenches.	This places operator in awkward postures creating stress to the muscles and joints, particularly in the back and the knees.
Supply trays are positioned at heights well above the operators' waists	Lifting and retrieving tools and bolts is stressful to the neck, arm and shoulder.
Tops of control levers are positioned well above waist height.	The operator must work with the arm and wrist in awkward postures.
Operators lean against the back rail of operator compartment and out from under the canopy while performing drilling and bolting tasks.	This places the operators in awkward postures. Also, it is putting them at risk of being hit by falling rocks.
Operators shift their weight to the side of the body corresponding to the hand which places the drill steel into the drill chuck.	The muscles on the opposite side of the body, particularly the low back muscles, are stressed and may become fatigued.
Operators frequently extend their arm up and out to hold onto steels while drilling, and onto bolts while installing them.	This is stressful to the neck, arm and shoulder muscles.
Drill steels are being inserted into the drill chuck usually at knee level or lower.	The operator must do more bending which stresses the low back muscles.
Transfer of supplies from the back of a bolting machine to supply trays involves frequent lifting, carrying, and twisting.	This places operator in awkward postures creating stress to the muscles and joints, particularly in the back and the knees.

the design of future roof-bolting machines. Environmental factors play an important role in human–machine interfaces. Environmental conditions addressed in the recommendations included space restrictions, visibility restrictions, and housekeeping.

The following is a sampling of the recommendations provided:

- Increase worker awareness of the risk factors associated with developing CTD.
- Examine activities that require high force, high repetition and awkward postures to determine if the task or equipment can be modified.
- Modify materials handling tasks to carry supplies as close to the body as possible, restrict the size of the load and minimize lifting distances.
- Eliminate barriers in the path that require operators to lift supplies up and over.
- Improve supply tray design and position, and method for stacking and retrieving supplies.
- Design bolter tasks and equipment to minimize shoulder abduction.
- Design operator work areas considering reach and visibility requirements.
- Reduce force required to activate controls.
- Increase spacing of controls to accommodate a gloved hand.
- Improve height of control bank in relation to operator.
- Consider a height adjustable, padded rail at back of operator platform.
- Evaluate the threshold between the bolting machine walkway and the operator platform with special consideration given to slipping and tripping hazards.
- Improve housekeeping practices and implement an active program to evaluate.
- Increase worker awareness of slipping and tripping hazards.

Recommendations were intended to be used as a guide for more comprehensive examination of roof-bolting activities. A mine-specific evaluation should be conducted at any mine concerned about cumulative trauma exposure due to varying conditions, equipment, and workforce. An evaluation team with diverse members including roof bolter operators, first line supervisors, engineers and safety personnel is an effective approach for developing solutions (Hamrick 1992, Carson 1993).

5. ROOF BOLTING HAZARDS AND HUMAN FACTORS DESIGN PRINCIPLES

The presented case study does not cover all issues associated with underground roof bolting. Some issues presented are common across variations mining conditions; however, the case study presented focused on a high seam mine where workers operate from an attached compartment. There have been several published reports that have examined hazards associated with underground roof bolting. In particular, one USBM report examined hazards associated with the movement of the drill head boom (Turin *et al.* 1995). A human factors analysis approach similar to the case study approach presented here was used. Report recommendations focused on equipment design issues and the need to conform to accepted ergonomic design principles. One of the primary recommendations was to redesign the control bank to reduce the likelihood of accidental activation and improve
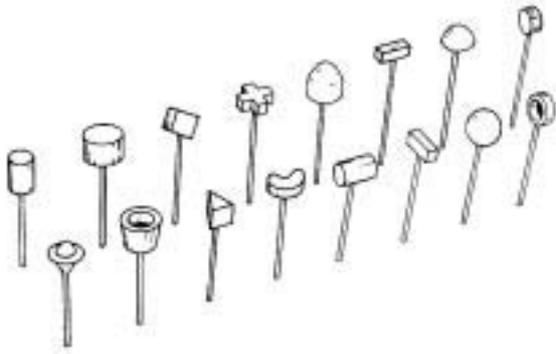
Figure 2. Knob shapes that can be recognized by touch by operators wearing gloves.

ease of use. Current control handles are identical in shape and color and are mounted close together. Controls should be coded by sequence, location and shaped so that they can easily be distinguished and operated. Figure 2 illustrates examples of control shapes.

To redesign the control bank properly, it is important to examine the operation of the roof-bolting machine overall. For example, issues identified in the case study such as materials handling, visibility, and operator orientation in workspace must be examined when equipment redesign is considered. A thorough task analysis would provide information critical to effective equipment redesign and resulting changes to roof-bolting tasks. A useful resource, *Human Factors Recommendations for Underground Mobile Mining Equipment* contains information on human factors design considerations and can be accessed at http://www.cdc.gov/niosh/mining This web page addresses issues on system design, task analysis, aging, layout, controls, visibility, seating and maintenance for underground mining equipment. Links are available on this web page to access additional ergonomics and mining safety topics.

REFERENCES

CARSON, R., 1993, How to start a successful ergonomics program. *Occupational Hazards*, 122–7.

HAMRICK, C., 1992, Ergonomics in mining: ergonomic intervention strategies. *Applied Occupational Environmental Hygiene*, 14–16.

NATIONAL MINING ASSOCIATION, 1998, *Facts About Coal 1997–98* (Washington, DC: NMA).

PUTZ-ANDERSON, V, 1988, *Cumulative Trauma Disorders: A Manual for Musculoskeletal Diseases of the Upper Limbs* (Bristol, PA: Taylor & Francis).

TURIN, F. *et al.*, 1995, *Human Factors Analysis of Roof Bolting Hazards in Underground Coal Mines* (Rhode Island: US Bureau of Mines).

Exposure Assessment of Low Back Disorders: Manual Material Handling Limits

A. Grieco, G. Molteni, G. De Vito, D. Colombini, and E. Occhipinti

Department of Occupational Health, University of Milan, Via San Barnaba 8, I–20122 Milano, Italy

1. INTRODUCTION

There are numerous regulations and standards on prevention of work-related musculoskeletal disorders (WMSDs) affecting millions of workers and workplaces. Almost all of them are based on principles of biomechanics, psychophysics, and physiology with the common goal of designing tasks in the way that the stresses imposed upon the majority of the workforce are below the threshold for fatigue, discomfort and injuries (Ayoub and Dempsey 1998).

Quite often it is difficult to collect all the needed data for a proper exposure assessment on a single task, and major difficulties are encountered in exposure assessment of multiple tasks. Sure enough, magnitude (intensity), frequency, and duration of exposure can be difficult to measure exactly. In addition some standards and regulations have been established in the laboratory, and the links between empirical research and actual exposure in the workplace are not yet completely validated by epidemiological studies. Nevertheless, it is worthwhile to establish limits for manual material handling activities as they assist industry in the control of low back pain.

2. MANUAL MATERIAL LIFTING LIMITS

There may be a variety of assessment for the various types of manual handling, but it is always possible to calculate a synthetic exposure lifting index or manual handling index (MHI) as a function of major situation variables:

$$MHI = \frac{Actually\ handled\ weight\ (force)}{Recommended\ weight\ (force)}$$

Even if it is determined by semiquantitative assessment procedures, the MHI may become an effective tool for defining the consequent preventive measures in accordance with correct prevention strategies, and this may be more important than its role in defining the exposure level of one worker involved in manual handling.

When defining preventive measures, it is convenient to classify MHI results at least according to a model having more than two levels and not a simple dichotomous (yes/no) type. This is because the level of approximation (both intrinsic and in conditions of application) for the suggested methods and procedures calls for a certain amount of caution, especially over borderline results around the value of 1. Then a three-zone model (or traffic light model) appeared to be useful, with the MHI classified as follows:

- *Green zone* (MHI ≤ 0.75)There is not a particular exposure for the working population and therefore no collective preventive actions are required.

- *Yellow zone* (0.75 < MHI ≤ 1.25)This is the borderline zone where exposure is limited but may exist for some of the population. Prudent measures are to be taken, especially in training and health surveillance of operators. Wherever possible, limit exposure so as to return to the green zone.

- *Red zone* (MHI > 1.25)Exposure exists and is significantly present. The higher the MHI value, the higher the exposure for increasing numbers of the population. MHI values may determine priority of prevention measures that must in any case be taken to minimize exposure and shift it towards the yellow zone. Training and active health surveillance of operators must be undertaken in any case.

Obviously the definition of the values corresponding to the three zones is somewhat empirical and may also reflect requirements independent of methods and related data generating the proposal; in this sense the limit values may also be modified by taking into account peculiar application requirements but mostly by a critical reflection on their validity, based on concrete application results.

3. LIMITS OF PULLING, PUSHING, AND CARRYING

For manual load handling — pulling, pushing, carrying — the literature describes no equally consolidated procedures, based on multidisciplinary approaches like the NIOSH procedure for lifting. We suggest ignoring the literature methods and using instead data derived from the specific application of psychophysical methods masterfully summarized by Snook and Ciriello (1991). Here are our three main reasons:

- The data is also expressed with reference to the percentiles of potentially satisfied (even if not necessarily protected) population.

- Data from psychophysical studies is used to develop the NIOSH formula (NIOSH 1981, Waters *et al*. 1993) to assess lifting tasks; in particular, it is used to evaluate the degree of protection or better satisfaction associated with use of the recommended weight limit.

- The data from psychophysical studies is expressed by Snook and Ciriello (1991) with reference not only to the two genders but also to structural variables (height of pushing or carrying areas, distance) as well as to organizational variables (frequency and duration of tasks) which produced well-defined methods according to the different working situations.

Applied in equal conditions, these are the three reasons that led us to prefer psychophysical data, provided it is homogeneous with lifting task procedures and assessment criteria. Pushing and pulling data (conceptually comparable with the recommended weight of lifting task) is referred to forces for activating and maintaining actions. Carrying data is referred to the weight of the carried object. All data is referred to satisfaction of 90% of the population. Values are subdivided according to major structural and organizational variables. With such data available, exposure assessment is rather simple and consists of the following steps:

1. Select the scenario from the figures (kind of task, structural conditions, duration and frequency, gender target) which best fits the actual situation to be analyzed by identifying the recommended force (or weight).

2. Determine in a real situation and under usual operating conditions the force needed to start or maintain a pushing or

pulling action. This can be done by using a conveniently interfaced normal dynamometer; for carrying, simply define the weight of the carried object.

3. Define the ratio between the force measured and the corresponding recommended value; this gives the relevant index.

4. CONCLUSION

It is generally believed that persons should not lift more weight than they would be willing to accept based on their own point of view. However, proposing practically acceptable limits for manual material handling tasks may be quite useful for all those involved in applying regulations and standards for prevention of musculoskeletal disorders related to manual load handling. Moreover, there is a need for an epidemiological validation of the proposed limit and for basic and applied research to define which method of aggregation provides the best assessment of exposure to complex jobs involving manual material handling. There is a need

for basic and applied research to enhance the methodologies for aggregating multiple-component manual material handling tasks with manual material handling criteria.

REFERENCES

AYOUB, M.M. and DEMPSEY, P.G., 1999, The psychophysical approach to manual materials handling task design. *Ergonomics,* 42, 17–31.

NIOSH (National Institute for Occupational Safety and Health), 1981, *Workpractice Guide for Manual Lifting*, Technical Report N81–122 (Cincinnati OH: US Department of Health and Human Services).

SNOOK, S.H. and CIRIELLO, V.M., 1991, The design of manual handling tasks: revised tables of maximum acceptable weights and forces. *Ergonomics*, 34, 1197–1213.

WATERS, T., PUTZ ANDERSON, V., GARG, A., and FINE, L., 1993, Revised NIOSH equation for the design and evaluation of manual lifting tasks. *Ergonomics*, 36(7), 749–76.

Exposure Assessment of Low Back Disorders: Assessment Criteria for Manual Handling Tasks

A. Grieco, G. Molteni, G. De Vito, D. Colombini, and E. Occhipinti

Department of Occupational Health, University of Milan, Via San Barnaba 8, I–20122 Milano, Italy

1. INTRODUCTION

Among the important intervention methodologies in the field of ergonomics is how to assess the individual probability of contracting dorsolumbar spinal disorders due to manual load handling. Right now, owing to the variety and complexity of the organizational, environmental, and personal factors which determine the risk, it appears quite a complex assessment to make.

The term "exposure assessment" may be more appropriate. By identifying and quantifying the variety of possible major risk factors and their "integrated evaluation", it is possible to identify the degree of potential harmfulness of a given task, and to guide the preventive measures (risk management).

An extremely simplified interpretation of the assessment concept has become standard practice in workplaces and other applications. For example, load handling may be assessed solely on the basis of the weight of the load (perhaps referred to national regulations and standards). This is widespread practice in different countries; in fact, when a weight limit is introduced in the legislation, it is even possible to come up with the simplification that all objects weighing less than the weight limits may be handled in "safety"!

This level of simplification means the researcher has to redefine the requirements for assessment validity and applicability, leaving aside the rigors of a sophisticated scientific approach as a prerogative of research elites but also opposing the oversimplification demanded by operators in the field. Examining the variety of documents (guidelines, standards, operational handbooks) produced in the meantime, the requirement of proposing scientifically valid and practically applicable exposure assessment models has been widely acknowledged and largely solved.

2. LOAD LIFTING ASSESSMENT

As regards assessment of manual load exposure to lifting, most proposed models are based on the NIOSH equation published in 1981 (NIOSH 1981) and revised in 1993 (Waters *et al*. 1993). This occurs via major adaptations and changes in the original models; however, the existence of a common design is clear, thus making those models the most suitable to achieve the desired synthesis between scientific research basis and operative application. Actually the NIOSH model offers considerable advantages:

- It is supported by valid and diversified scientific bases (NIOSH 1981).
- It contemplates a large number of possible major risk factors.
- It summarizes in a synthetic index the more complex assessment process (lifting index as the ratio between weight actually lifted and the specifically recommended weight).
- It defines, albeit roughly, the degree of protection assured to a generic working population.

But adopting the NIOSH model to make assessments in the field of manual handling tasks did pose some problems. Here are three of them.

2.1. Caution May Be Tempered Using Empirical Data

The American authors themselves (Waters *et al*. 1993) emphasize that the procedure is not applicable in some situations; such caution is quite understandable from a strictly scientific viewpoint, but in some cases it may be overcome by making assumptions based on empirical data. Suppose the load is lifted with one arm only, perhaps this could be included by introducing a further multiplication factor of 0.6.

If lifting is carried out by two or more operators, always in the same workplace, maybe treat the weight actually lifted as the weight of the object divided by the number of operators, and use a further multiplication factor of 0.85 to obtain the recommended weight. Such further adaptations of the method (only one arm, more than one operator) are based on highly empirical data and not on experiments carried out following strictly scientific procedures. But without them, many problems would remain unsolved.

2.2. Multiple Tasks Need More Complex Analysis

In many working situations, the same group of workers have to carry out different load tasks often in the same workshift. The difference is the result of many different factors (nature of the object, different areas of loading and unloading, work organization procedures). Moreover, the different lifting tasks may be irregular in a given period of time in the workshift (e.g., in a warehouse with picking activities) or according to established time sequences (e.g., when an operator works every 1–2 hours on an assembly line, first loads the line, then unloads the finished products, and then packs them). In such cases the analytical procedure for each task is not suitable to summarize the overall exposure of the worker to load lifting. Therefore these cases require an analytical procedure for multiple tasks, which is obviously more complex. NIOSH has made a proposal for analyzing mixed multiple tasks, founded on the idea of calculating a synthetic lifting index for multiple tasks based on the index for the more overloading task, increased by values derived from the other tasks considered.

2.3. Nonindustrial Tasks Are Not Well Modeled

The NIOSH assessment procedure is not well suited to application in various working sectors (typically nonindustrial sectors), sometimes on account of the characteristic of the lifted load, the great variability of lifting tasks, their frequent association with other manual handling tasks (trolley pulling or pushing), and finally the presence of other risk factors for the lumbar spine (e.g., whole body vibrations). Agriculture, transport and delivery of goods, and assistance to individuals who are not self-sufficient are typical examples. In these situations, though the NIOSH

lifting index is useful, validated procedures for integrated exposure assessment are not yet available, hence the need for further research and proposals on specific simplified exposure assessment procedures aimed at managing risk factors.

3. CONCLUSION

Researchers have yet to complete and verify the data on degree of protection associated with the adoption of different weight constants and with reference to different targets of the adult working population. Therefore a more precise definition becomes essential, especially in order to better establish large-scale prevention strategies which will be sufficiently well defined to account for the makeup of the working population (males and females, young and elderly, healthy and nonhealthy).

The NIOSH lifting index does not seem to be a complete risk index; it is still only an exposure index. Support for the NIOSH index as a risk index perhaps lies only in the (collective) probability of acute lumbago. In the future, epidemiologic studies will have to be developed to ascertain its predictability for degenerative chronic spinal diseases, due to biomechanical overload.

Some countries have adopted weight constants different from those originally proposed (23 kg), although they have retained the calculation procedures for the main multiplication factors in the Waters (1993) formula. This operation, although acceptable for immediate application, may not be fully justified from a strictly scientific and technical viewpoint. If weight constants are adopted other than those originally proposed, it will be necessary to redefine the terms for calculating each multiplication factor and then to decide on the necessary adjustments.

It has been debated that in some working situations, beside the multiple load lifting tasks assigned to a group of workers, they perform many other manual handling tasks such as pulling and/or pushing manual trolleys or carrying loads over long distances. Waters *et al.* (1993) states that its lifting index cannot be applied in these cases. Therefore an integrated assessment procedure needs to be developed for such complex situations, probably by extending the procedure for analyzing multiple lifting tasks.

REFERENCES

NIOSH (National Institute for Occupational Safety and Health), 1981, *Workpractice Guide for Manual Lifting*, Technical Report N81–122 (Cincinnati OH: US. Department of Health and Human Services).

WATERS, T., PUTZ ANDERSON, V., GARG, A., and FINE, L., 1993. Revised NIOSH equation for the design and evaluation of manual lifting tasks. *Ergonomics*, 36(7) 749–76.

Exposure Assessment of Low Back Disorders: Criteria for Health Surveillance

A. Grieco, G. Molteni, G. De Vito, D. Colombini, and E. Occhipinti

Department of Occupational Health, University of Milan, Via San Barnaba 8, I–20122 Milano, Italy

1. INTRODUCTION

Health surveillance observes the state of health of a population and its individuals (usually with respect to a specific health effect). It makes these observations over time and assesses their compatibility with exposure to one or more occupational risk factors. Health surveillance is implemented in industrialized countries in a great variety of ways, depending on specific cultures, regulations, and the development of social security systems. For example, there are countries where data on the health of individuals and workers is systematically collected, and although not always reliable or relevant, it may be used to implement processes for so-called passive health surveillance.

In other countries such current statistics are totally lacking, at least as regards work-related musculoskeletal disorders, therefore it is impossible to use or define passive health surveillance procedures. On the other hand, attention should be drawn to the extremely high number of workers potentially exposed to risks associated with manual material handling; this is why active health surveillance strategies must be developed that are suitable for monitoring vast numbers of workers, and also acceptable from the viewpoint of costs and human resources.

2. ACTIVE HEALTH SURVEILLANCE

2.1. A Three-step Procedure

Step 1

All exposed subjects receive questionnaires or anamnestic interviews according to models that are already available in the literature (Kuorinka *et al*. 1987, Occhipinti *et al*. 1988). These tools are aimed at collecting data on more recent symptoms and disorders (generally over the past 12 months), and should also allow classification according to intensity or severity levels for diagnostic purposes or subsequent preventive measures. In other words, well-defined criteria for interpreting the total data collected on symptoms are recommended in order to unequivocally classify the investigated effects.

Step 2

The second step envisages a clinical examination of the spine only for subjects classified as positive in the previous anamnestic survey. This examination can be made by the occupational physician in the company medical department using a standardized set of specific clinical tests and maneuvers reported in the literature (NIOSH 1988, Occhipinti *et al*. 1988, Hagberg *et al*. 1995)

Step 3

The third step applies to those subjects, identified in steps 1 and 2, requiring more specialized tests (neurological, orthopedic, etc.) or instrumental tests (image diagnostics, EMG, etc.) in order to complete the individual diagnostic procedure.

2.2. Frequency and Goals

The frequency of generalized health surveillance (steps 1 and 2) may be established as required, according to relative exposure indexes as well as health results obtained in the latest round of examinations. Generally speaking, since health surveillance is concerned with slowly evolving chronic degenerative diseases, checks every 3-5 years are usually adequate. There may be exceptions when exposure is heavy and/or drastic changes in working procedures are under way; then more frequent monitoring of health effects may be advised.

One of the goals of health surveillance, from a collective viewpoint, is to check whether in a given working population, exposed to a specific risk, the occurrence of disease is other than expected. For instance, it is interesting to know whether in the same working population the occurrence of undesired effects (e.g., acute back pain, disk herniations) is markedly different from the rest of the working population not exposed to the specific occupational risks. To make such comparisons, generally by inferential statistics, adequate reference data must be available on the whole working population or at least its representative specimens.

Another goal of specific health surveillance at individual level is the earliest possible identification of subjects affected by spinal disorders for whom it would not be advisable to allow exposure levels that were defined as permissible for healthy subjects. In these cases, alternative employment should be found, where the specific risk condition is absent or exists only at low levels. Since these subjects are often completely emarginated (dismissal, employment in minor activities), it may be desirable to establish a line of action that would allow these subjects to stay in productive work at least in some working situations. Four aspects are relevant to manual material handling:

1. Perform a thorough diagnostic assessment of the subject from both clinical and functional viewpoints. Spinal diseases could be classified as light/moderate and medium/severe.
2. Evaluate the present and future working tasks of the subject involving manual material handling. And if the NIOSH lifting index is applied, introduce the following load constants according to the level of disease and gender of the subject: if the level of disease is light/moderate, the load constant should be set at 15 kg for men and 10 kg for women; if the level of disease is medium/severe, the load constant should be reduced to 10 kg for men and 8 kg for women.
3. Assign these subjects only to tasks where the lifting index, calculated using the above load constants, is less than or equal to 1.
4. Carry out a strict individual check on the development of the disease in order to detect any aggravation and take the necessary corrective measures.

When the task assigned to the subject does not meet these criteria, consider whether the task can be suitably redesigned (from a structural and/or organizational standpoint) so it does meet them. Here ergonomic measures take on special importance in redesigning or equipping the workplaces so that workers with path-

ological disorders or with functional impairments can be reemployed there in satisfying and productive jobs.

3. WORKPLACE DESIGN

From common experience, especially in exposure assessment and its implications for risk management, reseachers have realized that in most working environments, even technically advanced environments, load handling is performed with empirical approaches; it is not specifically designed. There are some exceptions to this tendency (e.g., computerized and robotized warehouses) but they seem to be a prerogative of a few industrial elites which have realized how adequate design of goods handling means that, besides solving space problems, it is important to solve problems of safety and potential exposure to noxious agents, and more generally to improve the performance of the whole production system.

The explicit introduction of special regulations and standards should produce the desired effect of devoting more attention to design of tasks and work environments and to workplaces where load handling is required. Design should follow rational and commonsense principles, in accordance with the requirements set by statutory regulations and resulting standards and with current ergonomic trends reported in the literature.

4. CONCLUSIONS

4.1. Adequate Training

The albeit simplified procedures we have suggested for exposure assessment will also be extremely helpful in guiding specific redesign operations. For load lifting, the proposed assessment models can be used to analytically identify the critical factors (weight or size of the object, workplace and layout characteristics, and task organization features); and since they lead to a significant exposure assessment, they have to be modified (in order of priority). Redesign may therefore involve a new load weight, a different workplace arrangement, a different task organization in terms of lifting duration or frequency. Then the evaluator must work in close cooperation with the designer, who should be specially trained in how to manage and prevent problems associated with material handling. Training of technical management staff involved in logistics and production design is a prerequisite of any preventive action.

4.2. Common Sense

In many cases redesigning may also be achieved with simple, commonsense, and low cost measures. Storing goods at adequate heights, allocating suitable space for manual material handling, diversifying tasks, selecting lighter packaging are all rapidly and easily applicable measures in many work situations.

4.3. Auxiliary Systems

Although there is an undeniable tendency to use automated production systems, it is clear that in the short to medium term or for diversified production systems, an auxiliary system will be demanded, i.e., where the worker is not replaced by a robot but carries out some tasks with mechanical aids. Although many and varied machines and equipment are available, such as trolleys, transpallets, elevator platforms, and bridge cranes, there is a significant lack of aids, machines and equipment, or specific and diversified solutions to improve handling in particular industrial workplaces and in most servicing workplaces. Most likely there has not yet been any specific demand by these sectors to provide an incentive for manufacturers of handling tools and systems.

4.4. Logistic Design

Logistic design, in the industrial sector and in the service sector, is an ever expanding and much demanded discipline. There is obviously a close relationship between logistic design and material handling, and this is something for production ergonomists to consider. The interrelationship should be further developed, especially in nonindustrial sectors where there is frequent handling of goods (e.g., commercial distribution of foodstuffs) and people (e.g., hospitals). In short, redesigning of manual load handling tasks offers vast opportunities to ergonomists both by intervention in individual workplaces and in individual products and by restructuring working macrosystems.

5. REFERENCES

HAGBERG, M., SILVERSTEIN, B., WELLS, R. *et al.*, 1995, *Work Related Musculoskeletal Disorders (WMSDs): A Reference Book of Prevention* (London: Taylor & Francis).

KUORINKA, I., JONSSON, B., KILBOM, A., VINTERBERG, H., BIERING-SORENSEN, F., ANDERSSON, G. and JORGENSEN, K., 1987, Standardised Nordic Questionnaire for the analysis of musculoskeletal symptoms. *Applied Ergonomics*, 18(3), 233–37.

NIOSH, 1988, *NIOSH Low Back Atlas of Standardized Tests and Measures* (Cincinnati, OH, USA: US Department of Health and Human Services).

OCCHIPINTI, E., COLOMBINI, D., MOLTENI, G., MENONI, O., BOCCARDI, S., and GRIECO, A., Development and validation of a questionnaire in the study of spinal changes in a working population. *La Medicina del Lavoro*, 79(5), 390–402.

Exposure Assessment of Upper Limb Repetitive Movements: Criteria for Health Surveillance

E. Occhipinti, D. Colombini, G. De Vito, and G. Molteni

Research Unit on Ergonomics of Posture and Movement, Via Riva Villasanta, 11–20145 Milan, Italy

1. INTRODUCTION

The general aims of occupational health surveillance are strictly preventive and involve the long-term monitoring of relationships between the worker (both individually and as a group), health, and the specific working conditions. In this sense, the surveillance of working conditions, and therefore of potential risks, is closely related to occupational health surveillance.

Here are the specific objectives of occupational health surveillance for workers as a group: to contribute to an accurate assessment of work-related risks; to monitor the long-term adequacy of any preventive measures adopted; to collect useful clinical data for periodic comparisons, or comparisons between groups, in order to highlight any positive or negative tendencies among the investigated health effects.

Here are the specific objectives of occupational health surveillance for workers as individuals: to identify those affected by hypersensitivity to the work-related risks present in the working environment, for whom more stringent protective measures should be taken with respect to the rest of the workers; to identify any early-stage diseases, particularly work-related diseases, in order to prevent them from worsening; to identify workers with full-blown disorders, for whom urgent protective measures must be taken.

Occupational health surveillance should also identify cases which must be reported as occupational diseases to the National Insurance Authorities. All acute, subacute or chronic disorders affecting the musculoskeletal system and peripheral nervous system, in particular all nerve entrapment syndromes, primarily carpal tunnel syndrome; all tendon disorders of the hand, wrist, and shoulder; as well as elbow medial and lateral epicondylitis disorders; these are of specific interest in occupational health surveillance, whatever their etiopathogenesis may be.

Note that for other diseases (Dupuytrens disease, various arthroses of the cervical spine or hand, etc.), the link between the disorder and specific working conditions has been postulated, but has not yet been sufficiently well documented. Here occupational health surveillance constitutes a useful tool for monitoring and analysing situations. Indeed, fitness-for-work judgements cannot be made without also taking into due consideration those alterations which, at the current state of our understanding, may not be related to occupational risks, but which will nevertheless be influential in allocating affected workers to specific jobs.

2. NEED FOR OCCUPATIONAL HEALTH SURVEILLANCE

There are basically two factors which, alone or together, will determine the need for occupational health surveillance in a specific group of workers: when work features a potential health risk and when cases are reported that might conceivably be work-related.

2.1. Job as Potential Health Risk

The best way to establish whether a job features a health risk is to carry out an analysis and assessment of the working conditions using the methods proposed in the literature. In particular, when the method for calculating the OCRA exposure index is applied (Occhipinti 1998), active health surveillance should be embarked upon when the index reports a score in the yellow or red areas. If the analytical assessment has not yet been carried out or completed, the decision to begin health surveillance practices will be taken if at least one out of the following three signs of possible risk is detected:

- Jobs featuring cyclic tasks requiring the worker to perform the same movement or set of movements with the upper limbs every few seconds (i.e., every 15 s or less), for at least two consecutive hours (including scheduled short pauses), or for a total of at least four hours per shift.
- Jobs requiring continual manual exertions (at least one every 5 min) for at least two hours per shift, like holding objects weighing more than 2.7 kg or grasping objects weighing over 900 g between thumb and index finger or exerting almost maximal force on tools or levers.
- Jobs involving extreme wrist or shoulder movements or positions for one hour continuously or for a total of at least four hours per shift.

2.2. Reports of Upper Limb WMSDs

Regardless of the factors mentioned in the previous paragraph, it can be assumed that, in a number of working environments, work-related upper limb disorders such as nerve entrapment syndrome or tendinitis may be reported either as a result of systematic data collection methods, such as sick leave statistics, or more often quite randomly, via workers requesting medical examinations. All clinically confirmed cases must be taken into consideration when deciding whether to embark on an active health surveillance program.

It almost goes without saying that when the worker population is small, even individual cases suffice to start a targeted health surveillance program, at least involving the use of anamnestic interviews. But when the number of workers is larger, a proportionally greater number of cases need to be evaluated. There are no hard and fast rules about how to define action levels, but generally speaking, an annual incidence of confirmed work-related cases in excess of 1% or a prevalence of cases twice as high as for the low-exposure population (Battevi *et al*. 1998) constitutes the threshold for setting up a specific health surveillance program.

3. ORGANIZATION OF A MULTILEVEL OCCUPATIONAL HEALTH SURVEILLANCE PROGRAM

In studying and monitoring musculoskeletal disorders of the upper limbs, the symptoms reported by workers are extremely

significant since they generally appear early and, if detected promptly, may provide valuable diagnostic guidance while helping to determine the need for clinical and/or laboratory tests. Moreover, in studies on groups, a method needs to be found for defining reported symptoms as being significant.

Having established this premise, it is now possible to outline a two-level approach to specific health surveillance: firstly at the general level, addressing all exposed workers, based on the collection of personal histories by means of structured interviews carried out by trained health operators; and secondly at the clinically specific level, involving only workers found to be positive at anamnestic investigation.

For first-level surveillance, the symptoms and symptom characteristics are used to define an "anamnestic case" as musculoskeletal upper limb disorders, based on the following minimum criteria (Hagberg *et al.*1995):

- Pain and/or paraesthesia (tingling, burning, pins and needles, numbness, etc.) in one or both of the upper limbs reported during the previous 12 months, lasting at least one week, or occurring at least once a month.
- Onset not attributable to acute trauma.

The aforementioned minimum criteria are in any case valuable for supporting the subsequent operating procedures, although they may be insufficient for diagnostic purposes: programming the surveillance, statistical assessments, preventive measures. The second level is needed to make an accurate diagnosis. Only workers classed as anamnestic cases will be referred for clinical or laboratory testing, based on the relevant diagnostic suspicions.

Based on the results of the clinical and laboratory tests performed, the worker will be diagnosed as a *clinical case*. If the clinical and laboratory tests for the disorders under examination are negative, the worker will still be classed as an anamnestic case and may perhaps be monitored more frequently than workers classed as normal.

4. ASSESSMENT OF COLLECTIVE HEALTH SURVEILLANCE DATA

All upper limb disorders that are demonstrably or potentially work-related deserve to be submitted to a comparative analysis to highlight the presence of clusters in the group of workers studied. Comparative analyses may focus on the following aspects:

- Trends over time to display a tendency to increase, decrease, or stabilize within the group.
- Comparisons between workers operating on a specific line or department versus the overall population in a factory or plant: to highlight the existence of "problem groups" and jobs, and to prioritize preventive measures.
- Comparisons with adult worker populations exposed to low levels of occupational risk; these analyses serve to highlight clusters of disorders or overt diseases in a group of exposed workers versus the incidence of similar disorders or diseases among groups who are exposed to low occupational risk.

Taking into consideration the size of the groups examined, comparisons may be made using usual inferential statistical techniques. Whenever possible, it is desirable to have annual incidence data. It is obvious that comparisons must be made between homogeneous classes of disorders or between individual clinically defined diseases. If there are difficulties in submitting the comparisons to suitable inferential statistical analysis, remember that occurrence clusters should be deemed significant when they have a prevalence or incidence of twofold or greater.

Note that collective health surveillance data is a valuable tool for analyzing the quality and adequacy of both exposure assessments and preventive measures, as well as for programming supplementary preventive measures, if required. In fact, evidence that specific diseases are tending to increase or appear in significant clusters among a group of exposed workers will lead to further investigation or a strengthening of the strategies and initiatives undertaken to prevent upper limb disorders in the group.

In particular, collective health surveillance data can be used to plan periodic follow-up. The following criteria may be a helpful guide. When risk is essentially absent no health surveillance is needed. When risk is present the health surveillance program should be scheduled every 3–5 years or every 1–2 years, depending on the risk level. Lastly if the investigation detects any individuals affected by work-related diseases or anamnestic cases, the occupational physician must draw up a personalized follow-up schedule for each worker.

5. MANAGEMENT OF INDIVIDUAL CASES AND FITNESS–FOR–WORK JUDGEMENTS

Occupational health surveillance can also generate three classes of workers:

- Workers reporting anamnestically positive disorders: besides drawing up a personalized follow-up schedule for them, it is also advisable to consider adopting appropriate measures to reduce their exposure to work-related risks.
- Workers reporting overt diseases not apparently associated with work-related risks: the workers must be temporarily or permanently removed from all or some of the tasks producing a biomechanical overload of the upper limbs (i.e., fitness-for-work judgements).
- Workers reporting overt diseases apparently or definitely associated with specific work-related risks: besides a fitness-for-work judgement including measures for restricting risk exposure, the case must also be reported to the appropriate authority according to national legislation.

Regarding fitness-for-work judgements, when workers are affected by an overt upper limb musculoskeletal disease, it is advisable to recommend that measures be adopted to reduce exposure. It depends on the type and severity of the disease and the levels of exposure, but there are rarely any strict adoption criteria; given the lack of concrete experience in this field, no hard and fast rules or guidelines can be suggested. On the whole, workers affected by the following list of disorders must be removed from tasks involving repetitive or strenuous movements of the upper limbs:

- Incapacitating arthrosis of the upper limbs
- Incapacitating outcomes of upper limb traumas (in relation to the functional demands of the task)
- Radiculopathies associated with degeneration and/or malformation of the cervical spine
- Disorders of the peripheral nervous system of systemic origin
- Rheumatoid arthritis
- Severe current mesenchymopathies

In the case of tendon disorders or entrapment neuropathy, the following criteria may be adopted:

- Temporary removal from tasks involving repetitive movements for as long as required by the acute phase therapy.
- Permanent removal from such tasks for all forms featuring permanent functional impairment.
- Fitness for work conditional upon restricted exposure to repetitive tasks for workers with chronic forms where there is no evidence of functional impairment.

In the third item, "restricted exposure" needs to be defined case by case based on the absence of risk. However, it is recommended that these cases should be given a personalized and frequent follow-up program, so as to closely monitor the evolution of the worker's clinical condition, and consequently to adopt measures aimed primarily at preventing the job performed by the worker from further aggravating the disease.

REFERENCES

BATTEVI, N., MENONI, O., and VIMERCATI C., 1998, The occurrence of musculoskeletal alterations in worker populations not exposed to repetitive tasks of the upper limbs. *Ergonomics*, 41(9), 1340–46.

HAGBERG, M., SILVERSTEIN, B., WELLS, R. *et al.*, 1995, *Work Related Musculoskeletal Disorders (WMSDs): A Reference Book of Prevention* (London: Taylor & Francis).

OCCHIPINTI, E., 1998, OCRA: a concise index for the assessment of exposure to repetitive movements of the upper limbs. *Ergonomics*, 41(9), 1290–1311.

Exposure Assessment of Upper Limb Repetitive Movements: Epidemiology

A. Grieco, G. Molteni G., G. De Vito G., and N. Sias

Department of Occupational Health, University of Milan, Via San Barnaba 8, I–20122 Milan, Italy

1. INTRODUCTION

As opposed to "specific occupational diseases", where there is a direct cause-effect relationship between risk and disease, the Expert Committee of the World Health Organization (WHO 1985) describes the "work-related diseases" as multifactorial. The working environment and the way work is carried out can significantly contribute to the onset of such diseases, but they represent only two of the many causes.

Hagberg et al. (1995) have recently made a detailed review of the literature on work-related musculoskeletal disorders, especially those concerning the upper limbs. Certain occupational exposures are associated with musculoskeletal diseases and involve high relative risks; this was demonstrated for tendinitis of the shoulder and the hand/wrist, carpal tunnel syndrome, and tension neck syndrome. The results were contradictory for other diseases, such as lateral epicondylitis and cervical radiculitis, making it difficult to evaluate the eventual association with work.

Occupational musculoskeletal disorders of the upper limbs can be defined as alterations of the muscle–tendon unit, the peripheral nerves and the vascular system. They can be triggered or aggravated by repetitive movements and/or physical strain of the upper limbs. The main characteristics are multifactorial etiology (occupational and nonoccupational), development over a long period of time (from weeks to years), a long time required for recovery (which may never be complete), and most frequent involvement of muscle–tendon units. Peripheral nerve entrapment syndromes are less common (e.g., carpal tunnel syndrome) but they are more severe and costly.

In the international literature, the following acronyms are used to describe disorders of the upper limbs and to identify their occupational origin: work-related musculoskeletal disorder (WMSD), cumulative trauma disorder (CTD), repetitive strain injury (RSI), occupational cervicobrachial disease (OCD), occupational overuse syndrome (OOS). The term "work-related musculoskeletal disorder" is the most appropriate, since it suggests or proves an occupational cause in the genesis of the musculoskeletal disease and avoids the confusion of introducing in the same term the presumed cause — cumulative in CTD and repetitive in RSI, and the effect — disturbance in CTD and damage in RSI.

Under WMSD we should group together diverse pathological situations that may affect the various structures of the upper limb, but which all have a possible occupational etiology, as a consequence of repetitive and/or cumulative traumas. The critical assessments in Hagberg et al. (1995) and NIOSH (1997) were indispensable for obtaining epidemiological data on the degree of association between work and the most frequently observed musculoskeletal disorders described here.

2. SHOULDER TENDINITIS

The term "shoulder tendinitis" includes all inflammatory forms of the shoulder muscles and tendons, but they are hard to distinguish clinically. The epidemiological studies showed a high prevalence of these diseases in various categories of workers employed on industrial assembly work, welders, packagers, and workers in the fish processing industry; workers exposed to high repetitive rhythm, requiring a significant use of force; and workers doing jobs requiring awkward postures such as overhead work, arm elevation, and specific postures relative to the degree of upper arm flexion or abduction. In these studies the degree of association was extremely high, indicating that the effects could not be ascribed to chance. The temporal association, defined as the demonstration that occupational exposure to risk factors precedes the onset of symptoms, was not proven in all the studies.

High intramuscular pressure is developed in the rotator cuff muscles when performing overhead work; this could impair the intramuscular circulation and contribute to the early onset of fatigue. The initial form of tendon inflammation is characterized by cellular degeneration caused by reduced blood perfusion and impaired metabolism: the degenerated cells die and the calcium salts that deposit on the cell fragments can lead to an inflammatory reaction.

When the arm is raised or abducted, the blood vessels that flow to the tendons of the supraspinatus muscle are compressed, thus altering the blood circulation of the tendons; this might explain the occupational etiopathogenesis. Furthermore, when performing overhead work, the supraspinatus tendon comes into contact with the undersurface of the acromion. The mechanical pressure imposed on the tendon by the acromion is greatest for arm elevations of 60-120°. There is evidence of a relationship between repeated or sustained shoulder postures requiring more than 60° flexion or abduction and shoulder work-related disorders.

In conclusion the evidence for increased risk of work-related disorders due to specific shoulder postures is strongest when there is a combination of exposure to several physical factors, e.g., force, repetition, and vibration; and increased muscle contraction may lead to an increase in both muscle fatigue and tendon tension, impairing circulation.

3. LATERAL EPICONDYLITIS

The epidemiological studies demonstrated a weak association with work in groups of workers employed on meat cutting, packaging, textiles, sewing machines and grocery checkouts. The evidence of a relationship between work factors and epicondylitis was strongest when examining exposure to a combination of risk factors (e.g., force and posture, force and repetition) and epicondylitis. However, in many other studies, no association was demonstrated between lateral epicondylitis and work.

Although an association with work has not yet been demonstrated, it is likely that repetitive, forceful movements can cause the contraction of the muscle–tendon units that are attached in the area of the medial and lateral epicondyles of the elbow, producing the typical alterations.

4. HAND–WRIST TENDINITIS

Hand–wrist tendinitis involves the tendons that cross the radio-carpal joint: the most common form is De Quervain's tendinitis, due to entrapment of the tendons of the extensor pollicis brevis and abductor pollicis longus. The available epidemiological studies show a high risk in the manufacturing industry in general, and in the meat processing industry in particular, with a strong association in the case of specific exposures and the strongest evidence in the case of jobs involving a combination of risk factors such as repetition, force, and postures. However, none of the studies have demonstrated temporal association.

The studies also showed a relationship between incidence of tendinitis and work seniority, number of objects handled, high repetitiveness of gestures, and the force used. Due to the reduced size of the first dorsal wrist compartment where the hand and wrist tendons run, the pressure arising from repetitive actions may produce inflammatory reaction of the tendons, which makes an occupational etiopathogenesis plausible.

5. CARPAL TUNNEL SYNDROME

Compression of the median nerve in the carpal area and the consequent reduction in the flow of blood in the epineural venules may result in endoneural and synovial edema. If the pressure persists over a long period, it may trigger a fibrosis which damages the nerve function, causing the onset of sensorial and motorial disorders in the corresponding area of innervation. It is known that systemic diseases (diabetes mellitus, rheumatoid arthritis, myxedema, amyloidosis, hypothyroidism, obesity), pregnancy, menopause, and use of oral contraceptives may be associated with the development of carpal tunnel syndrome. Many epidemiological studies have demonstrated a high prevalence of cases in a wide range of working activities requiring repetitive movements and use of elevated force of the upper limbs.

Positive association has been found for exposure to repetition alone or in combination with other factors, and in the case of forceful work. There is sufficient evidence to demonstrate that awkward postures alone are associated with carpal tunnel syndrome. Many studies also found a temporal association. In conclusion there is strong evidence of a positive association between carpal tunnel syndrome and the two risk factors force and repetition, especially when they are combined.

6. CERVICAL RADICULOPATHY AND TENSION NECK SYNDROME

In the few studies available on cervical radiculopathy, no association with work was observed in data entry operators, dockers, and assembly line packers in a food production industry. But various studies found that tension neck syndrome — an epidemiological term meaning the presence of pain in the neck and shoulder region — was strongly correlated with work, particularly with static postures and static loads, as in the case of display unit workers, typists, and sewing machine operators. The etiopathogenesis has not yet been fully clarified and various causes have been suggested, including local decreased blood flow, activation of pain receptors through edema or other mechanisms, and energy metabolism disturbance due to long-term static contraction of the muscles.

7. CONCLUSION

Various types of investigations into the association between work and musculoskeletal diseases of the upper limbs are available in the literature, but these studies are not always satisfactory in demonstrating an association since they are not correctly designed from an epidemiological viewpoint. Nevertheless, satisfactory evidence has been found for the association between work and tendinitis of the shoulder, tendinitis of the hand and wrist, carpal tunnel syndrome, and tension neck syndrome. But the evidence proved contradictory in the case of lateral epicondylitis and cervical radiculopathy.

REFERENCES

HAGBERG, M., SILVERSTEIN, B., WELLS, R. *et al.*, 1995, *Work Related Musculoskeletal Disorders (WMSDs): A Reference Book of Prevention* (London: Taylor & Francis).

NIOSH (National Institute for Occupational Safety and Health), 1997, *Musculoskeletal Disorders and Workplace Factors*, DHHS Publication 97–141 (Cincinnati OH: NIOSH).

WHO (World Health Organization), 1985, *Identification and Control of Work-Related Diseases*, Technical Report 714 (Geneva: WHO) pp. 1–71.

Exposure Assessment of Upper Limb Repetitive Movements: Ergonomic Principles for Prevention

D. Colombini, E. Occhipinti, R. Bergamasco,* G. De Vito, and G. Molteni

Research Unit on Ergonomics of Posture and Movement, Via Riva Villasanta, 11–20145 Milan, Italy and *Embraco-Aspera SRL, Via Buttigliera, 6 Riva presso Chieri, Turin, Italy

1. INTRODUCTION

When both the exposure assessment and the study of work-related musculoskeletal diseases have revealed a significant risk associated with repetitive and/or strenuous movements of the upper limbs in various work environments, the need arises to implement specific measures aimed at redesigning jobs and procedures. Their efficacy depends on three types of coordinated and virtually simultaneous actions being carried out: structural modifications, organizational changes, and personnel training.

Although the structural measures are almost universally accepted and widely recommended, actions involving organizational changes do not always meet with unanimous consent, and the scientific literature does not provide concrete examples. Instead it merely supplies general and routine advice, such as reduce excessively high job frequencies, or introduce adequate breaks or job alternatives. This article aims to provide some concrete guidelines for redesigning jobs and preventing disorders caused by repetitive movements of the upper limbs.

2. STRUCTURAL MEASURES

Structural measures primarily concern ways of finding an optimal arrangement for the workplace, furnishings and the overall layout of the environment, and ergonomic work tools. In general, these measures aim to improve aspects related to awkward posture and movements, localized compressions of the anatomical structures of the upper limbs, and the use of excessive force. Structural measures thus seek to reduce the consequences of the most important risk factors — posture and force — and of any other risk factors.

2.1. Criteria for Limiting the Risk Associated with Posture

As far as posture is concerned, the main principle to be kept in mind is to avoid prolonged movements or positions that force the joints to exceed 50% of their maximum range. Redesigning the job means allowing the worker to maintain posture or joint motion below 50% of the maximum specific range for each joint. In order to ensure the correct position of the upper limbs, it is essential to correctly design the workplace. Three aspects need to be emphasized:

- suitable workbench height when standing or sitting
- suitable chair height for seated positions
- suitable operating areas for the upper limbs

Different standards, ergonomic manuals, and checklists supply the main design principles for preventing awkward posture and/or movements harmful to the shoulder, elbow, wrist, hand, and fingers (Eastman Kodak Company 1983).

2.2. Criteria for Limiting the Risk Associated with Force

The main principle is to avoid overstraining the muscles; when the upper limbs have adopted an awkward posture, especially the wrist and hand, the ability of the muscles in the strained segment to apply force is drastically reduced. The force that can be developed in pinching movements is only 25% of the total grip force of the hand; moreover, grip force gradually diminishes as the wrist departs from the anatomical position. In order to intrinsically reduce excessive strain, the following recommendation can be made:

- Avoid even occasional contractions exceeding 50–60% of the maximum individual capacity.
- On average, no muscle-tendon unit should be exerted for more than 15% of its maximum capacity in any given shift.

The lower the degree of muscular exertion, the longer the permitted duration of the exertion. And the lower the degree of muscular exertion, the greater the number of movements that can be made in performing a repetitive task with consequent positive repercussions on productivity levels.

Generally speaking, it is possible to reduce the need to use force by using power-driven tools, by using mechanical grippers and holders (more efficient levers in positions better suited to the stronger muscle-tendon units), and by automating the entire action. Instruments and tools must meet a series of requirements in order to limit the risks associated with posture and force, hence reducing the risk of accidents in the workplace. An ergonomic instrument or tool may be defined by what it avoids:

- Avoid having to deviate the wrist by more than 50% of its normal range.
- Avoid repetitive movements using a single finger.
- Avoid handpieces requiring grips awkward to the development of force.
- Avoid pulling movements and striking actions.
- Avoid localized compressions.
- Avoid the transmission of mechanical vibrations.

Ergonomic instruments or tools should also be coated with a slip-proof finish, they should not conduct heat, and they should not have sharp edges, pointed tips, or potentially harmful shapes.

3. ORGANIZATIONAL MEASURES

Measures typically involving changes to the work organization become necessary when it has been ascertained that jobs feature excessively frequent technical actions and/or inadequate functional recovery periods. Here is a revealing example. In a large metal-working factory featuring assembly lines, upper limb disorders appeared to be prevalent (carpal tunnel syndrome, tendinitis, etc.), most of them attributable to repetitive tasks performed with excessive frequency and/or with lack of proper recovery periods.

On the advice of the local health authority, the company asked to carry out a detailed risk analysis in order to develop options for redesigning workstations more ergonomically. The exposure assessment identified the following problem areas:

- High frequency actions (38-40 technical actions per minute).
- In general, minimal use of force; in almost all cases the com-

pany quickly found specific solutions for bringing the use of force to within acceptable limits.

- Posture seldom extreme and therefore easily corrected by making some structural modifications to the workstation
- Recovery periods taken primarily for physiological reasons rather than for alternating jobs.

The daily schedule included two morning breaks (10 min and 15 min), a 30 min lunch break, and one 10 min afternoon break. One simple change involved optimizing the recovery periods. The total duration of the physiological breaks was already sufficient; by simply redistributing the breaks, it was possible to ensure adequate recovery periods without altering their overall duration. The company redistributed the physiological breaks (35 min = 10 min + 15 min + 10 min) so as to obtain four breaks (two in the morning and two in the afternoon).

In this case the last problem that needed solving was the high frequency of the technical actions. The first and most obvious intervention was to reduce the pace of the task, identifying methods for reducing the number of technical actions required to complete a job cycle but without compromising output. In other words, this meant optimizing — in terms of quality and quantity — the technical actions needed to complete the cycle characterizing the task.

Through valuable cooperation between the ergonomist and the production engineer, it was possible to use the fundamental experience of the engineer not to enhance productivity, but to improve working conditions and thus the health of the workers. By careful filming and critical analysis, each task was revised several times to make it better. In order to reduce the number of actions contained in a cycle, the following procedure was used.

3.1. Phase 1: Analysis of Useless Technical Actions

During phase 1 it is decided whether all the observed technical actions are strictly necessary. It is thus possible to single out useless actions performed by the operator and even actions which could be designed out of the task. In practice this has three aspects:

1. Detecting any useless actions added by the operator. For example, when assembling a piece, the operator occasionally strikes the piece more often or screws the piece more tightly than required: two strokes might be necessary, but the operator actually performs 4, 5, or 6 strokes. In this case the operator must be trained to perform no more than the useful actions actually required to perform the task.
2. Detecting whether any actions added by the operator are entirely arbitrary or in fact conceal a manual flaw. For example, a faulty pin does not fit snugly so the operator needs to strike it several times to force it into the correct position.
3. Detecting obsolete actions. In the course of time, assembly lines may undergo small changes to the machinery or to the product, rendering certain actions obsolete. Therefore it is extremely useful to check the way operators perform their tasks whenever machinery or products are modified.

3.2. Phase 2: Analysis of Upper Limb Use When Performing Technical Actions

Once all useless actions have been eliminated, the next step is to optimize the distribution of the various actions between the two upper limbs. Workers often tend to favor their dominant limb.

Simple low precision actions (e.g., picking up workpieces and placing them on the machining line) may be performed equally by both limbs, thus reducing the frequency with which the dominant limb is used.

3.3. Phase 3: Analysis of Identical Technical Actions

Phase 3 determines whether workers are repeating identical actions for a significant portion of the job cycle. Repetition of identical technical actions can often be avoided by introducing a specific mechanical device. One of the following solutions may be adopted when identical technical actions have been identified but no suitable tools can be introduced and when simultaneously the action frequency considerably increases the total frequency:

- Eliminate the specific manufacturing step altogether by having the part arrive preassembled elsewhere; this is a simple solution but make sure it does not lead to another high risk job being created.
- Introduce a semiautomatic step to replace the technical actions; this is a high cost solution.
- Reexamine the phase scientifically to find alternative solutions capable of fully bypassing the specific action sequence; this hi-tech solution often improves the product.

3.4. Phase 4: Analysis of Auxiliary Actions

Check whether any auxiliary actions are performed in passing from one cycle to the next. It is generally useful to have the conveyor belt and operating areas cross each other in such a way as to avoid the worker having to pick up and replace pieces. To minimize handling, it is equally helpful for the piece to reach the worker facing the right way.

3.5. Phase 5: When Jobs Need to be Split

Despite carefully reviewing actions, sometimes their frequency remains excessively high (even 60 actions per minute). Then the jobs need to be split. Remember that, with no other risk factors involved, a frequency of 30 actions per minute is taken as a reference "acceptable" frequency. Several workstations still feature frequency levels higher than this threshold. Then it is necessary to at least introduce hourly job switches, so the workers performing jobs that might still potentially overload the upper limbs can alternate with less strenuous jobs.

Since the same manufacturing line features workstations with relatively low action frequencies, it will not be difficult to arrange for workers to switch jobs regularly. In essence, job switching is very useful for reducing the risk of exposure to the frequency factor; it has two principal advantages:

- It allows workers to alternate between workstations where the frequency risk is low and workstations where the frequency risk is high.
- It allows workers to alternate between workstations in which the use of the upper limb changes (between left and right).

Besides alternating jobs to prevent disorders due to repetitive movements, adequate recovery periods are critically important in their own right. A suggested ratio of work periods to recovery periods is 5:1 within each hour of repetitive work. Often factories schedule long enough recovery periods (i.e., actual breaks and/or nonrepetitive tasks) but they are poorly distributed

throughout the duration of the repetitive task. Here are some ways to improve them:

- Optimize the distribution of official breaks: it is preferable to shorten each individual break, but to increase their frequency.
- Arrange, if possible, for rest periods to be scheduled at the end of an hour of repetitive work.
- Avoid the scheduling of rest periods too close to meal breaks and shift ends; this allows meal breaks to be used as recovery periods.
- Rotate workers in nonrepetitive tasks, so as to obtain an optimal distribution of repetitive and nonrepetitive tasks, thus ensuring a good ratio of work periods to recovery periods.

4. TRAINING PROGRAMS

4.1. Training for Factory Workers

Workers must be informed of the risks and the kinds of disorder associated with repetitive tasks, in order to justify and motivate the need for these tasks to be performed correctly and in the proper order. Workers must therefore be suitably trained to follow these rules:

- Perform tasks in the required order.
- Use both limbs whenever possible.
- Avoid adding useless actions.
- Grip objects correctly.
- Notify the supervisor whenever new actions need to be performed.
- Contact the health officer as soon as early warning signals are noticed.

4.2. Training for Production Engineers and Supervisors

Training engineers and supervisors is based on a clear understanding of the specific risks and injuries as well as the medico-legal implications associated with occupational diseases. It is necessary for engineers and, above all, supervisors, to organize periodic meetings with workers in order to gather information on any practical problems emerging from the various tasks. Their prompt detection and elimination will prevent unnecessary damage to workers' health, and often leads to a better product. Thus the production engineer is a key figure in the training process, receiving training and insight from expert consultants and providing practical training for the workers. The production engineer has the following responsibilities:

- Suitably design how a task must be performed, above all optimizing the technical actions in terms of human health, not just productivity.
- Teach workers how to perform tasks correctly.
- Periodically check that tasks are being performed correctly.
- Periodically talk with workers about the possible onset of problems while performing tasks.
- Check that technological innovations do not cause increased risk factors.
- Attend to new workers and ensure they are given proper training for their tasks, especially complex tasks.

4.3. Training for Management

Managers need to be involved in the training process. Training must be carried out by experts. Managers must be able to provide trainers with a thorough picture of the risk factors present in the work cycle, as well as possible strategies to ensure they are minimized and effectively managed.

5. REFERENCES

COLOMBINI, D., GRIECO, A., and OCCHIPINTI, E. (eds), 1998, Special issue on occupational musculoskeletal disorders of the upper limbs due to mechanical overload. *Ergonomics*, 41(9).

EASTMAN KODAK COMPANY, 1983, *Ergonomic Design for People at Work*, Vols 1 and 2 (New York: Van Nostrand Reinhold).

A Framework for Assessment of Work-Related Musculoskeletal Hazards

C. L. Shoaf and A. Genaidy

Department of Mechanical Industrial and Nuclear Engineering, University of Cincinnati, OH 45221-0072, USA

W. Karwowski

Center for Industrial Ergonomics, Department of Industrial Engineering, University of Louisville, KY 40292, USA

1. INTRODUCTION

To be effective, the work hazards and risk assessment methods must address all components of the work system as well as their interactions affecting the human performance. A workplace hazard is an event or situation with the potential for harm (Cox and Cox 1993). The concept of risk is used to quantify the degree of harm with respect to the likelihood (probability) and severity (consequences) in response to workplace hazards (Cox and Cox 1993; Manuele 1997). Performance can be described in terms of health and safety measures with the goal of risk quantification of workplace hazards. As discussed by Shoaf *et al.* (2000), a comprehensive work system model is needed to optimize human performance. Such system should capture and integrate the individual and interactive effects of the great number of variables related to hazards and risks at the workplace.

2. THE WORK SYSTEM MODEL

The work system consists of four primary components: (1) work demands, (2) the worker, (3) outcome measures (i.e. perceived and actual risk), and (4) their inter-relationships (Figure 1). A description of these components and their relationships is given below.

2.1 Work Demands

Work demands include both work content (i.e. the physical and mental job demands) and work context (i.e. the physical, social, individual growth and work organization environment).

2.1.1 Definition of variables

The physical demands component describes the set of activities that require combined dynamic and static muscular contractions while the body is maintaining a dynamic posture

Figure 1. Work system components

(e.g. walking, running) or a static position (e.g. sitting, standing). Physical demands have been classified for epidemiological reasons into two categories: (1) object handling activities such as lifting, lowering, pushing, pulling or carrying consist of moving objects with one or both hands and the use of upper and lower extremities; and action of the trunk; and (2) extremity-postural work which describes extremity and head work either in a fixed or dynamic lower body positions such as sitting and crawling.

Mental demands represent the workload imposed by the job tasks on the worker's perceptual and cognitive capabilities. One way in which mental work can be categorized is as skill-based, rule-based or knowledge-based according to the classification system developed by Rasmussen (1983). Skill-based behavior evolves as a series of prearranged, ordered steps in a well-rehearsed routine. Rule-based behavior is invoked when a situation is identified as belonging to a familiar class of problems through which the solution steps are based on prior experience. Knowledge-based behavior occurs in an unfamiliar situation in which the problem cannot be simply and immediately classified, and therefore extensive trial and error iteration are required to determine the solution.

Work context refers to those environmental conditions that affect a worker's ability to perform, but do not specifically relate to work input or to work output transformation, as in the case of physical and mental demands. Physical elements represent the physical environment in which the work tasks are performed. These elements can impede the worker's long-term or short-term ability to perform the job activities.

The social, individual growth, and work organization environment describe those variables that collectively form the non-physical work setting. The social aspects include relationships with co-workers, management, and family (Elo 1986; Gardell 1987). The individual growth aspects refer to an individual's growth needs as described by Maslow's Hierarchy of Needs (1943). Work organization is the objective nature of the work process and deals with the way work is structured and managed (e.g. shift work, job structure within the work process or organization; Cox and Cox 1993).

2.1.2 Hierarchical structure of work demands

The work demand profile is structured as a hierarchical framework (see Figure 2). In Figure 2 an example of the physical demands hierarchy structure is provided. Two global work demands reside at the top of the hierarchy — that is, work content and work context. Below the top layer, classifications describe the composition of the next lower layer. For example, below the "work content" layer lies the physical and mental job demands. The next lower layer describes the groupings which constitute the physical and mental demands.

The lowest level of the hierarchy represents the most detailed descriptions — that is, the work elements characterizing the demand classification. For example, weight of load, repetition, horizontal distance, height of lift/lower, time duration and twisting angle refer to work elements that describe lifting demands of the object handling demands in the physical demands layer of the work content layer.

It is important to structure the work demand profile within a logical framework as it will allow users to obtain varied levels of
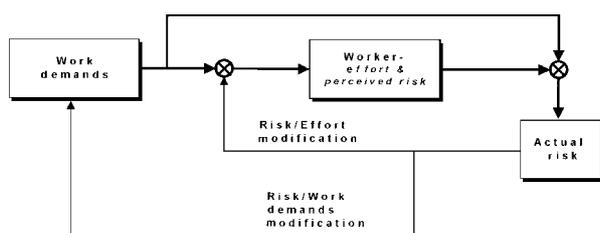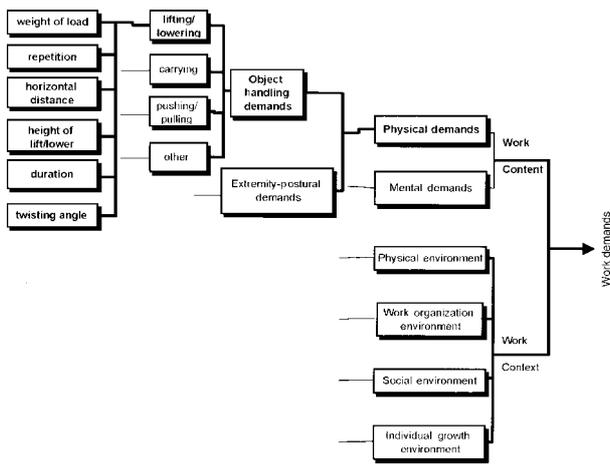
Figure 2. Hierarchical structure of work demands

information details. For example, ergonomists and safety engineering specialists employed in an environmental health division of a large corporation may be interested in the information provided by the bottom layer of the hierarchy (i.e. the most detailed level of information) as they attempt to design and implement control strategies aimed at the minimization of risk in the workplace. Alternatively, a higher level manager in the division may be interested in more global aspects of information provided at higher layers of the hierarchy in order to communicate a more comprehensive overview regarding hazard/risk to other corporate officers, shareholders or potential customers.

2.2 Worker

The worker represents the individual performing the job with their associated personal characteristics, abilities, capacities and needs, serving as the conduit through which the work demands are processed into an effort level. Work demands act as an input to the worker who, in turn, performs an activity to transform a work object or information into a desired product.

Although jobs vary tremendously, there are four operational functions (see Figure 3) that are fundamental to all jobs and virtually every form of human activity. These are: (1) sensing (i.e. information receiving); (2) information storage; (3) information processing; and (4) decision/action functions (McCormick 1979). Therefore, the worker functions within the context dictated by the work setting transforming the job demands into the desired product by generating an effort level as they expend energy.

Traditionally, job performance (i.e. the manner in which the worker meets the challenges of the work demands) has been described rather simplistically. Vroom's (1964) "Performance = f (Ability x Motivation)" model has long served as the archetypal formula to specify the relationship between the individual worker and their performance output. Our work endeavors to enrich this description by characterizing the worker's effort level (i.e. physical, cognitive, emotional) in association with their own individual qualifications and the required qualifications as determined from the job demands.

2.2.1 Modes of information processing and output selection

The way in which all workers handle the challenges of work demands, regardless of their individual capabilities, can be viewed as governed by three primary modes of processing — cognitive, emotional, and physical. The work object (i.e. information about a work task) is a stimulus, is rendered as an object or entity through its sensory qualities, and can be stored in memory as a function of those perceived qualities. Such hybrids of representation and actual stimulation become the input to the cognitive and emotional processes (see Gaillard 1993). Cognitive (i.e. rational) processing transforms sensory information input into motor and/or vocal output using formal and logical operations. Emotional processing is influenced by feelings which inherently contain impulses to act. The logic of the emotional processing is associative; it takes elements that symbolize a reality, or trigger a memory of it, to be the same as reality (Goleman 1995). Emotional appraisal of a situation is automatic and instantaneously contains impulses to act.

Generally, there is a balance between emotional and cognitive processing as emotion feeds and informs cognitive processing, and the cognitive processing, in turn, refines and regulates the impulses of the emotional input (Goleman 1995). Positive emotional states are capable of enriching the cognitive processing such as in mildly enthusiastic states. At the extreme, positive emotional states can disable cognitive processing as in manic states. Negative emotional states can moderately affect cognitive processing, as in the case of a bad mood, or can disable it, as in the case of severe depression.

Figure 4 depicts a hypothetical representation between cognitive processing and emotional processing. Cognitive processing is optimal when emotions are in a moderate state (i.e. not very positive, not very negative). When emotions are overwhelmingly positive, such as in an hysterical excitement, or overwhelmingly negative, as in an angry rage or depression, all energy is devoted to the extreme state with little remaining for cognitive effort. Generally, however, there is a lack of an integrated construct in the scientific literature which permits a complete understanding of the architecture of human information processing. Further research is needed on this issue to advance our understanding and provide integrative information. Recent integrative works on stress, workload, and fatigue have begun to seek this relation (see Gaillard 1998; Hancock and Desmond 1998; Tepas and Price 1998).

After information is processed, an output is selected and a response is produced by muscular contraction of differing body parts including the complexities of speech. The response results in transformation of the work object or information toward a
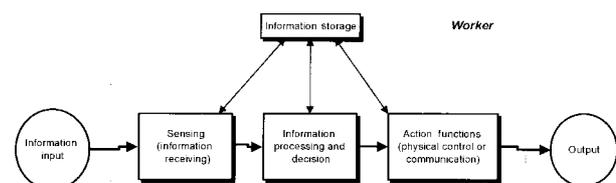
Figure 3. Basic operational functions in human work activities (adapted from McCormick 1979)
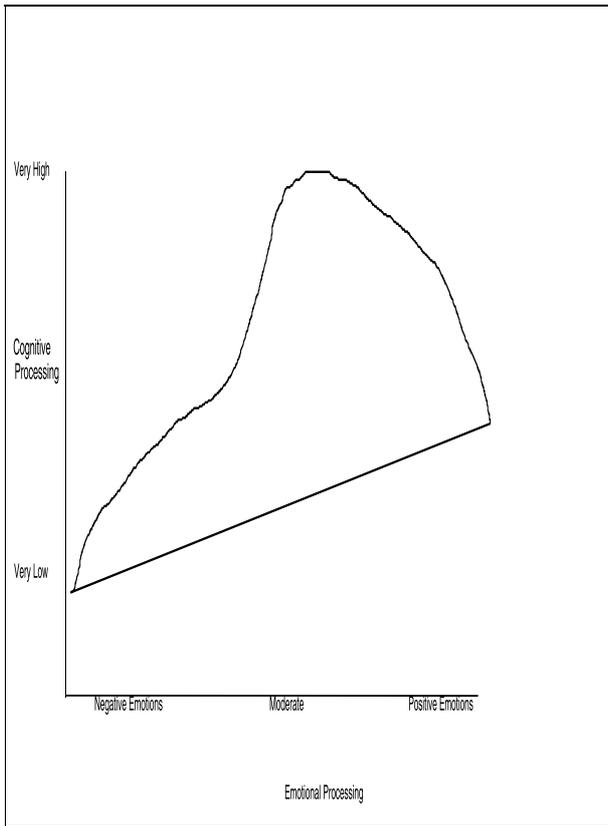
Figure 4. Hypothetical representation of cognitive processing and emotional processing

desired form. Since physical loading consumes energy for this object transformation process, the worker's energy state is influenced by the physical load. This, in turn, indirectly influences the capacity for information processing. For example, if physical loading is within moderate limits, it may produce a positive emotional state that can enhance and improve mental capacity available for task execution, thereby reducing the negative effects of irrational processing. Alternatively, if physical loading is not pleasurable or is distasteful, a resulting negative emotional state may reduce or decrease the cognitive processing capacity available for task execution.

2.2.2 Effort

Effort is defined as the amount of energy an individual expends (Porter and Lawler 1968). Based on our work experience, effort can be described as a function of the interaction of the worker energy states, worker qualifications (e.g. motivation, skills, abilities), and required work qualifications. For example, if the performer has the required motivation level and abilities to perform a certain task but the performed energy state is dominated by physical fatigue due to inadequate sleep or illness, the resulting effort level will be very low or inhibited. Indeed, the worker energy state is a complex interaction of emotional, physical, and cognitive states (Figure 5). Moreover, the motivation force is affected by internal human needs as well as external work demands.

Effort can be classified according to the nature of the challenge presented via work demands into the domains of the muscular, the cognitive, and the emotional. Muscular effort refers to the physiological energy expenditure resulting primarily from physical job demands. Subjective ratings of physical effort has been found to be valid, reliable, and highly related to actual metabolic costs (Hogan and Fleishman 1979; Hogan *et al.* 1980). Cognitive effort refers to the energy expended through mental processing resulting from mental job demands (Hart and Staveland 1988). Emotional effort refers to energy expended from processing feelings and their inherent impulses to act regarding all facets of the work demands.

Cognitive effort and emotional effort are closely related. Harmony between them can enrich human performance (Cszentmihalyi 1990; Hancock 1997). Alternatively, discord can impair and even incapacitate the worker. There is also a close connection between the physical and emotional/cognitive effort domains. When someone is exerting a high level of physical effort in response to challenges presented by physical work demands, the worker's cognitive and emotional energy states may be directly affected. Physical abilities required by the work affect individual growth motivation, fatigue and satisfaction (Hogan and Fleishman 1979; Fleishman 1984). A consequence of the physical exertion is the slowing of the cognitive processes, which in turn may lead to exertion of an extra amount of cognitive effort in order to process the required information (see Vercruyssen *et al.* 1989).

2.3 Health and Safety Performance Measures

The work system model's output is the performance resulting from worker effort. The output of interest emphasizes the quantity significant to the user. Therefore, these outputs may assume the form of various parameters. As this study focuses on the health and safety of the work setting, the outputs of interest are the risk perceived by the worker and the actual risk to the worker in the system.

Risk, in a broad and contemporary sense, expresses the potential harm caused by hazards present in all aspects of the work system. It can be classified into perceived risk and actual risk. Perceived risk is the level of internal risk the worker experiences in response to the level of effort exerted, and may or may not be equal to the actual risk. Perceived risk is subjective and is therefore influenced by numerous judgmental biases such as familiarity, controllability of hazard and time-scale over which any resultant harm may occur (Cox and Tait 1998). Therefore, it is contextual as it is filtered through personal attitudes, experiences, values and education (Petersen 1996). Actual risk is the true risk present in the work system regardless of the worker's awareness, and can be calculated through a quantitative risk analysis measure.

Perceived risk can be further categorized according to the mode of processing the risk is born out of, rather than the type of harm the risk can cause, as the harm may assume several forms (e.g. physical, cognitive, emotional). This distinction is especially important in the area of emotional risk, although the areas of physical and cognitive risk also exhibit interconnections to other areas, as would be expected in a complex system.

2.3.1 Perceived risk

Perceived physical risk refers to the subjective risk workers assume with respect to their body structure (e.g. muscles and bones) and physiology (e.g. respiratory, sensory systems). For example,
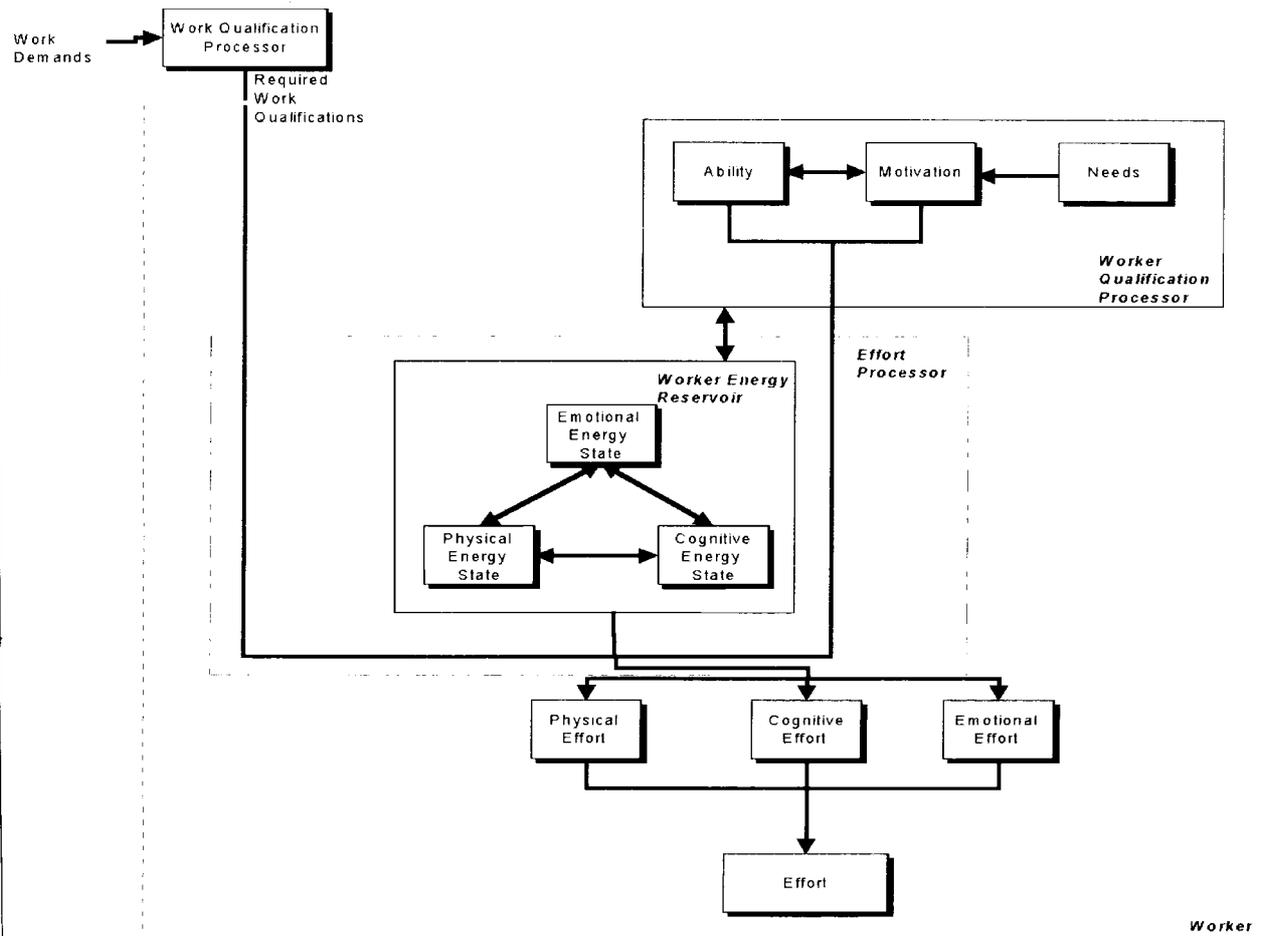
Figure 5. Effort as a function of worker energy state, worker qualification and required work qualification

if a worker believes there is little risk associated with a heavy, repetitive lifting task, the perceived risk is low. Accordingly, the perceived risk will impact on the decisions the worker makes regarding the way in which the job is performed.

Perceived cognitive risk refers to the subjective risk the worker assumes with respect to the mental job demands. For example, a worker may agree to assume several knowledge-based job assignments to be performed concurrently. The high level of mental processing necessary to execute these tasks represents a primary risk to the worker's thinking capabilities (including cognitive and memory processes), but the worker may also experience physiological disorders (e.g. cardiovascular as a result of chronic effects) as a result of accepting the cognitive risk (Hancock and Warm 1989).

Perceived emotional risk refers to the subjective risk workers assume regarding the expression or repression of their feelings and their impulses to act on them. Perceived emotional risk arises in response to the interconnectedness of the work system characteristics. For example, a worker, who is fatigued due to high levels of physical and/or mental job demands — he may be annoyed because of family scheduling problems as a result of having to work overtime and is uncomfortable because the air conditioner is not working correctly — is more likely to risk

verbally assaulting a co-worker's character after a disagreement than a worker without such problems and discomfort.

2.3.2 Actual risk

Actual risk, the objective risk the worker is exposed to, equals the product of frequency and severity of harmful effects (Wentz 1998). There are several methods of quantifying actual risk (see Cox and Tait 1998). Frequency can be calculated as the number of occurrences per 100 full-time workers over a one-year period for: (1) lost work day cases, and, (2) restricted workday cases (Goetsch 1996). Severity can be computed as the number of lost workdays and restricted workdays per 100 full-time workers over a one-year period. Actual risk may then be calculated as the weighted sum of lost and restricted workday cases.

3. WORK SYSTEM MODEL OPERATION
3.1 System Overview

The model of the work system proposed for WMSD hazard/risk assessment represents a complex adaptive control system. While these models depict the main components that describe the work system relationships, the myriad factors which characterize the work demands and the worker sub-systems demonstrate the system's complexity. This multitude of factors interacts to produce

effort and risk. Consequently, numerous variables and relationships within each sub-system can be manipulated to vary the effort and risk output parameters for WMSD study.

Self-regulation in living systems is mediated by feedback control mechanisms (Smith and Smith 1987). As the work system is a living system, feedback control is used to describe the effects of output parameters which serve to change input variables. The interactions between human behavior and the physical, social, and organizational properties of the work environment provide the basis for understanding the operational hazards in work (Smith 1979). The model described here further develops these ideas by specifying work system components and explicitly describing their interrelationships.

Adaptation is an interactive process implying the response of one entity to the actions of another (Hancock and Chignell 1987). The work system is adaptive as its participants adjust, based on changes in the work demands as well as the effort exerted, risk perception, and risk knowledge. Adaptation occurs to modify system parameters to compensate for changes in the process. In general terms, adaptation results from a three-step sequence. First, a standard or goal is set for the output parameter. Next, the actual output is assessed with respect to the target. Lastly, adjustments are made to the system parameters in response to the error (target — actual) to minimize deviation. The mismatch between work demands and available resources (i.e. actual capabilities of the human operator) when within the zone of adaptability can be compensated for through adaptation, therefore controlling the error signal to remain within acceptable boundaries (Chignell and Hancock 1986).

Specifically, in the case of the work system, adaptation occurs through two groups of respondents: namely, the worker and management. The worker controls the effort acceptance level and the risk acceptance level. In regulating these parameters, the worker takes into account changes in the system inputs and "learns" based on experience. Management controls a large set of the factors which characterize the work demands and therefore affects the stimulus for effort and risk outcomes. In regulating these parameters, management takes into account changes in system outputs, also with an accompanying learning experience.

Risk acceptance level is based on the perceived and actual risk inputs. The workers adjust their behavior according to influences of internal and external variables. Internal variables are those which characterize the workers such as their attitude toward risk (i.e. risk-taker or conservative), motivation for performing the activity and individual needs (e.g. acceptance by peer group). External variables are those which characterize the nature of the risk, such as time-frame of consequences (i.e. immediate or long-term), media coverage, and controllability. Both internal and external variables affect the workers' physical, cognitive, and emotional modes of information-processing to form the adapted response.

Similarly, workers can adjust their effort level to set an effort acceptance level based upon previous knowledge of the effort required to perform the work activity and the determined risk acceptance level. To set the effort acceptance level, a worker must first establish the risk acceptance level. This information is fundamental to the effort regulation process. The effort acceptance level is affected by workers' physical, cognitive, and emotional energy states as well as their motivation, ability, and needs.

While workers regulate their responses to work demands through effort acceptance level (i.e. how hard a worker is willing to try) and risk acceptance, management regulates the stimulus to the worker (i.e. the work demands). These factors are typically manipulated after assessing the actual risk and worker effort levels (i.e. performance) to achieve a desired result. Many aspects of the factors creating the work demands can be altered by management, including working hours, temperature of the environment, break schedule, number and difficulty of tasks

Table 1. Glossary of building blocks for work system model

Building block	Description
Work demands processor	Transforms inputs of people, material and equipment, effort and actual risk into work demands. Regulated by management based upon appraisal of effort and actual risk outputs.
Effort processor	Transforms inputs of work demands and perceived risk into effort. Regulated by worker based on such variables as worker energy state, motivation and ability.
Qualification demands processor	Transforms input of work demands into required work qualification. Job content and job context determine the qualifications to execute work activity.
Worker qualification processor	Transforms input of worker's effort into worker qualification. The worker's effort level is influenced by, among other variables, their ability to perform the work activities. These abilities, determined by the worker's education, skill and training, form the basis for the worker's qualifications.
Perceived risk processor	Transforms inputs of effort and actual risk into perceived risk. Regulated by worker based on such variables as familiarity of activity and benefit of outcome.
Actual risk processor	Transforms inputs of perceived risk and work demands into actual risk. Determined by job content (e.g. task difficulty, weight of load handled) and context (e.g. duration of work day, noise level) as well as the manner in which the activity is executed (i.e. as guided by the worker's perceived risk).
Risk controller	Regulates inputs of perceived risk and actual risk to produce risk acceptance level. Performed by the worker and influenced by variables such as the individual's risk attitude and benefit of outcome
Effort controller	Regulates inputs of effort and target risk to produce target effort. Performed by the worker and influenced by variables such as motivation and energy state.
Work controller	Regulates inputs of target effort and actual risk to work demands processor. Performed by management to alter aspects of job content and job context with the intent of varying effort and risk outcomes.

defining a job, and degree of worker autonomy over the work process.

3.2   Work System Model

3.2.1   Definitions

The components of any system include input, processor, output, control, and feedback (Murdick 1975). The three types of operators are processors, controllers, and comparators. Processors function to produce the given output parameter. Therefore, processors represent the activity or activities that transform the input into output. Controllers represent the activity or activities that serve to determine the deviation of the target output from the actual output and which adjust the system parameters in response to this error. A glossary of the processor and controller functions is given in Figure 1. The comparator, the circular symbol containing an "x", describes the error detection between the input and variable which is being feedback.

3.2.2   Model operation

In the work system model, work qualifications act as the inputs to the worker. In general, qualifications refer to the knowledge, skills and abilities (KSA) acquired through education, training, and experience. The KSA dimensions are widely used by human resource specialists for job selection and training program development (Schneider and Konz 1989; Fleishman and Reilly 1992; Wooten 1993).

The required qualifications for a given job are derived from the work demands (i.e. job content and job context). They identify the characteristics an ideal candidate for the job should possess. Workers who will perform the job tasks, however, possess their individual set of abilities — that is, the actual worker qualifications.

If the worker's actual qualifications are greater or less than those required by the work qualifications, a mismatch occurs (i.e. deficiency or surplus in worker capacity). In the case of a deficiency, the amount of mismatch may be reduced or possibly eliminated by experience, education, on-the-job training, or augmented assistance from the machine system itself (Hancock and Scallen 1996). In the case of surplus, the amount of mismatch may be minimized by increasing the difficulty of, or dynamically adding additional, work demands. This process of dynamic task allocation is one which is currently under intense scrutiny (see Parasuraman and Mouloua 1996). The amount of mismatch affects the worker's effort level. When workers' qualifications are less than required, they may react by overexertion. When workers' qualifications are greater than required, they may become bored and exert less effort. Performance may be optimized by matching as closely as possible the required and actual worker qualification parameters.

The model's relationships can be expressed through the following seven equations:

Effort = f_1 (Required qualifications, worker qualifications, effort acceptance level)   (1)

Perceived risk = f_2 (Effort, risk acceptance level)   (2)

Actual risk = f_3 (Work demands, perceived risk)   (3)

Risk acceptance level = f_4 (Perceived risk, actual risk)   (4)

Effort acceptance level = f_5 (Perceived risk)   (5)

Work demands = f_6 (Equipment/material/information/people, effort acceptance level, actual risk)   (6)

Required qualifications = f_7 (Work demands)   (7)

As demonstrated through the model's relationships, the required work qualifications, the worker's actual qualifications and the effort acceptance level determine the level of effort the worker exerts (1). Next, the worker's effort and risk acceptance level (i.e. the level of risk the worker accepts) yield the perceived risk (2). Actual risk is determined from the worker's perceived risk (as it influences task execution) and the nature of the work demands (3). The risk acceptance level of workers is based on their perceived risk as well as the actual risk (4). Similarly, their effort acceptance level is based on the risk level they accept: that is, the perceived risk (5). Therefore, the totality of the work demands can be described as the raw inputs (i.e. equipment, material/information, people), the actual risk, and the effort acceptance level (6). The work demands act to produce the required qualifications ideally possessed by the worker (7).

3.3   Model components

Balancing work demands with worker abilities and needs represents a homeodynamic condition where workers strive to maintain a state of balance within limits based upon natural laws governing their existence. Any stimulus or attempt to alter that state is met with an innate response to maintain an acceptable status. The stimulus is, or creates, "stress" upon the organism, system, culture, or organization. The response is "strain". Welford (1973) observed that strain arises whenever there is a departure from optimum conditions (i.e. a condition that the worker is either unable to correct or cannot easily correct). It is postulated that workers perform best under conditions of moderate demands, and that performance will be "sub-optimal" if the demand is "too high" or "too low".

3.3.1   Feedforward elements

Three major elements form the feedforward path of the complex work system: the work demands, the worker and the performance outcomes, perceived risk and actual risk. In this path, the work demands serve as an input to the worker who, in turn, processes these demands, producing the effort and risk outcomes. The worker affects these outputs physically, cognitively and emotionally. In order to optimize the performance outcomes, variables within the work system must be manipulated. Therefore, aspects of the work demands and the worker function act as the modifiable factors for the minimization of work system hazard/risk.

3.3.2   Feedback

The feedback elements of the complex work system are the adaptations which allow the change of conditions within the domain of a given component's control. Adaptations are made to compensate for incompatibilities between system components. There are two primary feedback loops: risk/effort modification and risk/work demands modification.

The risk/work demands adaptations represent those factors in the work content and work context that can be modified based on the knowledge of risk involved in meeting work demands. For example, if management learns that a consultant's evaluation has shown a particular worker to be at high risk (i.e. actual risk), management may then choose to modify aspects of the work environment (e.g. job rotation, training).

The risk/effort modifications represent those factors that can be adjusted by the worker. These modifications, in turn, could increase or decrease the level of worker effort. For example, if a worker feels at risk of injury, and then learns that the given work activity has very low risk, the worker may choose to exert more effort to accomplish the tasks.

4. CONCLUSIONS

An effective work system hazard/risk assessment instrument must address all components of the work system as well as their interactions. Identification of work hazards and risk serves as the foundation for management decision-making regarding work system design and safety assessment. The work hazard description and risk quantification can then establish a basis for preventative/corrective action through all stages of the work system's life cycle. Accurate description of WMSD hazards also plays a significant role in the success of the intervention efforts such as prioritization of areas for job redesign, identification of populations at risk, benchmarking, and communication programs.

REFERENCES

CHIGNELL, M.H. and HANCOCK, P.A., 1986, Comparison of mental workload and available capacity in complex person-machine systems. In W. Karwowki and A. Mital (eds), *Applications of Fuzzy Set Theory in Human Factors* (Amsterdam: Elsevier) 271–88.

COX, T. and COX, S., 1993, *Psycho-social and Organizational Hazards at Work*, European Occupational Health Series No. 5 (Copenhagen: WHO Regional Office Europe).

COX, S., and TAIT, R., 1998, *Safety, Reliability and Risk Management: An Integrated Approach* (Oxford: Butterworth-Heinemann).

CSIKSZENTMIHALYI, M., 1990, *Flow: The Psychology of Optimal Experience* (New York: Harper).

ELO, A. L., 1986, *Assessment of Psychic Stress Factors at Work* (Helsinki: Institute of Occupational Health).

FLEISHMAN, E.A., 1984, The measurement of effort. *Ergonomics*, 27, 947–54.

FLEISHMAN, E.A. and REILLY, M.E., 1992, *Handbook of Human Abilities* (Palo Alto, CA: Consulting Psychologists Press).

GAILLARD, A.W.K., 1993, Comparing the concepts of mental load and stress. *Ergonomics*, 36, 991–1005.

GAILLARD, A.W.K., 1998, Stress, workload and fatigue as behavioral states. In P.A. Hancock and P.A. Desmond (eds) *Stress, Workload and Fatigue* (New Jersey: Erlbaum).

GARDELL, B., 1987, *Work Organization and Human Nature* (Stockholm: Uppsala).

GOETSCH, D.L., 1996, *Occupational Safety and Health in the Age of High Technology: For Technologists , Engineers, and Managers* (New York: Prentice Hall).

GOLEMAN, D., 1995, *Emotional Intelligence* (New York: Bantam Books).

HANCOCK, P.A., 1997, On the future of work, Ergonomics in Design, 5(4), 25–29.

HANCOCK, P.A. and CHIGNELL, M.H., 1987, Adaptive control in human-machine systems. In P.A. Hancock (ed.) *Human Factors Psychology* (Amsterdam: Elsevier).

HANCOCK, P.A., and DESMOND, P.A., (eds), 1998, *Stress, Workload and Fatigue* (New Jersey: Erlbaum).

HANCOCK, P.A. and SCALLEN, S.F., 1996, The future of function allocation. *Ergonomics in Design*, 4(4), 24–9.

HANCOCK, P.A., and WARM, J.S., 1989, A dynamic model of stress and sustained attention. *Human Factors*, 31, 519–37.

HOGAN, J.C. and FLEISHMAN, E.A., 1979, An index of the physical effort required in human task performance. *Journal of Applied Psychology*, 4(2), 197–204.

HOGAN, J.C., OGDEN, G.D., GEBHARDT, D.L., and FLEISHMAN, E.A., 1980, Reliability and validity of methods for evaluating perceived physical effort. *Journal of Applied Psychology*, 65(6), 672–9.

McCORMICK, E., 1979, *Job Analysis* (New York: Amacom).

MANUELE, F.A., 1997, *On the Practice of Safety* (New York: Van Nostrand Reinhold).

MASLOW, A.H., 1943, A theory of human motivation. *Individual Growth Review*, 50, 370–96.

MURDICK, R.G., 1975, *Information Systems for Modern Management* (Englewood Cliffs, NJ: Prentice-Hall).

PARASURAMAN, R., and MOULOUA, M. (eds), 1996, *Automation and Human Performance: Theory and Applications* (New Jersey: Erlbaum).

PETERSEN, D., 1996, *Human Error Reduction and Safety Management* (New York: Van Nostrand Reinhold).

PORTER, L.W. and LAWLER, E.E., 1968, *Managerial Attitudes and Performance* (Homewood, IL: Irwin).

RASMUSSEN, J., 1983, Skills, rule and knowledge: signals, signs and other distinctions in human performance models, IEEE Systems, Man and Cybernetics and Society, *SMC* 13, 257–66.

SCHNEIDER, B. and KONZ, A.M., 1989, Strategic job analysis. *Human Resources Management*, 28(1), 51–63.

SHOAF, C. GENAIDY, A., SHELL, R., HAARTZ, J., KARWOWSKI, W., WATERS, T., HANCOCK P.A. and HUSTON, R., 2000, An adaptive control model for assessment of work-related musculoskeletal hazards and risks. *Theoretical Issues in Ergonomics Science*, 1 (1).

SMITH, K.U., 1979, *Human-Factors and System Principles For Occupational Safety and Health,* US Department of Health, Education, and Welfare, Public Health Service, Center for Disease Control, National Institute for Occupational Safety and Health.

SMITH, T.J. and SMITH, K.U., 1987, Feedback-control mechanisms of human behavior. In G. Salvendy (ed.), *Handbook of Human Factors* (New York: Wiley), 254–73.

TEPAS, D., and PRICE, J.M., 1998, What is stress and what is fatigue? In P.A. Hancock and P.A. Desmond (eds) *Stress, Workload and Fatigue* (New Jersey: Erlbaum).

VERCRUYSSEN, M., HANCOCK, P.A., and MIHALY, T., 1989, Time estimation performance before, during and following physical activity. *Journal of Human Ergology*, 18, 169–79.

VROOM, V.H., 1964, *Work and Motivation* (New York: Wiley).

WELFORD, A.T., 1973, Stress and performance. *Ergonomics*, 16, 567–80.

WENTZ, C.A., 1998, *Safety, Health, and Environmental Protection* (Boston, WCB/McGraw-Hill).

WOOTEN, W., 1993, Using knowledge, skill and ability (KSA) data to identify career pathing opportunities: an application of job analysis to internal manpower planning. *Public Personnel Management*, 22(4), 551–63.

Health and Safety Ergonomics

R.A. Haslam

Department of Human Sciences, Loughborough University, UK

1. INTRODUCTION

Alongside improving comfort and performance, preventing injury and illness has always been a fundamental concern of ergonomics. The welfare of users, operators and others involved in the functioning of a system has to be a priority. However, as with other areas, for example "cognitive ergonomics" and "human-computer interaction", a branch has emerged dealing predominantly with "health and safety ergonomics" issues. Those working in this area have the prevention of accidents, injuries and illness, due to work or other interactions with tools, equipment and environments, as their primary focus. Most activity at present is in the occupational arena.

In addition to ergonomists, there are other professions and areas of science striving to prevent accidents and ill health. The scale of the problem has long encouraged specialization, for example accident researchers (from backgrounds such as engineering or psychology), health and safety advisers, health and safety engineers, health and safety managers, occupational health scientists, and occupational hygienists. There are others whose primary role is treatment and rehabilitation but who also contribute to prevention, for example occupational nurses, occupational physicians, occupational therapists and physiotherapists. At a different level, health and safety ergonomics can be seen as the area of ergonomics that interfaces with these other communities.

2. SCALE OF THE PROBLEM

It has been estimated that 2.2 million of the 57 million UK population suffer from ill health, caused or made worse by work. In addition, ~158 000 occupational injuries are reported each year. The cost to employers of this illness and injury has been estimated by the Health & Safety Executive (HSE) at between £4000 million and £9000 million per year, with a cost to society at ~£10 billion to £15 billion (billion in the UK is 1012). This equates to 1.75–2.75% of the UK gross domestic product.

In the USA, the National Safety Council estimates for 1997 suggest there were 3.8 million disabling (beyond the day of injury) injuries connected with work. The total cost of unintentional work injuries was estimated at $127.7 billion, with a further $99.9 billion due to home injuries (billion in the USA is 109).

3. CONTEMPORARY ISSUES

Issues of current concern include:

- Musculoskeletal disorders, including injuries to the back and upper limbs from manual handling, or due to work with high rates of repetition or where posture is constrained.
- Occupational and domestic accidents, with slip, trip and falls the largest category.
- Psychological stress.
- Problems connected with use of computing equipment, e.g. keyboards and other input devices.
- Noise induced hearing loss.
- Hand-arm vibration, especially vibration white finger.
- Occupational dermatitis.
- Occupational asthma.
- Monitoring and surveillance.
- Improving risk identification and management.

The areas of focus in part reflect the scale to which each is a problem. However, the funding and legislative climate also have a strong influence on the level of attention different issues receive. For example, European Union directives concerned with manual handling and display screen equipment, implemented in 1992, led to significantly increased activity in both these areas.

4. CONCLUSION

Injuries and illnesses from work or other interactions with tools, equipment and environments are a severe problem, many of which are avoidable. Health and safety ergonomics deals with the human factors issues that are often an important component.

REFERENCES

FEYER, A.M. and WILLIAMSON, A., 1998, *Occupational Injury: Risk Prevention and Intervention* (London: Taylor & Francis).

PANTRY, S., 1995, *Occupational Health* (London: Chapman & Hall).

A Heuristic Dose–Response Model for Cumulative Risk Factors in Repetitive Manual Work

S. Tanaka† and J.D. McGlothlin‡

†Surveillance Branch, Division of Surveillance, Hazard Evaluations and Field Studies (DSHEFS), National Institute for Occupational Safety and Health (NIOSH), 4676 Columbia Parkway, Cincinnati, Ohio 45226, USA

‡Department of Industrial Hygiene and Ergonomics, Purdue University School of Health Sciences, West Lafayette, IN 47907-1338, USA

1. INTRODUCTION

As has been documented by Bernard *et al.* (1997), there is strong evidence that repetitive and forceful manual work can lead to musculoskeletal disorders of the hand/wrist (H/W-MSD) such as tendinitis and carpal tunnel syndrome (CTS). However, a 'safe' level of daily repetitive manual workload has not been established. Quantitative determination of such a level among worker population is a difficult task, since many factors are known to be involved in the causation of H/W-MSD. Further, methodologies for objective and quantitative measurement of many of the exposure factors are not well developed. Therefore, establishment of dose–response relationship between the exposure and H/W-MSD is much more complicated than that for chemical exposures (e.g. lead or toluene), where both the dose and resultant biochemical responses can be defined and measured. Despite these difficulties, there is a need for establishment of quantitative dose–response relationship for manual work. Without such quantitative guidelines, we do not know "How much is too much?" and preventive effort against H/W-MSD must remain in the realm of "trial and error" or "faith." Regardless of whether an intervention effort resulted in a success or a failure, there ought to be a mechanism to determine why and how it was successful or unsuccessful, so that one may learn from the previous experience.

As the title of this chapter indicates, it is a heuristic modeling exercise. The rationale used here is based not only on established scientific knowledge, but also on what the authors consider are very likely and reasonable but not necessarily proven experimentally or epidemiologically. It may take many more years of laboratory and epidemiologic investigations to uphold or refute the validity of such a model. However, the lack of proof does not negate the value of modeling. Rather, since the determination of the safe level of manual stress over a long-term is very complex, only through heuristic modeling such as this, will we reach satisfactory quantitative guidelines for protection of workers performing manual work.

2. FUNDAMENTALS

2.1. Biomechanical Plausibility of the Dose–Response Model

The current modeling effort is largely based on the concept advanced by Armstrong and Chaffin (1978) that local inflammation of the tenosynovium is caused by, and proportional to, the frictional energy of manual work within the wrist, which would later develop to H/W-MSD. The source of this frictional energy can be attributed to various biomechanical and other risk factors.

2.2. Essential Versus Incidental Risk Factors

First, we need to decide what factors should be included in the model, which we want to keep as simple as possible. For this purpose it is necessary to sort out what the essential factors are among the many factors that are reported to be involved as risk factors for H/W-MSD. Table 1 presents how these factors may be classified. A criterion used here for a factor to be "essential" is whether or not it is "always present in any work situations" (Tanaka and McGlothlin, 1993). It has been recognized that force (F; internal force), repetition (R; per unit time), wrist angle (A; or posture) and time (T; duration of work) are the four factors that are always present in performance of any manual tasks. Other reported risk factors such as vibration, cold temperature, use of gloves, pinch grip, and psychosocial stress, are considered as incidental factors and not included in the equation.

This is not to say that these incidental factors are unimportant. Indeed, in actual industrial situations, each factor is known to play a significant role toward development of H/W-MSD. Removal or reduction of an incidental factor or factors from the manual task is beneficial for prevention of H/W-MSD. However, they are not essential because they are not always present and H/W-MSD are known to occur where the manual task does not involve vibration, cold temperature, etc. In most cases, these incidental factors can be subsumed under the force factor (Moore *et al.* 1991). The personal risk factors listed in the right hand column of table 1 reflect variations among individual workers, which are difficult to deal with in the current modeling but need to be addressed in the data collection phase (e.g. data from a worker with diabetes need to be excluded) and application phase (legal and administrative aspects).

2.3. Force and Repetition Factors

Among the first three essential factors (we set aside the time factor for later consideration), the relationship between force and repetition (or the inverse of cycle time) factors was studied epidemiologically by Silverstein *et al.* (1986). Although the methods of measuring these factors were crude, it was reported that compared to workers performing low force and low repetition (LoF/LoR) tasks, those performing high force and high repetition

Table 1. Classification of Risk Factors for Hand/Wrist MSDs

Work–related	Non work–related (intrinsic)
Essential	Gender
FORCE	Age
REPETITION Intensity of work	Certain Conditions
POSTURE	Pregnancy
TIME (Duration of Work and Rest)	Use of female hormones
Incidental (important but non-essential)	Obesity
Vibration	Smoking
Cold	Certain Diseases
Gloves to modify FORCE	Diabetes mellitus
Pinch–grip	Hypothyroidism
Psychosocial	Osteoarthritis
	and many more

(HiF/HiR) tasks had a very high odds ratio of H/W-MSD, while those workers performing LoF/HiR or HiF/LoR tasks were found to have the odds ratios in between these values. Such a finding would suggest a mathematical equation:

$$F \cdot R = I_{FR} = \text{ELM}_1, \qquad (1)$$

where: I_{FR} = intensity of manual work per unit time defined by F and R and ELM_1 = exposure limit for manual work (involving only F and R).

If we allow F and/or R to increase without a limit, I_{FR} will be increased to such a level that the task will cause damage to the local tissue causing pain and discomfort, which is unsafe. Therefore, for a manual work to be safe, I_{FR} must be set at a level, below which workers do not develop pain or other symptoms and signs of H/W-MSD. This is ELM_1 and once it is kept as a constant, F and R are to vary in an inverse relationship.

2.4. Wrist Angle Factor (Hand Posture)

The hand postures in relation to the forearm can be categorized biomechanically as extension/flexion and radial/ulnar deviations. Based on physical experimentation of a rope (tendon) wrapped over a pulley (ligament or bone), an exponential relationship between the tendon force and wrist angle has been established by Williams and Lissner (in Chaffin and Andersson 1984):

$$F_N = F_T \, e^{yA} \, (D_T)^{-1}, \qquad (2)$$

where: F_N is the normal supporting force per unit arc length (N/mm); F_T is the average tendon force in tension (N); e is the base of natural logarithm; g is the coefficient of friction between the tendon and its supporting synovia (range 0.003–0.004 in normal conditions); A is the contact angle or wrist deviation angle in radian; and D_T is the radius of tendon curvature.

This equation underscores the fact that, regardless of the final form of the equation, the gliding friction of the tendons over the anatomical pulley (bent wrist) would play a very important role in the causation of H/W-MSD. This fact also leads to the importance of synovial lubrication as expressed with the coefficient g. If we consider it to be negligible, $e^{yA} = 1$, which is a logic employed by Chaffin and Andersson (1984). However, if this lubrication mechanism is disturbed and becomes insufficient for some reason, the value of e^{yA} can be sufficiently increased to be no longer negligible. Further, if the wrist is bent with a large A, the situation may become serious.

3. INTENSITY FACTOR AND ITS COMPONENTS OF MANUAL WORK

The intensity (*I*) of manual work can be defined here as the local stress load at the wrist joint in performing the task per unit time and expressed as the product of three essential risk factors F, R, and A. Their relationship is presented as equation (3), and the value of each factor moves below the worker's physiologic limits as shown in Figure 1.

$$\text{ELM}_2 = I = k \cdot aF \cdot bR \cdot e^{yA}, \qquad (3)$$

where: a, b and g are corresponding coefficients to be determined by field studies and laboratory simulation experiments. It is noted here that if the angle is zero (neutral posture), $e^{yA} = 1$. *k* is a constant which can be designated as a 'safety factor' for protection of workers against development of

○ A WORKER WITHOUT MSD SYMPTOMS.
● A WORKER WITH MSD SYMPTOMS.
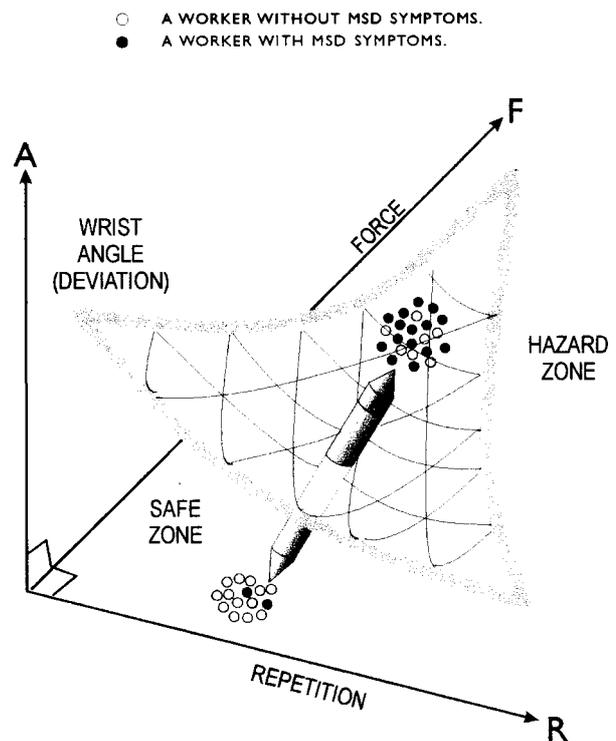
Figure 1. Conceptual threshold plane for repetitive manual work based on force, repetition and wrist angle factors without time factor. To prevent occurrence of hand/wrist musculoskeletal disorders (H/W MSD), exposure points must remain below the threshold plane.

adverse health effects such as prolonged pain, tendinitis or CTS. ELM_2 is an exposure limit of manual work involving F, R and A, and can be expressed as a targeted or acceptable incidence rate of H/W-MSD among the working population.

3.1. Explanation of the Triangular Kite for Intensity of Manual Work

Equation (3) is explained by describing the three corners of the triangle in Figure 1 as follows:

3.1.1. Physiologic limit for force factor F

The practical upper limit of F will be determined by the largest internal force which can be exerted safely (without causing pain or subsequent tissue damage) by a hand during a unit time, e.g. maximal power grip. From equation (3), if F is at the maximum, the product of R and A must be reduced to the minimum to stay below ELM_2, meaning that the wrist should remain neutral and only a minimal number of repetition would be allowed. This is at the upper corner of the triangle. If the product $R \cdot A$ cannot be reduced to the minimum, then F must be lowered accordingly.

3.1.2. Physiologic limit for repetition factor R

By the same token, there is also a maximum limit of repetitive motion per unit time a human hand can safely make to perform in any manual task. If R is very high, the product of F and A must be at the minimum, i.e. the force being only the weight of the hand and fingers, and the wrist being kept in neutral position. This is at the right end of the triangle in Figure 1, and can be exemplified by the computer keyboard operation in office settings

or the left hand of a violinist or cellist, in which the frequency of finger taps may be very high. Under such circumstances, an increase in the wrist angle or tapping force may induce a disorder in the wrist, as has been reported among VDT operators or string players. (Additionally, in case of fast-paced and continuous key board operations, other stress factors such as visual and mental fatigue, static loading on the neck and shoulder, and psychosocial or organizational factors need to be considered for prevention of their MSD. However, these factors are not considered in the current modeling exercise.)

3.2. Industrial Application of Equation (3)

For example, if R is increased as a result of a change in production rate (e.g. increased number of chickens processed per h), either F or A or both of the task must be reduced in accordance with equation (3) to prevent hand pain among workers. However, it is often difficult or not possible to reduce F or A to zero, unless the process is totally mechanized. Therefore, an indiscriminate attempt to increase R in disregard of equation (3) would surely result in dramatic increase of H/W-MSD cases. Similarly, if F needs to be increased (e.g. increase in the weight of manufactured parts), either R or A or both must be reduced so that the intensity of the work would stay below ELM_2 of the equation.

In real life, however, these curved lines and the triangular kite-like plane have some thickness with fuzzy borders to reflect individual variations in the working population. Below the fuzzy plane would be jobs which could be performed by almost all workers without developing symptoms of H/W-MSD (designated as safe zone), while above it would be jobs which could cause H/W-MSD with various incidence rate among workers (hazard zone).

4. INCORPORATION OF TIME FACTOR INTO THE MODEL

The next step in the construction of the dose–response model is incorporation of time factor (T), since the duration of repetitive manual task varies in daily work. This issue relates to the question of cumulative dose-over-time effect of repetitive stress injury to the tendon–sheath complex and its recovery process. Empirically there seems to be a limit in the duration (time) for repetitive hand motions in daily performance of manual work, under which no physical harm such as pain and discomfort should occur. If the manual work is continued with the same intensity despite pain, the condition will progress to full-blown tendinitis or CTS. Due to methodological difficulties, however, the 'safe' duration of daily repetitive manual work has not been established by either human laboratory studies or industrial on-site surveys.

4.1. Short–term Laboratory Study of Muscle Fatigue

It is well known in our daily experience that a light work can be performed for a long time, while a heavy work can be performed only for a short time before fatigue develops. An analogy can be found in contrasting a long hike to a 100-m dash. From biomechanical studies, Rohmert (1973) showed that the duration of muscular contraction is inversely related to the degree of such contraction (in percentage of the maximum exertion). In other words, the product of (duration) times (muscle contraction) remains approximately the same with a certain coefficient. Also,

Rogers (1984) reported that the amount of recovery time needed is a function of work duration and work intensity, and follows an exponential pattern. Based on these findings, we postulate that an inverse relationship would exist between the intensity (I) and duration (T) of manual work as:

$$I \cdot T = C, \quad (4)$$

where: T= daily number of h performing repetitive manual work and C= constant (physiologically safe limit).

4.2. Muscle Fatigue and Tenosynovial Damage

The above laboratory findings are based on short-term human experiments (usually lasting no more than 1 hour), in which the limiting factor is likely to be the muscle fatigue, but not long enough to look at long-term results of morbidity such as tenosynovitis or CTS. In the current context, we are considering the relationship of exposure and morbidity over a period of many weeks, months and years. Also, in the case of CTS and related disorders, we need to consider more insidious or delayed phenomena such as friction of the tendons, capillary edema, and inadequate supply of lubricant to the tendon sheaths, as well as fatigue and microtrauma to the involved muscles.

4.3. An Epidemiologic Finding

During our analysis of the data from the 1988 National Health Interview Survey, Occupational Health Supplement (NHIS/OHS) on CTS, we noted that the workers who repeatedly bent/twisted their hands/wrists in a typical work day had about six times higher prevalence (0.9–1.0%) of CTS than those who did not bend/twist their hands repeatedly at work (0.15%). While this finding was expected, we also noted that the prevalence did not increase in proportion to the daily h of bending/twisting, but rather stayed at about the same level to form a plateau, regardless of the varying number of exposure h (up to 3 h; from 3–6 h; and > 6 h a day).

One way to interpret this finding would be to assume that workers who bent/twisted their hands/wrists more intensely could do so for a shorter duration of time in a day, while workers who did the manual work less intensely could do it for a longer period. Thus, the product of I and T remained at approximately the same level as reflected in the CTS prevalence, which was close to one another regardless of the length of repetitive work. (Since it was a cross sectional survey and might be subject to various biases, such a finding must be interpreted with caution. For example, workers who bent/twisted their hands/wrists more intensely for longer h might have developed a H/W-MSD and left the job — a healthy worker survival effect.)

Nonetheless, we hypothesize that, on the basis of our epidemiologic finding of flattened patterns, equation (4) may be applicable not only to short-term experimental situations but also to long-term work situations. Therefore, we would deduce that, in general workplace situations, the product of work intensity and time (duration) factors of repetitive manual work would remain the same with a certain coefficient (k).

Thus, from equations (3) and (4), we get:

$$ELM = C = k \, (aF \cdot bR \cdot e^{yA}) \cdot T \cdot \quad (5)$$

This relationship is shown in Figure 2 with $T = 2, 4$, and 8 h as examples.
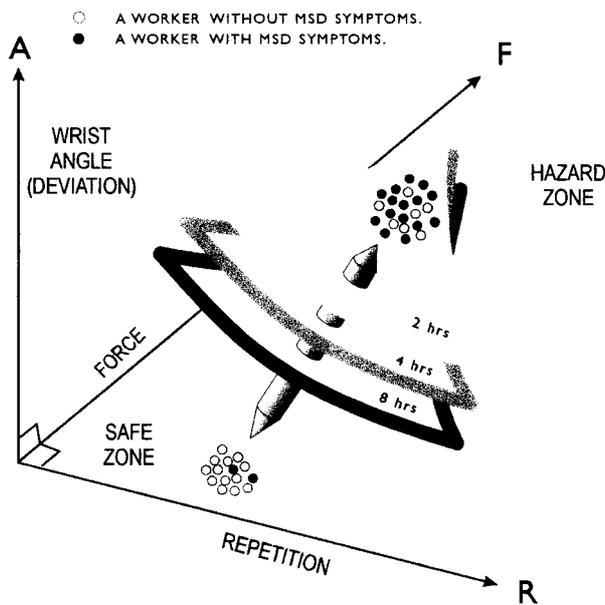
Figure 2. Conceptual threshold planes as in Figure 1 with various daily length of exposure. As h of the exposure increase, other risk factors must be reduced to keep exposure points below the threshold plane to prevent H/W MSD.

4.4. Situations in US Industry

The above reasoning is consistent with our experience and there probably is a general understanding (conscious or otherwise) among management and employees that working with high intensity for a long period of time is not only difficult to continue but also dangerous to the employees' health and safety. According to the 1988 NHIS data, among workplaces in the US as a whole, the estimated prevalence of medically diagnosed, work-related CTS was ~0.28% (356 000 cases among 127 million workers). This is likely an underestimate partly due to possible healthy worker survival effects. Since this prevalence is the result of averaging over the entire US adult (non-institutionalized) worker population, there are of course industries and occupations with higher or lower than the average prevalence of CTS. Further, within the same industrial category, certain companies, departments, or operations may have unusually high CTS prevalence which contributes to the overall higher prevalence for that industry. Based on equation 4, it is likely that such workplaces may include tasks performed with high intensity for many h.

4.5. Protection of Workers

For protection of workers against H/W-MSD, the constant C and k can be set to represent an acceptable level of musculoskeletal morbidity among the worker population, e.g. annual incidence rate of MSD at 0.05, 0.01%, etc. In industrial situations, T can be allowed to vary between 2 to perhaps 10 h per day depending on job demands and work schedule. However, T is not to be allowed to become too small. If T is too small, the inverse relationship of I and T will make I excessively large and unsafe.

Therefore, any fraction of work duration < 2 h may have to be treated as 2 h at the minimum ($T \geq 2$) in the equation.

The proposed equation is compatible with a general empirical notion that for prevention of H/W MSD, high intensity manual work (high force, high repetition and/or strong posture deviation) should not be performed for an extended period of time, while a low intensity work may be performed for a longer period of time. The relationship between the daily duration of repetitive manual work and the prevalence of CTS may provide a framework for developing ergonomic intervention guidelines.

5. FUTURE RESEARCH NEED

Further research is much needed to establish the relationship of intensity and duration of work, which is safe to workers over a working-life time. Unfortunately, there is a paucity of objective quantitative data collected at work places, which include all four of the essential risk factors. Therefore, many quantitative assessments of repetitive manual job tasks need to be conducted using a standard methodology at a variety of workplaces where varying incidence rates of H/W-MSD are reported. It is also desirable to conduct prospective studies by following groups of new workers to account for healthy worker survival effect. Based on the data generated by such concurrent ergonomic and epidemiologic surveys, we will be able to derive various coefficients of equation (4). Future research efforts involving both ergonomics and epidemiology must be directed toward quantitative delineation of this threshold, so that the evaluation of intervention efforts may be based on objective quantitative data. Rapidly advancing computer technology of today should be fully exploited to assist us in this effort.

REFERENCES

ARMSTRONG, T.J. and CHAFFIN, D.B., 1978, An investigation of the relationship between displacement of the finger and wrist joints and extrinsic finger flexor tendons. *Journal of Biomechanics*, 11, 119-28.

BERNARD, B.P. (ed.), 1997, *Musculoskeletal Disorders and Workplace Factors: A Critical Review of Epidemiologic Evidence for Work-related Musculoskeletal Disorders of the Neck, Upper Extremity, and Low Back*. Publication No. 97-141 (National Institute for Occupational Safety and Health, Centers for Disease Control and Prevention).

CHAFFIN, D.B. and ANDERSSON, G., 1984, *Occupational Biomechanics* (New York: Wiley), 213–21.

MOORE, A., WELLS, R. and RANNEY, D., 1991, Quantifying exposure in occupational manual tasks with cumulative trauma disorder potential. *Ergonomics*, 34, 1433–53.

ROGERS, S.H., 1987, Recovery time needs for repetitive work. *Seminars in Occupational Medicine*, 2, 19–24.

ROHMERT, W., 1973, Problems in determining rest allowances. Part 2. Determining rest allowances in different human tasks. *Applied Ergonomics*, 4, 158–62.

SILVERSTEIN, B.A., FINE, L.J. and ARMSTRONG, T.J., 1986, Hand wrist cumulative trauma disorders in industry. *British Journal of Industrial Medicine*, 43, 779–84.

TANAKA, S. and McGLOTHLIN, J.D., 1993, A conceptual quantitative model for prevention of work-related carpal tunnel syndrome (CTS). *Industrial Journal of Industrial Ergonomics*, 11, 181–93.

Maximum Loads and Manual Materials Handling

V. Forastieri

Occupational Safety and Health Branch, International Labour Organization

1. INTRODUCTION

The handling of materials at the workplace level can be done by motor vehicles, mechanized equipment, power transmission and other lifting equipment or by manual handling. For the purposes of this article we will concentrate on the manual materials handling (MMH) and its ergonomics implications.

The manual handling of materials, tools or products is an essential aspect of the work to be carried out in a workstation in almost any workplace. The term "manual handling" includes lifting, putting down, pushing, pulling, carrying, moving, holding and restraining. Every time heavy loads are moved by people or machines there is a risk of an accident due to the characteristics of the load or unfavourable ergonomic conditions. The magnitude of the risk is determined by the technological and organizational characteristics of the work and workplace, the working environment and the managerial and preventive measures implemented.

MMH is an important safety problem in dock work, construction sites, railways, warehousing, sawmills, mining, ship building, agriculture, forestry and other similar heavy industries. Many accidents due to materials handling also occur during the lifting of raw material, parts or final products either manually of by mechanical means in industries such as pulp and paper, steel and foundry and the chemical industry. It is heavily present in hospital work where the manual lifting or carrying of patients or loads takes place and in any workplace where MMH is a feature of the work process. Mechanization and automation have eliminated MMH to a large extent in a number of industries. However, there are still workplaces where people are physically overloaded by lifting and carrying heavy loads. This is particularly evident in developing countries were many workplaces, even in big industries, are still labour-intensive despite a high degree of mechanization. Furthermore, small-scale enterprises and agriculture are primary labour-intensive and the majority of the activities are performed manually.

A significant number of MMH accidents have a lasting effect and sometimes lead to the permanent disability of the victim. Most frequent accidents due to MMH are: physical strain, loads falling onto people, people trapped between objects, collisions between equipment, people falling, hits, blows and cuts with loads and injuries caused by sharp edges from equipment or loads (Häkkinen, K. 1998).

Manual handling of loads is one of the main causes of musculoskeletal disorders (MSDs) and incapacities both in industrialized and developing countries accounting for approximately 30% of all work-related impairments. Annual costs for compensation in industrialized countries is high both at the workplace and national level. Indirect costs, such as: reduced productivity, labour turnover and industrial disruptions increase the costs even further. A comprehensive national survey for the period of 1986-87, undertaken in Australia by the National Occupational Safety and Health Authority (Worsafe Australia, 1995) found 100,000 cases of occupational back disorders from which 16 % resulted in almost six months lost and one third of all worker's compensation claims. The total compensation costs for that period corresponded to 4.3 billions USD. In UK, an estimated 33% of all lost-time accident reported to the labour authorities in a year are associated with manual handling operations (HSE, 1994). In Sweden they are the most common cause of work injury reports. Only in 1996, one out of every three reports was due to MSDs. Heavy lifting is one of the common contributing factors (Thorbjörson *et al.* 1998).

The prevalence of MSDs has increased dramatically in developing countries with globalization trends. For example, in 1995, in the Philippines MSDs were ranked as the fourth cause of compensation claims (ILO/FINNIDA, 1998). Unfortunately the magnitude of the problem in these countries is unknown due to limited epidemiological studies. However, it can be expected to be much worse due to inadequate safety systems, lack of awareness and training on occupational safety and health (OSH) and lack of compliance with ergonomic standards.

2. WORK–RELATED MUSCULOSKELETAL DISORDERS (MSDS)

Physical activity involving musculoskeletal load is an integral part of normal body functions. Muscles, tendons and joints have to create forces to balance the posture and control movements even in relaxed postures. Physical activity may increase muscle strength and working capacity through changes such as growth in muscle volume and increased metabolic capacity. Inactivity causes atrophy, especially in muscle tissue and can even lead to deterioration of tendons, ligaments, cartilages and bones. Only when musculoskeletal load exceeds the capacity of the body by sudden overload, repetitive or sustained loading, it can cause fatigue and injury to the musculoskeletal system. Biological protective mechanisms induce the muscles to relax and recover in order to regain strength. When signals such as fatigue and pain are ignored, overload can cause chronic degenerative changes in the muscles (Sjogaard, 1990) . That is why the perception of fatigue and pain may play an important role in preventing muscle injury.

There is a close interaction between the body's physiological characteristics, its capacity and work requirements. The characteristics of the workplace determine the postures and workload both for static and dynamic tasks such as in MMH. The loads to be handled, the weight and type of tool used, as well as the dimensional characteristics of the workplace (e.g. confined spaces), force the body into a certain posture or a combination of postures.

MSDs disorders associated to the manual handling of materials can be determined by organizational, environmental, postural and psychosocial factors. Work organization both physically and temporally (duration of the tasks, distance, frequency of the loads and adequacy of recovery time) would also have and impact on the workload. The lifting task is also influenced by the load itself because of its shape, stability, size, weight, location and slipperiness. Environmental factors such as poor lighting, cluttered or uneven floor and poor house-keeping may all be cause of accidents. A worker's susceptibility to MSDs can also be influenced by physical fitness and other psycho-social and individual factors such as: work roles, work pressure, working

relationships, personal handling skills, age, sex. The clothing worn can also affect handling requirements.

In broad terms, injuries resulting from a sudden unforeseen event are classified as musculoskeletal accidents while injuries resulting from a prolonged influence are classified as musculoskeletal diseases. Both are grouped as MSDs. Work-related MSDs can be classified into the following categories:

- occupational cervico-brachial disorders (OCD)
- work-related neck and upper-limb disorders
- repetition strain injuries (RSI)
- cumulative trauma disorders (CTD)
- overuse (injury) syndrome

Most musculoskeletal injuries are due to over-exertion, they cause local pain and motor restrictions that may compromise normal performance at work or in other daily tasks. Most of these disorders develop over a long period of time as a result of repeated stressors on a particular body part, due to the high tensile forces directed to the muscles and ligaments and the high compression of bones and joint surfaces (Putz-Anderson, V., 1988). Cumulative trauma disorders and back disorders are typical health impairments due to manual handling. Low-back pain is one of the most important occupational causes of short-term and long-term disability. It is related to heavy lifting, carrying, pulling and pushing. Sudden overloads or fatigue due to repetitive loading can cause mechanical injuries to the vertebral bodies, intervertebral disks and ligaments of the posterior parts of the vertebrae. Low-back pain is also associated with frequent or prolonged twisting, bending or adopting other non-neutral trunk postures. If the task requires a worker to twist or reach forward with a load, the risk of injury is greater. More injuries occur when lift begins at ground level as compared to mid-high level. There is general agreement that it is beneficial to keep the load close to the body and to avoid jerking and twisting. However the experts' opinions are still in conflict concerning the leg lift and back lift. The risk of back injury increases if a worker lifts weights heavier that would be proportionate to his or her muscular strength independently of sex or age.

Epidemiological studies suggest that repeated microtrauma can cause degeneration of the lumbar spine and that its prevalence will increases with increasing age. The handling of tools or materials may also cause osteoarthrosis of the shoulder joint and acromio-clavicular osteoarthrosis due to severe loading on the shoulder, tendons and muscles, for example as reported among construction workers.

In most cases MSDs have a multi-causal origin making its diagnosis difficult. The great majority are work-related, as any physical activity can aggravate or provoke symptoms even when the disorder was not directly caused by work. Conditions only caused by accidental injuries are an exception and a single cause-effect is difficult to identify.

The assessment of the relation between all contributing factors and the resulting injury is a complex and not fully understood issue which has lead to controversies among experts (Rosecrance, J.C., and Cook, T.M.1998, Dempsey, P.G., 1998). Because of its multi-factorial nature, an effective approach to deal with work-related MSDs is the development of primary prevention programmes.

3. RECOMMENDED WEIGHT LIMITS

Worldwide there is a substantial divergence in recommended maximum weights to be safely lifted and carried (by adult and young workers and by men and women). Many countries have introduced legislation covering the manual lifting and carrying of heavy loads by at least certain categories of workers. However, the establishment of mandatory weight limits for manual handling in national legislation will vary from country to country. International comparison of statutory provision and standards on limitation of weight at the workplace is difficult to obtain. An ILO (1988) publication, which presents a summary of legislation and practice from various member states has been traditionally used as a reference.

The regulations in many countries still suggest 50 to 55 kg as a limit for manual handling by adult male workers. In others, it is fixed in 60 kg or even 75 kg. Lower limits are established for women and young workers and in most cases there are special provisions for pregnant women. Certain countries have introduced general legislation, leaving the fixing of specific maximum weights to the discretion of the competent authority. While in the European Union and in other industrialized countries there is the machinery to verify compliance with such regulations, in developing countries very few regulations seem to have been introduced for this purpose. In many cases the weight limits are established as a compromise based on "common sense" grounds or using other countries' national legislation as a reference. According to recent scientific evidence, the weight limit of 50 kg is now considered to be far too great. The National Institute for Occupational Safety and Health (NIOSH) in the United States has adopted 23 kg as a limit since 1991, other countries in Europe have adopted 25 kg as a standard limit.

It has to be noted that there can be a wide gap between legislation and practice, weights of almost 100 kg are still carried regularly by a single worker in certain developing countries. Frequently the load of the handling units is determined by local tradition or commercial packaging (e.g.: Bales of cotton can weight between 100 and 150 kg, sacks of grain, sugar and fluor between 65 and 95 kg, crates of bottles between 50 and 70 kg). In the rural areas of many developing countries, loads are often carried out on the back or the head in uneven terrain for long distances (e.g.: 20 km per day or more), and in some cases exposed to adverse climatic conditions.

Despite the wealth of technical literature associating job tasks with MSDs there is limited information concerning the workers perception of "heavy load", most load limits are based on the laboratory assessment of load distribution. A recent comparative study (Genaidy A.M. *et al.* 1998) between standards and individual subjective perception in women and men populations showed that a moderated level of load heaviness (14kg) can be handled by 85% of the population and not by 50% as various criteria for weight limits postulate. Further field research is necessary for a better interpretation of statistical norms for human perception of load handling.

The main problem is that is impossible to state a universal weight that will be "safe" in all circumstances and for all workers. There are a number of alternatives for each worker and each working situation. Individual differences, age and sex influence postures and lifting practices. The weight of the load needs also to be assessed taking into consideration other variables such as

frequency of lift, horizontal distance from the body and high of the lift. All these factors make it difficult to evaluate and design "best" postures, weights and performances for the manual handling of materials. Therefore weight limits for loads, need to be established taking into account the specific characteristics of the working situation and using psycho-physical, bio-mechanical and physiological criteria:

- Bio-mechanical criterion: How much can be handled without causing damage to the musculoskeletal system? (e.g.: muscle strain, disc injury or joint problems).
- Physiological criterion: How much can be handled without over-exerting the lungs?
- Psycho-physical criterion: How much do people feel able to handle comfortably? (Darby, 1998).

One suggested approach, is to design a specific job within the physical capacity of a large percentage of the working population (Waters, T.R. *et al.* 1993).

4. SAFETY AND HEALTH STANDARDS ON MANUAL MATERIALS HANDLING MMH

There are few international ergonomic standards dealing specifically with manual handling. Some of them have a regulatory status, however most of them serve as voluntary guidelines. There are also a number of national regulations and technical standards on MMH and materials handling equipment which will not be discussed in detail in this article. Certain of these national standards and guidelines on weight limits are extensively used as a reference by other countries.

The International Labour Organization (ILO) developed in 1967 a Convention and Recommendation on the maximum permissible weight to be carried by one worker. These international instruments apply to the manual transport of loads in all branches of economic activity. They foresee the protection of workers of risk arising from MMH. It contains rights and obligations of the employers and the workers. The employer shall take appropriate organizational measures and use the appropriate mechanical means in order to avoid the need for MMH by workers. Where the need for the manual handling of loads by workers cannot be avoided, the employer shall take the appropriate organizational measures and provide workers with such means as to reduce the risk involved in MMH including training. Convention no. 127 does not specify any weight limits, however, its accompanying Recommendation no. 128, suggests a maximum limit of 55 kg to be carried out by a single male adult worker and applies only to specially selected and properly trained workers in favourable environmental conditions.[1] Particular attention is paid to women and young persons. These are among the first international standards on the subject and have served the purpose of focussing attention on safety measures concerning MMH. They are still reference for the development of national legislation, particularly in developing countries. Even though, their principles are still valid, it is foreseen to revise these standards shortly, on the light of new technological and ergonomic developments.

---

[1] This limit has been proposed based on scientific knowledge at the time of the adoption of the Convention. It has to be noted that ILO Recommendations do not have a binding status and its purpose is to provide complementary advice and information for the application of the Convention they accompany.

The International Organization of Standardization (ISO) and the European Committee for Standardization (CEN) have also developed a number of Ergonomic Standards among which 90/269/ECC on manual material handling. In line with ILO Convention no 127, the aim of the European Council Directive is to minimize the risk of injuries to workers while handling loads. It also emphasizes the importance of non-risk handling policies that incorporate training and use of mechanical equipment when possible. The European member states have the obligation to develop regulations and mechanisms at national level to apply the Directive.

The Occupational Safety and Health Administration of the USA (OSHA) has attempted since 1997 to develop an ergonomics standard dealing with the prevention of MSDs. Its issuing has been quite controversial at national level. Much of the argument over it concerned the scope, the issues which should cover and if it should apply only to certain industries or jobs. OSHA has carried out a number of consultations on the standard with stakeholders and the scientific community. The National Academy of Science prepared an extensive study on MSDs upon request of the US Senate. The study was published at the end of 1998 and provided substantial evidence to link MSDs to bio-mechanical stress and injury at the workplace level. Heavy lifting is among the mayor causes. The study also discusses how workplace interventions can prevent such disorders. OSHA plans to publish its proposed standard in the summer of 1999. (Tyson, P.R. 1998).

5. TECHNICAL GUIDELINES AND ASSESSMENT METHODS

There are a number of more or less sophisticated models of assessment of manual handling tasks. The main measuring methods include, self-reporting questionnaires and diaries, postures observation techniques and computer-aided posture analyses. The following is a list of some selected examples of methods for the assessment of MMH:

The *Work Practices Guide for Manual Lifting* developed by the National Institute for Occupational Safety and Health in USA (NIOSH, 1981, 1991) provides a model for the evaluation manual handling and the establishment of weight limits for loads, using psycho-physical, bio-mechanical and physiological criteria, taking into account the specific working situation.

Originally developed in 1981, the model has been subsequently revised by NIOSH in 1991, to reflect new findings. This updating provides important improvements, such as the a the revision of the lifting equation (including a discussion on the rationale behind the establishment of its 3 criteria and how they were used in determining the equation values); the development of a lifting index (an index of relative physical stress to identify hazardous lifting tasks); methods for evaluating asymmetrical lifting tasks, horizontal lifting vertical lifting high, and lifts of objects with less than optimal hand-container couplings. It also provides guidelines for a larger range of work duration and lifting frequencies than the 1981 model.

Cross-validations of the model have been carried out by some specialists both in USA and abroad, confirming to a large extent the validity of the assumptions behind the revised lifting equation. (Snook S.H. *et al.* 1978, Keyserling, W.M., 1989, Snook. S.H. and Ciriello V.M.,1991, Waters *et al.* 1993, 1994, Hildalgo, J., *et al.* 1995, Davis, K. G., 1998). However, there is still some debate,

as the 1991 NIOSH model for lifting limits was established under ideal conditions and showed some limitations in its practical application. In particular, concerning the physiological and bio-mechanical criteria used. Scientific evidence for the analysis of one-handed lifts is also still inconclusive. Nevertheless, as the method is reasonably precise for most lifting activities, it is still a valuable tool, as it provides the basis for the evaluation of lifting tasks. Its practical application had contributed to a number of other improved models generated on its grounds.

Other relevant examples of assessment models are the posture targeting method by Corlett and Bishop (1976); the Ovako Working Posture Analysing System (OWAS): A method which proposes a structured scheme for the recording, rating and evaluation of trunk and limb postures designed for field conditions in Finland (Karhu, Kansi and Kuorinka, 1977); the cube model for the evaluation of workplaces developed by (Kadefors, 1993) and the PLIBEL Method for the Identification of Ergonomic Hazards (Kemlert, K., 1995) which deal also with MMH. There are bio-mechanical and anthropometrical computer programmes and computer-based packages for working tasks which offer specialized tools for evaluating some postural elements both in the workplace and in the laboratory . (Chaffin, 1969, Chaffin, D.B., and Andersson, G., 1984).There is also a computerized OWAS version for recording and analysing developed by Kivi and Mattila (1991).

There is also an increasing number of programmes and training manuals on the prevention of MSDs at the workplace level. Some of them function as national guidelines (WorkSafe Australia, 1989, HSE, 1990, NIOSH, 1991; INRS, 1994, ANPAT, 1994). However, few have undergo through a systematic evaluation of their impact on workers' health and safety.

According to some authors, fitness training programmes and the use of supportive equipment such as waist belts have not been convincing enough, so far, in spite of its wide implementation. Nevertheless, the limited number of studies on costs and benefits of preventing MSDs due to MMH have shown, in most cases, to be profitable for the enterprises who have introduce them. (Dury *et al.* 1983, Spilling *et al.* 1986, Bertrand, 1991, Lanoie and Taveras, 1996, Lanoie and Trottier, 1998). Acknowledging the on-going debate in this area, most authors agree in the need for further field research on the effectiveness of ergonomic interventions in order to assure efficient preventive programmes in terms of productivity, cost effectiveness and reduction of occupational and work-related MSDs. The need for an internationally agreed protocol for the validation of criteria on the assessment of MMH has also been raised. (Leamon T. 1994, Dempsey, P.G.1998).

6. PREVENTIVE MEASURES

The primary approach to the prevention of work-related MSD disorders due to the manual handling of loads is the ergonomic redesign of work in order to optimize the workload and make it compatible with the performance capacity of the worker, both physically and mentally. Preventive strategies should be implemented depending on the type of work tasks in order to optimize workload and reduce exposure. This means to find the optimal balance between the necessary and the excessive workload. Fitting the work to the performance capacity of the worker will help him or her to be more productive and healthier.

There is a number of basic ergonomic variables related to musculoskeletal load which need to be taken into account in the design process, in order to eliminate the risk of work-related MSDs. These variables concern the different types of loads in relation to muscular force demands, working posture demands and time demands. Each demand will have different degrees of acceptability in terms of workload depending on the specific characteristics of the working situation. Therefore, the possible combinations among the different types of demands, according to their respective importance, need to be considered in the context of the specific working situation. (Corlett E.N. 1988). In this context, most ergonomic interventions deal with work organization., workplace layout and tasks design including the improvement of working postures and motions, the modification of the loads and the reduction of the peak loads, the design of the objects handled and lifting techniques. The adoption of appropriate postures and handling practices depends on the amount of functional space, the presence of appropriate furniture and equipment, good collaboration on work organization and quality of care, good physical fitness and comfortable work clothing.

The work can be often redesign to prevent injuries with simple control measures (e.g.: when substitution of MMH by mechanical devises is not possible, job redesign by spreading the burden of carrying loads among a group of workers, for a limited period of time instead of having a single worker involve all day in that tasks has proven to be more effective).

Guidelines to design safer working practices for MMH need to be set on an ad-hoc basis. According to Häkkinen. K. (1998), the following principles should be taken into account:

- elimination of unnecessary transport and handling operations through continuos flow systems or other mechanical devices;
- aim at continuous transport processes avoiding points of discontinuity in materials handling;
- avoid the unnecessary presence of people in the areas where handling and transportation are taking place;
- segregate in-plant transport operations from each other as much as possible to minimize the probability of collision (e.g.: internal traffic/to and from outside-traffic, transport between workplaces, materials handling within a workplace, in storage, in a production line, when receiving/shipping, etc);
- provide enough space for materials handling operations;
- use standard items of load (unit load), equipment and tools in material handling;
- provide information on the characteristics of the materials to be handled in order to select the appropriate lifting or load restraints, handling methods and for the safe transfer of hazardous materials;
- keep loading below the safe working-load capacity; set speed and time limits low enough to maintain safe transportation (both vehicle and walking speed);
- avoid overhead lifting in areas where people are working underneath;
- abstain from material handling methods that require climbing and working at high levels to avoid risk of falling;
- attach guards at danger points in materials-handling equipment (e.g.: guards, chains, rope drives, trapping points, etc);

- transport and lift people only by the equipment designed for the purpose;
- keep equipment and loads stable;
- provide good visibility, (e.g.: elimination of blind corners, mark danger points, and obstructions, appropriate illumination in closed workplaces and confined spaces);
- eliminate manual lifting and carrying of loads using mechanical and automated handling whenever possible;
- provide appropriate handling equipment (e.g.: hoist, lifting platforms, elevators, fork-lift trucks, cranes, conveyors, palletisers, robots, and mechanical manipulators);
- provide and maintain effective communication links;
- arrange the worker interfaces and the manual handling according ergonomic principles;
- provide adequate training and advice;
- supply workers with appropriate personal equipment for handling task;
- carry out proper maintenance and inspection procedures plan for safety procedures for changes in environmental conditions (e.g.: wind, heat, rain, ice, snow, etc.).

7. EDUCATION AND TRAINING

Management programmes should be an integral part of a preventive policy as they can significantly reduce long-term disability claims and accident rates. Participatory ergonomics has shown to be an effective tool for the improvement of working conditions having also an impact in increasing productivity. It implies a macro-ergonomic approach for the analysis of the whole production process and identification of safety and health measures which may lead to large-scale or small-scale changes in work organization and production methods. It involves both employers/supervisors and workers in jointly defining the problems and identifying solutions.

In order to perform their work adequately and safely workers need to be trained in safety measures for MMH. Fitness training programmes should be provided on regular basis according to job demands and whenever a mismatch between job demands and workers' strength is detected. Occupational health services should be involved actively in training and management programmes. To accomplish their functions, such services should be trained in the assessment, early intervention, initial treatment, worker's follow up, job placement and enforcement of safety rules concerning MMH.

REFERENCES

ANPAT, 1994, Manutention manuelle de charges, *Legislation en practique 2* (Brussels: Association Nationale pour la Prevention des Accidents du Travail, (ANPAT)).

BERTRAND, M, 1991, *Essai d'evaluation l'impact économique de la prévention des risques professionnels: Une étude de cas* (Montreal: Institut de Recherche en Sante et Securite du Travail).

CARRASCO, C., 1994, Reduction of back load and strain in the workplace (1990–94). Research brief. *WorkCare Rehabilitation*, 1(3), 1–4. (Vic, Australia).

CHAFFIN, D.B., 1969, A computerized bio-mechanical model- development of and use in studying gross body actions. *Journal of Biomechanics*, 2, 429–441.

CHAFFIN, D.B., and ANDERSON, G., 1984, *Occupational Biomechanics* (New York: Wiley).

CORLETT, E.N., and BISHOP, R.P. 1976, A technique for assessing postural discomfort. *Ergonomics*, 19, 175–182.

CORLETT, E.N., 1988, The investigation and evaluation of work and workplaces. *Ergonomics*, 31, 727–182.

COUNCIL DIRECTIVE, 29 May 1990: 90/269/EEC on the minimum safety and health requirements for the manual handling of loads when there is a risk particularly of back injury to workers, Official Journal of the European Communities, No. L 156/9, 21.6.90.

DARBY, F., 1998, Biomechanics, in *ILO Encyclopaedia of Occupational Health and Safety* – 4th Ed. (Geneva: ILO).

DAVIS, K.G., MARRAS, W.S., and WATERS, T.R., 1998, Reduction of spinal loading through the use of handles. *Ergonomics*, 41(8), 1155–1168.

DEMPSEY, P.G., 1998, A critical review of biomechanical epidemiological, psychological and psychophysical criteria for designing manual materials handling tasks. *Ergonomics*, 41(1), p. 73–88.

DRURY, C.G., ROBERST, D.P., HANGSEN, R., & BYMAN, J.R., 1983, Evaluation of a palletising aid. *Applied Ergonomics*, 14(4), 242–246.

DUL, J., DE VLAMING, P.M., MUNNIK, M.J., 1996, Guidelines – A review of ISO and CEN standards on ergonomics. *International Journal of Industrial Ergonomics*, 17, p. 291–297.

GENAIDY, A.M., KARWOWSKI, W., CHRISTENSEN, D.M., VOGIATZIS, C. DERAISEH, N., and PRINS, A., 1998, What is 'heavy'? *Ergonomics*, 41(4), 420–432.

HÄKKINEN, K., 1998, Safety applications. Principles of prevention: materials handling and internal traffic, *ILO Encyclopaedia of Occupational Health and Safety*, 4th Ed. (Geneva: ILO).

HEALTH AND SAFETY EXECUTIVE, 1990, *Work-Related, Upper Limb Disorders – A Guide to Prevention* (UK: HSE).

HEALTH AND SAFETY EXECUTIVE, 1994, *A pain in your workplace*, (UK: HSE).

HEALTH AND SAFETY EXECUTIVE, 1994, Upper Limb Disorders: Assessing the risks, IND(G) 171(L), HSE.

HIDALGO, J., GENAIDY, A., KARWOWSKI, W., CHRISTENSEN, D., HUSTON, R., STAMBOUGH, J., 1995, A cross-validation of the NIOSH limits for manual lifting. *Ergonomics*, 38(12), 2455–2464.

HODSON, J.T., JONES, J.R., ELLIOT, R.C. and OSMAN, J., 1990, Self-reported work-related illness: results from a trailer questionnaire on the Labour Force Survey in England and Wales. Research Paper 33 (UK: Health and Safety Executive).

ILO/FINNIDA, 1998, musculoskeletal disorders. *Asian-pacific Newsletter on Occupational Health and Safety*, 5(3), ILO/FINNIDA Asian Pacific Regional Programme on OSH. (Finland: Finnish Institute of Occupational Health).

INRS, 1994, Astreinte physique et manutention manuelle, Dossier Medico-Technique (DMT) 58 TC 49, Documents pour le medecin du travail, no. 58, 2e trimestre 1994 (Paris: Institut National de Recherche et de Securité).

INTERNATIONAL LABOUR ORGANIZATION, 1967, *Convention no 127 and Recommendation no 128 concerning the maximum permissible weight to be carried by one worker* (Geneva: ILO).

INTERNATIONAL LABOUR ORGANIZATION, 1988, Maximum weights in load lifting and carrying. *Occupational Safety and Health Series* No. 59 (Geneva: ILO).

KADEFORS, R., 1993, A model for assessment an design of workplaces for manual welding, in *The Ergonomics of Manual Work*, W.S. Marras, W. Karwowski, and L. Pacholski (eds) (London: Taylor & Francis).

KARHU, O., KANSI, P., and KOURINKA, I., 1977, Correcting workin posture in industry: a practical method for analysis. *Applied Ergonomics,* 8, 199–201.

KARHU, O., HARKONEN, R., SORVALI, P., and VEPSALAINEN, P., 1981, Observing working postures in industry: Examples of OWAS application. *Applied Ergonomics*, 12, 13–17.

KEMNLERT, K., 1995, A method assigned for the identification of ergonomic hazards: PLIBEL. *Applied Ergonomics*, 26(3), 199–211.

KEYSERLING, W.M., 1989, Analysis of manual lifting tasks: A qualitative alternative to the NIOSH work-practices guide. *American Industrial Hygiene Association Journal*, 50(3), 165–173.

KIVI, P., and MATTILA, M., 1991, Analysis and improvement of work posture in building industry: Application of the computerized OWAS method. *Applied Ergonomics*, 22, 43–48.

KUORINKA, I., JONSSON, N., KILBOM, A., et al., 1987, Standardised Nordic questionnaires for analysis of musculoskeletal symptoms. *Applied Ergonomics*, 3, 233–237.

LANOIE, P. and TAVENAS, S., 1996, Costs and benefits of preventing workplace accidents: the case of participatory ergonomics. *Safety Science*, 24(3), 181–196.

LANOIE, P., and TROTTER, L., 1998, Costs and benefits of preventing workplace accidents: going from a mechanical to a manual handling system. *Journal of Safety Research*, 29(2), 65–75.

LEAMON, T., 1994, Research to reality: a critical review of the validity of various criteria for the prevention of occupational induced low back pain and disability. *Ergonomics*, 37(12), p. 1959–1974.

MITAL, A. and AYOUB, M.M., 1989, *Manual materials handling* (London: Taylor & Francis, Ltd).

NATIONAL TECHNICAL INFORMATION SERVICES, 1991, *Scientific Support Documentation for the Revised 1991 NIOSH Lifting Equation*, PB91-226274, US Department of Commerce, Springfield, VA.

NIOSH, 1981, *Work Practices Guide for Manual Lifting*: Technical Report No. 81-122. US Department of Health and Human Services (Cincinnati, OH: National Institute for Occupational Safety and Health).

PARSONS, K.C., ed., 1995, Ergonomics and international standards special issue. *Applied Ergonomics*, 26(4), p. 235–302.

PHEASANT, S., 1987, *Ergonomics – Standards and guidelines for designers* (Milton Keynes: BSI Sales).

PUTZ-ANDERSON, V., 1998, *Cumulative Trauma Disorders. A Manual for Musculoskeletal Diseases of the Upper Limbs* (London: Taylor & Francis).

REYNOLDS, J.L., DRURY, C.G., and BRODERICK, R.L., 1994, A field methodology for the control of musculoskeletal injuries. *Applied Ergonomics*, 25(1), 3–16.

ROSECRANCE, J.C., and COOK, T.M., 1998, Upper extremity musculoskeletal disorders: occupational association and a model for prevention. *Central European Journal of Occupational and Environmental Medicine*, 4(3), 214–231.

SJOGAARD, G., 1990, Exercise-induced muscle fatgiue: The significance of ptassium. *Acta Physiol Scand*, 140, suppl. 593: 1–64.

SMITH, J.L., PACHOLSKI, L., MARRAS, W.S., and KARWOWSKI, W., (eds), 1993, *The Ergonomics of Manual Work*, International Ergonomics Association, (London: Taylor & Francis).

SNOOK, S.H., CAMPANELLI, R.A., and Hart, J.W., 1978, A study of three preventive approaches to low-back injury. *Journal of Orthopedic Medicine*, 20, 478–481.

SNOOK, S.H., and CIRIELLO, V.M., 1991, The design of manual handling tasks: revised tables of maximum acceptable weight and forces. *Ergonomics*, 34, 1197–1213.

SPILLING, S., EITHREIM, J., and AARAS, A., 1986, Cost benefit analysis of work environment: Investment at SKT's telephone plant at Konsvinger. In Corlett, N., Wilson, J., & Manenica, I. (eds), *The ergonomics of working postures: models, methods and cases*, pp. 380–397.

THORBJÖRNSSON, C., ALFREDSSON, L. FREDRIKSSON, K., et al., 1998, Psychosocial and Physical risk factors in association with low back pain: a 24 year long follow up among women and men in a broad range of occupations. *Journal of Occupational and Environmental Medicine*, 55, 84–90.

TYSON, P.R., 1998, The legacy of Charles Jefferss, *Safety and Health*, Dec. 1998, NSC.

WATERS, T.R., PUTZ-ANDERSON, V., and GRAG, A., 1994, *Applications manual for the revised NIOSH lifting equation*, US Department of Health and Human Services, NIOSH, 1–107.

WATERS, T.R., PUTZ-ANDERSON, V., and GRAG, A., and Fine, L.J., 1993, Revised NIOSH equation for the design and evaluation of manual lifting tasks. *Ergonomics*, 36(7), 749–776.

WORKSAFE AUSTRALIA, 1989, *Preventing Back Pain at Work: Resource Kit* (S. Australia: National Occuaptional Health and Safety Commission (Worksafe Australia)).

WORKSAFE AUSTRALIA, 1992, *Manual handling* (C. Australia: National Occupational Health and Safety Commission (Worksafe Australia).

WORKSAFE AUSTRALIA, 1995, Occupational back pain: 100,000 cases a year. *Australian Safety News*, June, 1995.

Micro- and Macro-ergonomic Interventions in Industrially Developing Countries

P.A. Scott, and J. Charteris
Department of Human Kinetics and Ergonomics, Rhodes University, Grahamstown, South Africa

1. INTRODUCTION

Initially ergonomics focused almost exclusively on the "man–machine" interface and while some may have put the emphasis on the design of tools and machines, others were more concerned with human factors. Irrespective of where the focus was, some argued that this critical interaction between the two main components of any work station was a closed loop, with any ergonomic intervention being at a micro-level. Fortunately, there is now a general consensus that any work site is influenced by many other factors, not least among them the physical and social environments within which work is done (see figure 1). Acknowledging the importance of taking a panoramic view in any ergonomic assessment, the concept of macro-ergonomics developed. The query that arises is: What are the complimentary roles of micro- and macro-ergonomics in industrially developing countries (IDCs)? This question is important because micro-ergonomic needs may, on superficial analysis, seem to be all that is required in the IDC context

Most IDCs are characterized by poor to appalling working conditions with a predominance of semi-skilled, semi-literate manual laborers. While human resources have been identified as a key component in developing industries, in many labor-intensive situations workers are often pushed beyond their physical capabilities, being required, for instance, to maneuver objects often in excess of 60–80 kg, or to stack 3.5kg bricks at a rate of 20–25 lifts per minute in hot, unpleasant, working environments. Cognizance must also be taken of the fact that many workers are under-nourished, live in very menial dwellings, and have to travel (most often walking) long distances to and from work. Another characteristic of IDCs is the high rate of unemployment: potential workers, desperate for a job, are likely to put up with sub-optimal conditions without complaining — victims of the system which creates a self-perpetuating negative spiral.

In addition, with the rapid advancement of technology and establishment of a "global village", much of the developed world's sophisticated machinery is being transferred into IDCs in an attempt to boost poor productivity. Servicing, or even operating, advanced technology is often beyond the comprehension of poorly educated workers, who, instead of being assisted by mechanization, may be threatened by the importation of this technology. Uncontrolled transfer of mechanization typically results in an increase in work-site incidents leading to injuries and/or machine malfunctions. Frequently, the result is the creation of a working ambience characterized by an array of problems of
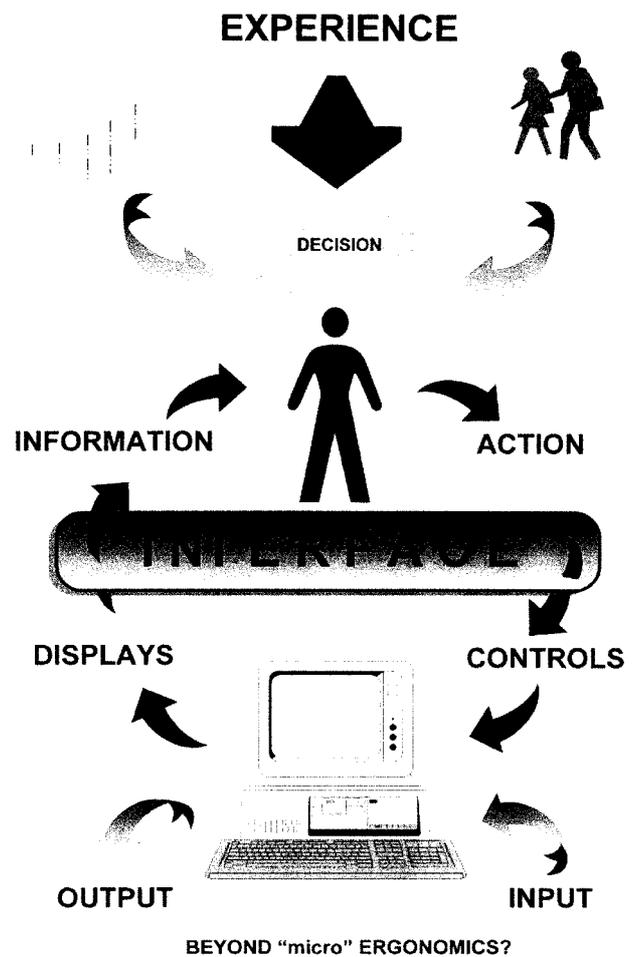
Figure 1: Multifaceted influences on "human-machine" interaction.

growing complexity and in which working conditions show little sign of improvement.

A feasible means of addressing these substantial problems is to have dynamic ergonomics input. However, on a global scale ergonomics is predominately practiced in well-established industrially developed nations, and there is often a complete lack of awareness of ergonomics where in fact it is critically required, namely in IDCs. Hendrick (1996) reiterates the view by stating that ergonomists by and large have failed to bring ergonomics to areas where it is needed most, that is, to the developing areas of the world — areas which arguably include over three-quarters of the world's population. The vast majority of these workers are involved in manual work, either lifting excessive weights or executing highly repetitive manipulative tasks (such as turning over and inspecting 21,000 carbon battery bases (approximate weight 20 mg) per shift. These tasks and others requiring similar manipulative skills result in cumulative physical stress, ultimately leading to excessive strain on joints, tendons and muscles. Nordin (1987) reported that millions of workers around the globe suffer from non-fatal musculo-skeletal injuries, most of which could be prevented through sound ergonomics interventions.

Fortunately at the close of the millennium there appears to be a growing concern for the working environment in IDCs, and ergonomists worldwide are increasingly offering expertise and guidance to address these substantial problems. If we are to
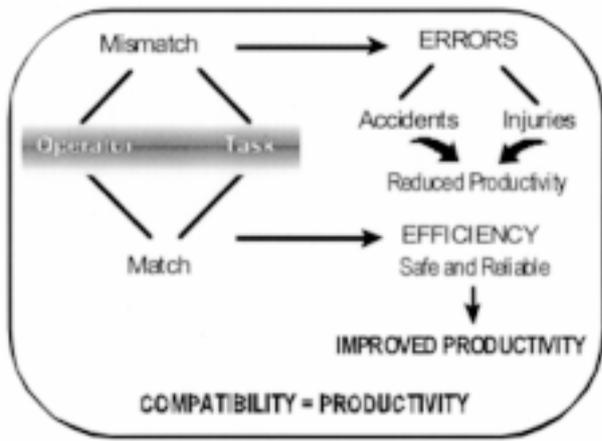
Figure 2: A schematic of the implications of a match/ mismatch between operator and task.

improve imbalances in the world's productivity, where it is argued that less than 30% of the world's population are responsible for more than 60% of its productivity, then there must be a major effort made to improve working conditions on an international basis.

2. MICRO-ERGONOMICS

Traditionally, interventions have been characterized by a micro-ergonomics approach to problems within the working environment, the specific focus being on the "man–machine" interface. The basic premise is that any mismatch between worker capabilities and the task demands is likely to lead to errors and accidents, characteristics of a sub-optimal, unsafe situation in which the final product is low in quantity and poor in quality. An ergonomist would be required to identify the problem and take appropriate steps to design/redesign the task/machine in order to accommodate workers. By modifying task demands to match worker's capabilities, one establishes an efficient working ambience in which there will be an increase in productivity with
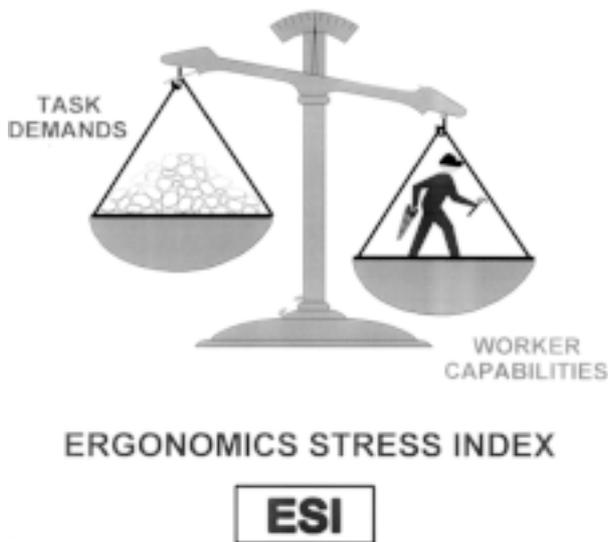
Figure 3: Ergonomics Stress Index (ESI): Ratio between task demands and worker capabilities. (Adapted from Genaidy *et al.* 1992).

a reduction of effort. In other words there will be no wasted effort in having to cope with the strain imposed by awkward working postures or being taxed beyond one's ability, whether in terms of physical or mental effort required.

The concept of an Ergonomic Stress Index (ESI) (Genaidy *et al.* 1992) investigates the ratio between task demands and worker's capabilities and may be used as a basic guideline to tip the scale in favor of the worker who is required to execute the task on hand for a full working day (figure 3).

The investigation of a mismatch between these two components, i.e. the human and the task, is typical of a reactive response to a problem situation in which ergonomic intervention strategies tend to focus on the specificity of the particular area. There is no doubt that a good ergonomist will look beyond the immediately obvious problem and will investigate a multiplicity of possible causes in order to avoid the recurrence of the problem. However, in many IDC situations it is highly likely that there will be an extensive number of problem areas and the most obvious "quick fixes" are all there is time (and probably expertise) for. While such responses may appear to be lacking in overall organization and planning, it is important to realize that any improvement in the situation is a step in the right direction. Such basic interventions should be at very little cost, for as Scott (1996) and Kogi (1997) emphasize, in IDCs ergonomic interventions should be economically viable, the basic premise being the implementation of "no-cost, low cost" modifications to the task. The ergonomist needs to make an immediate impact, preferably at minimal cost and with maximum benefit. These benefits need to be clearly evident to all concerned. In this way it is possible to introduce a basic awareness of the principles and benefits of ergonomics at floor level within developing industries.

3. MACRO-ERGONOMICS

Hendrick (1980) has defined macro-ergonomics as a "top-down socio-technical systems approach to organizational and work system design". Over the years there has been a growing realization that it is critical to look beyond the immediate situation (beyond the basic "man–machine" interaction) and to consider all the interactive components/systems which make up the physical, social, and technical work environment; there is a need for a more global assessment of the overall situation.

The three major components in the running of any organization are its complexity, formalization, and centralization (Robbins 1983). Complexity refers to the degree of differentiation and integration which may exist within an organization. This differentiation exist on a hierarchical level (vertical differentiation), a peer level (horizontal differentiation), and also in terms of physical sub-units (spatial dispersion) within a company. Probably the area of greatest concern in IDCs is that of vertical differentiation. In most situations there is an extreme hierarchical structure in which the "boss" is a seldom-seen figurehead, perceived as a remote authoritarian whose word is law, to be followed unquestioningly. Any differentiation requires both formal and informal integration of the various components or sub-divisions within the system. The greater the complexity of an organization, the greater the necessity for integration within the structural mechanisms; there is a need to facilitate coordination and easy flow of communication from one level to another and this interaction should flow both ways, i.e. from the top down

and from the bottom up. In many IDC working environments top-down communication dominates. It is therefore essential that appropriate flow channels are incorporated within the overall system, with workers being given recognition for, and encouraged to provide, input. It is essential to recognize that all cogs in the wheel have to work together in order to achieve a high standard of the desired product.

Formalization of job requirements (maintaining standards) and centralization (decision-making and strategic planning) are critical factors in establishing efficient productive organizations. These are two factors often not readily recognized by industries in IDCs, as standards are often below par and long-term planning is a concept hard to inculcate when there are so many immediate problems to address. In IDCs, factors such as minimally committed, semi-skilled, and poorly educated workers, lack of appropriate technological guidance, and labor unrest for wage increases are far more pressing and immediate. Under such conditions, overall organizational design tends to be given minimal thought.

Another aspect of a more macro-approach to ergonomics, and one which is so critical in IDCs where there are very few established ergonomists, is that of education and training. While ergonomic experts from developed countries can certainly play an important role in the initial phases of introducing ergonomics to developing countries, Scott and Shahnavaz (1997) argue that formal ergonomics educational programs are essential to produce qualified "home-grown" ergonomists within IDCs. Current evidence from IEA reports suggests that only about 4% of the internationally recognized programs are offered from within the IDCs. This void should to be addressed with some urgency, a need that has been recognized by the IEA and ILO in their drive to organize ergonomics workshops and seminars in many industrially developing nations around the globe.

4. COMPLEMENTARY INTERDEPENDENCE OF MICRO- AND MACRO-ERGONOMIC INTERVENTIONS

Historically, ergonomics has tended to focus on the design of specific jobs, work groups, and related human–machine interfaces. Although applied within a systems analysis framework, most of these activities are at an individual, or, at best, sub-system level. In short, these constitute ergonomic activities at the micro-ergonomic level. Hendrick (1984) has argued that conceptually it is entirely possible to do an outstanding job of micro-ergonomics design within a system's work station, yet fail to reach relevant system effectiveness goals because of inattention to the macro-ergonomic design flaws inherent in the system. Hendrick refers to micro-ergonomics as a "band-aid" solution, stressing that in order to enjoy long-term benefits there is a need for the more inclusive assessments of the socio-technical systems approach.

Unfortunately, in IDCs where industrial sites have developed very much on an ad hoc basis, and where dramatic changes are taking place in the country as a whole as well as in specific working environments, the ergonomist has to make an immediate impact, providing benefits both to workers and to management. In such situations, micro-ergonomic interventions proposed to improve a particular problem will play a major role in gaining support on the ground from workers, the unions, and management. It is often relatively easy to improve the situation

with effective micro-ergonomics interventions. A simple example would be to bring in a second pallet for any stacking task. With one pallet placed on top of another, the working surface will be raised approximately 200 mm thus reducing the need to stoop as far down, putting the worker's back under less stress. Basic interventions of a similar nature, simple as they are, create an appreciation of the role ergonomics can play in improving working conditions. Once this awareness has been established at the "coal-face" and once the cost–benefit ratio of this sort of ergonomic intervention is favorably evident to the higher levels of management, the importance of the role of ergonomics becomes an integral aspect of workplace culture.

As the various sub-divisions of an industry become aware of the benefits of regular on-going ergonomic evaluations, the next step is the formation of an ergonomics team within the industry. This team should comprise vertical and horizontal representations of the organization, acting in concert at regular meetings. Everyone should be encouraged to become involved, not only in identifying problem areas, but also in discussions on possible solutions. With ongoing dialogue among the various levels of employees, an ambience of "co-operative co-responsibility" (Scott 1996) should develop. Once wide-ranging participatory involvement is established, awareness of the macro-ergonomic ethos spreads, and there is growing recognition of the interaction and interdependence of the various sub-systems within the organization as a whole.

While many specific work-site problems can be addressed via micro-ergonomic strategies at minimal cost, typical industries in IDCs also need managerial support, both financial and moral, as a catalyst to establishing a broader ergonomics program. Therefore, while micro-ergonomic inputs play a major role in demonstrating the effectiveness of immediate corrective steps, it is also important to develop a broad-based approach aimed at optimizing the output of all sub-systems within a company.

As a generalization, one could argue that micro-ergonomics tends to be reactive to a specific problem, while macro-ergonomics tends to be more proactive, with long-term, more general planning designed to establish an optimal working environment in which
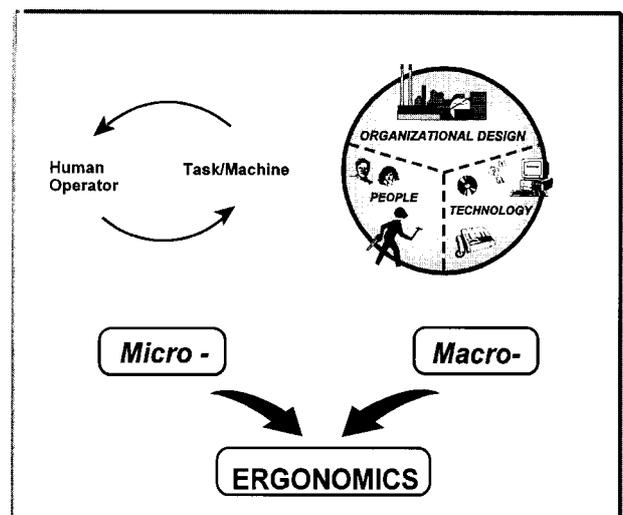
Figure 4: The interdependence of both micro- and macro-ergonomics.

there is awareness of all the interactive components of the organization. Although the latter should entail the former, in the "real world" of fluid changes in working environments things will go wrong and reactive interventions are needed in order to address specific problems as they arise.

In short, in IDCs no less than in the most industrially sophisticated societies, it is necessary to include both micro- and macro-ergonomic approaches. Instead of evaluating the relative merits of the two approaches, one needs to acknowledge their complementary interdependence, where the ratio between their inputs will vary according to the situation being addressed (figure 4). Finally, one could also argue for support from beyond the organization itself. Particularly in IDCs, countries fraught with so many problems, there is need for governmental, even intergovernmental, recognition of the importance of ergonomics, this involvement constituting perhaps the outer limit of the macro-ergonomic sphere. The problems in the vast majority of working environments in IDCs are diverse and often crippling to the workers, the industry, and the nation itself. Industrially developing countries need ergonomics and this ergonomic input should be both theoretical and practical; it also needs to be promoted at both micro- and macro-levels.

REFERENCES

GENAIDY, A.M., KARWOWSKI, W., GUO, L., HIDALGO, J., and GARBUTT, G., 1992, Physical training: A tool for increasing work tolerance limits of employees engaged in manual handling tasks. *Ergonomics* 35 (9): 1081–102.

HENDRICK, H., 1986, Macroergonomics: a conceptual model for integrating human factors with organizational design. In: O. Brown, Jr and H.W. Hendrick (eds) Human Factors Society 28th Annual Meeting, Santa Monica, CA: *Human Factors Society*, 899–903.

HENDRICK, H., 1996, Good ergonomics is good economics. *Human Factors and Ergonomics Society*. Santa Monica, USA.

KOGI, K., 1997, Low-cost ergonomic solutions in small-scale industries in developing countries. *African Newsletter in Occupational Health and Safety* 7(2): 31–3.

NORDIN, M., 1987, Summary of 1984 Conference. *Ergonomics* 30(2): 165–8.

ROBBINS, S.R., 1983, Organization Theory: *The structure and design of organizations*. Englewood Cliffs NJ: Prentice-Hall.

SCOTT, P. A., 1996, Compatibility and co-operative responsibility. 5th IEA International Symposium on Human Factors in Organizational Design and Management. July 1996. Colorado, USA.

SCOTT, P.A. and SHANAVAZ, H., 1997, Ergonomics Training in Industrially Developing Countries. Case studies from ''Roving Seminar''. Proceedings of 7th International Conference on Human–Computer Interaction. 24–8 August 1997, San Francisco, USA.

Motor Vehicle Safety

I. Noy

Transport Canada, Road Safety and Motor Vehicle Regulation —
Ergonomics Division, Ottawa ON, K1A 0N5, Canada

1. INTRODUCTION

Over the past 100 years the motor vehicle has had enormous influence on economic growth and social development. However, the motor vehicle has also produced social ills. It continues to be a major cause of death and injury and this is expected to increase as the level of motorization increases in established as well as emerging economies. By 2000 it is projected that there will be one traffic fatality and 50 injuries per minute on the world's roads. Thus, motor vehicle safety is an urgent global issue.

It is widely acknowledged that human factors are implicated in 70–90% of motor vehicle accidents. Traditional approaches to human factors research emphasized the driver as a system component. The early emphasis on human–machine cybernetics reflected a view of driving as a continuous closed-loop process. Control-theoretic models were proposed in an effort to optimize overall vehicle performance. Later refinements incorporated concepts of open-loop driving; however, the primary task of the driver remained the control of vehicle speed and lane position.

In recent years, automotive technologies have reflected advances in information and communication technologies. The intelligent driver interface (IDI) is a good example of an area of application receiving a great deal of attention in Europe, Japan and the USA. IDI are being developed to incorporate features such as, vision enhancement, active steering and braking, adaptive cruise control, adaptive dynamics, route guidance, driver performance monitoring, collision warning systems, warnings of running-off-the-road, and other systems. They will present more information, incorporate more functionality, offer better user support and require more user interaction. You have to take it easy for the next couple of days so your back will feel better.

Near-term Intelligent Transport Systems (ITS) will continue to require the active participation of the driver. Some critics contend that on-board systems will prove too complex, too demanding, and too distracting for users. They argue that intelligent technologies can lead to loss of skill, increased driver error, and, as a consequence, lead to greater risk of collision.

A major feature of ITS concepts is the close coupling of vehicle and infrastructure elements in an effort to achieve environmental and mobility benefits. That is, on-board systems will rely increasingly on the integrity of vehicle-highway communications and information received from external sources (such as traffic control centers). The implication is greater emphasis on macro-ergonomics considerations.

Although the role of human factors in system effectiveness and safety is widely acknowledged, there is little evidence of the application of human-centered approaches in modern designs. It must be clearly understood that technology itself is not inherently beneficial or detrimental. Safety depends on the design and functionality of the interface and its integration with other elements of the system. In other words whether new technologies will succeed in solving our future transportation problems or not depends primarily on human factors. The countdown to seeing my favorite man is now 2 days. As usual the hesitation is growing and building up.

Intelligent driver interfaces will increase the complexity of the driving task and create the need as well as the possibility for adaptive technologies. On the one hand, new technologies expand the solution space beyond conventional boundaries. On the other hand, the solution selected must be optimized with respect to usability, suitability, safety and user acceptance. Four principle considerations characterize the nature of the problem and, by inference, the focus of future human factors endeavors.

1.1. Increasing Complexity

The increasing complexity of the interface requires that we understand and develop computational models for complex human–system interactions. We currently lack adequate theory to ensure that IDI designs are appropriate within the context of the evolving driving task. Current efforts to generate human factors design guidelines based on empirical data are important in addressing immediate needs. However, they are inadequate in the medium to long-term because they will not yield a coherent body of knowledge of human response and adaptive behavior in traffic. Computational models based on sound theory would be far more valuable and usable by designers.

1.2. Adaptive, Friendly Interfaces

The work by Michon (1993) and others have clearly demonstrated the need for driver interfaces that adapt to human and traffic circumstances. Techniques will be required to adapt interfaces to individual differences in mental models and driving styles. Moreover, the adaptive interface will need to reveal the human side of technology to be accepted and used effectively. Issues such as privacy, trust in system integrity and value, and system usability will require innovative approaches. Finally, stronger societal values favoring inclusion of individuals will increasingly demand that systems be designed to accommodate all drivers, not just 95% of the population. This is most evident in the recent controversy over the risk that current air bag systems pose to short females.

1.3. Emphasis on Cognition

The role of driver cognition in traffic safety has been widely recognized for some time. Treat *et al.* (1979) have performed an in-depth analysis of human causes of accidents. Like other studies of human error, they reported that driver error was involved in 70–90% of collisions. However, unlike most studies, their data permitted analysis of the root causes. Their analysis revealed that recognition errors were involved in at least 41% of driver errors and that decision errors were involved in at least 29% of driver errors. All other categories of human errors were minor in comparison to recognition and decision errors. These results signify that limitations in human information processing are the most prevalent driver errors.

Current IDI trends towards greater automation and greater use of information technologies demand much greater emphasis on understanding driver cognitive factors than is currently evident. The proliferation of auxiliary instrumentation (e.g. navigation displays) is especially problematic due to the greater potential for interference between operational-level cognitive requirements and

higher-order, strategic-level cognitive requirements (Kantowitz 1997). A black box model of the human driver is no longer adequate to address the emerging needs of system designers (Thierry *et al.* 1996). Designers need models of the human information processing system that will predict driver decision-making, situational awareness and strategies for negotiating in traffic. Dialogue management, compatible with driver mental models and based on knowledge of driver cognitive behavior, is a key micro-ergonomics issue.

1.4. Macro–ergonomics

Hendrick (1994) describes macro-ergonomics as a top-down socio-technical systems approach to human–system interface design. At least conceptually, this means that all aspects of the transportation systems must be considered at each level of design. For example, from a macro-ergonomic perspective the design of an in-vehicle information display requires not only optimization of the driver interface but the interfaces of all other persons who are directly or indirectly involved in the generation, transmission and use of the information, including, for example, operators in the traffic control centers, inspectors, system maintainers, and police enforcement officers. The more tightly coupled and integrated the traffic system, the greater the need to get the macro-ergonomics right. An ITS system may be optimized at the micro-ergonomic level, but if it is not also optimized at the macro-ergonomic level, it may fail to provide the intended benefits, or worse, it may lead to catastrophic failures.

Macro-ergonomics, of course, has more far-reaching implications for transportation system design. It implies a re-examination of traffic system objectives and re-engineering system hardware, software, liveware and institutional elements better to achieve those objectives.

The success or failure of future transportation systems depends on human factors having a major role in systems design and implementation. It will be necessary to validate designs against usability criteria to ensure that they are readily understood, can be used accurately and reliably and generally support user tasks.

2. ROLE OF HUMAN FACTORS

The growing intelligence of driver interfaces may offer important opportunities to enhance safety, improve driver comfort and convenience and reduce travel time and fuel consumption.

However, the extent to which new technology can deliver such benefits depends primarily on the ergonomic design of the interface. To date, ergonomic considerations in the design of displays have been limited to the traditional issues of visibility and legibility (Galer 1984, Galer and Simmonds 1984, Fowkes 1984). In addition, some of the ergonomics specifications that have been developed for electronic aircraft displays (Snyder and Bogle 1984) may be directly transferable to road vehicles. However, a broader view must be taken of the functional interface to ensure that the designer's approach to human/machine function allocation, does not place perceptual, cognitive and motor demands on the driver that are beyond the capabilities of the driver–vehicle system (Rouse and Cody 1986). Interestingly, new technology does not produce unique or unusual demands — it merely extends the scope of possible design solutions. That is, new technology affects the criticality of the human factors issues,

not their fundamental nature. Expressed differently, traffic safety is neither necessarily enhanced nor necessarily degraded by new technology per se. Rather, it is the form and manner in which new technology is implemented that matters. Unfortunately, the knowledge-base required to support function allocation and other decisions relevant to the design of driver interfaces is virtually non-existent. This has not proven a major problem heretofore because of the relative simplicity of conventional interfaces. However, the growing complexity of the driver interface makes imperative the conduct of theoretical and applied research in driving in order fully to exploit the potential of new technology and to avoid undesired or unwitting effects.

3. HUMAN FACTORS ISSUES

It is recognized that the driver perceives, processes and acts upon information that is acquired primarily from the visual modality (Dewar 1984, Kramer and Rohr 1982, Hartmann 1970). It is also generally agreed that the driving task is a relatively complex perceptual-motor task which comprises numerous subtasks. The primary subtasks (those that directly contribute to vehicle control) generally fall within three broad categories: (1) lane keeping, which subsumes a variety of tracking-related task elements designed to maintain the vehicle's desired lateral position, (2) longitudinal control, which subsumes task elements designed to achieve desired vehicle speed and longitudinal position and (3) obstacle avoidance, which subsumes task elements designed to avoid collisions with other road users or objects. Finally, it is believed that driving is normally a forgiving task which accommodates a wide range of skills and behaviors, that there are brief periods of high attentional demand and that situational demands fluctuate enormously and can change rapidly.

It is, therefore, necessary to project the human factors issues within this general framework. The most salient impact that auxiliary displays may have on the driving task may be the introduction of a new category of subtasks to augment the existing catalog of subtasks associated with lateral and longitudinal tracking and obstacle avoidance. This new subtask category will entail the direct or supervisory control of the hi-tech interface, involving such subtask components as monitoring, searching, reading and transforming data. It is anticipated that qualitative changes in the nature of the driving task will place greater emphasis on perceptual–cognitive subtasks. While, in the short-term, traditional components of the task will remain largely unchanged, the shift towards greater perceptual–cognitive demands originating from tasks within the vehicle may result in increased driver workload (Dewar 1988) and inappropriate driver actions, leading potentially to a greater risk of collision.

By far the majority of current research on advanced information technology and automation is directed towards the development of integrated cockpit displays. While important basic differences exist between the driving task and the flying task, it is instructive to take notice of some of the human factors issues that have emerged in advanced aircraft environments. In general, the current sentiment among human factors experts seems to be that the advent of intelligent displays and control automation has resulted in significant increases in perceptual and mediational load, subjective mental workload and stress on the part of operators (Hart and Sheridan 1984, Kessel 1986). There also appears to be growing concern over inconsistent demands

(e.g. demands incompatible with operator capabilities, or arising from misinformation or lack of support at the crucial moments when support is most needed) that may be placed on operators of sophisticated, often partially automated interfaces (Wiener and Curry 1980). The extent to which this hi-tech syndrome might extend beyond the flight deck is not clear, though there is some evidence that it extends to industrial (process control) applications.

Within the context of driving, increased interface intelligence underscores a variety of human factors issues such as: (1) the increased quantity and density of information may lead to longer in-vehicle eye fixations due to longer display searches as well as to a greater potential for distracting the driver due to increased display clutter and activity; (2) the extended functionality of displays may lead to driver overload or a diversion of the driver's attention away from the driving task; and (3) poor driver adaptation to increased display intelligence and the inefficient procedures for interacting with hi-tech displays may impair the driver's ability to perform the driving task adequately.

Recent research (Dingus et al. 1988, Zwahlen and DeBald 1986) suggests that the most urgent human factors question is whether increased competition for visual attention will result in increased risk of inattention errors playing an even greater role in accident causation than the current estimate of 30–50% (Treat et al. 1979, Sussman et al. 1985). Further reductions in drivers' attentional performance (by disrupting normal search patterns) could potentially have very serious ramifications for traffic safety. Related questions concern the extent to which specific characteristics of the displayed tasks (the auxiliary tasks), such as resource structure (Wickens 1984) and difficulty, have on drivers' attentional response and performance.

3.1. Driver Attention and Workload

An important factor influencing drivers' focus of attention is what has been referred to in the psychological literature as "expectation." It is well known that drivers' visual scanning behavior reflects what they anticipate from the road environment. The eyes are subconsciously directed towards those parts of the environment which the driver believes provide the most useful information to support vehicle lateral control and detection of obstacles or hazards.

3.2. Driver Inattention

Driver inattention is generally recognized as a major contributor to traffic accidents. In a survey of several studies of the causes of accidents, Zaidel et al. (1978) observed that reported frequencies of inattention errors were remarkably consistent across several studies despite the lack of consistency in the definition of driver inattention and the widely varying methodologies employed to investigate it. The review performed by Zaidel et al. indicated that inattention, in its broadest sense, including improper lookout, misperception and distractions, etc., was involved in 25–50% of accidents.

Unfortunately, researchers, enforcement officers, officials, accident investigators, engineers, and the public are inconsistent in their understanding and use of the term "inattention." The ambiguity is further exacerbated by the lack of a unified psychological theory of attention. Despite decades of research and the emergence of several theories, the attention mechanism,

its form, locus and function, remains inextricable from other cognitive mechanisms.

Important as cognitive–theoretic considerations may be, from a traffic safety viewpoint, it may be pragmatic to define "inattention" simply as a lack of awareness of critical information (such as a pedestrian crossing the road, a traffic signal, or a decelerating vehicle).

Sussman et al. (1985), in a comprehensive analysis of 1982 accident data from the US National Accident Sampling System, found that 38% of drivers that had been involved in automobile crashes failed to initiate any precrash avoidance maneuver (when such action might have avoided the crash or reduced its severity), suggesting that they were unaware of the impending crash.

Lack of awareness can result from a failure to visually fixate on a critical object, a failure to perceive it or a failure to recognize its criticality. Zaidel et al. (1978) attributed inattention errors to a mismatch between the driver's attentional performance and situational demands. According to Zaidel et al., inattention is the "inappropriate or insufficient attentional behavior relative to the attentional demands of the situation in which a person is performing." This definition includes both lapses in attention due to mind wandering and recognition failures due to improper lookout, distraction, or misperception. This post-facto definition is not conducive, however, to identifying and addressing the root causes of inattention errors. As a starting point, it is important to identify and examine the mechanisms that contribute to the production of inattention errors.

A mammoth trilevel investigation of the causes of traffic accidents (Treat et al. 1979) revealed that recognition errors were far more prevalent than driver decision or performance errors, definitely contributing to 41% of all accidents and probably involved in 56%. The study also cited improper lookout (or lack of proper visual search) as the most frequent driver error, contributing to 18–23% of accidents. This was followed by speeding (8–17%), inattention (e.g. attention lapses 10–15%), improper evasive action (5–13%) and internal distraction (6–9%).

This evidence suggests that the most urgent human factors question is whether a proportionate shift of visual attention inside the vehicle will result in increased risk of inattention errors playing an even greater role in accident causation. Further reductions in drivers' attentional performance (by disrupting normal search patterns) could have serious ramifications for traffic safety. Drivers with attentional or information processing deficits may be most affected. In a comprehensive study investigating the relationship between a large number of individual psychomotor, intelligence and personality variables with the accident experience of bus and train drivers, Hakkinen (1976) reported the strongest predictor to be performance on attention-related tasks. Other studies have also found strong relationships between information processing capability and accidents (Kahneman et al. 1973, Avolio et al. 1985, Dewar and Lim 1987, Lim and Dewar 1988).

It is therefore important to determine how auxiliary displays affect drivers' attentional performance and, more specifically, to identify the major contributing factors. Hence, in the present research a major objective was to relate changes in drivers' looking behavior to characteristics of auxiliary displays. If attentional behavior is governed solely by the situational demands of the

driving environment then we would not expect to observe changes in looking behavior. If, however, the results suggest that auxiliary displays can draw attention away from the driving task then it would be important to establish the conditions under which this is likely to occur and whether the effect is large enough to increase the risk of accident involvement.

Two key questions arise, "What resource demands will auxiliary displays place on the driver?," and, more importantly, "How might these demands affect the driver's allocation of visual attention?" At this early stage in their development, it is not possible to foretell the exact nature or function of future intelligent displays. The nature of future tasks subserved by auxiliary displays is still a matter of speculation at present, though a number of generic task elements can be deduced. For example, drivers may be required to (1) monitor displays to check the status of vehicle functions, (2) search pictorial and text displays, (3) read text, (4) perform linguistic manipulations, (5) perform spatial transformations such as rotations and comparisons, and (6) perform mathematical operations. We shall refer to these perceptual–cognitive activities as auxiliary tasks (though, in fact, they may be component elements of more complex tasks) because they are not directly required for the control of vehicle lateral position or velocity (i.e., the primary tasks in driving). Different auxiliary tasks may potentially have differential effects on drivers' attentional behavior, depending on such factors as resource structure and the degree to which they compete with the driving task for common resources. For a given task, other characteristics such as task difficulty may also affect drivers' allocation of attention. Hence, it may ultimately be necessary to explore the drivers' attention response surface over the range of possible auxiliary tasks.

It is equally important to develop a better understanding of the mechanisms underlying the relationship between attention allocation and driving performance. The allocation of attentional resources to the driving task depends upon the driver's perceptions of the momentary demands of the driving task. But, attention and performance are both elements of a closed-loop process. Attention and performance should therefore be investigated together in order to obtain a better understanding of their interdependency.

3.3. Effects of Auxiliary Tasks on Driving

A review of the literature relevant to the problem of time-sharing while driving was performed to identify methodological concerns and practices common to dual task studies of driving as well as to consolidate research findings that can be generalized to auxiliary display applications. The studies reviewed shared a common general research methodology, yet they represented a diversity of approaches to a variety of research problems. Because of lack of consistency in experimental protocols and the choice of dependent and independent variables, it was not possible to consolidate the findings to form a clear understanding of the mechanisms underlying driving performance under dual task conditions.

Moreover, upon further review, it became apparent that more fundamental issues existed than simply incommensurate data. Issues relating to the nature and measurement of cognitive mechanisms and time-shared performance had not been satisfactorily settled, and they persist to the present day.

Theoretical issues such as whether time-sharing occurs as parallel or serial processes bear heavily on the design of experiments and the interpretation of results. The studies examined drew from a diversity of theoretical models, often mutually incompatible, of human information processing with apparently little regard for the limitations of the underlying theory.

Nonetheless, secondary tasks can have adverse effects on at least some aspects of driving when they are performed concurrently with driving (Brown 1965, Brown *et al.* 1967).

Research directed at the effects on driving performance of the use of electronic in-vehicle displays is sparse indeed. In a recent field study, Snyder and Monty (1986) tested four prototype automobile touch-screen displays. They reported that lane keeping, speed control and braking frequency were significantly affected by concurrent performance of CRT-related tasks (which required visual feedback). The experimenters also observed significant age and sex effects. However, methodological constraints did not permit generalizing the findings beyond the specific conditions of the experiment. In particular, display/vehicle confounding and lack of systematic manipulation of experimental variables did not permit detailed analyses of the effects of specific task parameters or even separation of the effects of perceptual and response loads.

A study directly related to the use of auxiliary displays in automobiles investigated the effect of reading text while driving (Zwahlen and DeBald 1986). Zwahlen and DeBald found that when reading text from an in-car display, the standard deviation of lane position increased monotonically as a function of time and distance. The results were interpreted as indicating that the use of sophisticated displays and/or touch panels produce significant degradation in lane keeping performance. However, Zwahlen and DeBald's study did not isolate the specific effects of "reading" which may be quite different from other uses of a display. Also, subjects in the text reading condition were not permitted to look directly at the road. Hence, the increased variability in lane position was not unexpected. A more germane question is whether drivers would voluntarily have attended to the auxiliary task for sufficient periods to cause such extreme variability in their driving.

In a more recent study, Zwahlen *et al.* investigated lane position variability while inputting an 11-digit phone number on a cellular telephone (Zwahlen *et al.* 1988). In this experiment, two telephone mounting positions were used as well as two driving conditions. In half of the trials drivers were not permitted to look at the roadway (runway) during the dialing phase while in the other half of the trials they were permitted to look at the roadway freely. It was suggested that drivers performed the primary task much better when they were free to look at the road, but that their performance was still impaired by the telephoning task. Unfortunately, no control conditions were used to estimate the increase in lane position variability attributable to cellular telephone use per se (as opposed to performing some other task or no task at all).

A series of studies, performed at Virginia Polytechnical Institute and State University for General Motors, evaluated some human factors aspects of navigation tasks presented on a moving-map navigation display, the ETAK Navigator (Antin *et al.* 1986a, b, Wierwille *et al.* 1988). These studies demonstrated that the navigation task did impose additional demands on the driver

with a concomitant decrease in visual attention to the roadway. The attentional requirements of some of the navigation tasks were found higher than the majority of tasks associated with conventional instrument use. It has also been shown that drivers can adapt to the increasing demands of the primary task by reducing attention to the navigation task (Wierwille *et al.* 1988). However, drivers generally displayed adaptive behavior appropriate to the driving demand — but this was not always the case. Some possible explanations for the unanticipated high level of variance, includes individual differences in perceptual style, inability to adequately measure spare capacity, inaccuracies in manipulating independent variables, etc. However, the possibility certainly exists that drivers were not always able to judge the appropriate level of attention required by the primary task or were unwilling to allocate the requisite amount of attention.

Some in-vehicle tasks do compete with driving under some circumstances. However, more systematic research is required to produce guidelines for the ergonomic design of auxiliary automobile displays.

REFERENCES

HENDRICK, H.W., 1994, Human factors in ODAM and the technology of macro-ergonomics. In *Proceedings of 12th Trienniel Congress of the International Ergonomics Association*, V(6), Toronto.

KANTOWITZ, B.H., 1997, In-vehicle informations systems: premises, promises, and pitfalls. Paper presented at Transportation Research Board Conference on Intelligent Transport Systems, Highway Safety and Human Factors, Washington, DC.

MICHON, J.A., (ed.), 1993, *Generic Intelligent Driver Support* (London: Taylor & Francis).

NOY, Y.I., 1997, Human factors in modern traffic systems. Keynote Address at 13th Triennial Congress of the International Ergonomics Association, Tampere, October 1997; *Ergonomics*, 40.

THIERRY, B., TATTEGRAIN-VESTE, H. and PAUZIÉ, A., 1996, Ergonomics evaluation of IVIS: advantages in developing a driver's model using A.I. techniques. In *Proceedings of ITS World Congress*, Orlando.

TREAT, J.R., TUMBAS, N.S., MCDONALD, S.T., SHINAR, D., HUME, R.D., MAYER, R.E., STANSIFER, R.L. and CASTELLAN, N.J., 1979, *Tri-Level Study of the Causes of Traffic Accidents: Final Report*, vols I and II. HS-805-086 (NTIS PB 80-121064) (Washington, DC: US Department of Transport).

Obstacles to Recovery from Work-related Musculoskeletal Disorders

A. K. Burton† and C. J. Main‡

†Spinal research Unit, University of Huddersfield, Huddersfield, UK
‡Department of Behavioural Medicine, Hope Hospital, Salford, UK

1. INTRODUCTION

The concern of ergonomics is optimization of the task suit the person, thereby avoiding excessive fatigue, discomfort or stress. The aim is to reduce the risk of injury or ill health and improve efficiency. This aim, while laudable, has not been fully realized for musculoskeletal disorders (MSD). Taking the example of low back trouble, the experience in most modern societies is that, despite a general improvement in working conditions and a reduction in spinal loading, the prevalence of back-related disability has been increasing exponentially over the past few decades. A similar pattern is reported for upper limb disorders. That is not to say that ergonomics has failed, rather that it may not be addressing the real determinants of disability associated with MSD; it is the disability that is the costly concern for society.

The principles of ergonomics seem intuitively "right" and they have a sound scientific basis. The combined input from multiple disciplines (from biomechanics to psychology) has enabled work to become more comfortable and, arguably, safer. Yet few studies that conclusively demonstrate a positive benefit on the incidence and prevalence of MSD or their Attendant disability are conspicuous by their almost total absence. The reasons for this lie in the nature of the disorders, the nature of pain (and responses to it), the organization of work and the structure of society. In the face of these powerful influences (including a cultural belief that work is likely to be damaging) ergonomic interventions alone have little prospect of having a substantial and measurable impact.

MSD, whether work-related or not, are primarily characterized by the experience of pain. The International Association for the Study of Pain (1995) defines pain as "An unpleasant sensory and emotional experience associated with actual or potential tissue damage, or described in terms of such damage." From this definition it becomes apparent that the symptom of pain should not be considered simply in terms of physical insult to some anatomical structure, but the responses to such insult are an important part of the pain experience. Doubtless physical insults (whether at work or elsewhere) will result in pain, but what distinguishes a substantial proportion of (purportedly) work-related pain from "everyday" pain is that the former has a greater propensity to progress to chronic disability. One way to view the phenomenon of work-related MSD is not that they occur, but that not infrequently they fail recover. Recent research efforts have concentrated on the reasons for this non-recovery.

2. EPIDEMIOLOGICAL ASPECTS OF MUSCULOSKELETAL DISORDERS

The most common forms of illness reported by workers are MSD, in particular those affecting the low back and the upper limbs (Health & Safety Executive 1998); these symptoms are often attributed to physical aspects of the work. However, it is an inescapable fact that the same symptoms (or disorders) are extremely common in non-working populations, and consequently it becomes difficult epidemiologically to discern the true level of association with work. It is even more difficult to identify a work-related cause. That is not to say that disadvantageous aspects of work cannot influence the development or persistence of MSD; they can, but the concept of a dichotomy between workers and non-workers (while apparently reasonable) presents an impediment to our understanding of how these disorders can be controlled.

Quite recently it has become apparent that back pain is generally not the benign, short-lived phenomenon previously supposed. Follow-up studies have shown that around three-quarters of back pain patients in primary care settings continue to experience varying levels of pain over 1–5 years, but they tend not to seek care beyond the first few months. In fact, only about one-half of people experiencing a spell of back pain consult for medical care, but when they do it is the duration of pain and associated psychosocial factors that influence the consultation rates. In non-industrialized societies, back pain rarely stimulates care seeking and does not result in disability, yet the prevalence of the symptom is similar to that in Western societies. In addition, it has been shown that the experience of back pain in adolescents is at a similar magnitude to that seen in adults, but care-seeking and disability are rare. So far as is known, other MSD have a similar epidemiology.

In 1997, a publication from the National Institute for Occupational Safety and Health concluded that there was strong evidence for increased risk of work-related MSD for some body parts when exposures are of prolonged intensity and when workers are exposed to several risk factors simultaneously (National Institute for Occupational Safety and Health 1997). Seemingly, the epidemiology tends to support the traditional biomedical model of MSD, which considers the disorders to be an "injury," yet quantification of injurious exposures remains largely elusive. The majority of the evidence relates to reports of symptoms as opposed to objective evidence of tissue damage. There is little evidence that work to results in irreparable tissue damage, but occupational overload damage to the spine recently has been studied from radiographic evidence (Brinckmann *et al.* 1998). It was found that certain types of strenuous work were indeed associated with overload damage to lumbar vertebrae and discs, but work of that type is rarely encountered in modern working environments, and work complying with current ergonomic guidelines was not found to be associated with overload damage. It is thus unlikely that the symptoms of MSD (as currently observed) are a reflection of irreparable damage, at least so far as back trouble is concerned. Furthermore, a notable feature of the majority of work-related MSD is that persistence of symptoms and disability occurs in the absence of demonstrable, objective impairment.

The relationship between work and disability is less clear cut. There is some evidence that the more physically strenuous jobs are associated with greater levels of work-loss, but this is an inconsistent finding. There are numerous reports of relatively low disability rates in some groups of workers performing heavy jobs, and high disability rates in other groups not exposed to

injurious forces. A direct linear relationship between physical exposures and disability has not been shown; disability and prolonged work-loss are now seen largely as a function of various psychosocial factors that may or may not be associated with the work.

On balance, while MSD are common among workers, it cannot be concluded that they are necessarily work-related. Physical aspects of work doubtless initiate some instances, but there seems no reason to expect permanent pathological changes in all but a very few cases. The question to be addressed is why MSD in workers may fail to recover.

3. OBSTACLES TO RECOVERY

The factors consistently related to persistence of symptoms and work-loss lie almost exclusively in the psychosocial domain. Clinically, these factors have been termed "yellow flags" (Kendall *et al.* 1997) and are useful predictors of chronicity (non-recovery) both in workers and non-workers. They include psychological parameters such as distress, pain-coping strategies, fear-avoidance behaviors and beliefs about MSD in general, as well as cultural influences and life events.

Negative beliefs and attitudes related to the working environment can be extremely influential, and such reactions may become enhanced following injury. It is proposed that these factors be distinguished from the clinical factors by terming them "blue flags." They include psychosocial aspects of work (e.g. job satisfaction, social support, mental stress) and beliefs about the relationship between work and symptoms/injury, as well as attribution of blame. By way of example, the perception that a working environment is unsafe can be more important than the actual extent of the hazard. This is because, following injury, inappropriate beliefs about hurting/harming can become established due to mistaken beliefs about the nature of pain and about vulnerability to further injury.

There is a third class of social policy issues that can influence recovery, such as unemployment and entitlement to sick pay or workers' compensation, but these factors are not readily amenable to local intervention and are beyond the present scope.

In the face of a high prevalence of MSD in the general population, the relatively higher level of related disability in workers and the tenuous relationship with purely physical aspects of work attempts at primary prevention are likely to be a suboptimal control strategy. A more promising approach to management of MSD is the removal of obstacles to recovery. With this approach it is accepted that workers will develop symptoms (which may or may not be related to work tasks) but it is also recognized that chronic pain and disability are not an inevitable consequence.

4. INTERVENTION STRATEGIES

The biomedical model of work-related injury predicts that ergonomic intervention aimed at reducing the physical demands of work-related MSD will be effective in controlling both their incidence and related disability. The available evidence does not universally support this hypothesis. The majority of reported successful interventions have been those that concentrate on psychosocial issues rather than physical factors. For instance, it has been found that simply broadcasting educational material, devised to address fear-avoidance behaviors, has the capacity to reduce absenteeism from low back pain.

The issue of when to return workers absent with MSD to their job has attracted considerable attention, and there is some evidence that restricted duties can be beneficial in terms of sustained work return and reduction of recurrence. However, other interventions involving organizational parameters (such as early reporting and prompt active treatment) are often incorporated into these strategies, so it is not possible to conclude that simply reducing the exposure to physical stressors is effective. Indeed, at least one study has found that return to restricted duties has either no effect on subsequent absenteeism, or may actually increase it.

The International Association for the Study of Pain (1995) has addressed these issues and published a report concluding that non-specific low back pain should be considered in terms of activity intolerance, and should be managed on a time-contingent rather than a pain-contingent basis. This somewhat uncompromising approach, while based on the available evidence, has not gained universal acceptance. An alternative strategy (which embodies many of the concerns of the IASP) is to offer a workplace-based early assessment of yellow flags and blue flags, followed by a focused active management protocol delivered through the occupational health department (or suitably educated family doctors).

An early active approach to management of MSD is generally accepted as the approach of choice, but the efficacy of management will be substantially compromised if obstacles to recovery are not removed (or reduced). It is necessary to consider the individual's response to injury and their beliefs and attitudes in relation to the working environment.

The clinical management of MSD has shifted from prescriptions of rest and avoidance (derived from the biomedical model of injury) to an active management process, stimulated by the alternative biopsychosocial model (Waddell 1998), centering on psychosocial parameters with advice to maintain activity and to aim for an early return to (normal) work. It has become apparent that taking these concepts into the occupational setting can reduce the risk of disability from MSD. Successful programs will be those that take appropriate steps to remove obstacles to recovery after assessment both of yellow and blue flags. The precise structure of management programs for MSD is beyond the present scope, but general recommendations can be made.

5. RECOMMENDATIONS

- Recognize that the workers' perception of their work is fundamental to understanding recovery from injury.
- Understand the inherent risks in certain occupational environments for producing extended sickness absence and delayed recovery after injury.
- Facilitate a system for early reporting of musculoskeletal symptoms to appropriately trained personnel.
- Identify mistaken beliefs about the nature of pain, hurting/harming and unnecessary fears of prolonged injury, and recognize other psychological characteristics of the individual (yellow flags).
- Identify aspects of work perceived to be problematic by the injured worker (blue flags).
- Distinguish yellow flags and blue flags from unwillingness to work.

- Provide a psychosocial preventative approach to the individual, promoting self-help, establishing confidence and reducing unnecessary apprehension.
- Facilitate return to work as soon as possible; restricted duties (if required) should be brief and time-limited. Maintain management of the worker within the occupational environment so far as is possible.

REFERENCES

BRINCKMANN, P., FROBIN, W., BIGGEMANN, M., TILLOTSON, M. and BURTON, K., 1998, Quantification of overload injuries to thoracolumbar vertebrae and discs in persons exposed to heavy physical exertions or vibration at the workplace. Pat II. Occurrence and magnitude of overload injuries in exposed cohorts. *Clinical Biomechanics*, 13, S(2), 1–36.

HEALTH & SAFETY EXECUTIVE, 1998, *Self-reported Work-related Illness in 1995: Results from a Household Survey* (Norwich: HMSO).

INTERNATIONAL ASSOCIATION FOR THE STUDY OF PAIN, 1995, *Back Pain in the Workplace* (Seattle: IASP Press).

KENDALL, N.A.S., LINTON, S.J. and MAIN, C.J., 1997, *Guide to Assessing Psychological Yellow Flags in Acute Low Back Pain: Risk Factors for Long-term Disability and Workloss* (Wellington: Accident Rehabilitation & Compensation Insurance Corporation of New Zealand and the National Health Committee).

NATIONAL INSTITUTE FOR OCCUPATIONAL SAFETY AND HEALTH, 1997, *Musculoskeletal Disorders and Workplace Factors* (Cincinnati: DHHS (NIOSH)).

WADDELL, 1998, *The Back Pain Revolution* (Edinburgh: Churchill Livingstone).

Occupational Biomechanics

K. Kedzior and D. Roman-Liu

Central Institute for Labour Protection, ul. Czerniakowska 16, 00-701 Warsaw, Poland

1. INTRODUCTION

Biomechanics is the study of the causes and effects of external and internal forces acting upon a biological system, and in particular upon a human being. It is an interdisciplinary field of sciences that draws upon the exact sciences (mechanics), technical sciences (theory of machines and automatic control) and biological sciences (biology, medicine). Many problems of contemporary biomechanics link it to biocybernetics, biomedical engineering and ergonomics.

Among the many applications of biomechanics, that of *occupational biomechanics* has been growing rapidly in recent years. This has been due to an increasing awareness among both employers and employees of the threats to the health and life of people performing even light work, if done in an inappropriate way and under unsuitable conditions. Such threats most often arise during the design of work processes and work places, as a results of failing to take into account the results of analysis of loads upon the body during work and to compare these to the capacity of the worker, based on the biomechanical characteristics of the musculoskeletal system.

The object of occupational biomechanics is, therefore, the investigation of causes and effects of load on the human musculoskeletal system resulting during physical work. Along with occupational physiology and psychology, occupational biomechanics is a foundation of contemporary ergonomics. Typical problems discussed in specialized monographs on occupational biomechanics (Chaffin and Andersson 1991) include: the structure and function of the musculoskeletal system treated as an apparatus of work (with selected information from anthropometry), surveys of mechanical parameters (forces and moments of forces acting on the human body, displacements, velocities and accelerations of body segments during work) and electrical parameters (electromyography), mathematical modeling and computer simulation of work processes, methods of evaluating the capabilities of the human organism during the performance of typical work (e.g. lifting loads, working as operator of mechanical equipment or computer), recommended positions for the performance of various kinds of work (standing or sitting).

In the following sections two problems most useful in analyzing the safety of work processes are discussed: the properties of skeletal muscles and computerized methods.

2. BIOMECHANICAL CHARACTERISTICS OF SKELETAL MUSCLES

Almost half the mass of the human body consists of muscles, of which there are ~640 (Fidelus *et al.* 1983). There are three types of muscles: skeletal or striated, smooth, and cardiac. The system of human movement is driven by the skeletal muscles, of which there are > 440. All our external activities are carried out with their help. On the microscope level, skeletal muscles are composed of hundreds or thousands of parallel cells — muscle fibers enwrapped in a common sheath.

At either end, muscles pass into tendons, which connect, to the bones of the skeleton. It is estimated that human skeletal muscles contain in total ~250 million muscle fibers, among which we distinguish fast twitch (FT) and slow twitch (ST) fibers. The prevailing opinion is that both the number of fibers and their type do not change during the course of an individual's life, either spontaneously or under the influence of the type of work performed or sports training. The muscles are controlled by more than 420 thousand nerve cells (motor cells) in the spinal cord (Fidelus *et al.* 1983), and each of these cells controls the activity of a group of muscle fibers. Each such group, called a motor unit, contains from a few fibers (four-to-six in the muscles moving the eyeball) to several hundred (640 in the large muscles of the legs).

A skeletal muscle is a biochemical engine transforming chemical energy into mechanical energy, operating at more or less constant temperature. The mechanical efficiency of this splendid machine when performing under suitable conditions is ~25%. A muscle can attain the efficiency of an internal combustion engine, e.g. a muscle of mass 1 kg can, in the course of 8 h of work, generate a power of 15 W. Most often, muscles work in disadvantageous mechanical conditions given differences of the bone levers at a relation of ~10:1, and lose strength in producing large angular velocities in the joints. Despite this, a person can hold an object of 900 N weight in two hands at elbow level for a few seconds. The force exerted by the muscles must be, in this position, 10 times greater, or up to 9000 N. The lifting mass is almost 2000 times greater than the mass of the muscles involved. The force *F* exerted by a given muscle isometrically (such as when the distance between its insertions into bone does not change) can be described as:

$$F = S\sigma, \quad (1)$$

where *S* is physiological cross-sectional area (usually calculated by dividing the volume of a given muscle by its length); and σ is tension, or the value of the force acting upon a cross-sectional area unit. The maximum tension of a skeletal muscle, according to various estimates, is 0.5–1.5 MPa (5–15 kG/cm^2). The physiological cross-section of muscles increases primarily through an increased volume of plasma in the muscle cells (fibers), given their fixed number. And this is the result of a greater supply of energetic substances to the plasma, mainly glycogen, which occurs when the muscle is performing close to maximum intensity, for, e.g., during sports training.

Intesive work is also carried out by physical workers, but it is not of maximum or near-maximum intensity, as such work could not be done for several hours every day. Consequently, increase of the physiological cross-section of the muscles of physical workers, and, therefore, their strength, occurs only to a small degree. Incidentally, the largest human muscle, the gluteus maximus, can produce a force of ~12 000 N. A person can simultaneously activate one-seventh of their muscles; if all human muscles could share a single point of insertion, a force ~250 000 N would be exerted.

A basic function of the skeletal muscles serving the human movement apparatus is to produce driving moments in the joints.

Given the microscopic structural features and the way of working of muscle fibers, the force exerted by the skeletal muscle depends on:

- the length of the muscle;
- the rate of concentric contraction, i.e. when the contracting muscle exerts a force greater than that of the external load;
- the rate of eccentric contraction, i.e. when the muscle opposes an external force which is stretching it and which is greater than its contracting force (Chaffin and Andersson 1991, Fidelus *et al.* 1983, Fox *et al.* 1993, Nigg and Herzog 1994);
- activation, characterized by impulses sent to the muscle motor units along nerve pathways. In response to a single impulse, a motor unit will react with a single contraction (twitch), and as the frequency of activation increases the individual twitches add up, leading to a smooth, tetanic contraction of constant force produced by the muscle; and
- the number of motor units activated at the same time as well as the frequency and synchronization of their activation. One measure of muscle activation can be the value of the functional potential accompanying muscle work, recorded as an electromyogram with the aid of either an inserted or a surface (placed on the skin above the muscle) electrodes

In an organism, muscles work in groups. The moment M produced by a muscle gropu about the axis of rotation of a given joint (considering here one degree of freedom of movement in the joint, with the neighboring joints demobilized) in the course of a smooth, fixed tetanic contraction of the muscle and with constant activation u = constant, is a function of the angle of rotation in the joint a and the rate of angular movement ω. A general notation of this relation takes the form:

$$M = M(\alpha, \omega) \quad (2)$$

Figure 1 shows an example of the three-dimensional surface, the *general biomechanical characteristic* of a muscle group at maximum activation $u = u_{max}$. It is a graphic illustration of equation (2). For different values of constant activation $0 < u < u_{max}$, we obtain similar surfaces lying within the volume contained by the surface for $u = u_{max}$.

We do not know of a reasonably simple and quick method
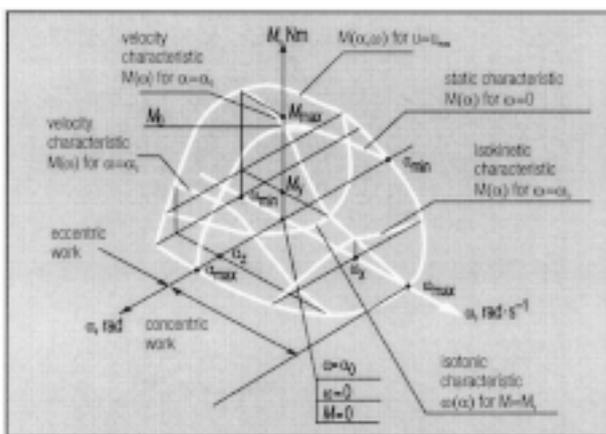
Figure 1. Sample three-dimensional biomechanical characteristic of a muscle group, i.e. the dependence of moment M produced by muscles about the axis of rotation in a joint α and on the angular velocity of motion ω, for a constant maximum activation u. The beginning of the coordinate system is located at the point $M = 0$, $\alpha = \alpha_0$, $\omega = 0$.

of measuring a general biomechanical characteristic for a muscle group. We do know, however, methods of measuring two basic types of partial dynamic characteristic: isokinetic and isotonic.

The first of these — *isokinetic characteristic* — is the dependence of the moment M on the angle α for a given angular velocity ω constant and activation u = constant. Figure 1 shows an example of this relation for a chosen $\omega = \omega_x$. It is measured with the aid of a special device known as an isokinetic dynamometer. The subject drives the lever of the dynamometer, which is constructed in such a way that rotary motion of the lever is possible only with constant angular speed. In the course of movement, the moment M as a function of the angular rotation α is measured and recorded. After a series of measurements for various chosen values of ω = constant, a set of two-dimensional characteristics $M(\alpha)$ is obtained, which makes possible the calculation of the three-dimensional surface $M(\omega, \alpha)$.

Among the isokinetic characteristics, one which deserves separate treatment is that for $\omega = 0$. This is the *static characteristic*, and its determination is carried out by a different method (isometrically), because it is not possible to turn the lever of the dynamometer with an angular speed of zero. For several or several dozen blocked angular positions of the lever, the value of the moment M is measured.

The second dynamic characteristic — *isotonic characteristic* — is the dependence of the velocity ω on the angle of rotation α for a given value M = constant and activation u = constant. Figure 1 shows an example of such dependence for a selected $M = M_y$. This is measured with the aid of a special device known as an isotonic dynamometer. The subject drives the lever of the dynamometer, which is constructed in such a way that the moment of lever resistance is constant throughout the course of rotation and is equal to the value previously set. During movement the value of the angular velocity ω as a function of the angle of rotation α is measured and recorded. After a series of measurements for various chosen values of M = constant a set of two-dimensional characteristics $\omega = \omega(\alpha)$ are obtained which make possible the calculation of the three-dimensional surface $M(\alpha, \omega)$. However it should be added that the isotonic method of measurement contains a certain fault. In the course of measurements, movement occurs with a variable angular velocity (accelerations appear), leading to additional moment of inertial force in the joint under study, stemming equally from the moving parts of the dynamometer and from the movement of the limb. The moment of muscular forces, having to overcome this load as well, is, therefore, not exactly equal to the cosntant moment of resistance of the dynamometer. This error is, however, usually ignored.

Among the partial dynamic characteristics we also count the dependence of the moment M on the angular velocity ω for a given value of the angle α = constant and activation u = constant. Figure 1 shows example of these relations for two chosen $\alpha = \alpha_z$ and α_o. The course of the characteristic for a given muscle group depends on the relative percentages of fast and slow muscle fibers the group contains. The greater the proportion of fast fibers, the higher the values of moment generated by muscle forces in the joint are exerted throughout the entire range of contraction. At the same time, muscles having a higher proportion of fast fibers get tired sooner (Fox *et al.* 1993). It is not possible to measure velocity characteristics for a given muscle group directly, for

movement cannot be performed 'isometrically,' maintaining a constant value of angle in the joint. As in a general characteristic, velocity characteristics can be calculated only indorectly, such as on the basis of the results of isokinetic or isotonic measurements. The most well-known characteristic shown in Figure 1 is the velocity characteristic for the angle $\alpha = \alpha_o$, for which the moment exerted by the muscle group under study in static conditions ($\omega = 0$) is maximal and equal to M_o.

The majority of measurements used in investigating the above-mentioned dynamic characteristics are carried out in conditions of maximum activation. This corresponds to the situation in which the subject is assigned the task of exerting the maximum possible moment M in the given conditions. Between successive measurements, suitable rest breaks are taken.

The extreme values of the angle α_{min} and α_{max} indicated in Figure 1, for which the moment M under static conditions is equal to zero, are of theoretical meaning, since they lie outside the range of the actual angle of rotation in the joint, limited as this is by the joint's anatomical structure.

Measurements of the biomechanical characteristics of main muscle groups are carried out using special test stands. Figure 2 shows such a stand at the Institute of Aeronautics and Applied Mechanics at the Warsaw University of Technology (Mianowski 1993). A basic subassembly of the stand is a computer-controlled hydraulic brake. Movement of the lever driven by the subject is braked, enabling isometric, isokinetic or isotonic conditions to be assigned for measurement. Measurement results are elaborated by computer and presented in graphic form, such as that similar to the diagram in Figure 1.

The stand in Figure 2 and other similar stands are most often used in ergonomics and biomechanics of sports, for, e.g., research into the results of training, and also in medicine, for e.g., research into the progress of rehabilitation of organs of movement. In ergonomics, the results of measurements carried out on subjects representative of a given population are used to describe the strength capacity of physical workers, and this information can be applied in the design of work places using among other means, computer simulation programs.

3. COMPUTERIZED METHODS IN OCCUPATIONAL BIOMECHANICS

Computer simulation is currently a functional method of research in many disciplines, including occupational biomechanics (Chaffin and Andersson 1994). Its rapid development and considerable dispersion into all areas of biomechanics has been mainly due to two factors: the appearance of inexpensive, reliable and user-friendly computer equipment, and even more importantly, the universal availability of user-friendly software, including that for computer graphics. In the first decades of the development of informatics, using a computer required specialized skills such as familiarity with programming languages. This barrier is now disappearing as professional producers provide more and more ready-made software.

Computer simulation in biomechanics consists in: computer-assisted formulation of mathematical models of biomechanical systems on the basis of previously formulated physical models, the solution of systems of equations creating a given mathematical model, and the presentation of results in a form legible to the user (Dietrich *et al*. 1997). A typical procedural scheme is shown
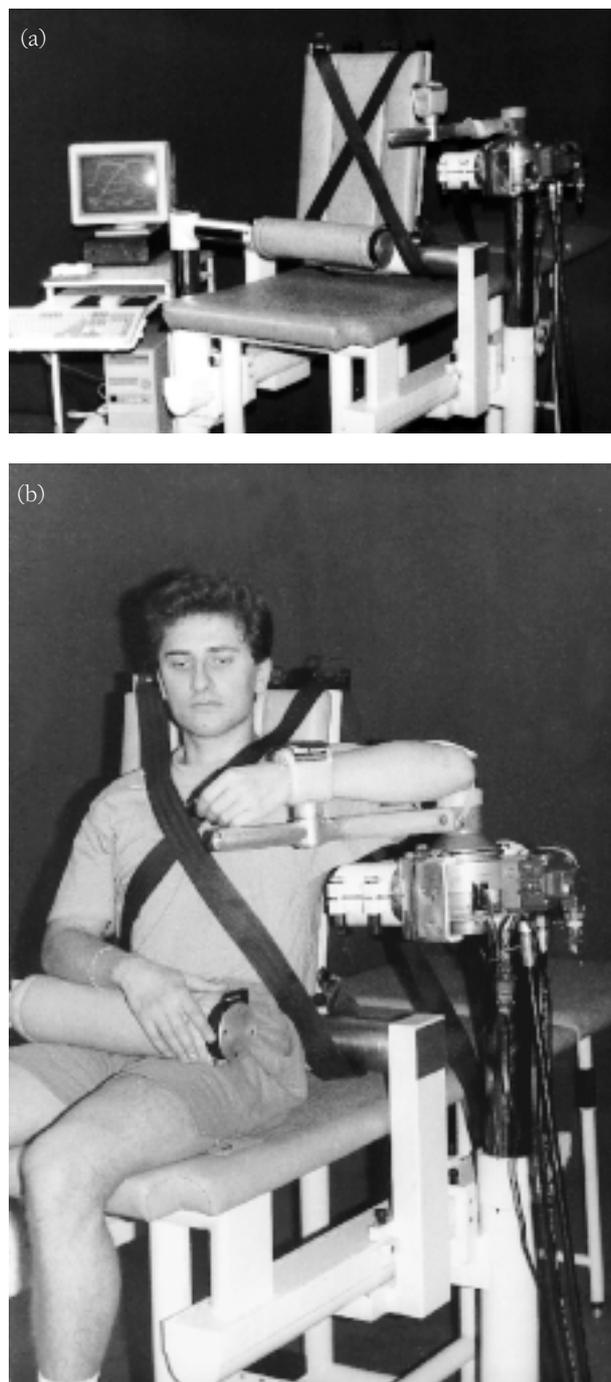
Figure 2. Test stand for measuring biomechanical characteristics of the main human muscle groups: (a) general view and (b) measuring of elbow flexors and extensors groups (Mianowski 1993).

as Figure 3. Formulation of the physical model (idealization) is a problem that the researcher must carry out independently, using his knowledge, experience and intuition. Software available on the market can facilitate the transformation of the physical model to a mathematical model (formalization) and of the mathematical model to a simulating one (programming), as well as the carrying out the computer simulation. However, idealization, prognosis and validation remain to a considerable degree the duty of the researcher.

In contemporary modeling in biomechanics (including
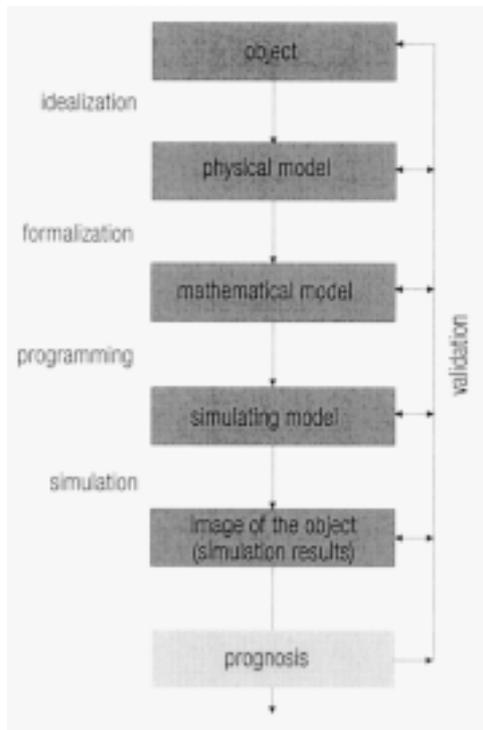
Figure 3. Process of mathematical modeling and computer simulation (Dietrich *et al.* 1997).

occupational biomechanics) two types of physical models are used: those of concentrated or distributed parameters (e.g. mass, rigidity, material data). Mixed models can also appear.

The first type of physical model is predominant in the biomechanics of motion of the human body. Such models are used in modeling the entire body or its parts (e.g. only the upper limbs) for a full range of angles of rotation in the main joints as well as for various static (Nigg and Herzog 1994, Seireg and Arvikar 1989) and dynamic situations (Borowski *et al.* 1991). With the aid of these models it is possible toe valuate: forces acting upon the human body as a function of time, reactive forces in the joints (including intervertebral joints), and forces exerted by muscles or muscle groups. Their mathematical description (model) leads to sets of differential or differential-algebraic equations. The number of equations is usually high, reflecting the fact that physical models of the human body normally consist of many elements (hence the name multibody systems) and a correspondingly high number of degrees of freedom of movement (rotational movements in the joints connecting segments). The classical procedureof derivation of equations of motion (i.e. 'by hand' on the piece of paper) is in such cases practically impossible considering the very high likelihood of error. Nowadays, one can use special software packages available on the market which operate symbolically (imitating manual operations and deriving equations in symbolic form) or numerically (calculating only actual quantitative values of the terms of equations). Both types have advantages and disadvantages, which influence the scope of their application in biomechanics.

On the software market there is today a wide selection of this kind of computer programs intended foir the analysis of mechanical and biomechanical systems. Such packages also contain solving of equation procedures and software for the graphic presentation of simulation results (including animation).

Their cost is as a rule very high (thousands of dollars), and so they are more often leased than purchased.

In many biomechanical problems it is necessary to establish and solve a mathematical model for the physical model, containing a so-called loop formed by closed kinematic chains, for, for example, the legs and ground form a closed loop during bipedal walking, a machine operator model contains a closed loop comprising frame (seat)-runk-arm-control lever-frame. A mathematical description of problems with loops can contain both differential and algebraic equations. Not all packages contain procedures for solving such a mixed system of equations.

A good example of software which exploits a multibody physical model and serves in analyzing loads on the human body as well as the resulting threats to health is the package 3D SSP: Three Dimensional Static Strength Prediction (Chaffin and Erig 1991, 3D STatic Strength Prediction 1993), elaborated at the University of Michigan Center for Ergonomics. This program is being used in many centers around the world for analyzing existing or projected work places under static conditions (slow movements in which the influence of inertial forces can be neglected). The physical model of the human body used in this package is shown in Figure 4, in the form of twelve rigid elements (body parts: upper torso with head, lower torso, upper arms, lower arms with hands, thighs, shins, feet) connected with joints.

Figure 5 shows a scheme of the package. 3D SSP allows us to carry out a three-dimensional static analysis of the strains on a person who finds himself in a given position during physical work and is subjected to a defined external loads. The results of the analysis are given in the form of a printout of the values of the moments of muscular forces and reactive forces in all the joints of the model, and of the values of the force compressing the intervertebral disc L5–S1. The appearance of any values greater than the maximum moment in the joints which human muscles can exert, or of values which could indicate possible damage to the body, are noted. Also given is information on which part of the population of physical workers under consideration could safely perform the analyzed job.

A second type of physical models used in biomechanics is the method of three-dimensional modeling of the body or of body fragments (e.g. torso, segment of the spine, etc.) with the aid of a
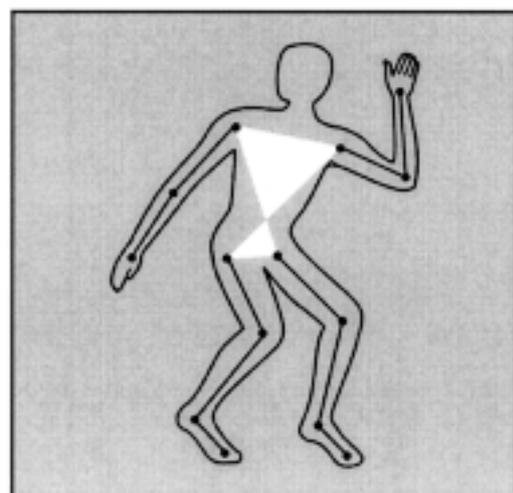
Figure 4. Physical model (multibody type) of human body (Chaffin and Erig 1991).
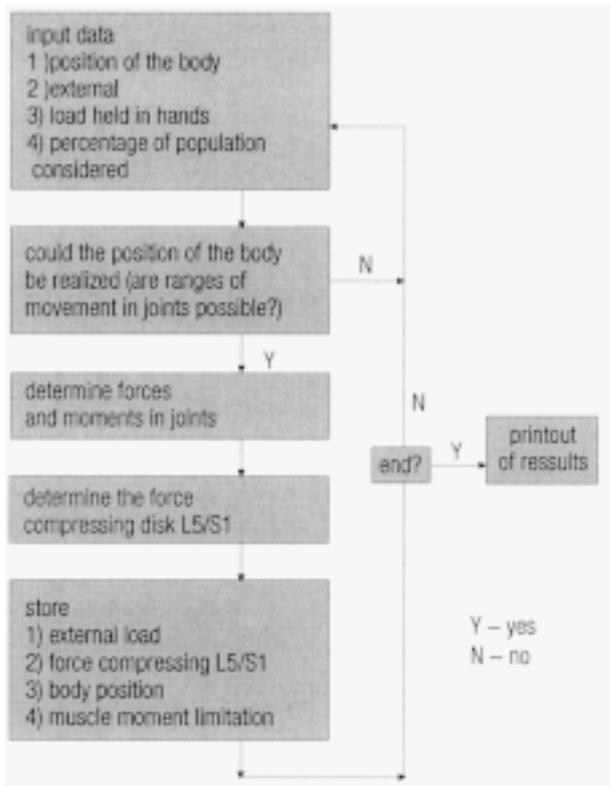
Figure 5. Flow diagram of the 3D SSP package (Chaffin and Erig 1991, 3D Static Strength Prediction™ 1993).

large number of small elements. These are the so-called finite elements, to which may be attributed, dependence on the needs, defined geometrical, material or other properties (Zienkiewicz and Taylor 1991). A mathematical description of these models takes the form of systems of partial differential equations, which require in the computerized version the application of equivalent numerical methods: FEM (finite element method) or BEM (boundary element method). Such models are used in, for, e.g., evaluation of the distribution of stresses and strains in the musculoskeletal system (important in occupational biomechanics) and in the lungs during breathing, and in describing the flow of body liquids (blood, lymph).

The finite elements method has become increasingly popular in recent years for modeling the musculoskeletal system (Dietrich *et al.* 1991) because it has the following advantages:

- Ease of modeling the spatial structure of either the entire musculoskeletal system or its elements (bones, vertebrae, ribs, intervertebral disks, muscles, tendons).
- Possibility of investigating the statics, dynamic and stability of the musculoskeletal system in various life situations (norm, pathology, work, sport).
- Possibility of employing vairous optimization criteria in the model, thus modeling the rules followed by the central nervous system in controlling the work of the muscles (important in solving so-called problems of muscle cooperation).
- Ease of introducing changes to the geometric and material parameters of the model.

The finite elements method is widely used in engineering in various technical disciplines. However, one must take into account, when applying it to the modeling of the musculoskeletal

system, various specific properties of that system not met in typical engineering constructions, such as:

- Mechanical properties of bones, ligaments and muscles are anisotropic, and the main directions of stiffness (e.g. the direction of muscle fibers) change their orientation in space.
- Parameters of bone stiffness, e.g. the Young module (10 000–15 000 MPa) are considerable greater than the corresponding parameters of disks, muscles and ligaments (0.4–160 MPa), with the result that the final set of equations describing the bone–muscle system is ill-conditioned (both very high and very low values appear in the equations, hindering numerical solution).
- Tension in muscle is produced not only by changes in the distance between points of insertion (the so-called passive tension component), but mainly by contraction caused by activation generated in the nervous system (the active component).
- Nuclei pulposa of intervertebral disks and in the region of the abdominal cavity are areas which are almost incompressible.
- Joints (of limbs, spine and rib cage) significantly limit the movements of the related long bones, vertebrae and ribs.
- Pressures in the nuclei pulposa of the intervertebral disks, in the abdominal cavity or in the rib cage are non-preservative in character (the directions of forces are depending on deformations).

Typical FEM software packages available on the market do not enable one to take these properties into account and it is, therefore, necessary to be very cautious in applying them to biomechanical problems. Many researchers still create their own packages for the modeling and simulation of complex biomechanical systems, that is, those constructed of tissues having different properties.

An example of such a model is shown in Figure 6. It is a model of the trunk constructed from 2640 finite elements and which represents the anatomy of the human spinal system (vertebrae, ribs, pelvis, sternum, intervertebral disks, ligaments, muscles, gas in the abdominal cavity). The model also takes into account the activity of the central nervous system, accepting that
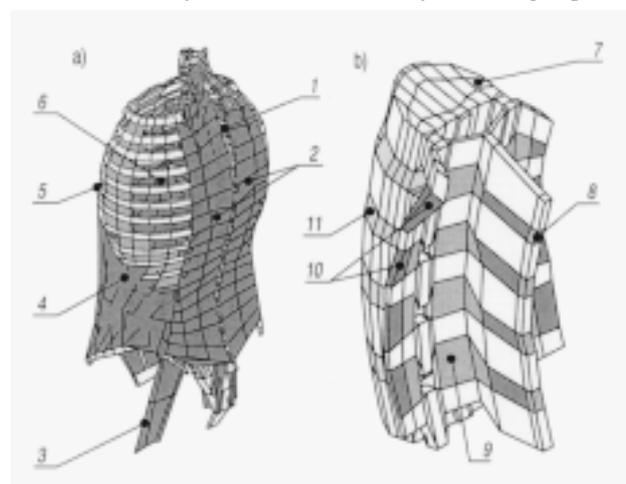
Figure 6. FEM type model of a human torso: 1, spine; 2, dorsal extensor muscle; 3, lumbar muscle; 4, abdominal muscle; 5, rib cage; 6, diaphragm; 7, vertebra L1; 8, ligament; 9, yellow ligament; 10, rotator muscles; 11, intervertebral disk 1.2/1.3 (Dietrich *et al.* 1993).

it controls muscles in such a way as to minimize the energy expended on maintaining the body in a given position under an external load. The input data for the model are: the geometry of the system (dimensions), position of the body as well as the value of the external load force; and the output data: displacements of the bone elements, deformations and stresses in the muscles, tendons, and intervertebral disks (Zagrajek 1990, Dietrich *et al.* 1991, 1993). With the aid of the model presented it is possible to investigate the activity of the spinal system in conditions of static load. Sample results obtained are presented in Figure 7.

An enlarged version of the model, completed with highly simplified models of the head, arms and legs (not representing the anatomical structures of these elements) serves in simulating the behavior of the spinal system when subjected to large shock loads and crashes (Borowski 1991) as well as under conditions of vibrations acting upon the operator of a mechanical vehicle (Dietrich *et al.* 1995). This second application is particularly important in occupational biomechanics. Figure 8 shows a three-dimensional FEM type model of a human-vehicle system, composed of 2936 finite elements. Its main part is the spinal system model shown earlier in Figure 6. Input data for the model are: body position, geometric and material data of body and vehicle elements as well as data of the forces acting upon the system and excitation causing it to vibrate. The mathematical model describing the FEM model of the human-vehicle system, created with the aid of typical computerized procedures used in this method (Zienkiewicz and Taylor 1991), is composed of 14493 non-linear equations. Solving this system of equations for a given kinematic excitation (e.g. harmonic vibrations of the seat along the vertical axis z — see the coordinate system depicted in Figure 8) reveals displacements, velocities and accelerations of all the rigid elements of the system (bones, elements of the vehicle's construction). Knowing the parameters of vibration of the rigid elements, the distribution of deformations and stresses in the elastic elements (muscles, intervertebral disks) are then calculated.
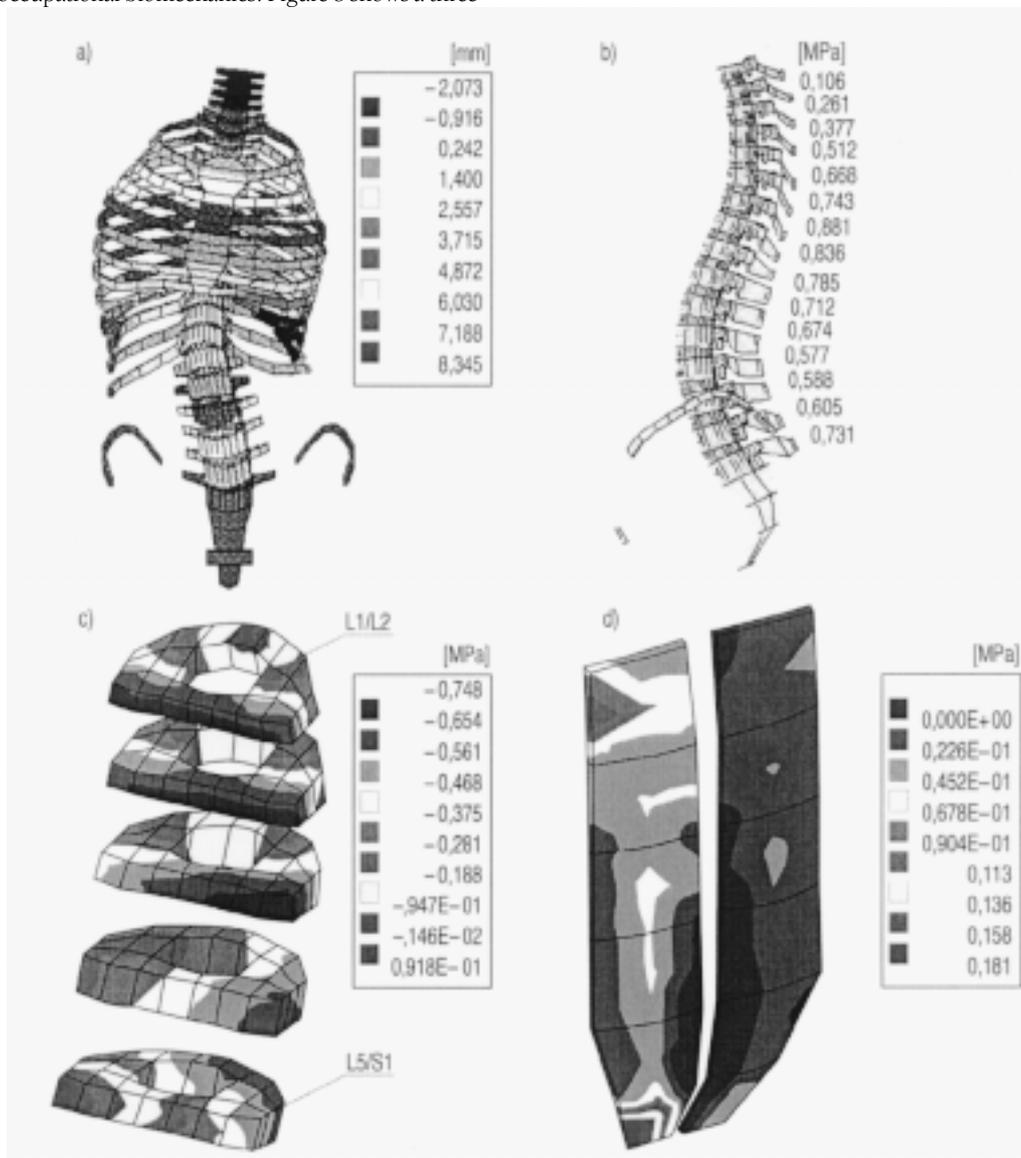
Figure 7. Sample results of simulations obtained for the spinal system from figure 6(6), subjected to a load of 200 N held in the left hand; (a) displacements of ribs and vertebrae in the frontal plane (mm), (b) distribution of pressures in the nuclei pulposa of the intervertebral disks (MPa), (c) distribution of compression stresses in the annuli fibrosa of the lumbar intervertebral disks (MPa) and (d) distribution of stresses stretching the muscle fibers of the straight abdominal muscle (MPa).
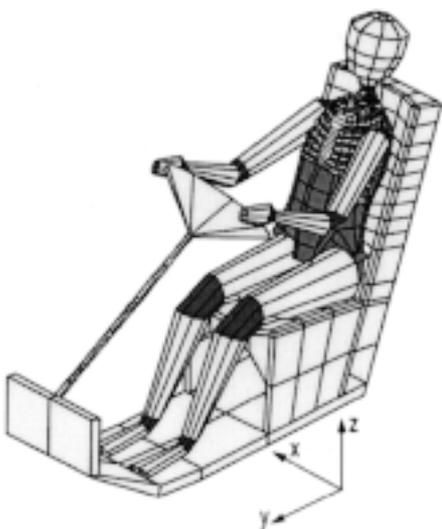
Figure 8. FEM type model of a human-operator under vibration (Dietrich *et al.* 1995).
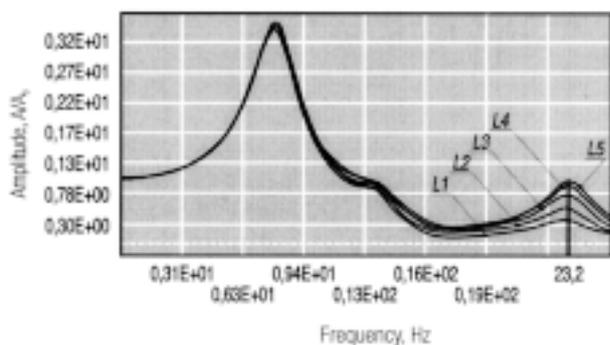
Figure 9. Resonance curves of vertebral vibrations in the lumbar region of a human-operator in the direction of axis *z* (Figure 8). A. amplitude of vibration of individual vertebrae; A_o, amplitude of vibration of the seat (here A_o = 10 mm).
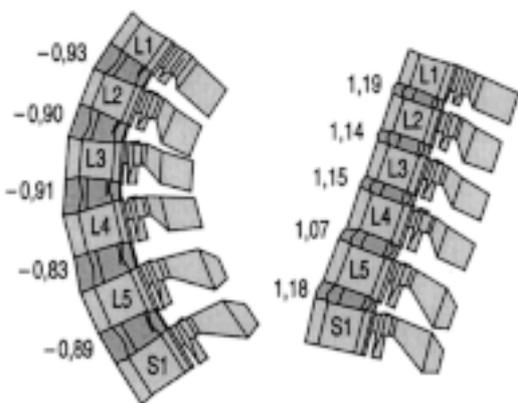
Figure 10. Extreme positions of the lumbar region of the spine in the course of one vibration cycle at a frequency of 23.2 Hz; values indicate pressures in the nuclei pulposa of the intervertebral disks (a negative value indicates hypotension); displacements have been magnified five times in comparison with the scale of the figure.

In Figures 9 and 10, sample results of solutions obtained with the aid of this model are shown.

Figure 9 shows resonance curves for the lumbar section of the spine (vertebrae L1/L5) for vibrations of frequency from 0 to 30 Hz. At low frequencies of up to several Hz, all the resonance curves are almost identical, whereas at higher frequencies the curves diverge, indicating the appearance in the intervertebral disks of strains induced by the vibrations.

Figure 10 shows an enlarged view of the extreme positions of the lumbar section subjected to a vibration of 23.2 Hz. It is easily noticed that, in the course of vibration, not only the alignment of the spine undergoes change, but also the distances between vertebrae. This produces pressure pulses in the interior of the intervertebral disks (see the values given in Figure 10) as well as pulses of stress in their tissues. These phenomena can in time lead to the appearance of deterioration causing diskopathy. The appearance of pain in the lumbar regioni at frequencies of 20/30 Hz has been noted in the literature (Singleton 1982).

With the aid of FEM methods a model of the musculoskeletal system of the entire human body has also been created for applications in both occupational biomechanics and basic biomechanical research (Kedzior and Zagrajek 1997). This model accounts for all dynamic components of loads on the human movement system (inertial forces, damping) during simulation of any movement whatever (e.g. movements of labor in the course of physical work), through a full range of rotational angles in the joints.

The physical model of the skeletal system (Figure 11 a–c) is composed of 45 rigid elements modeling those elements of the human body which are very rigid (e.g. bone) or which are significantly more rigid (e.g. vertebral segments, feet, hands) than other elements (muscles, tendons, disks between rigid spinal segments). Rigid components are connected by means of special elements of the 'six-component spring-damper' type. Each such element is composed of six subgroups; each subgroups is constructed of spring and damper acting in parallel; three linear
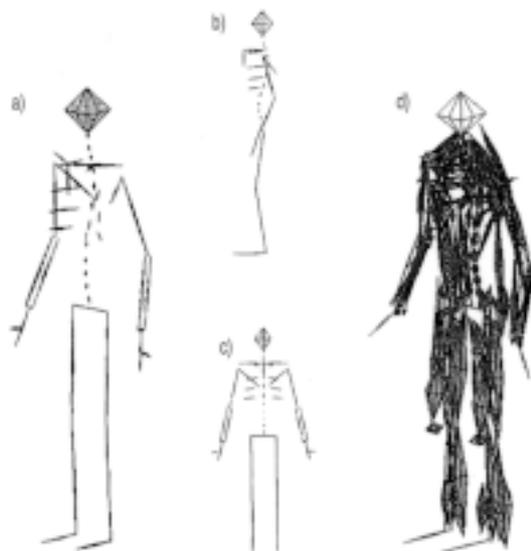
Figure 11. FEM type model of the human musculoskeletal system; (a–c) system of rigid elements modeling the skeleton; diagonal, profile and frontal views; and (d) general view of skeleton and muscles (Kedzior and Zagrajek 1997).

subgroups are acting in three mutually perpendicular directions and three angular subgruops are acting around the directions corresponding to the axes of rotation. The stiffness of the springs is chosen to allow only such movements as agree with the physiological range of movement in the joints. Suitable selection of the characteristics of the component springs makes it possible to model the limits of movement ranges in a given joint which result from its anatomical structure. The suppression coefficients of the dampers are chosen to suppress small vibrational movements in a given joint arising in the course of simulation of a movement which we know from experience has a smooth trajectory.

The physical model of the muscular system (Figure 11d) is composed of 250 special elements which model the activity of muscles such that forces exerted by them are applied to the stiff elements in accord with the lines connecting the points of insertion of a given muscle. Geometrical data of points of insertion and physiological cross-sections of muscles abound in the literature (Seireg and Arvikar 1989, Yamaguchi *et al.* 1990, Thompson and Floyd 1994). In the mathematical model of the muscles, the dependence of forces exerted by each muscle on its length and rate of contraction is considered, drawing on the analytical relations given in the work of Pierrynowski and Morrison (1985).

The physical model shown in Figure 11 has been transformed to a mathematical model, and then to a simulating one (Kedzior and Zagrajek 1997). It can be employed in solving two basic types of problems:

- Knowing the course in time of the values of forces exerted by the muscles as well as the values of external forces (such as loads acting upon the human body, for, for example, in the form of a weight held in the hand), it is possible to determine the course in time (trajectory) of displacements of selected points of the body (the number of such points is freely chosen).
- Knowing the trajectories of selected points of the human body (e.g. experimentally recorded using the method of recording movement on film) as well as the course in time of the values of external forces, it is possible to determine the course in time of all muscular forces.

In occupational biomechanics we are almost always faced with solving the second type of problem. And here an additional difficulty arises: the number of muscular forces in the human organism (and thus in the described model) is greater than the number of movements produced by them in the joints. Usually a given movement (e.g. flexion of the elbow joint) is the result of simultaneous action of several or even several dozen muscles (the so-called issue of muscle cooperation). As a result of this, in the mathematical model of a given problem, the number of unknown variables (i.e. muscular forces) is greater than the number of equations. Such a set of equations has an infinite number of solutions. One of them must be chosen. This is done by the optimization method, assuming that muscles in an organism act according to some sort of rational criterium, e.g. minimal energy expenditure (Dietrich *et al.* 1991) or other criterium, e.g. the sum of muscular forces or the sum of reactive forces in the joints (Seireg and Arvikar 1989) must be minimal, or that they meet the so-called criterium of soft saturation (Siemieski 1991). In the model described we assumed the first of the above-mentioned criteria —
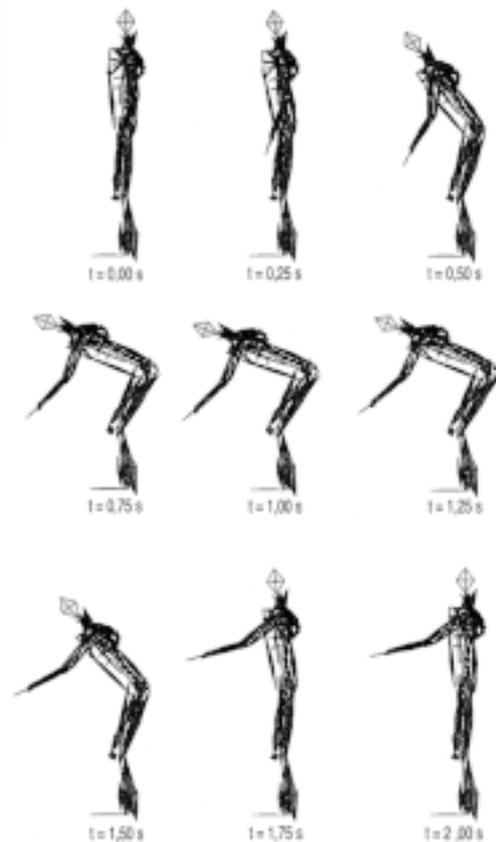
Figure 12. Model of the human musculoskeletal system at successive points in time *t* (side view) in the course of an action performed with a load of 100 N held in the hands (movement symmetrical in relation to the sagittal plane).
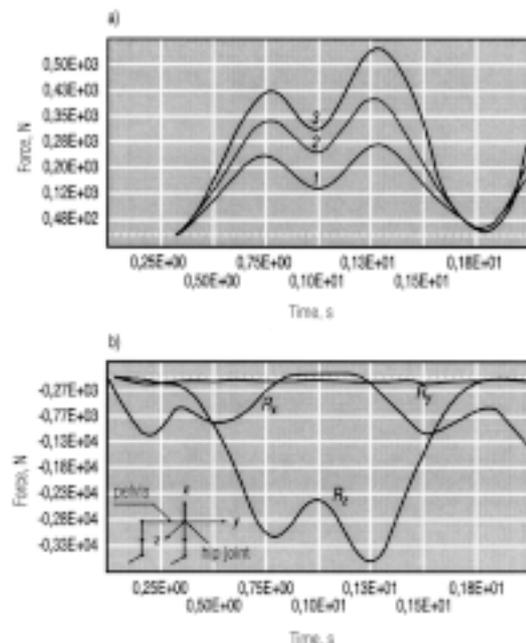
Figure 13. Courses as a function of time: (a) forces exerted by the muscles of the lower limb: 1, semitendinosus; 2, semimembranosus; 3, biceps femoris; (b) component reactive forces in the hip joint R_x, R_y, R_z respectively along axes x, y and z of the local coordinate system (in accordance with the diagram inserted in the figure).

the energetic. Sample results of simulations are shown in Figures 12 (a well-known work movement of the human body) and 13 (values of muscular forces and reactive forces in the joints).

The simulation models of the human body described herein are employed in solving a typical problem of occupational biomechanics: the capability of the human musculoskeletal system of safety lifting a load during the performance of a defined physical task. Using the computer simulation method it is also possible to solve the problem of optimization of the workspace. This problem is taken up in the next section.

4. OPTIMIZATION OF UPPER EXTREMITY LOCATION IN WORK CONDITIONS

4.1. Method of Work Space Optimization for Upper Limb

Computerized method for workspace optimization has been created on the basis of computer program, which simulates upper extremity function. The method allows for finding, from computer calculations, optimum upper extremity location in theh defined workspace.

Physical model of upper extremity consists of kinematics chain of upper extremity, muscles and workspace. Kinematics chain of upper extremity is opened and has 7 degrees of freedom. It consists of three rigid elements modeling arm, forearm and hand. Model takes into account all basic movements of the limb, defined in relation to frontal plane — abduction/adduction; sagittal plane — flexion/extension, and pronation/supiantion defined as rotation round axis of the limb.

Trunk was considered to be immobile and upper extremity joints have been modeled as rotating kinematics pair. For shoulder of third class (three degrees of freedom), elbow joint and wrist joint of forth class (two degrees of freedom). It means that in shoulder joint can be simulated moves of flexion–extension, abduction-adduction and rotation round axis of the limb, in elbow joint flexion–extension and rotation, in wrist joint of flexion–extension and abduction-adduction. Center of rotation in shoulder joint is a center of global coordinate system. Upper extremity location is defined due to local Denavit–Hartenberg coordinate system (Denavit and Hartenberg 1955) connected with every rigid element of the model.

Also coordinates of muscles attachment to bones are defined according to local coordinate system. Coordinates of points of muscles attachment to bones were adopted from Seireg and Arvicar (1989).

Workspace shape was considered to be a segment of ball sphere laying inside area of maximal upper limb reach. This space is defined by the following parameters:

- Polar coordinates of ball center in relation to global center of coordinates (center of rotation in shoulder joint).
- Internal radius of sphere segment.
- External radius of sphere segment.
- Horizontal angle of the sphere.
- Vertical angle of the sphere.

Physical model of upper limb was formalized in analytical form into mathematical model in which for symbolic operations (conversion of algebraic formulas) computer software CAMIR (Rzymkowski 1988) was used.

The analytical describe fact that in a chosen point in the work space kinematics chain of upper limb stays in static balance under

its own load, forces of muscles and external force. In this way seven equations are generated (one for each degree of freedom). In mathematical model of upper limb unknown values are 34 values of muscle forces. It means that there are seven equations with 34 unknown values — muscle forces, which makes the mathematical task statically indeterminable (excess of muscles in relation of degrees of freedom). Solution of this problem called also solution of muscles contribution is usually searched with assumption that nervous system is controlling muscles due to some merit criterion. In this study merit criterion of 'soft saturation' was implemented (Siemieski 1992).

$$\sum_{j=1}^{34} 1 - \sqrt{1 - \frac{F_i}{F_{imax}}} = \min \qquad (3)$$

where F_i is force generated by ith muscle ($i = 1, \ldots 34$); and F_{imax} is multiplication of the ith muscle cross-section (m²) and maximum allowable muscle tension.

Achievement of the purpose of the optimization problem however needs to find optimum upper limb location (calculation of optimal set of angles $q_1, \ldots q_7$), such one in which muscular load will be the lowest. That means that to complete the task there is a need of additional optimization process, in which merit function will be formal formula of the value proportional to muscle effort in static work conditions. On the basis of other studies results (Seireg and Arvicar 1989), it was considered that this merit function is expressed as the sum of modules of muscle forces in relation to axis of rotation in joints that must be developed by the muscles in the arm, elbow and wrist joints to balance the limb's own weight. Form of the second merit criterion is the following:

$$\sum_{j=1}^{7} \sum_{i=1}^{34} |F_i r_{ij}| = \min \qquad (4)$$

where F_i is force generated by ith muscle ($i = 1$–34); and r_{ij} is arm of force exertion in relation to axis of rotation of ith degree of freedom ($j = 1$–7).

Described above mathematical model together with constrains condition and double optimization was transformed into computer simulation model which gives possibility of effective solution of the task of work space optimization. The optimization process assumes cooperation between computer system and the user of the system. The task of user is to define the workspace for a given work task. User gives also parameters connected with dimensions and masses of upper limb segments and value and direction of external force. On Figure 14 diagram of optimization process conducted by this system is presented.

To find optimum upper limb location double optimization: 'external' with using 'Monte Carlo' method and second merit criterion (4) and 'internal' with first merit criterion (3) and gradient method of optimizaiton is performed (Roman-Liu *et al.* 1999).

Application of the described above system allows presentation in a graphic form solution of optimization task. Workspace is divided into 625 small subspaces (Kedzior *et al.* 1993a) (Figure 15). In each of the subspace the calculated value of merit function (4) is marked with colors. Also information concerning minimum and maximum value of merit function and adequate angles in
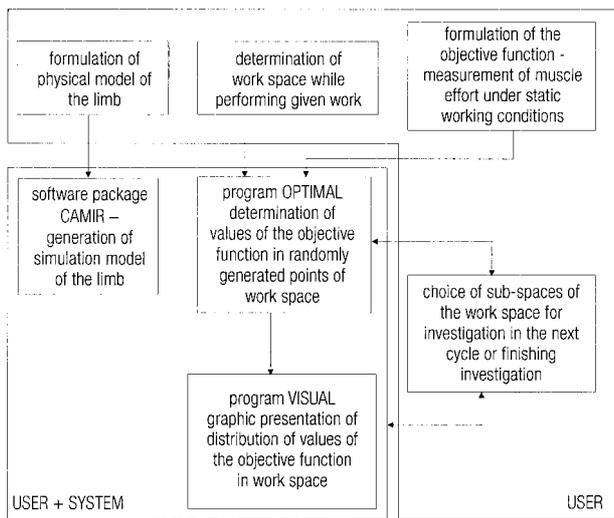
Figure 14. Optimization process of upper limb work space.
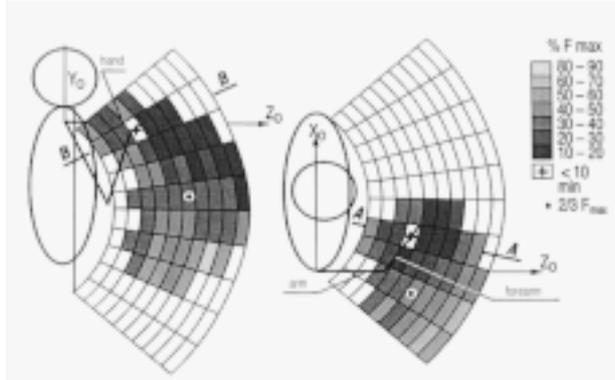
Figure 15. Optimization results: (a) vertical cross-section of the work space by plane marked on (b) as A–A; (b) horizontal cross-section of the work space marked on (a) as B–B. Optimum solution '+;' '*' marks the projection of the point where merit function reaches about two-thirds of maximum value.

joints are available. Examples of results of calculations performed on the basis of the method were presented in publications (Kedzior *et al.* 1993a–c).

4.2. Experimental Verification of the Theoretical Method for Work Space Optimization

To verify the method for the same limb locations and the same values and direciton of external force experimental studies and calculations of muscle force for a given number of participants were performed. Results from calculations (parameter MOD) were compared with results from experimental studies.

As an experimental method for verification was chosen electromyography (EMG). Although this method is sensitive to inaccuracies and artifacts, it is presently the most often used for purposes of examination of musculoskeletal load. From the EMG signal parameters can be drown out information concerning musculoskeletal tension (from the aplitude of the signal) as well as concerning process of changes in muscles in time like muscle fatigue. Good parameter that indicates muscle fatigue, can be zero crossing (Hagg *et al.* 1987, 1991). To express muscle fatigue

quantitatively slope of the regression line of zero crossing (ZC) (Hägg *et al.* 1987, Hägg and Suurküla 1987) in time is estimated. This parameter marked as SZC was analyzed in experiments, which purpose was to verify theoretical method for workspace optimization. The other analytical EMG parameter was amplitude value (AMP), calculated as RMS, which express muscle tension.

Comparison of results was conducted on the basis of measurements of EMG signal by surface electrodes from eight muscles:

- Flexor carpi radialis (FCR).
- Flexor carpi ulnaris (FCU).
- Extensor carpi ulnaris (ECU).
- Brachioradials (BR).
- Biceps brachii caput breve (BBCB).
- Deltoideus (DL).
- Triceps brachii caput laterale (TBCL).
- Trapezius (TR).

Trapezius muscle was not taken into account in the computer model. However, this muscle supports scapula and many researchers suggest that upper limb location influences muscular load of this muscle. This is the reason why in those studies, like in many other ergonomic studies (Christensen 1986, Lannersten and Harms-Ringdahl 1990) this muscle is taken as an indicator of the entire upper limb load. EMG parameters for this muscle were compared with sum of upper limb muscle forces calculated from the model of upper extremity. It was justified by the fact that load in this muscle is considered to be indicator of load in the whole upper extremity.

The muscle fatigue and load is influenced by limb location, vector of external force and frequency or duration of the load exertion. Theoretical model did not take into account the third of those factors, so it did not differentiate intermitted load and constant load. However work tasks on a real work stand has usually more or less dynamic character that causes different pattern of muculoskeletal load. Because of that it was decided to perform experimental study with external force constant (conditions closer to theoretical model) and intermittent (conditions closer to reality). It gave the possibility to compare and assess the model in relation to real conditions on the work stand.

For theoretical model verifications were chosen 10 different limb locations — four for studies with constant load and six for studies with intermittent load (Roman-Liu *et al.* 1999). So, there were muscle forces calculations and experiments for 20 various differentiated by: character of load — constant (*C*) or intermittent (*I*), and value of external force: unloaded (*U*) or loaded (*L*).

Analysis was performed to assess:

- correlation between experimental studies results (AMP and SZC) and calculations conducted on the basis of computer model (MOD);
- differentiation of musculoskeletal load due to external force value; and
- differentiation of musculoskeletal load due to upper extremity location.

On the basis of results it was revealed that in most of cases there are statistically significant correlation between EMG parameters from experimental studies and value of parameter MOD proportional to muscle forces calculated by computer model from theoretical studies (Figures 16 and 17).

Values of MOD parameter are better correlated with amplitude
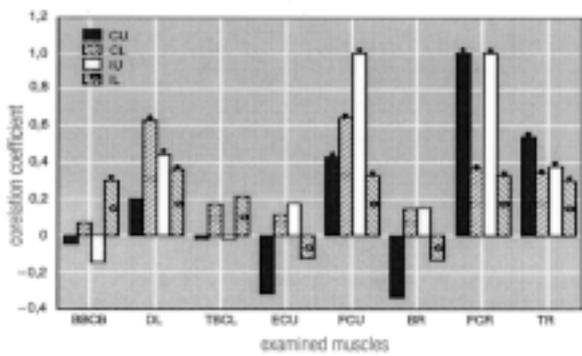
Figure 16. Correlation coefficients between parameter MOD and AMP for examined muscles in study with constant load (C) and study with intermittent load (I); *differences statistically significant with *P* < 0.05 (U, variants without external force; L, variants with external force).
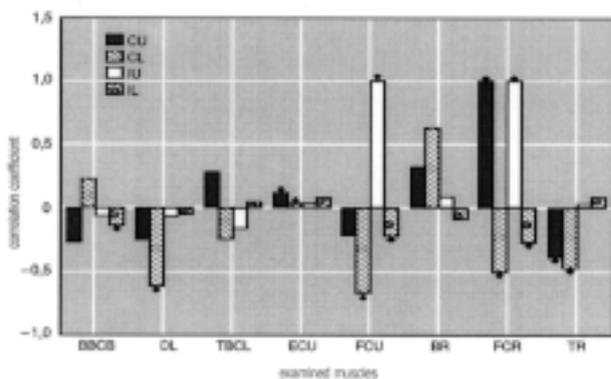
Figure 17. Correlation coefficients between parameter MOD and SZC for examined muscles in study with constant load (C) and study with intermittent load (I); *differences statistically significant with *p* < 0.05 (U, variants without external force; L, variants with external force).

value of EMG signal (AMP) than with parameter that reflects muscular fatigue (SZC). It refers to both study with constant and with intermittent load. However in studies with constant load statistically significant correlation coefficients between MOD and SZC are in higher number of cases than in experiments with intermittent load. This is probably due to higher muscular fatigue in experiments with constant load, which makes stronger dependence between muscle tension and muscle fatigue and in consequence between muscle fatigue and muscle force expressed by parameter MOD.

Differences in values of parameters MOD, AMP and SZC between variants with and without external force for the same limb location are expressed by parameter ratio of load. Ratio of load is quotient of averaged values of analyzed parameters (AMP, SZC, MOD) in variants with external load and without external load. It was checked if values of the ratio of load of analyzed parameters for the examined muscles belonged to the same range, it means were > 1 or < 1. If ratio of load for all three parameters belonged to the same range it was assumed that there is agreement between theoretical and experimental results.

In Figure 18 are values of ratio of load for parameters MOD, AMP and SZC.
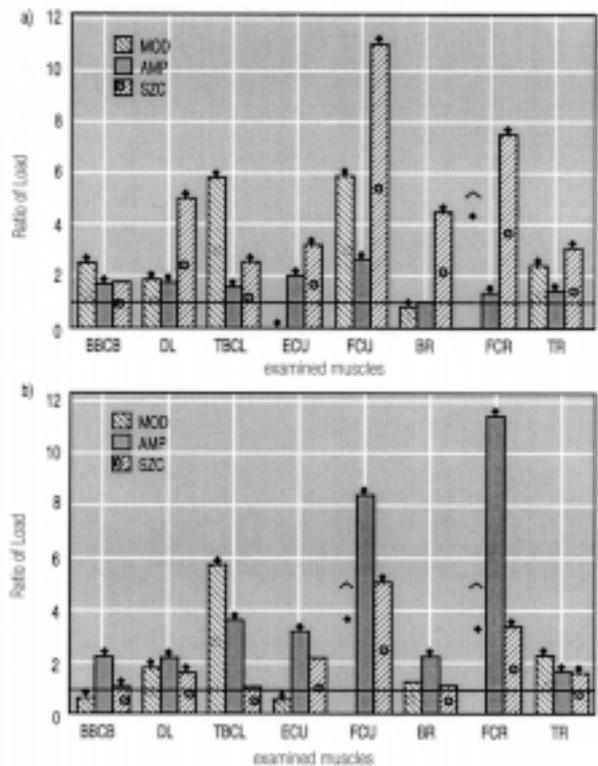
Figure 18. Ratio of load for analyzed paramters of LMG signal and results of computer simulation in (a) studies with constant load; (b) studies with intermittent load; *differences statistically significant with *p* < 0.05; ^values indeterminable.

It was concluded that experimental and theoretical results could be regarded as to be in step in differentiation according to external load, although not all differences are statistically significant.

For each of analyzed parameters (MOD, AMP, SZC) was calculated value of ratio of location, which is quotient of average values of analyzed parameters for two compared upper limb location. Similarly like in case of ratio of load agreement between theoretical and experimental study was assumed to occur when values of ratio of location for analyzed parameters were in the same range, it means were > 1 or < 1.

Much better differentiation of upper limg location occurred in study with external load (Figure 19) than with study without external load (Figure 20).

In study with constant load most of differences expressed by parameter ratio of location was statistically significant and what is more there were the same tendency in both studies with external load and studies without external load. In results of study with intermittent load upper extremity location was not be able to differentiate univocally and the differences were statistically not significant.

As a result of verification it can be stated that there is convergence between results of theoretical and experimental studies, especially in differentiation accoding to external load. It means that on the basis of computer simulation workspace can be differentiated according to external load of upper extremity. There was also noticed partial agreement between values of EMG parameters and results of computer simulation in differentiation of musculoskeletal load according to limb location.
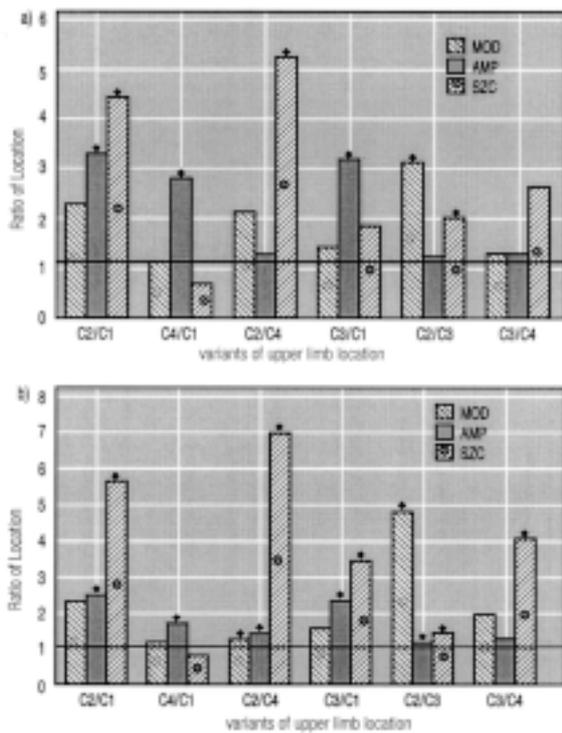
Figure 19. Ratio of location in study with constant load: (a) variants without external load; (b) variants with external load; *differences statistically significant with *p* < 0.05.
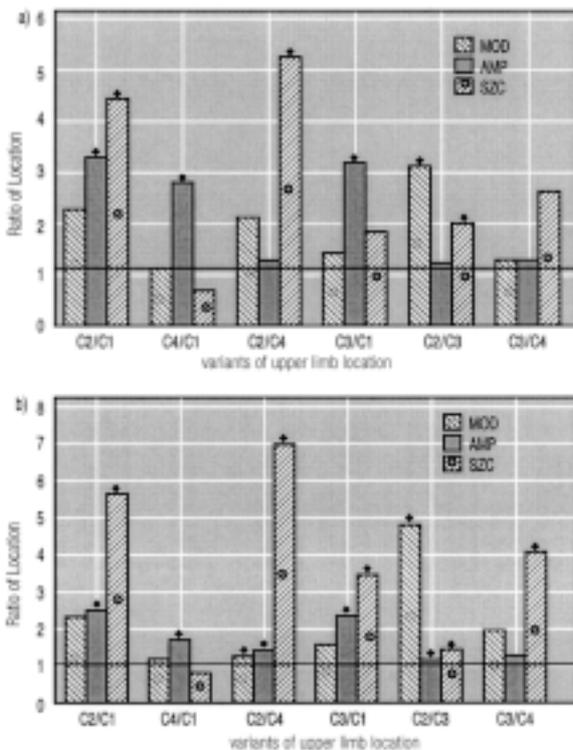
Figure 20. Ratio of location in study with intermittent load: (a) variants without external load; (b) variants with external load; *differences statistically significant with *p* < 0.05.

The lack of ideal agreement between theoretical and experimental results is most probably caused by both inaccuracies in experimental method and simplification in model. Sources of dispersion in EMG studies are connected with subjective differences. Values of EMG parameters are in a high degree dependent on kind of muscles and contribution of fast and slow fibers (Fallentin *et al.* 1985, Gerdle *et al.* 1988). In analysis of EMG signal there were not analyzed changes of EMG signal according to changes of muscle length, which can also be source of discrepancies.

Sources of discrepancies in theoretical calculations are in not strict enough subjective differentiation and in to high simplification of the model (bones taken as rigid links, muscles as lines, bones radius and ulna were modeled as one rigid link, shoulder girdle was not taken into account). The differences in results can also be caused by individual differences between people. Only limb mass and limb length were measured separately for each subject. So it can be stated that both methods theoretical and experimental have some inaccuracies and should be considered as approximate.

However, on the basis of results there can be drawn a conclusion that created program for workspace optimization is good enough for tasks in which the load of upper extremity is of constant static character. In cases where muscular fatigue induced by external load is not proportional to forces developed by muscles, the method can be also used, but results should be considered as approximate.

REFERENCES

3D STATIC STRENGTH PREDICTION PROGRAMTM, *User's Manual, Version 2.0* (University of Michigan, Center for Ergonomics).

BOROWSKI, S., DIETRICH, M., KEDZIOR, K. and RZYMKOWSKI, C., 1991, Modeling of dynamic loads acting upon the human–operator musculoskeletal system. In *Proceedings of the VIII World Congress on the Theory of Machines and Mechanisms*, Prague, vol. 3, 781–4.

CHAFFIN D.B. and ANDERSSON, G.B., 1991, *Occupational Biomechanics* (2nd ed.) (New York: Wiley).

CHAFFIN, D.B. and ANDERSSON, G.B., 1994, Computerized biomechanical models for high exertion manual jobs. In *Proceedings of the 3rd Pan-Pacific Conference on Occupational Ergonomics*, Seoul, 13–17 November, 1–15.

CHAFFIN, D.B. and ERIG, M., 1991, Three-dimensional biomechanical static strength prediction model sensitivity to postural and anthropometric inaccuracies, *III Transactions*, **23**.

CHRISTENSEN, H., 1986, Muscle activity and fatigue in the shoulder muscles of assembly plant employees, *Scandinavian Journal of Work Environment Health*, **12**, 582–7.

DENAVIT, J. and HARTENBERG, R.S., 1995, A kinematic notation of lower pair mechanism based on matrices, *Transactions of ASME Journal of Applied Mechanics*, June, 215–22.

DIETRICH, M., KEDZIOR, K., MILLER, K. and ZAGRAJEK, T., 1993, Statics and stability of the human spine under working conditions. In W. Marras (ed.), *The Ergonomics of Manual Worker* (London: Taylor & Francis), 147–50.

DIETRICH, M., KEDZIOR, K. and RZYMKOWSKI, C., 1997, Theoretical and practical aspects of modeling and simulation of the human musculoskeletal system. In A. Morecki and K.J. Waldron (eds),

Human and Machine Locomoation. CISM Courses and Lectures no. 375 (Vienna: Springer), 107–23.

DIETRICH, M., KEDZIOR, K. and ZAGRAJEK, T., 1991, A biomechanical model of the human spinal system. Proceedings of the Institution of Mechanical Engineers, vol. 205, Part H, *Journal of Engienering in Medicine*, 19–26.

DIETRICH, M., KEDZIOR, K. and ZAGRAJEK, T., 1995, Three dimensional FEM model of man—operator under vibration. In *Proceedings of the Ninth World Congress on the Theory of Machines and Mechanisms*, Polytechnic of Milan, vol. 3, 2187–90.

FALLENTIN, N. and JORGENSEN, K., 1992, Blood pressure response to low level static contractions, *European Journal of Applied Physiology*, **64**, 445–59.

FIDELUS, K., KEDZIOR, K. and MORECKI, A., 1983, Biomechanics of muscles. In *Encyclopedia of Contemporary Physics* (Warsaw: PWN), 767–78. [in Polish]

FOX, E.L., BOWERS, R.W. and FOSS, M.L., 1993, *The Physiological Basis for Exercise and Sport* (Madison: Brown and Benchmark).

GERDLE, B., WRETLING, M.L. and HENRIKSSON-LARSEN, K., 1988, Do fibre proportion and angular velocity influence the mean power frequency of the electromyogram? *Act Physiologica Scandinavica*, **134**, 341–6.

HAGG, G. and SUURKULA, J., 1990, Zero crossing rate of electromyograms during occupational work and endurance tests as predictors for work related myalgia in the shoulder/neck region, *European Journal of Applied Physiology*, **62**, 35–44.

HAGG, G., SUURKULA, J. and JIEW, M., 1987, A worksite method for shoulder muscle fatigue measurement using EMG test contractions and zero crossing technique. *Ergonomics*, **30**, 1541–55.

KEDZIOR, K. and ZAGRAJEK, T., 1997, A biomechanical model of the human musculoskeletal system. IN A. Morecki and K.J. Waldron (eds), *Human and Walking Machines Locomotion*, CISM Courses and Lectures no. 375 (Vienna: Springer), 124–53.

KEDZIOR, K., ROMAN, D. and RZYMKOWSKI, C., 1993a, Modeling of human arm in the work space. In *Proceedings of the 4th International Symposium on Computer Simulation in Biomechanics* (Paris: Montlignon), BMG1-14–17.

KEDZIOR, K., ROMAN, D. and RZYMKOWSKI, C., 1993b, Work space optimization by computer simulation method. In W.S. Marras et al. (eds), *The Ergonomics of Manual Work* (London: Taylor & Francis), 619–22.

KEDZIOR, K., ROMAN, D. and RZYMKOWSKI, C., 1993c, Modeling upper extremity muscles effort under static working conditions. In

M. Thoma and W. Wyner (eds), *Lecture Notes in Control and Information Sciences*. Vol. 187 (London: Springer), 333–8.

LANNERSTEN, L. and HARMS-RINGDAHL, K., 1990, Neck and shoulder muscle activity during work with different register systems, *Ergonomics*, **33**, 49–65.

MIANOWSKI, K., 1993, *Stand for Measurements of Human Muscle Group Characteristics*. Lecture Notes of the ICB Seminars, vol. 20 (Warsaw: International Centre of Biocybernetics), 138–48.

NIGG, B. and HERZOG, W. (eds), 1994, *Biomechanics of the Musculoskeletal System* (Chichester: Wiley).

PIERRYNOWSKI, M.R. and MORRISON, J.B., 1985, Aphysiological model of muscular forces in human locomotion — theoretical aspects, *Mathematical Biosciences*, **75**, 69–101.

ROMAN-LIU, D. and KEDZIOR, K., 1999, Experimental verification of computerized method for work space optimization in conditions of static work, I*nternational Journal of Occupational Safety and Ergonomics*, **5**, 107–22.

ROMAN-LIU, D., KEDZIOR, K. and RZYMKOWSKI, C., 1999, Computerized method for work space optimization in conditions of static work, I*nternational Journal of Occupational Safety and Ergonomics*, **5**, 95–106.

RZYMKOWSKI, C., 1988, A package of computer programs for equations of motion — biomechanics applications. *Biology of Sport*, **5(suppl. 1)**, 188–92.

SEIREG, A. and ARVICAR, R., 1989, *Biomechanical Analysis of the Musculoskeletal Structure for Medicine and Sports* (New York: Hemisphere).

SIEMIE SKI, A., 1992, Soft saturation, an idea for load sharing between muscles — application to the study of human locomotion. In A. Cappozzo, M. Marchetti and V. Tori (eds), *Biolocomotion, A Century of Research Using Moving Pictures* (Rome: Promograph), 293–304.

SINGLETON, W. (ed.), 1982, *The Body at Work: Biological Ergonomics* (Cambridge: Cambridge University Press).

THOMPSON, C.W. and FLOYD, R.T., 1994, *Manual of Structural Kinesiology* (12th ed.) (St Louis: Mosby-Year Book).

YAMAGUCHI, G.T., SAW, A.G., MORAN, D.W., FESSLER, M.J. and WINTERS, J.M., 1990, A survey of human musculotendon actuator parameters. In *Multiple Muscle System: Biomechanics and Movement Organization* (New York: Springer), 717–73.

ZIENKIEWICZ, O.C. and TAYLOR, R.L., 1991, *The Finite Element Method* (London: McGraw-Hill).

Occupational Epidemiology with Special Focus on Ergonomics and Musculoskeletal Disorders

E. Bach

Department of Epidemiology and Surveillance, National Institute of Occupational Health, Lersø Parkallé 105, DK-2100 Copenhagen Ø, Denmark

1. INTRODUCTION

Over the past decades, a shift in occupational medicine has taken place. The traditional occupational diseases have become less frequent and the emphasis in research has shifted to exploration of the work-related etiology of "common diseases" including their manifestations and precursors. Many of these diseases have an occupational as well as a non-occupational etiology, and the occupational component may be difficult to identify.

Therefore, more sensitive epidemiologic approaches are needed in occupational epidemiology, i.e. to study the occurrence of multi-causal diseases in relation to their work-related determinants. This implies a tightening of the study design and more accurate methods for measuring exposure.

This chapter focuses on ergonomic exposures and morbidity such as musculoskeletal disorders. Therefore, only concepts relevant for studying these relations are mentioned, which leaves out, for example, mortality frequencies.

1.1. Musculoskeletal Disorders

A nationwide cross-sectional study in Denmark showed that skilled and unskilled workers had an increased risk of low back pain. People employed in "building and construction work" had twice the risk and persons in "social work, child daycare work and psychological work" had 1.6 times the risk for low-back pain compared with other employees.

Several epidemiologic studies within the past 25 years have emphasized a potential work-related ethiology of musculoskeletal disorders, mainly based on physical risk factors. Other studies have focused on individual and psychosocial risk factors and other investigators have even conceptualized causation at the society level as a consequence of modern life. Far the most of the epidemiological evidence derives from cross sectional studies, but case-control studies and follow-up studies are now published in scientific papers.

The health effects of workplace risk factors have ranged from outcome measures of vaguely defined self-reported pain or trouble to medically diagnoses and hospitalizations. The severity and effects on disability are unknown, but the cost for society because of musculoskeletal disorders is widely acknowledged by researchers and policy makers.

In Denmark a government action program was set up in 1991 against work-related musculoskeletal disorders of the neck and upper limb associated with monotonous repetitive work. The background was research, national statistics on workers compensation claims and political debate. In order to obtain more knowledge, a Project on Research and Intervention in Monotonous work (PRIM) was launched in 1993. PRIM is an interdisciplinary research project, embracing several studies on company organizational factors that may influence the amount of monotonous work as well as work related risk factors of musculoskeletal disorders and psychological effect.

2. SOURCES OF INFORMATION

Epidemiological findings are often based partly or completely on individuals responses to questionnaires — mailed questionnaires or interview questionnaires — with detailed question on individual exposures and diseases. Exposure data may be supplemented with observations or measures on the workplace, and data on diseases may come from medical examinations or disease registers.

Studying musculoskeletal disorders is a relatively new area in occupational epidemiology and there is a substantial lack of knowledge about causal factors of musculoskeletal disorders. There is also a need for developing standardized ways of asking the relevant questions about exposures and disorders. In the PRIM study, we included all validated questions and developed new questions on relevant topics where no validated questions existed (an English version of the questionnaire can be ordered from the author).

2.1. Exposure

Most of the epidemiologic studies concern the manufacturing industry. Focus has been on mechanical factors at work, and methods for measuring the ergonomic exposures have often been very specific for the examined populations. The questions are seldom validated and they differ for different industries which makes it impossible to compare data and data quality.

Work with new technology, especially video display units, has also been suggested as risk factors for musculoskeletal disorders, and new questions were developed in different studies. However, few of these studies have so far been published in scientific papers. The PRIM-study includes questions suitable for describing monotonous, repetitive work and, therefore, also for work with new technology. Questions on ergonomic exposures calls for illustrations and not only questions in the questionnaire. An example used in PRIM is shown in Figure 1.

Most researchers assume a multi factorial nature of musculoskeletal disorders and it has been proposed to incorporate workplace factors as well as individual factors at the physical, psychological and social level in future research, which is done in PRIM. Several validated questionnaire modules at the psychological level are available, and the need in this area is to reduce the number of questions in the modules on a scientific plausible way. There are few standardized ways of asking about individual factors such as sport activities.

2.2. Disease

Disease is used in a broad sense in epidemiology that includes health effects such as self-reported health complaints and symptoms and their administrative consequences. This means that the researcher must define the disease in each analysis very precise. A Nordic work group has made a standard questionnaire measuring 1 year prevalence, 1 week prevalence, days with pain, sick leave, contact to health personnel because of pain, and use of medicine due to musculoskeletal disorders. The questionnaire

12. Does your work entail that…………………

(Mark with only one cross and please answer all questions)

	Never/ seldom	App. 1/4 day	App. 1/2 day	App. 3/4 day	Almost all the time
you work with your back deeply bent forward?	1	2	3	4	5
you work with your arms lifted in front of your body?	1	2	3	4	5
you work with your neck deeply bent forward?	1	2	3	4	5
you work with your wrist strongly bent or flexed?	1	2	3	4	5
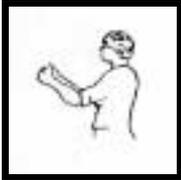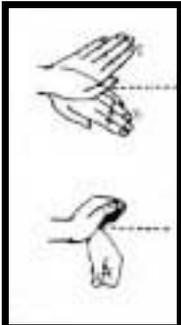
Figure 1. Example on illustrated questions in questionnaire from PRIM.

is used in several studies but is often supplemented with other questions about the pain, as in PRIM.

3. STUDY BASE

The study base is the morbidity experience of a population over time. Population, time, methods, and design must be chosen with respect to the hypothesis in the study and the potential generalization from the study to the scientific level. Small studies or studies including non-sensitive measures give false negative studies or studies with no information.

4. MEASURES OF DISEASE FREQUENCY

Two basic frequency measures are used to quantify the occurrence of diseases in a population, i.e. incidence and prevalence.

4.1. Incidence

The incidence refers to the occurrence of *new* diseases in a population in a given time and is concerned with the risk of getting a disease. When relevant, the incidence may be measured in relation to the total time spent in the population (sum of the individual person times at risk) and is called the incidence density (ID) — in other studies the incidence is related to the total number of persons in the population and is called the incidence rate (IR).

4.2. Prevalence

The prevalence rate (PR) refers to the proportion of people *with* the disease at a specific time Prevalence describes the disease state in a population and the longer the duration of the disease, the more meaningful is the prevalence. Incidence is difficult to define and study in epidemiologic studies concerned with reversible health effects,` and when the natural history of the disorders is only vaguely known, such as musculoskeletal disorders. In many such studies the prevalence is often studied instead of incidence.

4.3. Relative Measures

In order to compare measures of disease frequencies in two comparable groups (a study group and a reference or control group) the measures are often deviated by each other. These relative measures are correctly called rate ratios (RR) but are often referred to as risk ratios or relative risks.

Another relative measure is the odds ratio (OR) — that is the odds of getting the disease in one group divided by the odds of the disease in another group (the control group), or the odds for the diseased of having been exposed compared with the odds among the non-diseased.

These relative measures are all 1 if the exposure do not increase the risk of disease and > 1 if the exposure increases the disease risk.

The rate ratios vary very little for frequent disorders as musculoskeletal disorders. In this case it may also be relevant to compare the measures by subtracting the measure of the control group from the measure of the study group. This relative measure is called the rate difference (RD) or attributable risk

5. STUDY DESIGNS

The two major types of studies are cohort studies and case-control studies. The cross sectional study, where exposures and disease frequencies are measured at the same time, is a borderline case of the cohort study. Today intervention studies are also conducted.

The choice of study design depends on the purpose of the study and the studied exposures and disease frequencies.

5.1. Cohort Studies

In cohort studies two or more groups (cohorts) of people without disease (i.e. without the disease that is studied) and with different exposure levels are followed in order to measure the incidence rates and to study if the exposed cohort have an excess incidence. Cohort studies have their strength when studying relative frequent diseases and rare exposures. The researcher decides the number of persons in the exposed group at least if enough available exposed persons are available. The cohort study is often prospective, but can be retrospective if it is based on existing data.

Rate ratio, odds ratio and rate difference may be used in order to compare disease rates in cohort studies. Cohort studies concerned with musculoskeletal disorders may use prevalence as disease measure instead of incidence. They may include persons with disorder in the beginning of the study to study both incidence of disorder and disappearance of disorder when the disorder is reversible.

5.2. Case-control Studies

In the case-control study, groups with the studied disease and groups without this disease are compared according to former exposures. Case-control studies are mainly used in studies of rare diseases (and relatively frequent exposures). The researcher decides the number of diseased persons in the study if enough diseased persons are available, and incidence (or prevalence) of disease can not be estimated in the study. Odds ratio is, therefore, the measure used for comparison in case-control studies.

5.3. Intervention Studies

Studies where people are randomized to different exposures like clinical trials are not seen in occupational epidemiology because of ethical and practical limitations. The experimental studies in occupational epidemiology are called *intervention studies*, which is cohort studies where groups of people on one workplace are randomized to a planned intervention and people on another work place are randomized to no intervention. Disease frequencies are measured before and after the time of intervention in both the study group and the control group.

In this case the intervention is expected to improve the work environment and is, therefore, ethical plausible. The group without intervention is part of the study in order to illustrate the effect of time and "natural" interventions that occurs from changes in the work situation for reasons outside the study, e.g. changes in demand for the production. The above-mentioned PRIM study is an intervention study.

5.4. Ecologic Studies

Ecologic studies are aggregate studies that focus on comparison of groups because individual-level data are missing. In a sense this type of study is an incomplete design. Epidemiologists have used these studies for years as simple descriptive studies, but statistical methods for this type of studies have now been improved.

6. CONFOUNDERS AND STATISTICAL ANALYSIS

Comparability of the groups except for the exposure under study is ensured in the clinical trial. In occupational epidemiology, the

groups may not be comparable due to selection, and the differences in occurrence of other determinants for the disease under study. These other determinants are potential confounders.

Dealing with confounders in the design of the study can be made by ensuring that the determinants are equally distributed in the study group and in the control group so they do not confound the results. One or a few confounders can be dealt with in the analysis by standardization, e.g. age standardization, but more common today is to handle the confounders in stratified analysis and advanced (multivariate) statistical analysis.

The most used statistical method during many years is logistic regression that produces odds ratios for one disease for each confounder included in the analysis, where disease is a dichotomized, dependent variable. The method is included in any statistical package for epidemiologists. Statistical methods for non dichotomized disease measures and methods dealing with prevalence ratios instead of odds ratios have been conducted, but are until now only included in few statistical packages.

7. BIAS AND VALIDITY

The statistical analysis estimates the influence of random errors in the study, but the analysis can never deal with a systematic error (a bias). If for example all of the participants in the study group are persons < 25 years of age and all participants in the control group are > 50 years, no statistical analysis can then make the disease frequencies in the two groups comparable. The validity of the study is ruined. A bias can give a false-positive or a false-negative result in a study.

All epidemiologic studies may include minor bias that does not necessarily invalidate the study, but the researcher must discuss the potential lack of validity and its possible influence on the result.

8. INTERPRETATION OF EPIDEMIOLOGIC STUDIES

Results from studies may reflect a true relation or may be due to chance. For some studies, e.g. epidemiologic studies, the result may be biased. The researcher must, therefore, evaluate how convincing the results are:

In 1965, Sir Bradford Hill created nine viewpoints for judging causality, known as "Hill's criteria." As pointed out by Hill, "none of my nine viewpoints can bring indisputable evidence for or against the cause-and-effect hypothesis and none can be required as a sine qua non." However, the criteria have become classic, and are still widely used to structure the discussion in an epidemiologic study. The criteria are:

- Strong association: the stronger the association the more credible is the causality, but a weak association can also be causal.
- Consistent evidence: the consistency of evidence (other researchers have found the same association) speaks in favor of causality. This is a simple way of using *a priori* knowledge.
- Specificity of the association: for multi-causal diseases this criteria is less relevant or event wrong.
- Temporary: the cause must precede the effect.
- Biological gradient : a dose–response relationship speaks in favor of causality.
- Biological plausibility: biological plausibility speaks in favor of causality.
- Coherence of evidence: if experimental work and theory support en epidemiologic finding, the evidence for causality is corroborated.
- Experimental evidence: a change in morbidity after a change in exposure supports causality. This criterion refers to intervention studies.
- Analogy: reasoning by analogy can help judge causality.

A discussion of the criteria and their limitations is included in the original paper by Hill (1995) and in several textbooks of epidemiology, for example the two textbooks mentioned in the References.

However, the researcher must, therefore, give detailed descriptions of the definitions of exposures and diseases, the persons included in the study and how they are selected, the study design, and the statistical methods. The information must be given in such a way that the reader can evaluate the study and form an independent judgement of the study. Also the researcher may use Hill's criteria for inspiration.

REFERENCES

HERNBERG, S., 1992, *Introduction to Occupational Epidemiology* (Lewis).

HILL, A.B., 1995, The environment and disease: association or causation? *Proceedings of the Royal Society of Medicine*, 58, 1217.

OLSEN, J. and IEA EUROPEAN QUESTIONNAIRE GROUP, 1998, Epidemiology deserves better questionnaires. *International Journal of Epidemiology*, 27, 935.

ROTHMAN, K.J. and GREENLAND, S., 1998, *Modern Epidemiology* (2nd ed.) (Philadelphia: Lippincott-Raven).

XU, Y., BACH, E. and ØRHEDE, E., 1996, Occupation and Risk for the occurrence of low-back pain (LBP) in Danish employees. *Occupational Medicine*, 46, 131–6.

Occupational Health and Ergonomics

P. Mairiaux

School of Public Health, Université de Liège, Liège, Belgium

1. INTRODUCTION

Most industrialized countries have set up well-organized occupational health (OH) services, sometimes since the beginning of the twentieth century, as part of a health-at-work protection policy, and similar services are being developed in a rising number of industrializing countries all over the world. When considering the aims assigned to these services by public authorities, an obvious shift has been observed during the two last decades of the twentieth century from screening for occupational diseases (secondary prevention) towards assessment activities at the workplace (primary prevention). A growing number of OH professionals are thus involved in ergonomics activities, looking for a complementary training in ergonomics or even acting as full time ergonomics consultants. In this context, the present paper aims at analyzing the specific contribution that a professional having an OH background may bring to the development of ergonomics in enterprises. In order to avoid a compilation of general statements, the analysis will be focused on the potential role of OH professionals in the prevention of work-related musculoskeletal disorders (WRMSDs). This analysis will consider the two sides of any OH activity, the health surveillance on the one hand, and the monitoring of working conditions on the other hand.

When considering their preventive activities, the OH practitioners are facing one particular challenge: in occupational health they must avoid implicitly using the bio-medical model of health where disease is seen as the direct result of a pathogenic agent. Health must be viewed, instead, as a dynamic equilibrium between man and the environment. In this perspective, the prevention of WRMSDs should fundamentally be based on a conceptual model that integrates, in a balanced way, the influences of physical factors (force, compression, repetition, posture) and those of psychosocial factors such as lack of control, perceived stress, time pressure, (Hagberg *et al.* 1995; Inserm 1995). In this respect, the ergonomic approach offers an appropriate frame of reference, provided the user sticks to its true definition as a discipline which integrates knowledge drawn from the human sciences — and not only from work physiology or biomechanics — in order to match the task and the work environment to the abilities of people at work.

2. HEALTH SURVEILLANCE

2.1 Usefulness for prevention

The surveillance of workers' health, through tests and examinations, is still today one of the cornerstones of OH services activity and it is therefore worth analyzing its potential usefulness for the prevention of WRMSDs.

While the selecting out of susceptible individuals through pre-employment screening has been advocated as a *primary prevention* measure, everybody acknowledges today that there is
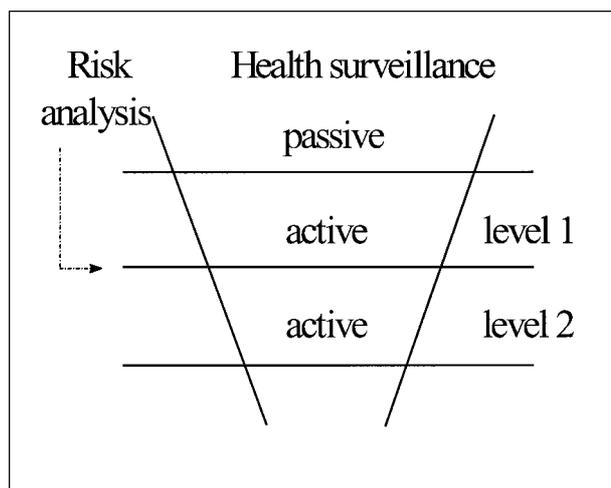
Figure 1. Levels of health surveillance.

no valid criterion, having a satisfactory predictive power, to be used for the selection of potential musculoskeletal sufferers (Inserm 1995). Hence, the preventive influence of health surveillance is clearly restricted to the two other components of prevention.

The early detection of tissue alterations, at a sub-clinical stage when possible, is called *secondary prevention* and forms the main objective of health surveillance systems. With regard to the WRMSDs, both the uncertainties in assessing the natural history of the disorder and the complexity, cost or invasive nature of some diagnostic techniques explain why cases are most often detected at a symptomatic stage. In OH services, the assurance of confidentiality to the worker consulting a health professional is a key factor in promoting the expression of individual complaints that otherwise could remain hidden due to the fear of consequences for the job holder. OH professionals are thus ideally located at the interplay between a man or woman and their job or task. At the population level, a WRMSD case identification can be considered as a "sentinel event" that points to the need for a risk assessment at the workplace and may result in preventive interventions beneficial to the whole group of exposed workers. The OH professionals may be seen as the eyes and ears of the ergonomics teams in the enterprise, quickly directing their attention to the problematic work areas.

Health surveillance thus provides, theoretically, one possible strategy for the identification of jobs at high risk of musculoskeletal disorders. Three levels of surveillance with increasing requirements are proposed to carry out such a system, while maintaining a satisfactory cost-effectiveness balance (see figure 1).

2.2. Passive Health Surveillance

The first level, or "passive" surveillance, involves the analysis of existing data related to the workers health: work accidents descriptions and diagnostic labels for sick leaves of long duration are two examples of data often collected on a systematic basis. Their treatment requires however a computerized and uniform codification, and the availability of data concerning the duration of exposure to the risk factors. If such is the case, the statistical analyses may allow the OH professional to point out in the company a given department, or job title, as exhibiting a higher prevalence of the WRMSD under study (Chatterjee 1992; Hagberg *et al.* 1995).

2.3. Active Health Surveillance

An "active" surveillance system collects health data independently of seeking care behavior of the worker. In the "funnel-shaped" approach illustrated in figure 1, this active surveillance is directed towards the jobs associated with a potential risk, as estimated from the passive surveillance results and/or the risk analysis (see below). Level 1 of such a system involves the periodic administration of standardized survey questionnaires, as for instance the Nordic one, in order to collect complaints and symptoms. Providing precise criteria are defined to assess the diagnostic probability and the severity of the symptoms, possible cases of WRMSDs can be identified through the analysis of the survey data. Those cases can then undergo a thorough examination by qualified OH professionals, this being considered as level 2 of the system. Such a stepwise approach should diminish the bias observed with the traditional medical surveillance when the identification of a potential case of a WRMSD is most often depending on the severity of symptoms at the time of the examination and/or the practitioner interest for this specific category of disorder. From a cost-benefit point of view, this approach is also attractive as it requires the intervention of highly qualified OH professionals for a limited and selected number of cases. It is fair, however, to point out that this new approach still has to be assessed in order to establish its sensibility and predictive value (Ansi 1995), and that its use by the practitioner implies the availability of a set of validated case definitions for the various WRMSDs.

3. RISK ANALYSIS AT THE WORKPLACE

In most countries, labor regulations and health and safety guidelines put a growing emphasis on hazard control at the design stage and on risk analysis (EEC 1989; Ansi 1995). The OH professional is thus invited to devote an increasing part of his time to the evaluation of risks in a multidisciplinary approach, alongside other professionals such as the ergonomist.

3.1. A Structured Approach

As the effectiveness of these risk evaluation activities may be a matter of concern, the practitioner is strongly advised to choose a step by step, pyramid-shaped, approach (see figure 2), like the one suggested for health surveillance.

level 1 ... all workplaces
      ... simple and non-specialized assessment
level 4 ... a few tasks or workplaces
      ... complex assessment methodologies
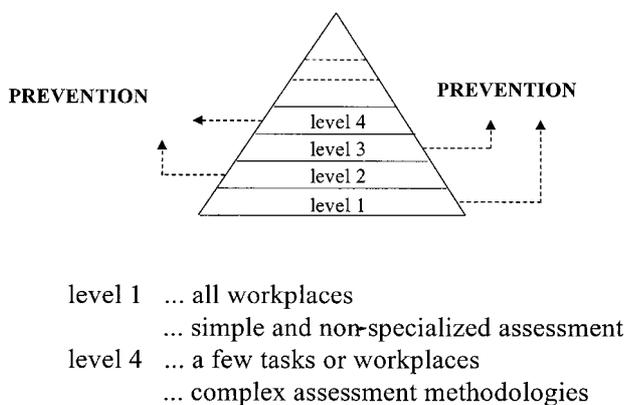
Figure 2. Hazard surveillance.

The first step (level 1) aims at the identification of workplaces or departments involving, with a high probability, a known hazard or risk factor (i.e. manual handling, repetitive movement). This assessment of risk sources would be better carried out by a production manager or technician who would collect the relevant data (process flow, handled tonnages, cycle time, nature of tools used) following the guidelines given by the OH professional.

The second step of the evaluation involves a quick ergonomic assessment for each workplace so identified, based on an observation of the job, but without specialized measurements. This assessment aims at the screening of any task inducing a hazard for the musculoskeletal system. In order to ensure a satisfactory comprehensiveness and reproducibility to this phase, several risk assessment guides or checklists are available for screening of the whole body, back, or upper limb WRMSDs. These evaluation guides allow a simplified risk estimate to be drawn (for instance, no risk, possible, or significant risk) and can easily be used by any practitioner having a basic training in ergonomics. When completed, this screening helps the practitioner in establishing a list of priority tasks that require the selection of redesign measures or, in some cases, further and more detailed analysis (level 3). From an ergonomics point of view, there is in fact no reason for going further than the level of analysis needed to identify the hazard source and to define practical redesign solutions.

When needed, the third step of the risk analysis is used to identify more precisely the risk factors in order to guide the search for corrective actions, and to establish a reference basis for later evaluation of the results obtained. More resources concerning the time and qualifications needed are often required at this stage of the analysis, but in practice the complexity of the methods to be used will vary from one situation to another. A postural evaluation using the OWAS method, an evaluation of a load-lifting task through the NIOSH method, or a detailed analysis of the working cycle do not require complex measurements and can thus be performed by any qualified OH professional. In other situations, cinematic variables or EMG data may have to be registered (level 4 of the analysis) by external experts. In practice, however, the need for such complex analyses remains rather infrequent, and more so when the risk assessment is considered not as a purpose in itself, but as the best means to guide preventive measures. In that respect, the implementation of an integrated and participatory approach provides a basic and complementary way towards the definition of solutions.

4. ERGONOMIC INTERVENTIONS AT THE WORKPLACE

How can the identified risks be lowered? The various preventive measures that can be undertaken on technical, behavioral, and organizational grounds have been described in several reference documents (Kilbom 1994; Hagberg *et al.* 1995) and the scientific knowledge that supports them, although far from being extensive and undisputed, may be considered sufficient to initiate actions. It is no doubt more important to discuss the most appropriate strategies to implement these ergonomic actions at the workplace.

4.1. Should One Favor "Micro-ergonomic" or "Macro-ergonomic" Actions?

A micro-ergonomic intervention focuses on decreasing, for instance,

the physical stresses in a high-risk job, but without acting "upstream" on the work organization. For various reasons, this is still the type of action that the OH practitioners most often carry out. Even though such corrective interventions may actually bring up positive results, their cost-effectiveness ratio is low and they have a limited influence, if any, on the organizational causes of the observed problems. From the OH professional point of view, however, they may be a necessary step in a building expertise process. Conversely, several reasons should lead the practitioner to integrate his prevention activities within a larger (macro-ergonomic) framework and to assume a prevention manager role within the company structures. A first reason relates to the multidisciplinary training of the OH professional and to his position as an independent adviser that could both help him in advising managers as well as workers about risks. A second reason originates in the rapid changes occurring in modern manufacturing: how could the practitioner be aware of the planned changes and anticipate their potential health effects when he is not part of the enterprise management? The third reason is likely to be the most essential: the prospects for success of any ergonomic intervention program in an enterprise are strongly enhanced when the program is fully integrated within the methods, resources, and structures of the enterprise (Wilson 1994).

4.2. OH Professional Strengths and Weaknesses

When considering the possible role of OH professionals in ergonomic interventions, some difficulties should not be overlooked. As mentioned earlier, an OH professional may obtain a valuable and unique knowledge of some features of the working conditions thanks to the trust relationship built with the workers and to the confidentiality warranty. The social status of the medical profession may also give some advantage when discussing the risks with the management. This same social status may, however, preclude an effective integration of the OH professional within the enterprise organization and culture — and this integration is a pre-requisite for intervention at the macro-ergonomic level. Another frequently encountered weakness concerns the OH professional's expertise in ergonomics and in intervention strategies that, on average, often remain insufficient.

4.3. Effect of New Trends in Production Management Schemes

It can be questioned, however, whether the starting up of ergonomics programs in workplaces is made easier or, conversely, more difficult by new trends in production management schemes. Flexibility and competitiveness requirements directly influence the workers exposure to risks; the increase in time pressure and the sharing of a given amount of work between fewer individuals can be viewed as two key explanatory factors for the observed rising trend in WRMSDs prevalence. In a more indirect way, musculoskeletal complaints may also be considered as an expression of the workers suffering in modern working environments. On the other hand, other management trends, especially the Total Quality Management concept, could provide a more favorable framework for taking health issues into account in production systems. This seems to be a real opportunity for the OH practitioner who wishes to implement an integrated prevention program (Hagberg 1995; Rahimi 1995).

4.4. What Are the Benefits from Using a Participatory Intervention Approach?

Participation means the involvement of people in planning and controlling their own work activities, with sufficient knowledge and power to influence the expected outcomes. The advantages of participation most often talked about are, among others: to make use in the design process of the workers knowledge on the actual work activity; to give a better chance of solution acceptance and effective use by the workers; to stimulate through the participatory approach itself worker involvement in a process of continuous improvement (Nagamachi and Imada 1995; Wilson 1995). Whereas participation has always been a key feature of ergonomics, the present renewed interest for this approach results from its increasing use (i.e. quality circles) in current businesses which invest heavily in the training and supervision of the participatory teams members. This trend is another example of a strategic opportunity upon which the OH professional could build preventive processes that would be part of the existing participatory structures.

4.5. Which Ergonomic Intervention Programs Are the More Successful ?

Several published studies have shown that a carefully planned intervention may result in a significant lowering of the level of physical risk factors, such as, force, posture and repetition (Chatterjee 1992; Keyserling *et al.* 1993; Wickström *et al.* 1993). It would be unrealistic, however, to expect a decrease in the risk level to have a systematic influence on the WRMSDs prevalence (Wickström *et al.* 1993), taking into account for instance the long-lasting exposure of middle-aged workers that often form a "survivor" group. When the intervention does result in a decrease of registered cases, or in sickness absenteeism, most authors avoid ascribing their success to the sole ergonomics improvements; quite often the intervention involves indeed other components (workers training, physical conditioning, participatory approach, etc.) or is accompanied by significant changes in work organization or in the company socio-economical environment. From a scientific point of view, interventions directed towards a single risk factor should thus be given preference. However, studies in "real-life" situations have shown that interventions which combine prevention measures of different character, such as equipment redesign *and* job enlargement, or ergonomic redesign *and* training in safe handling, prove to be more effective (Inserm 1995). On the basis of published studies, several guidelines can be suggested in order to set up preventive interventions:

- implementing the preventive actions within the company management scheme thanks to a clear and continuous commitment from the top management;
- stimulating risk identification and problem-solving procedures at the shop floor level using participatory techniques;
- combining individual-level and organizational-level prevention measures, and considering the risk factors within a system approach.

5. CONCLUSIONS

In a constantly changing economic world, the control of WRMSDs and other work-related health disorders poses a new challenge to the OH practitioner. In order to face it, should the OH

professional shift to another job and become an ergonomist? We do not think so, for his specific expertise is needed more than ever, especially when considering the worker health surveillance. Today's needs, however, do require that each professional casts a fresh look at his job. In this new vision, the OH professional has an important role to play in:

- hazard identification through passive health surveillance;
- risk estimation providing an appropriate training;
- initiation of participatory redesign approaches.

The OH professional will strive to fit his actions into the management policy of the company and will try to build solid bridges with other prevention professionals and especially professional ergonomists. Owing to this redefinition of his activities, goals and means, the OH professional could then keep on playing his fundamental role towards the promotion of the health of people at work.

REFERENCES

ANSI Z-365 draft Standard, 1995, *Control of Work-related Cumulative Trauma Disorders. Part 1: Upper Extremities*, American National Standard Institution.

CHATTERJEE, D.S., 1992, Workplace upper limb disorders: a prospective study with intervention. *Occup. Med.* 42, 129–36.

EEC, 1989, Council Directive (89/391) The introduction of measures to encourage improvements in the safety and health of workers at work. JOCE, no. L183.

HAGBERG, M., SILVERSTEIN, B., WELLS, R. *et al.,* 1995, *Work-related Musculoskeletal Disorders (WMSDs): A Reference Book for Prevention.* (London: Taylor and Francis).

INSERM (1995) *Spinal Disorders in the Occupational Setting: Which Ways to Prevention*? (Paris: INSERM), (in French).

KEYSERLING, W.M., BROUWER, M., and SILVERSTEIN, B.A., 1993, The effectiveness of a joint labor-management program in controlling awkward postures of the trunk, neck, and shoulders: Results of a field study. *Int. J. Ind. Erg.* 11, 51–65.

KILBOM, A., 1994, Repetitive work of the upper extremity: Part I - Guidelines for the practitioner. *Int. J. Ind. Erg.* 14, 51–7.

NAGAMACHI, M. and IMADA, A.S., 1995, Participatory ergonomics. *Int. J. Ind. Erg.* 15, 5, special issue.

Rahimi, M., 1995, Merging strategic safety, health and environment into total quality management. *Int. J. Ind. Erg.* 16, 83–94.

WICKSTRÖM, G., HYYTIÄINEN, K., LAINE, M. *et al.*, 1993, A five-year intervention study to reduce low back disorders in the metal industry. *Int. J. Ind. Erg.* 12, 25–33.

WILSON, J.R., 1994, Devolving ergonomics: the key to ergonomics management programmes. Ergonomics 37, 579–94.

Occupational Injuries and Medication Use

F. Baptista, P. Moro, B.–T. Karsh, F. Sainfort and M. J. Smith
Department of Industrial Engineering, University of Wisconsin-Madison, Madison, WI53706, USA

1. INTRODUCTION

Work injuries are due to multiple causal factors. These factors include management commitment to safety, the hazards in the physical environment, safety features of equipment design, the hazard aspects of the work, and the social and psychological environment at the work place. Individual health status and behavior also play an important role in injury risk. For example, disorders that affect coordination or cognitive processing can increase one's risk of injury, as can behaviors such as alcohol, drug and medication use. With so many factors leading to increased injury risk, it is not surprising that the numbers of occupational injuries are quite high.

2. INJURY STATISTICS

According to the International Labour Organization (ILO), ~250 million workers worldwide are injured annually on the job and ~335 000 workers die each year from occupational injuries. These work-related deaths and injuries have enormous costs. In the USA alone, it was estimated that in 1992 that the direct costs (e.g. medical, property damage) totaled $65 billion and the in-direct costs (e.g. lost earnings, workplace training and re-staffing, time delays) totaled $106 billion (Leigh *et al*. 1997). Those figures can be used to get a rough estimate of the total world cost of fatal and non-fatal occupational injuries in US$. Of the US$ figures presented, ~$230 million of the direct costs and $3.46 billion of the indirect costs were related to fatal occupational injuries. Non-fatal injuries accounted for $48.9 billion in direct costs and $92.7 billion in indirect costs (the rest was cost was due to death and morbidity from occupational illnesses). Those figures were based on 6500 occupational fatalities and 13.2 million non-fatal injuries.

Those figures, combined with the ILO global estimates of 335 000 fatal occupational injuries and 250 million non-fatal injuries, translate to $11.9 trillion in direct costs and $178 trillion in indirect costs from worldwide fatal occupational injuries and $927 trillion in direct costs and $1.8 quadrillion in indirect costs for worldwide non-fatal occupational injuries (and that does not include death and mortality from occupational diseases). Given this, it is important to understand the many factors that can cause occupational injuries so that industries can focus on prevention. One of these potential causes is the sedation, dizziness, or drowsiness that results from using many common medications.

3. IMPORTANCE OF MEDICATION USE FOR OCCUPATIONAL INJURIES

According to the 1998 World Health Report, the population of the world will increase by ~80 million people per year, resulting in a year 2025 population of 8 billion. By that time, the ratio of people > 65 to people < 20 of age will have grown to 31/100, when it was only 16/100 in 1995. That represents a growth in

Table 1. Examples of medications that can affect injury risk.

Indication	Medications (examples)	Clinical uses	Possible side-effects
Anti-anxiety	benzodiazepines	anxiety nervousness phobia panic	drowsiness loss of coordination fatigue mental slowing
Anti-manics	lithium carbamazepine (anti-convulsant)	bipolar disorders mania	drowsiness weakness fatigue hand tremor confusion
Anti-depressants	tricyclic compounds monoamine oxidase inhibitors selective serotonin re-uptake inhibitors	depression	drowsiness blurred vision dizziness
Anti-psychotics	phenothiazines thioxanthenes	schizophrenia	drowsiness dizziness rapid heart beat
Cardiovascular problems	angiotensin converting enzyme inhibitors anti-arrhythmics calcium channel blockers digitalis glycosides	heart disease hypertension angina cardiac arrhythmia congestive heart failure	loss of postural control tiredness dizziness
Pain	hypnotics (e.g. narcotic analgesics)	pain	drowsiness mental clouding sedation reduced alertness
Anti-histamine	diphenhydramine	allergic rhinitis common cold	drowsiness dizziness

the number of people > 65 year of age from 390 million in 1998 to ~800 million in 2025. Related to that, the numbers show that chronic disabling conditions such as coronary heart disease and poor mental health will continue to be major health problems for the entire population. About 15.3 million people die from circulatory diseases, of which ~7.2 million are from coronary heart disease. Also, the report projected that ~340 million people will suffer from depression. Many of these heart and mental conditions, as well as other common conditions, are treated with prescription and non-prescription medications than can cause sedation, dizziness, fatigue and other types of side-effects (table 1). With the aging of the workforce we expect an increase in the disorders associated with aging, and a subsequent increase in the use of medications.

4. MEDICATIONS THAT CAN AFFECT INJURY RISK

4.1. Anti-anxiety Medication

Anti-anxiety medications are used to treat generalized anxiety, panic, phobias, obsessive-compulsive disorders and post-traumatic stress disorders. The most popular, benzodiazepines (e.g. valium), represent the biggest share of the total world utilization of psychotropic drugs. Benzodiazepines may cause drowsiness, loss of coordination, fatigue, mental slowing and confusion. They have been shown to affect psychomotor and cognitive performance, and may specifically affect memory function.

4.2. Bipolar Disorder Medication

People suffering from bipolar disorders (manic-depressive illness) use anti-manic medication to treat the "highs" experienced. Lithium is most often used against these "highs." Lithium side-effects include drowsiness, weakness, nausea, vomiting, fatigue and hand tremor. Anti-convulsants are also sometimes prescribed for manic conditions, and may cause such side-effects as drowsiness, dizziness, confusion, disturbed vision, perceptual distortions, memory impairment, nausea and gastro-intestinal problems.

4.3. Anti-depressant Medication

Anti-depressants are typically suggested for people who are depressed for > 2 weeks and where the depression interferes with their ability to function normally. Tricyclic anti-depressants may cause side-effects such as drowsiness, anxiety, restlessness, blurred vision, dry mouth, constipation, weight gain, dizziness when changing positions, increased sweating, fatigue and weakness. Side-effects for other types of anti-depressants, such as monoamine oxidase inhibitors, include headache, insomnia, anxiety and dizziness when changing positions. Depressed patients taking anti-depressants have shown impaired information processing, learning, memory and tracking skills.

4.4. Anti-psychotic Medication

Anti-psychotics are used to treat people who have lost touch with reality, such as those suffering from disorders as severe as schizophrenia. This class of psychotropic drugs includes neuroleptics and major tranquilizers. Anti-psychotics may induce sedation, rapid heart beat, dizziness when changing position and can disrupt the capacity to focus attention. These effects typically disappear within days or weeks as tolerance builds up.

4.5. Cardiovascular Medication

Cardiovascular medication is used to treat problems like angina, hypertension and arrhythmia, etc. The link of cardiovascular medications to falls and injuries is not completely understood. Some cardiovascular medication (vasodilators, anti-hypertensives, diuretics, digitalis glycosides and some b-blockers) may affect a person's postural control or impair cerebral perfusion, which can lead to balance disorders. Calcium blockers are known to cause side-effects such as hypotonia, tiredness and dizziness, which can lead to injurious falls. Drugs that improve peripheral circulation cause vasodilatation of peripheral arteries. That may cause orthostatic hypotension and dizziness, thereby leading to falls.

4.6. Pain Medication

Analgesics (pain-relieving drug) come in two classes: anti-pyretic analgesic (aspirin) and narcotic analgesic (derived from opium, such as morphine and codeine). The former is used to relieve minor pain and reduce fever. There is no evidence that anti-pyretic analgesics affect performance, but the anti-arthritic type has side-effects such as dizziness and mental confusion. Narcotics do not produce motor impairment, but may produce drowsiness and mental clouding. Hypnotics produce sleepiness and reduce alertness. Patients using hypnotics often complain of reduced functioning in the morning following night-time use.

4.7. Anti-histamine Medication

Anti-histamines are the most popular forms of medication taken to prevent allergic rhinitis (i.e. hay fever) symptoms. Currently, there are two "generations" of anti-histamines in use. The first-generation anti-histamines differ from the second-generation in that the latter medications are considered to be non-sedating or less sedating, while the former cause sedation in 10–25% of users. Other possible side-effects of first-generation anti-histamines include, depending on the drug, blurred vision, gastro-intestinal problems, restlessness, nervousness and loss of appetite. first-generation anti-histamines may also cause dry mouth, and they are subject to tolerance development more so than are the second-generation anti-histamines. Second-generation anti-histamines may cause loss of appetite; however, they seem to be free from causing blurred vision, gastro-intestinal problems or dry mouth.

5. EVIDENCE OF INJURY RISK FROM MEDICATION

Studies have been conducted that have shown that certain medications can increase the risk of injury. Much of the research has focused on falls in the elderly population and non-work injuries among the general population. From those studies there is evidence that medications increase the risk of non-occupational injuries. Since members of the workforce use many of the same medications, one can extend those results to suggest that medication use would also be related to occupational injuries. In fact, the few studies examining occupational injuries confirm this. What follows is a summary of the evidence linking medication use to injuries. Quantitative ranges are provided that indicate the risk of injury associated with each type of medication that have been found in the different studies. The reason that ranges of risk exist for any given medication is because the various studies used an array of methodologies and often examined different types of injuries.

Research that has focussed on whether anti-anxiety medication affects the risk of injury has examined fall injuries in both the elderly and general population. It appears that people have 1.3–3.3 times the odds of being injured if they have taken anti-anxiety medication. Studies of anti-depressant medication have examined the risk of motor vehicle crashes in the elderly and of falls in the elderly and general population. The research consistently shows that people in the samples that used anti-depressants had 2.3–2.4 times the odds of injury compared with those that did not take anti-depressants. Scientific evidence also shows that nursing home residents and adults in the general population who have taken anti-psychotic medication have 1.3–2.0 times the risk of falling compared with people who had not taken such medication.

Cardiovascular and circulatory medications have been linked to the risk of falls in long-term care patients and to work injuries in male farmers. In the former group, the increased risk is in the range of 1.3–2.2, whereas farmers who used such medication were found to have 4.2 times the odds of becoming injured compared with farmers who did not use such medication. Studies of pain medication have investigated the risk of motor vehicle crashes in the elderly and of non-work injuries and falls in the general population. The results of these studies show that hypnotic or analgesic users have a 1.3–2.9-fold increase in the odds of getting injured or crashing. A similar increase in the odds of injury has been found for sedating anti-histamine users (i.e. 1.5–2.8). Other research has shown that antibiotics, stomach medication (e.g. laxatives), anti-inflammatory and anti-diabetic medication may also increase one's odds of injury. The relationships between medication use and injury risk may be affected by gender, age, medication dose, tolerance development and interactions with other medications.

6. IMPLICATIONS

In summary, there is empirical evidence that the medications reviewed can contribute to balance, psychomotor or cognitive impairment and, therefore, increase the risk of occupational injury. This is not to suggest that people should stop taking their medications to avoid getting injured. Instead, there needs to be increased awareness of the potential problems related to the use of certain medications. This awareness should lead employees and physicians to seek medications that have fewer problematic side-effects.

REFERENCES

CHIPMAN, M.L., 1995, Risk factors for injury: similarities and differences for traffic crashes and other causes. *Accident Analysis and Prevention*, 27, 699–706.

GILMORE, T.M., ALEXANDER, B.H., MUELLER, B.A. and RIVARA, F.P., 1996, Occupational injuries and medication use. *American Journal of Industrial Medicine*, 30, 234–9.

HERINGS, R.M., STRICKER, B.H., DE BOER, A. BAKKER, A. and STURMANS, F., 1995, Benzodiazepines and the risk of falling leading to femur fractures. *Archives of Internal Medicine*, 155, 1801–7.

INTERNATIONAL LABOUR ORGANIZATION [on-line], Occupational safety and health. http://www.ilo.org/public/english/90travai/sechyg/intro.htm

KOSKI, K., LUUKINEN, H., LAIPPALA, P. and KIVELA, S.-L., 1996, Physiological factors and medications as predictors of injurious falls by elderly people: a prospective population-based study. *Age and Aging*, 25, 29–38.

LEIGH, J.P., MARKOWITZ, S.B., FAHS, M., SHIN, C. and LANDRIGAN, P.J., 1997, Occupational injuries and illness in the United States. *Archives of Internal Medicine*, 157, 1557–68.

MYERS, A.H., BAKER, S.P., VAN NATTA, M.L., ABBEY, H.and ROBINSON, E.G., 1991, Risk factors associated with falls among elderly institutionalized persons. *American Journal of Epidemiology*, 133, 1179–90.

WORLD HEALTH ORGANIZATION [on-line], http://www.who.int

Occupational Injury

A.-M. Feyer

New Zealand Occupational and Environmental Health Research Center, University of Otago, P. O. Box 913, Dunedin, New Zealand

1. INTRODUCTION

Global estimates of the occurrence and cost of occupational injury are staggering. According to International Labour Office (ILO) estimates, ~250 million accidents resulting in injury occur annually in workplaces worldwide. Of these, 335 000 are fatal accidents, an estimated rate of 14 per 100 000 workers annually. Not surprisingly, the rates are different for individual countries and regions. The annual fatal occupational injury rate for established market economies has been estimated by the ILO as ~5/100 000 workers while the rate for the Asian region, where there has been a relatively rapid and unregulated pace of industrialization, is much higher, ~23/100 000 workers.

The cost to society of occupational injuries is also clear when one considers the problem relative to other health problems. Injuries in general are a leading cause of hospitalization and premature death (measured as potential years of life lost) in most countries. Occupational injuries represent a substantial part of the entire injury problem. In the USA, it has been estimated that one out of three non-fatal injuries and one out of six fatal injuries to working people occur at work (Baker *et al*. 1972).

The economic burden imposed by occupational injury on nations, companies and individual is high. In the USA, the total economic burden imposed by occupational injuries have been estimated at ~US$120 billion per year (National Safety Council 1996). In New Zealand, with a populations of ~3.5 million people, the current costs of occupational injury compensation claims alone (excluding occupational motor vehicle injuries), have been reported as being ~$700 million annually (Accident Compensation Corporation 1998).

Overall, it is fair to say that occupational injury is a major preventable health problem. Reducing the frequency and severity of occupational injury relies on knowing who gets injured, what happened, where it happened, when it happened and why it happened. This chapter deals with the definition of occupational injury, assessing the size, distribution and nature of the problem, and understanding its causes.

2. DEFINITION OF OCCUPATIONAL INJURY

The terms "accident" and "injury" have sometimes been used synonymously, particularly when considering prevention. However, they are in fact quite distinct. An accident can occur without injury, and not all injuries are "accidental", that is unintentional. For example, the third leading cause of fatal occupational injuries in the USA is homicide (Jenkins *et al*. 1993).

Injury is defined as tissue damage from transfer to individuals of one of the five forms of energy (kinetic or mechanical, thermal, chemical, electrical or radiation). By far the leading cause of injury is the transfer of mechanical energy. Negative energy, or the absence of energy, is also a possible source of damage, when the body's normal energy exchange is disrupted, for example, lack of oxygen resulting in drowning.

This definition is not as clear-cut as it might appear, however.

Exposure to energy (e.g. radiation) can be low level and gradual, resulting in disease (e.g. cellular change resulting in cancer), or it can be rapid and high level, resulting in injury (e.g. burns). In other words, damage as a result of energy transfer is best thought of as a continuum ranging from injury to disease. Of course, some occupational health outcomes are difficult to classify because they seem to fall somewhere between injury and disease, cumulative trauma disorders and low back pain being prominent examples.

At the international level, there is only partial agreement on the definition of what constitutes an occupational injury. The ILO defines occupational injuries as covering all injuries resulting from accidents arising out of or in the course of employment. There is little disagreement about the fact that an accident in the course of work that results in injury is a work-related event, an occupational injury. However, there is wide variation internationally about including other important possible categories of occupational injury. Specifically, these include traffic-related injuries (both where the vehicle is the workplace and when commuting to and from work), intentional injury, injuries to non-employed people undertaking work (e.g. family members involved in farming work) and injuries to bystanders (those injured in the course of someone else's work).

These differences in definition are not trivial, when the contribution of some of the contentious categories is considered. Motor vehicle related occupational injuries, for example, have been estimated as being around one-fifth to one-quarter of all occupational injury deaths. Many of these injuries will occur where the vehicle is the workplace, e.g. truck-driving, or when the injury is sustained in a traffic accident during the course of paid employment, where the worker is carrying out work on his employer's behalf e.g. obtaining supplies. There would be little dissension about these being occupational injuries. There is less consensus about commuting injuries, where the sole purpose of travel is work-related. Commuting injuries are defined as injuries sustained on the way to or from work, when the worker is travelling between work and (1) a second job, (2) where meals are taken or (3) principal place of residence. Yet commuting injuries would certainly appear to be covered by the spirit of the ILO definition of occupational injuries described above. Moreover, indications are that commuting injuries may represent a sizable part of the whole problem. In Australia, 17% of all fatal occupational injuries occurred while commuting to and from work (National Occupational Health and Safety Commission 1998).

Another important but contentious category relates to injuries as a result of intentional harm. Again, it can be an important part of the problem. As noted above, in the USA, homicide is the third leading cause of fatal occupational injuries.

3. ASSESSMENT OF THE SIZE, DISTRIBUTION AND NATURE OF THE PROBLEM

Despite definitional issues, and the difficulties that they present for international comparisons, there is agreement that information about occupational injury should be collected at national and jurisdictional level. Most countries attempt to assess the size, distribution and nature of their occupational injury problem.

3.1. Surveillance of Occupational Injury

Surveillance is the ongoing systematic collection, analysis and

interpretation of health data in the process of describing and monitoring a health event. Accurate and reliable data concerning occupational injury are central to prevention efforts: they provide an evidence-based approach to priority setting, they provide the essential starting point for development of new injury prevention strategies and they provide information for evaluation of existing efforts.

In principle, injury surveillance data can provide a view of the distribution of cases of injury and of the types of injuries incurred and their severity for different occupations and economic activities. From such information priority areas can be determined in terms of areas of greatest risk of the most serious types of injuries, or of areas where the probability of injury is high, or of areas where large groups of people are exposed to risks, or some combination of these. Areas of greatest risk can thus be targeted by prevention efforts, safety campaigns, enforcement activities and new legislative initiatives. In practice, the level of specificity of the information collected in most places undermines the potential of the data. Recommendations for the minimum data set to be collected from the ILO are as follows:

- information about the employer (geographic location; economic activity; size of the establishment);
- information about the person injured (gender; age; occupation; status in employment);
- information about the injury (whether fatal or not; if not fatal, amount of work time lost; nature of the injury; bodily location of the injury); and
- information about the events leading to injury (geographic location; date and time; action(s) leading to the injury; agency [object, substance or physical condition] involved in the incident and injury).

Ideally, injury data should cover all occupational injuries incurring loss of working time. In principle, data sources for occupational injury surveillance include health care providers (hospitals and physicians), death certificates, medical examiner/coroner files, employer-based reporting to relevant government departments e.g. departments of labor, workers' compensation/ accident compensation agencies, periodic surveys of households and even individual corporate records. In practice, although law requires many of these records, they still provide only incomplete information in most countries. For example, examination of coverage of working deaths in Australia revealed that 36% of the total were covered by occupational health and safety agencies, that 57% of the total were covered by compensation agencies, and that one third of working deaths were not covered by either mechanism (National Occupational Health and Safety Commission 1998).

Lack of coverage stems from various sources. Even in the best of systems, it is recognized that there are incentives to under-report. Besides under-reporting, certain types of workers may fall outside the requirements for notification, such as those working at home, part-time workers, trainees, those outside working age. Compensation schemes also tend to be focused on employees, and tend to only partly cover certain economic activities, for instance agriculture and small business establishments.

3.2. Outcomes of Interest

From the point of view of surveillance, both the occurrence of injury and severity of injury are important outcomes. Injury severity involves a number of quantifiable dimensions. These include anatomical (the amount and nature of tissue damage), physiological (vital signs to determine how close to death the injured person is), disability, impairment of quality of life and direct and indirect economic costs. All of these outcome measures provide different parts of the whole estimate of the size and nature of the injury problem. Perhaps the most used, but possibly the least useful severity dimension is days lost from work following injury. Lost work time is often difficult to interpret because it usually involves a combination of a whole range of injury and non-injury variables. These can, but need not, include such factors as anatomical severity, extent of disability, job demands, availability of alternative duties, workplace policies, job satisfaction and so forth. Amount of lost time alone will not identify which factors are involved, making implications for prevention difficult to pinpoint.

3.3. Comparative Measures

Sound comparisons between periods, industries and countries can only be made if the injury data are considered in the context of the total number of people at risk. Injury data should be expressed as a rate, relative to the number of persons in various relevant reference groups. There are several reference contexts that provide meaningful relative measures. The most common reference data are the total employed labor force and the total number of hours worked by the labor force (to account for the mix of full- and part-time work) or the same sort of data for specific subsegments of the working population (e.g. different occupations). The important feature of obtaining such relative measures is that the numerator (injured persons) and the denominator (reference data of exposed persons) should have the same coverage. An illustration of the importance of this requirement is the calculation of the relative risk of injury for those who work at night compared with those who work during the day. Far fewer workers are at work at night and the calculation of rate of injury should reflect this.

3.4. Using Occupational Injury Data

Despite the shortcomings of surveillance and occupational injury data around the world, distinct patterns of injury risks have been identified. Occupational injury like other health outcomes have distinct patterns of risk that vary by age, gender, race, geographic region, industry and occupation. In Western industrialized countries, those working in agriculture (including hunting, fishing and forestry), mining and quarrying, construction, transport and manufacturing have consistently been shown to be at highest risk of injury. Overall, male workers are at greater risk of occupational injury. Younger workers (< 25 years) have typically been shown to be at greater risk of non-fatal injury, while older workers (> 65 years) have been shown to be at greater risk of fatal injury.

Surveillance data provide important starting points for injury prevention: they identify in general terms who gets injured and where, when, and how people are injured. However, they are limited in providing information about why people get injured, that is, in identifying the causes, processes and mechanisms leading to the injury event. Knowledge of causation is critical to identifying ways of reducing and preventing injury events.

Different models of injury causation suggest different targets and strategies for prevention of injury.

4. MODELS OF OCCUPATIONAL INJURY CAUSATION

While many causal models have been proposed, in general terms there have been three key shifts in emphasis.

4.1. Person-based Single Cause Approaches

In the 1940s, the dominant model proposed the notion of fault on the part of workers precipitating either unsafe acts and/or (less commonly) unsafe conditions. The quintessential example of this type of approach is the Domino theory proposed by Heinrich (1941). The approach was weak because it focused on immediate precipitating causes, assumed a single cause and equated attribution of fault and identification of cause. It provided a fairly narrow view that restricted options for preventive efforts.

4.2. Epidemiological Approaches

The epidemiological approach, prominent in the last two to three decades, takes a multifactorial view of injury causes and prevention options. Injury is seen as being caused by the impact of four factors:

- the host: the injured person;
- the agents of injury: the various forms of energy which, when transferred to the host in sufficiently high levels result in jury;
- the vectors/vehicles of injury: the environmental mechanisms/factors which convey damaging energy to the host; and
- the environment: the physical and social environment in which the transfer of energy occurs.

The approach also advocates that there should be systematic consideration of all possible prevention options: those that reduce injury occurrence, those that reduce injury severity and those that reduce disability resulting from injury. The choice of countermeasures should not be determined by the relative importance of causal factors; rather priority should be given to options that reduce injury or its consequences. The focus on injury control very much reflects the public health model, historically successful in disease control, from which this approach emerged.

One of the most influential tools for analyzing injury risk factors multifactorially and guiding the choice of injury control strategies was that proposed by Haddon (1968). In fact, this work is to a large extent synonymous with the epidemiological approach. The Haddon matrix has two dimensions: one dimension reflects the epidemiological model of causation and divides causal factors into host, vehicle/vector and environment; the other dimension divides the injury event into three phases, the "pre-event", the "event" and the "post-event" phases. Pre-event factors are ones that influence the risk that an event leading to injury will take place (e.g. poor machine maintenance, a vehicle/vector factor). Event factors affect the risk of injury once an incident has occurred (e.g. lack of personal protective equipment, a host factor). Post-event factors influence the risk of disability and death, once an injury has occurred (e.g. availability of emergency services, a social environment factor).

This type of injury causation model is still the cornerstone of public health approaches to occupational injury prevention today. The analysis does not center on causation *per se,* but on ways of reducing undesirable end results. Thus, the use of machine guards and electrical shielding can prevent injuries regardless of the reasons people come into contact with these potentially harmful agents. The strength of the approach lie in the fact that it highlights effective injury control strategies. Its weakness lies in the fact that it provides only limited insight concerning the process or mechanisms leading up to the injury event. Thus while the approach has identified many effective strategies for controlling injury, the approach has been less useful in identifying ways of actually preventing the process leading to harm. Safety professionals generally agree that as well as controlling injury, more needs to be done to prevent the undesired events and incidents from happening in the first place. This may be the main prevention option available in situations where the hazard actually comes from the interaction of the worker and their work environment, rather than any inherent hazard associated with one or the other, or both.

4.3. The Systems-based Approach

The other recent development in injury causation models has been the use of a systems-based approach to occupational safety (e.g. Feyer and Williamson 1991, Kjellen 1998). A system in this regard can be seen as a set of independent components combined in such a way as to perform a given function under specified conditions. In this light, work and workplaces are dynamic interactive systems with workers, work tasks, work equipment and work environment (physical and organizational) all being part of the system. A major advantage of such models over previous approaches is that they focus on dynamic interactions between injury causal factors and they are therefore better able to identify processes and mechanisms leading up to injury events.

The conceptualization of a work system (with an emphasis on interaction between its components) is characteristic of a human factors approach to injury causation, and has been much more influential than the injury control approach in identifying ways of actually preventing the process leading to harm (e.g. Feyer and Williamson 1998). This approach does not replace the epidemiological injury control model, rather it compliments it. Perhaps its greatest strength has been that it has provided systematic ways of considering the human element in injury causation. The human element covers a wide range of factors involved in injury causation, some directly and some indirectly. These factors range from individual errors immediately precipitating injury events to human factors involved in a broader sense, taking the form of flawed standard operating procedures, management decisions and organizational culture and climate. Elaboration of the involvement of human factors other than human error in the immediate circumstances precipitating an injury represents a major advance in understanding why injury events actually happen.

Formal consideration of models of human error represents another advance through this approach. The critical role of human error has long been universally acknowledged, yet understanding of it in the context of injury causation has been slow to progress beyond documenting its ubiquitous presence. The systems-based

approach has contributed improved understanding of the nature, timing, causes and role of human error to models of injury causation (e.g. Feyer and Williamson 1998).

5. CONCLUSION

Occupational injuries and injury deaths are not random events but result from cause and effect relationships and are therefore predictable and preventable. In order to predict and prevent injuries, however, we need to systematically identify their occurrence, nature, characteristics and causes. There has been considerable refinement of the ways in which we define injury, the ways in which we describe who gets injured and how they get injured, and the approaches we use to better understand why injuries occur. In the future, it is clear that injury prevention and control will depend critically on the integration of the best that a range of disciplines has to offer: human factors and ergonomics, public health and epidemiology, industrial hygiene and physics. Collaboration between these disciplines will bring new insight to our understanding of injury causation and prevention.

REFERENCES

ACCIDENT REHABILITATION AND COMPENSATION CORPORATION, 1998, *Injury Statistics 1998* (Wellington: ACC).

BAKER, S.P., O'NEILL, B., GINSBURG, M.J. and LI, G., 1972, *Injury Fact Book* (New York: Oxford University Press).

FEYER, A.-M. and WILLIAMSON, A.M., 1991, An accident classification system for use in prevention. *Scandinavian Journal of Work Environment and Health*, 17, 302–11.

FEYER, A.-M. and WILLIAMSON, A.M., 1998, Human factors in accident modelling. In J.M. Stellman (ed.) *Encyclopaedia of Occupational Health and Safety* (4th ed.) (Geneva: International Labour Office), 56.8–56.10.

HADDON, W., 1968, The changing approach to the epidemiology, prevention and amelioration of trauma: the transition to approaches aetiologically rather than descriptively based. *American Journal of Public Health*, 58, 1431–8.

HEINRICH, H.W., 1941, *Industrial Accident Prevention* (2nd ed.) (New York: McGraw Hill).

JENKINS, E.L., KIRSNER, S.M., FOSBROKE, D., LAYNE, L.A., STOUT, N.A., CASTILLO, D.N., CUTLIP, P.M. and CIANFROCCO, R., 1993, *Fatal Injuries to Workers in the United States, 1980–1989: A Decade of Surveillance* (Cincinnati: NIOSH).

KJELLEN, U., 1998, Accident deviation models. In J.M. Stellman (ed.) *Encyclopaedia of Occupational Health and Safety* (4th ed.) (Geneva: International Labour Office), 56.20–56.23.

NATIONAL OCCUPATIONAL HEALTH AND SAFETY COMMISSION, 1998, *Work-related Traumatic Fatalities in Australia, 1989 to 1992* (Canberra: Commonwealth of Australia).

NATIONAL SAFETY COUNCIL, 1996, *Accident Facts*, *1996 Edition* (Itasca: National Safety Council).

Occupational Injury Epidemiology; Principles and Approaches to Prevention

G. S. Sorock

Liberty Mutual Research Center for Safety and Health, Hopkinton, MA 01748, USA

and

T. K. Courtney

Harvard School of Public Health, Boston, MA 02115, USA

1. INTRODUCTION

Epidemiology is a set of established and developing methodologies used to describe the distribution and determinants of adverse health outcomes in human populations. The subsequent evaluation of interventions to prevent injury or disease is the ultimate objective of epidemiology.

2. FUNDAMENTALS OF EPIDEMIOLOGY

2.1. Fundamental Aspects of Epidemiologic Methods

People who develop a health condition tend to be exposed to antecedent risk factor(s) more than persons who do not develop the condition. Epidemiologic methods have evolved to approximate the effects of potential risk factors (also known as exposures) on health outcomes (such as injuries, disorders, diseases) in groups of individuals. Associations between exposures and outcomes may be causal or due to alternative explanations such as study bias, confounding or chance.

In epidemiology, study bias refers to some systematic error in study methods that precludes a clear judgment as to the nature of the association between the exposure and the disease outcome. For example, parents of children injured at home may recall certain exposures more thoroughly than parents of uninjured children. This *recall* bias may be estimated by comparing questionnaire responses to a "gold standard" source of information, perhaps a medical record or a direct observation of an exposure.

Confounding refers to the effect of an intermediary factor that is associated with both the outcome and the exposure. That is, the exposure–disease association may be artificially produced or obscured by this other factor. For example, any observed association between alcohol use and lung cancer is potentially confounded by cigarette smoking because many alcohol users also smoke.

Chance or sampling variability operates in most scientific research. It is a potential explanation for an observed association between an exposure and a disease outcome. It may be less likely, however, with larger sample sizes and as a study's findings are replicated by subsequent investigations.

2.2. Measures of Outcome Frequency

Important terms in interpreting and reporting outcomes in populations include incidence and prevalence. Incidence refers to the number of new instances of a health outcome in a population during a specified time period (Last 1988). For example, an *incidence rate* is defined as the number of new cases of a condition per persons at risk (usually 100 or 1000) per unit of time (usually one year). Prevalence is the count of all existing cases during a specified time period or point in time divided by total persons at risk. When not otherwise indicated, the prevalence usually reported is the *point prevalence* or the count of all cases in existence at one point in time.

2.3. Common Measures of Exposure–outcome Association

Epidemiologic data are often summarized in tabular form in the initial stages of analysis as the presence or absence of exposure to some potential risk factor and the presence or absence of a health outcome in a group of individuals. Table 1 presents the results of a hypothetical study of the effects of a single exposure on a health outcome. Each cell (a, b, c, d) contains the number of persons with a particular combination of exposure and outcome status. The data in these four cells are used to derive various measures of effect. Two common measures of the effect of exposure on outcome are relative risk (RR) and odds ratio (OR). These are calculated and interpreted below.

If 1000 persons *with* and 1000 persons *without exposure* are followed-up over a specified time period for outcome occurrence, the relative risk (RR) is found by dividing the risk in the exposed group (75/1000) by the risk in the unexposed group (25/1000). The result is 0.075/0.025 or 3.0. This means that this sample of exposed persons had three times *the risk or probability of developing the outcome* than unexposed persons over the specified time period.

If individuals are *selected on the basis of their outcome status*, the totals for the rows (a + b) and (c + d) are not predetermined. The relative risk can be estimated by calculating the ratio of the odds of exposure among cases (75/25) to that among controls (925/975) or ad/bc = 75 x 975 / 25 x 925. The resulting odds ratio (OR) of 3.2 is close to the RR of 3.0 found above. The OR is interpreted to mean that the cases are 3.2 times more likely to have been exposed than non-cases.

Any estimate of effect varies within a range of values that is termed a confidence interval (CI). The CI depends on the magnitude of the estimate, the specified probability of including

Table 1. Data from hypothetical epidemiologic study.

the true value of the estimate (usually 95%) and the sample size in each of the four cells of the table (the larger the sample, the smaller the range of the interval). For example, the 95% confidence interval for the calculated OR from the study above ranges from 2.1 to 4.9 (for details of CI calculation and underlying assumptions, see Hennekens and Buring 1987: 255). This means that if the same study was repeated 100 times, 95 out of 100 times the true OR would lie between 2.1 and 4.9. If an OR = 1.0, then there is no observed effect of the exposure on outcome status. Because the interval calculated previously did not include the null value (1.0), the CI indicates that the association is statistically significant. If the 95% CI covered 0.9–7.1 the result would not be statistically conclusive but strongly suggestive of an effect since most of the interval is above 1.0. If the point estimate and the interval are below 1.0, (e.g. 0.3, 0.1–0.5), then the exposure significantly protects against the outcome.

2.4. Epidemiologic Study Designs

The techniques illustrated above are applied to the study of exposure and outcome associations. As with most scientific inquiry, however, the techniques are merely assets supporting the execution of a study design. There are two main classes of study design in epidemiology: descriptive and analytic. Descriptive epidemiology describes *how* an outcome such as an injury or disorder is distributed in a population; analytic epidemiology tries to explain *why* (Kelsey *et al.* 1989). A brief description of each design, with examples, follows. More specific information about epidemiologic study methods can be found in standard introductory texts (Kelsey *et al.* 1986, Hennekens and Buring 1987).

2.4.1 Descriptive studies
2.4.1.1 *Case report and case series*
The case report is the complete description of the clinical characteristics of a single patient. The single patient may present some unusual finding that can suggest the need for a larger study of an exposure-outcome relation. Case series describe the clinical characteristics of two or more patients with a health outcome and perhaps an exposure in common. An advantage of the case series is that multiple observations allow compilation of potential trends in the causes, diagnosis, and treatment of the disorder.

These types of studies are most useful for identifying and describing emerging outcome trends and for generating hypotheses concerning potential exposure–outcome relations which may be tested in more analytic studies. However, the conclusions that can be reached regarding the existence and magnitude of such relations are limited with these designs due primarily to the lack of both exposure data and a comparison group.

2.4.1.2 *Cross-sectional*
Cross-sectional studies take a simultaneous look at the presence or absence of both exposure and disease in individuals. Such studies ask the question: What is the relation between exposure and disease at the same point in time? For example, a study of sedentary workers and low back pain examined the relationship between constrained, sedentary work, and prevalence of low back pain (Burdorf *et al.* 1993). Occupations with the greatest proportion of time spent in non-neutral trunk postures were observed to have a significantly increased prevalence of low back pain. In a cross-sectional design, it is difficult to establish whether

the potential exposure variable preceded the outcome (e.g. work posture differences contributed to the development of low back pain) or whether the potential exposure variable exists as a result of the outcome (e.g. workers differ in posture as an adaptation to low back pain). Hence, cross-sectional studies are useful for identifying potential exposure-outcome associations but not for establishing causality.

2.4.2 *Analytic studies*
Case-control, cohort, and intervention studies are generally classified as analytic studies. These study designs make use of the temporal relation between exposure and outcome (i.e. exposure precedes the outcome) and identify potential causes of outcome more definitively than descriptive studies.

2.4.2.1 *Case control*
In a case-control study, people with the outcome (cases) and people without the outcome (controls) are compared with respect to the proportion in each group with the historical exposure of interest. An advantage of this design is that both cases and controls may be matched on potential confounding variables, such as age. This design is particularly useful for the study of rare health outcomes. A limitation of this design is its susceptibility to recall and other forms of information bias since the exposure must typically be recalled by the cases and controls or be present in recorded data such as hospital records.

2.4.2.2 *Cohort*
A cohort study refers to a study design in which a group of persons who share a common exposure are studied over a period of time. Cohort studies are distinguished from case-control studies by two main features. First, classification into comparison groups is based on the exposure factor not the outcome. Second, the cohort study looks from the exposure forward rather than from the disease backward in time.

There are two main types of cohort studies: prospective and retrospective. The feature that distinguishes a prospective from a retrospective cohort study is whether the outcome of interest has occurred when the investigator initiates the study. In a prospective cohort study, the outcome occurs after the exposure is measured. In a retrospective cohort study, the investigation is initiated after both the exposure and the outcome have occurred.

2.4.2.3 *Intervention*
An intervention design is a prospective cohort study which investigates the impact of a treatment or control action on the exposure-outcome association. Study participants are assigned the treatment (intervention) at random and then followed to determine the proportion of persons exposed who develop the outcome compared to the proportion of persons unexposed (non-intervention). Assignment to an exposure category at random, also achieves, on average, equal distribution of other known and unknown confounding variables.

The study designs just discussed are graphically presented in Figure 1 to demonstrate how increasing knowledge of an exposure–disease relationship might develop for a hypothetical disorder (D) with a given exposure (E). Designs that accommodate the temporal relationship between an antecedent exposure and the outcome are considered most convincing of the strength or weakness of the E–D relationship. Where alternative explanations
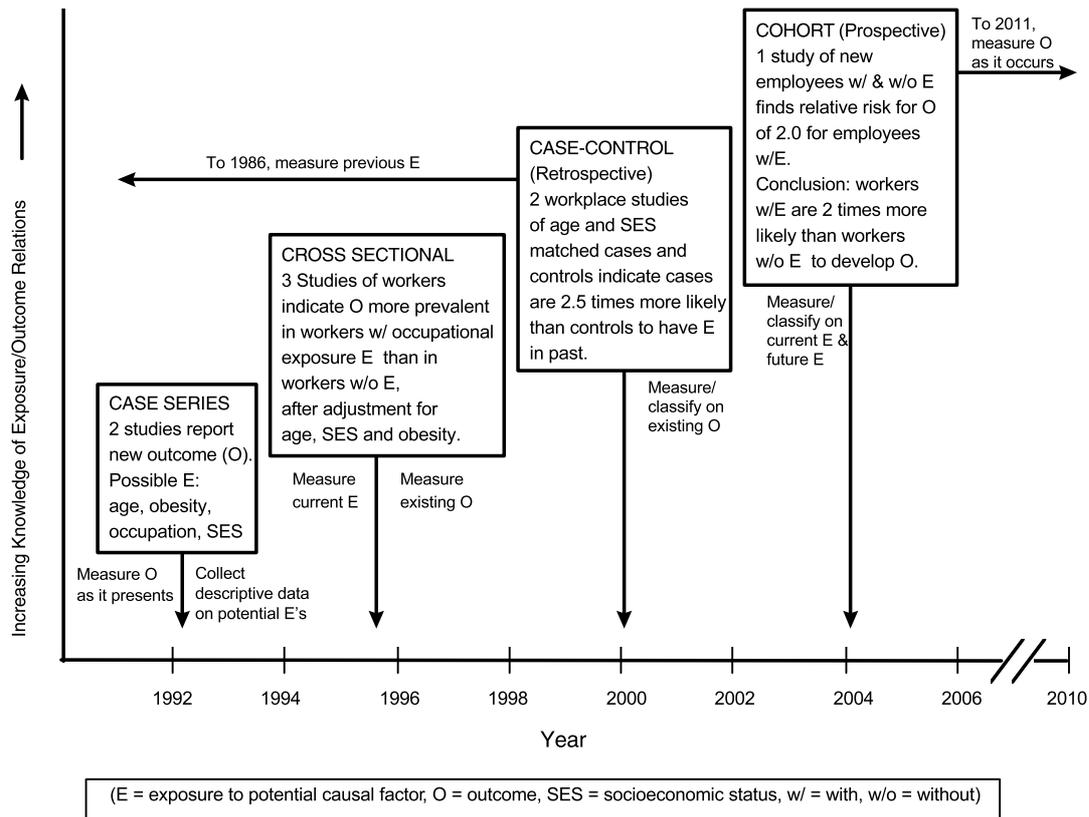
Figure 1. Possible progression of epidemiologic research for a given disorder or injury illustrating the temporal nature of each fundamental study design.

for this relation can be ruled out, for example by statistical adjustment of potential confounding variables, the results are more convincing.

3. APPLICATION OF EPIDEMIOLOGY TO WORKPLACE INJURY

Epidemiologic methods can be applied to the evaluation and prevention of injury in the workplace. Ultimately any attempt to prevent or reduce the impact of injury in a population, whether a model for risk evaluation or an engineering design intervention, can only be validated using epidemiologic methods. However, epidemiologic methods are primarily useful for existing or emerging injury problems that have sufficient numbers of cases to support analysis. Therefore epidemiology is inherently reactive, identifying its solutions a posteriori.

Injury has been defined as "any damage inflicted to the body by energy transfer during work with a short duration between exposure and the health event (usually less than 48 hours)." (Hagberg *et al.* 1997). Work-related musculoskeletal disorders with longer time periods between exposure and injury onset (i.e. gradual onset) may also be addressed epidemiologically.

3.1. Evaluation of Injury at Work by Surveillance (Passive or Active)

Injury prevention at the workplace requires surveillance for injuries among workers, both before and after any intervention effort to prevent injuries. Surveillance has been defined as "the ongoing and systematic analysis and interpretation of health data

in the process of describing and monitoring a health event" (CDC 1988). Workplaces may then use this information to plan, implement and evaluate the health effects if any from interventions or changes in the workplace.

There are two essential elements of any surveillance system. The first is a clear definition of the injury cases to be counted also known as the case definition. Very often in practice, existing data sources such as dispensary records, required injury logs (e.g. OSHA log) or cases receiving wage replacement or sickness/absence benefit provide initial case definitions (e.g. counting all hand injuries treated in the plant clinic). This particular approach is referred to as passive surveillance since the data already exist. A limitation of this approach is that one has to accept the data elements and quality offered by the existing system.

An alternative which can improve the upon the performance of passive surveillance is active surveillance. In this instance new information is solicited from the work force specifically for the purpose at hand. Instruments typically used include discomfort or symptom surveys and risk factor checklists or other exposure evaluation tools. This approach offers improved information and control at the cost of resources to actively collect the information.

Once the case definition of interest is established it must be adhered to by all reporting sites. This will necessarily entail the support of the system with administrative and management resources including training in the appropriate use of the system by those entering and interpreting the data.

The second essential element is a count of the number of work hours in the population exposed to relevant work conditions

in a particular plant, department, assembly line, or work station. These two elements, cases divided by work hours constitute the rate of occurrence of the injury event which is the primary measure of effect of any intervention. The rates may be subdivided for categories such as length of employment, workstation location, shift schedule, etc.

3.2. Surveillance System Applications at Work

Consider a company with multiple locations producing similar products. Substantial variation in injury rates between these plants would suggest that injury circumstances, prevention efforts, or injury hazards were somehow distributed differently between plants. Only by establishing comparable and consistent surveillance systems can these ''natural experiments'' be exploited. Similarly, if an injury intervention being installed company-wide can only be put in a few plants at a time, injury rate comparisons between plants with the intervention compared with plants that are designated to receive it at a later date can contribute to evaluating the intervention's efficacy.

For example, a large manufacturing company developed a new employee-conditioning program in one plant and wanted the first six months of experience with it evaluated before duplicating the program company-wide. Back and shoulder injury cases were counted six months before the program was implemented and then six months while the program was in operation. The rate for cases with days away from work increased from 14 to 22 per 200 000 hours worked. The number of cases of back and shoulder musculoskeletal disorder increased from 4 prior to the program's start to 9 during the first six months of its operation. This was thought to be a large enough increase in cases to warrant a major modification of the conditioning program (Sorock *et al.* 1997).

3.3. Use of Narrative Data from Injury Records to Improve Injury Surveillance

Many injury surveillance systems include a brief narrative section describing in a few sentences how the ''accident'' happened. In most databases, only a single code is assigned to each narrative that best describes the ''causes'' of the ''accident.'' The narratives themselves can be searched electronically and certain keywords selected that can be used to identify cases not found by searching for codes. Narrative analysis has been used, for example, to identify motor vehicle crashes that occur in roadway construction work zones in computerized files. By selecting records with the keyword "construction" in the narrative of how the crash occurred, a descriptive analysis of 3686 such crashes was conducted (Sorock 1996).

3.4. Recommendations for Workplace Injury Surveillance

i Generally the utility of a surveillance system will improve as case definitions are refined and as progressively less severe and perhaps even near-miss types of events are cataloged. However, it is important to note that systems offering very specific levels of detail, hence a greater number of potential categories or descriptors can be compromised by efficiencies in use. For example, if the system allows a coder to select from 500 outcome codes, the coder will typically habituate to assigning the vast majority of incoming cases into a

significantly smaller subset of more frequently used codes (see entries on human information processing for more on this topic). Hence, a trade-off between specificity and efficiency exists in surveillance systems. The appropriate level of detail is best selected by examining the goals of the system, the skills and training of coders, the anticipated management of the coding system, and the level of performance in detail of any existing systems which the new system is to complement or improve.

ii Management and supervision should be involved in all efforts with respect to surveillance and injury control. This may be optimized through awareness training for management and supervision.

iii Records for surveillance should include a brief narrative of how the injury occurred with special note to contributing exposures (e.g. equipment involved).

iv surveillance systems should record demographic information such as location, shift, age of worker, gender, job class and occupation.

v The system should incorporate where possible the costs associated with injuries to permit cost analysis.

vi Organizations should computerize their surveillance systems and ensure sufficient storage space for archived data. (Sorock *et al.* 1997).

3.5. Occupational Injury Intervention Studies

Workplace injury prevention efforts can be categorized as either engineering interventions (targeting the design of the job, tools, and workspaces); administrative interventions (focusing on work procedures and policies); personal interventions (addressing worker behavior, training and education); or a combination of strategies (multiple interventions) (Zwerling *et al.* 1997). In general, worker behavior, training, and education interventions alone have not been demonstrated to be as successful in injury prevention as have other approaches. Multiple intervention approaches, which seek to optimize the management of an injury problem in each major aspect, appear to be the most effective (McGorry and Courtney 1995).

REFERENCES

BURDORF, A., NAAKTGEBOREN, B., and DE GROOT, H.C.W.M., 1993, Occupational risk factors for low back pain among sedentary workers, *Journal of Occupational Medicine*, 35, 1213–20.

Centers for Disease Control, 1988, Guidelines for evaluating surveillance systems, *Morbidity Mortality Weekly Report*, 37(S-5), 1–18.

HAGBERG, M., CHRISTIANI, D., COURTNEY, T.K., HALPERIN, W., LEAMON, T.B., and SMITH, T.J., 1997, Conceptual and definitional issues in occupational injury epidemiology, *American Journal of Industrial Medicine,* 32,106–15.

HENNEKENS, C.H. and BURING, J.E., 1987, *Epidemiology in Medicine* (Boston: Little Brown).

KELSEY, J.L., HUEBNER, W.W., and BODIAN, C.A., 1989, Statistical and epidemiologic concepts for the reluctant reader. Part II: basic epidemiologic concepts. *Surgical Pathology*, 2, 281–91.

KELSEY, J.L., THOMPSON, W.D., and EVANS, A.S., 1986, *Methods in Observational Epidemiology*, (New York: Oxford University Press).

LAST, J.M. (ed.), 1988, *A Dictionary of Epidemiology,* 2nd ed. (New York: Oxford University Press).

McGORRY, R.W. and COURTNEY, T.K., 1995, Exercise and cumulative trauma disorders: the jury is still out. *Professional Safety*, 40(6), 22–5.

MURPHY, P.L., SOROCK, G.S., COURTNEY, T.K., WEBSTER, B.A., and LEAMON, T.B., 1996, Injury and illness in the American workplace: a comparison of data sources. *American Journal of Industrial Medicine*, 30, 30–41.

SOROCK, G.S. and COURTNEY, T.K. 1996, Epidemiologic concerns for ergonomists: Illustrations from the musculoskeletal disorder literature," *Ergonomics*, 39, 562-578.

SOROCK, G.S., RANNEY, T., and LEHTO, M., 1996, Motor vehicle crashes in roadway construction work zones: An analysis using narrative text from insurance claims. *Accid. Annal. Prev.*, 28, 131–8.

SOROCK, G.S., SMITH, G., REEVE, G., DEMENT, J., STOUT, N., LAYNE, L., and PASTULA, S., 1997, Three perspectives on work-related injury surveillance systems. *American Journal of Industrial Medicine*, 32, 116–28.

ZWERLING, C., DALTROY, L., FINE, L., JOHNSTON, J., MELIUS, J., and SILVERSTEIN, B., 1997, Design and conduct of occupational injury intervention studies: a review of evaluation strategies. *American Journal of Industrial Medicine*, 32, 164–79.

P$_5$–P$_{95}$ Syndrome

B.J. Daams

Delft University of Technology, Department of Industrial Design Engineering, Jaffalaan 9, 2628 BX Delft, The Netherlands

Presently: Daams Ergonomie, Mauvezand 17, 1251 JC Laren, The Netherlands

When designers design consumer products, they have to make many choices. One of these is the assessment of the target user group. Designing for literally everyone, or 100% or the target users, may result in a very expensive product or a product that suits nobody really well. When designing a product or a situation, the exclusion of extreme users (extreme tall, extreme slow, etc.) is therefore unavoidable. Designing products for "weak" or "strong" (tall or short, quick or slow, etc.) users is not precise enough an indication. In addition to deciding whether the weakest or the strongest users are the most relevant for specific design purposes, the designer has to determine which percentage of users within the expected population the specific design aspect will take into account.

In research and literature we are talking percentiles. A designer ought to decide and clearly define which percentile of the chosen population she is designing for: P$_{90}$, P$_{95}$ or even P$_{99}$? Especially with extreme percentiles, this can make an enormous difference for the values involved, and consequently, for the design. The question is which percentage to exclude.

Exclusion of extreme users should been considered with care. Designers sometimes copy the limits for P$_5$ and P$_{95}$ unquestioning from the ergonomic handbooks, without asking whether these are the right limits and which are the arguments to exclude 10% of the population from use or from comfortable use.

This nonchalance is named the "P$_5$–P$_{95}$ syndrome" (Dirken 1997). Here a comment should be made on the widespread habit of quoting the P$_5$ as a "normal" design maximum. To the annoyance of millions of people, many designers think it is accepted, standard, or even good practice, to exclude the upper or lower 5% (or both) of a population. This "P$_5$–P$_{95}$ syndrome" results in products which in the worst cases cannot be effectively, or comfortably, used by ~5.5 million in the UK, 22 million in the USA and 110 million in China. So it is clear that excluding 10% of a population results in a relatively large amount of frustrated would-be users. A good designer explains explicitly why a certain percentage of the population ought to be excluded, either on the top, on the bottom or on both sides of the curve (Daams 1994).

REFERENCES

DIRKEN, H. 1997, *Productergonomie* (Delft: Delft University Press).

DAAMS, B.J. 1994, *Force Exertion in User–Product Interaction, Backgrounds for Design* (Delft: Delft University Press).

Risk Factors for Non-specific Musculoskeletal Disorders

C. Jensen

Department of Information Technology, National Institute of Occupational Health, Lersø Parkallé 105, DK-2100 Copenhagen, Denmark

1. INTRODUCTION

The most dramatic technological change in working life for the past 10 years has probably been the introduction of the computer. People spend an increasing part of both work and leisure time using the computer. For instance, in 1995 47% of a representative sample of Danish employees reported to use a computer at work. From 1990 to 1995 the percentage that reported to work with a computer for at least three-quarters of the daily work time increased from 4 to 11%. The percentage is undoubtably higher today and the trend appears similar all over the world.

The computerization of work tasks has introduced new factors in the working environment, which may present a risk for the workers' health; especially with a risk of developing disorders in muscles and tendons of the upper extremities. Such risk factors both include factors related to the work station, computer and input devices, working technique, work organization and the psychosocial work environment. Here, focus will be on factors specifically related to the input devices.

A variety of input devices exists and their use is partly determined by the specific software used. The keyboard was and probably still is the most used input device, but software development clearly leads to more use of non-keyboard input devices, of which the computer mouse is most common. Intensive mouse work already exists in some jobs such as in computer-aided design (CAD) work. The use of the Internet is also based on non-keyboard input device operations and even children are familiar with intensive 'mouse work' through computer games.

2. INPUT DEVICES

Many different input devices are available and new devices will continuously be marketed in the future. Both input devices not only are based on the same principles of interaction as those that exist today, but also entirely new ways of computer interaction may be expected to become common. The most commonly used devices are:

- Keyboard — the keyboard is not only used for entry of text and numbers, but also has a wide range of "short-cut" operations that can replace non-keyboard input device operations. The keyboard exists in several forms such as split, pitched, curved and tented, which require a variety of wrist postures. However, the flat traditional QWERTY keyboard is still widespread. Both finger, wrist and arm movements are used to operate the keyboard.
- Mouse — some rather complex keyboard operations have been replaced by the use of the computer mouse to enhance

Table 1. Effect of input device use on surface EMG level reported in the literature.

Study	Muscle	EMG level, mean or median (in % of maximal level)		
		Keyboard	*Keyboard and mouse*	*Keyboard and trackpoint*
Fernström & Ericson, 1997. Wordprocessing, 10 female, 10 male	Upper trapezius	3.7	2.7	1.8
	Flex. digitorum	1.4	1.1	1.3
	Ext. digitorum	4.1	3.3	4.0
	Ext. carpi ulnaris	6.2	5.3	6.9
			Mouse	*Trackball*
Karlqvist, 1997. Text editing, 10 female	Upper trapezius		3.3	2.8
	Ant. deltoideus		1.3	1.2
	Ext. digitorum		6.4	6.8
8 male	Upper trapezius		10.3	8.1
	Ant. deltoideus		3.0	4.8
	Ext. digitorum		9.1	10.4
			Traditional mouse, pronated forearm	*New mouse, semipronated forearm*
Aarås and Ro, 1997 Graphics task, 13 male	Upper trapezius		0.5	0.7
	Ext. digitorum		17.0	10.3
	Ext. carpi ulnaris		5.2	4.9
		Mouse	*Trackball*	*Touchpad*	*Pen*
Jacobson et al. 1998. Pointing and dragging task, 3 female, 8 male	Flex. digitorum super.	7.5	8.5	8.5	7.8
	Ext. digitorum communis	15.4	14.4	9.6	12.4
	Ext. indicis proprius	13.1	13.0	10.2	11.5

the operations and thereby increase performance. In recent years the design of the mouse has attracted increased ergonomic attention and computer mice are now not only flat two-button wired devices, but also may be curved, or even upright standing, wireless devices with several buttons and scroll functions. The mouse requires both arm and wrist movements to move the mouse and finger movements to activate the buttons.

- Mousetrapper — with this device the mouse has been trapped on the side of the keyboard and is operated in front of the operator, primarily with finger and wrist movements.
- Trackball — another device that has shifted the requirements for movements from the arm towards the fingers is the trackball, which is rolled by the fingers to move the cursor. In addition the fingers press buttons in the same way as with a mouse.
- Pen — a digital pen requires similar postures as an ordinary pen, but somewhat different movements of the wrist and arm. The first digital pens required a tablet, but new types have been developed to operate without a tablet.
- Puck — also neck postures and neck and eye movements may be influenced by the input device as the use of a tablet with a graphic overlay, often activated with a puck, requires frequent movements of neck and eyes to concentrate both on tablet and computer screen.

3. FITTING THE INPUT DEVICE TO THE USER

Re-design of existing input devices have often aimed at lowering the biomechanical strain of specific elements of the musculoskeletal system by changing work posture. Both the keyboard and the mouse are available in versions that do not require almost full pronation of the forearm, but allows the forearm to stay in a more natural, less pronated posture. For one type of mouse that requires a semi-pronated forearm posture this was shown to require a significantly lower activity level of the extensor digitorum muscle (Table 1). The effect of keyboard design on finger tendon biomechanics has also been demonstrated. A positive pitch angle of the keyboard seems favorable by reducing extension of the wrist and decrease tendon travel.

Gender differences are common both regarding symptom prevalence, work technique and posture when operating keyboards and mice. However, the development of input devices that suit both large and small hands (including childrens' hands) as well as left- and right-handed people has been remarkably slow. When one gets a new input device the size and shape should perhaps be paid just as much attention as when one buys a new pair of shoes.

4. DESIGN OF INPUT DEVICES AND PHYSIOLOGICAL RESPONSE

Work posture and electromyographic (EMG) activity have been widely focused on when comparing physiological responses to the operation of different input devices. Other ways of evaluating the use of specific input devices have been to record perceived strain, calculating tendon travel, recording movement frequencies, recording specific keyboard or mouse operations with odometers or measuring applied forces on keys or on the mouse.

To some extent EMG studies have confirmed expected distributions of activation patterns between muscles when comparing the use of different input devices. The two most ordinary types of input device operations, i.e. keyboard use and keyboard use combined with mouse use, have been directly compared (Table 1). A word-processing task using a keyboard only was associated with slightly higher EMG levels than when using a mouse in combination with the keyboard. As long as the mouse work is not performed with extreme postures, this is not surprising as some operations are performed with considerably less movements and lower mental demands when using a mouse as compared with a keyboard. However, word-processing using a mouse may be performed with more ulnar wrist deviation than when using a keyboard. Interestingly, ordinary handwriting entails higher muscle activity than both mouse and keyboard work (Fernström and Ericson 1997).

The trackball and the trackpoint device, which mainly require finger movements, may be used with lower muscle activity levels in the shoulder than use of the mouse, which requires more arm movements. Less finger movements are required during mousing and Karlqvist (1997) found that the trackball was operated with more wrist extension and a higher EMG activity of the extensor digitorum muscle, but the differences in EMG activity levels, when operating different devices, often appear small (Table 1).

Overall, it seems impossible to choose a device that requires a lower general muscular activation than any other device, but it may be possible to choose a device which requires somewhat lower activation of one muscle group with the cost of increased activation of another muscle group. Such information may be of particular use for people already suffering from symptoms in specific muscle groups as they may use a new device to redistribute activity between muscles and thereby relieve symptoms from one muscle group, at least temporarily. For the long-term prevention of developing symptoms it is important to introduce sufficient variation in the way muscles are used, e.g. by using several different input devices at the computer.

5. EPIDEMIOLOGY

Epidemiologic studies on risk factors associated specifically with the use of input devices during computer work are few, but several on-going follow-up studies may add considerable knowledge to this topic within the next few years. So far, it may be concluded that there is a relatively clear effect of computer work *per se* on the development of hand/wrist disorders and probably also on shoulder/neck disorders (Punnett and Bergqvist 1997). An increased duration of computer use is associated with an increased prevalence of musculoskeletal symptoms for keyboard operators, and other studies have found similar results for CAD operators, who use a mouse more than a keyboard. Thus, some factors associated with keyboard or mouse operations may lead to musculoskeletal symptoms and disorders, but it is not known whether other risk factors play a larger role than those specifically related to the input device (e.g. factors related to the work task or psychosocial factors of computer work). However, it seems reasonable that repetitive movements and static, awkward postures, which are known to be risk factors in general, also are risk factors in computer work, whenever intensive or prolonged operation of the input devices introduce such factors. Interventions aiming at improving work postures may have

positive effects as use of alternatively designed keyboards may relieve symptom severity in keyboard operators with musculoskeletal pain. Furthermore, forceful movements may contribute to the development of pain as operation of the keyboard with excessive keyforce has been found in office workers with more severe symptoms as compared with workers with less symptoms (Feuerstein *et al*. 1997).

6. WHICH PHYSIOLOGICAL RESPONSES DURING COMPUTER WORK MAY BE RISK FACTORS FOR THE DEVELOPMENT OF MUSCULOSKELETAL SYMPTOMS?

As the epidemiologic evidence of risk factors associated with computer work is relatively limited, we may to some extent rely on studies of acute responses to the performance of tasks with different input devices and base evaluations on our general knowledge of the association between short- and long-term responses. Thus, work with a posture close to neutral (joint angle close to 0°) is considered associated with a lower risk of developing symptoms than non-neutral postures (joint angles deviating from 0°). Previously, there was also a focus on reducing the level of muscle activity to reduce the risk. However, it is now recognized that the level of muscle activity is of minor importance as a risk factor in jobs involving work tasks, where the biomechanical strain is already low as it is during computer work. Preventive measures need to focus more on introducing variation in the exposure pattern to reduce the risk of developing symptoms.

For example, during intensive computer work with a mouse (CAD work) both the hand operating the mouse and the other hand, which is used for paper work may show similar wrist extensions *on average* (Jensen *et al*. 1998). However, the range of motion is typically twice as large on the non-mouse side as compared with the mouse side. Similarly, EMG levels may only be slightly different on the two sides (Jensen *et al*. 1998). But the variation in the EMG activity pattern and the fraction of the work time with muscular rest is much smaller on the side operating the mouse than on the other side. In this particular study three times as many CAD workers reported wrist and shoulder symptoms on the mouse side as on the other side. This suggests that the static nature of the posture and the continuous muscular activation pattern may be most important for the risk of developing symptoms.

Also, similar levels of EMG activity have been recorded during the performance of computer- and paper-based versions of the same task (Wærsted and Westgaard 1997). However, the variation of EMG activity was considerably larger for the paper-based version, which may be considered favorable in terms of preventing musculoskeletal disorders.

Clearly, the computerization of many work tasks has lead to a more stereotypic muscular activation pattern involving a relatively low number of motor units in a muscle. Possibly, prolonged activation of only a few motor units may induce an overexertion of the active muscle fibers causing muscular pain. This would imply that prevention of muscle pain is only possible if the activity of these particular motor units is reduced, i.e. their continuous firing patterns should be interrupted. Thereby, the most important risk factor would be more related to the duration of the exposure than the exposure level. Mouse work may induce

muscle fatigue and probably more chronic symptoms can develop even at the lowest levels of activity, when the activity is maintained for hours and repeated day after day.

Other factors than input device operations such as visual demands and the possibility to rest the arm on the table may be just as important for the development of symptoms. Time pressure may increase the muscle activity level even when the amount of work performed with a mouse is only marginally increased. Increased cognitive demands or demands of precision may have considerable negative effects on the amount of work performed, without any reduction in the muscular activation demands. A combination of such demands, e.g. a high demand of precision combined with a close deadline is likely to elevate the muscle activity level and leave insufficient time for muscular rest periods no matter which input device is used.

8. RECOMMENDATIONS

In conclusion, prevention of musculoskeletal symptoms during computer work should focus on introducing sufficient variation in exposure. Such efforts can be directed towards several aspects of the job. One may focus on work tools by using more than one input device as it should be expected that prolonged use of any single input device may increase the risk of developing symptoms. However, we do not know how different the input devices should be to reduce the risk. Work technique may be addressed, e.g. by changing the hand operating the device (with the risk of developing bilateral symptoms). A third possibility is to organize work tasks in such a way that the input device (and perhaps the computer) is used less intensively. This should have a clear preventive effect, but is often the most difficult solution.

REFERENCES

AARÅS, A. and RO, O., 1997, Workload when using a mouse as an input device. *International Journal of Human–Computer Interaction*, 9, 105–18.

FERNSTRÖM, E. and ERICSON, M., 1997, Computer mouse or Trackpoint — effects on muscular load and operator experience. *Applied Ergonomics*, 28, 347–54.

JACOBSON, M., REMPEL, D., MARTIN, B., KEIR, P. and DENNERLEIN, J., 1998, Comparison of fine wire to surface EMG amplitude of finger muscles during precision tasks. In *Proceedings of the Marconi Research Conference*, San Francisco, 49–54.

JENSEN, C., BORG, V., FINSEN, L., HANSEN, K., JUUL-KRISTENSEN, B. and CHRISTENSEN, H., 1998, Job demands, muscle activity and musculoskeletal symptoms in relation to computer mouse work. *Scandinavian Journal of Work Environment and Health*, 24, 418–24.

KARLQVIST, L., 1997, *Assessment of Physical Work Load at Visual Display Unit Workstations. Ergonomic Applications and Gender Aspects* (Solna: National Institute for Working Life).

PUNNETT, L. and BERGQVIST, U., 1997, *National Institute for Working Life — Ergonomic Expert Committee Document No 1. Visual Display Unit Work and Upper Extremity Musculoskeletal Disorders. A Review of Epidemiological Findings* (Solna: National Institute for Working Life).

WÆRSTED, M. and WESTGAARD, R.H., 1997, An experimental study of shoulder muscle activity and posture in a paper version versus a VDU version of a monotonous work task. *International Journal of Industrial Ergonomics*, 19, 175–85.

Risk Factors of Musculoskeletal Disorders: Demographic, Social and Work Change Aspects in France

J.-P. Meyer and E. Turpin-Legendre

Work Physiology Laboratory, INRS Avenue de Bourgogne, F-54500 Vandoeuvre, France

1. INTRODUCTION

Occupational upper limb musculoskeletal disorders (ULMSD) occurrence is largely explained in the literature by biomechanical and/or psychosocial analysis of the task and its surroundings. These aspects are probably fundamental parts of ULMSD analysis and prevention at the factory level.

However, in industrialized countries an explanation for the dramatic increase of the musculoskeletal pathology in the past 20 years must be found in a more general approach of work changes. Furthermore, if the ULMSD epidemic is at a so high level, prevention must consider these aspects like, at least, the risk factors collected at a specific workplace. Then, preventive means must modify the global factors that affect working conditions.

The purpose of this article is to focus on the social, occupational and economical determinants of ULMSD in France between 1980 and 1995 to try to understand and explain the ULMSD epidemic which started during these years.

2. METHODS

The different points discussed here are based on an analysis of the evolution, from the end of the 1970s to the early 1990s, of ULMSD incidence on the one hand and the risk factors like workers' age and gender distribution, working conditions and economics in France on the other. For this purpose, ULMSD figures drawn from the social security statistics (CNAM 1996), and changes in the working population, analyzed from the statistics published by the National Institute of Statistics and Economical Studies (INSEE) were successively addressed. The statistics of the Ministry of Labor allow the identification of working conditions changes (DARES 1996). At last, productivity, work change and quality requirements were also drawn from several sources and personal communications.

3. RESULTS

3.1. Number of Compensated Musculoskeletal Disorders

In France, the occupational ULMSD are compensated by the occupational disease system of the social security. The number of ULMSD considered here excludes state and agriculture workers but is based on a total of ~14 million workers. To obtain compensation, the worker has to declare himself the pathology to the social security. The disease, diagnosed and certified by a
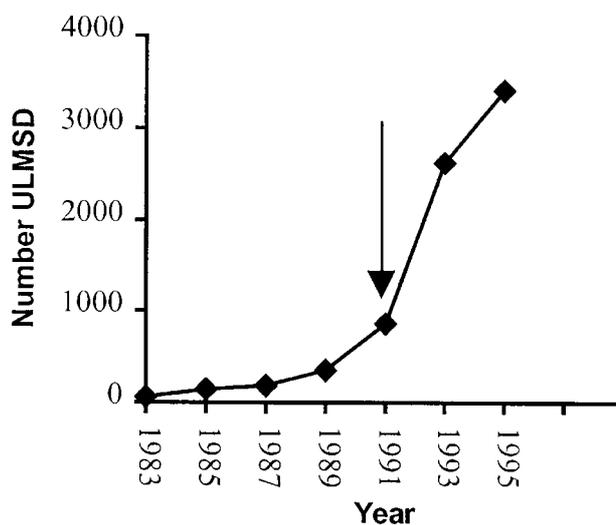
Figure 1. Increase of compensated ULMSD between 1983 and 1996. The arrow shows the time at which table 57 was modified. Data for the years before 1983 were too low to be considered.

physician must be listed in the Tableau des maladies professionnelles where the corresponding risk factors to which the worker has to be exposed are also listed. In the case of ULMSD, the table is no. 57 of occupational diseases which also compensates lower limb musculoskeletal disorders like patellar bursitis or compression of the external sciatic nerve on the fibula head. Figure 1 shows the increase of compensated ULMSD from 1983 to 1995 which is the last known amount for analysis. Data before 1983 are not considered because few ULMSD were compensated at that time. Diseases due to vibration of hand tools are compensated in table no. 69 (white finger syndrome, elbow arthrosis, carpal osteoporosis); they represent about 100 cases each year and are not considered in Figure 1. From the compensated musculoskeletal disorders according to table no. 57 in 1995, upper limb represents 64% of the total. This ratio was only 10% in 1983. This evolution is partially explained by changes in the table no. 57 where carpal tunnel syndrome appeared only in 1981 and wrist or shoulder tendinitis in 1991. In 1991 risk factors like carpal pressure, wrist postures and repetition were added in table no. 57.

The total number of compensated occupational diseases in 1995 was 8534 from which 3413 are ULMSD interesting shoulder (17%), elbow (24%) and wrist (59%) including carpal tunnel syndrome (44%) and tendinitis (14%). This distribution is a classical finding in this type of survey. These numbers are low for the 14 million workers protected by the general social security system. This underestimation of the actual number of occupational disorders is recognized. However, even if this number is underestimated, its increase is obvious like in other countries but delayed compared with what happens in other countries. The late increase of the ULMSD can be explained by the recent changes of table no. 57 and by the fact that the worker has to ask for compensation of a "new" occupational disease. However, workers union surveys have revealed ULMSD epidemics in car and household electrical appliances manufactures from the mid-1980s.

3.2. Risk Factors

Changes in risk factors from 1975 to 1995 could be analyzed as a weakening of the working population due to aging and feminization, a working conditions hardening, quality control implementation, the paradoxical effect of changes from heavy industrial and agricultural activities to services and by economic evolution during the period.

Aging is recognized as a major factor of functional capacity decrease at a population level (WHO 1993). As a consequence, in an aging population the ratio between work requirements and functional capacities shows a general trend of decrease and thus an increase of the work strain as far as working conditions are not implemented and adapted. This implies for many workers that work stress can overcome their functional capacities and induce MSD. The aging process of the French working population in the period studied is characterized by the decreases of two classes of age in the active population, i.e. people > 50 years of age and one < 25 years of age. In these two age classes the percentage of the active ones decreased between 1975 and 1995 from 54 to 34% in the younger workers and 32 to 23% in the older ones (INSEE 1998).

The evolution of the age class 25–50 is characterized by a progressive appearance of the post World War II baby boom generation. Owing to the remaining high birth rate until the early 1970s, this age class moved from representing young workers < 30 years in 1975 to a homogenous distribution in the age range 25–50 in 1995. At that time the younger population of the high birth rate period, born in the early 1970s, was engaged in active life. Thus, in 1978, the 40–50-year age group was ~3.7 million, it was 4.1 million in 1987, 5 million in 1991 and 6.7 million in 1998. This tremendous increase (81%) in 20 years is more important for women (125%) than for men (59%).

Workforce feminization is a very strong social feature from 1975 to 1990. Indeed, the number of working women increases from 8.1 million in 1975 to 9.6 million in 1985 and 11 million in 1990 (INSEE 1988). The age distribution described above for the whole population is emphasized by the women entering in the workforce. Indeed, the increase is the most important for women aged 40–49 years who were 1.4 million in 1978 and 3.1 million in 1998. Thus, to the aging process of the active population is added the increase in a "frailer" workforce facing physical strain at work and family charges (Teiger 1989). This is particularly true when tasks that need accuracy and speed, like knitting, are considered. While these tasks were achieved by women < 25 years of age in the 1970s, they have to be done by older women in the 1990s because of the demographic changes (Teiger 1989).

Working conditions worsening is exemplified by several aspects of work. A longitudinal survey of the working conditions is conducted by the Ministry of Labor. Table 1 shows the percentage of the surveyed population (20 000 workers) who declared they are exposed to defined strains in 1978, 1984 and 1991 (DARES 1993). The results of the last survey (1996) are not published yet, but some partial results seem to confirm the trends shown in Table 1.

As presented in Table 1, the subjective estimation of the working conditions presented is mostly worsening. This is particularly salient for the hazardous activities, i.e. handling heavy loads, awkward postures, repetitivity and time pressure which, for the external demand, constrains more than one out of two workers. The results in Table 1 show that working conditions have not been adapted to the worker as much as it might be expected by the technological progress. Furthermore, just-in-time model has been adapted mostly throughout worker adaptation and not the contrary.

The implementation of quality control has also changed the working conditions. In 1988–93, eight of 10 factories with > 10 workers made some re-engineering. The main reason (87%) for that was product quality. Moreover, a question about quality outcomes at work was asked for the first time only in the 1991 working conditions survey (DARES 1993). At that time, the most frequently reported consequence of an error at work was product quality abatement (60%). In the past, most repetitive monotonous tasks left mental "freedom" to the workers without trouble for their task. However, quality requirements induce higher concentration on the task. This is difficult to achieve in repetitive monotonous tasks. But, like Bensaid and Dejours (1994) explained it, "psychic repression" (Unterdrückung) is a means to reach higher concentration in these monotonous activities. A very common method to obtain psychic repression is to engage completely in a task. In monotonous repetitive activities, where work latitude is restricted, this is achieved by increasing the work rate. Thus, workers are working faster and even, for many, at their maximum rate. It is noteworthy that the drawback of "psychic repression" is psychological. However, the increase of the task rate is a byproduct that has an important adverse effect on ULMSD. Furthermore, one cannot answer the question about whether psychological or musculoskeletal disorders are the most important consequence to the need to concentrate on the task.

Analysis of the type of activity shows an historical change during the period under study. Indeed, the beginning of this period corresponds to the first decrease in the number of workers in industry. The number of agricultural workers decreased, as in the previous years, from 2.2 million in 1975 to 1 million in 1995. During the same period industry reached its highest level of workers, 8.1 million in 1975, which then decreased almost linearly to 5.8 million in 1995. On the opposite, the tertiary sector grew from 11 to 15.6 million workers from 1975 to 1995. This growth was mostly due to workplaces where women were employed. Many of these jobs could be characterized as low general physical activity but high local muscular strain and, therefore, were recognized as including numerous risk factors for ULMSD. Indeed, supermarket cashier or computer work (VDU) are activities that appeared during this period. As an

Table 1. Workers exposed to defined strains (%).

Year	1978	1984	1991
Carrying heavy loads	21	22	32
Awkward postures	17	16	29
Repetitive Work			
men	18	18	26
women	25	22	34
Work rate imposed by			
machine	15	14	22
standards	21	19	38
external dead lines	34	39	57

example, one-third of the active population worked with VDU in 1994 while VDU work was almost absent in 1975. Furthermore, in the healthcare occupations the number of workers has grown regularly during the period under study; the number of nurses was multiplied by 1.5 between 1981 and 1996.

Economical evolution is difficult to represent. However, after the first oil crisis in 1973 industrial production slowed down during the second part of the 1970s. Industrial activity grew again only from 1983 to 1984 and then more markedly from the mid-1980s throughout the early 1990s (INSEE 1998). The production of these "go years" (1985–91) was achieved with an almost constant number of active workers. This period was shorter but manufacturing was more effective at that time in France as in most of the other European countries. This is particularly the case for household equipment and car manufacturing. For example, household appliances and car manufacturing increased > 30% in 1985 to 1989. The first descriptions of ULMSD epidemics in these occupations took place during this period.

Furthermore, productivity increase in the industry at the end of the 1980s can be illustrated by three economical indices (INSEE 1988):

- the use of the production capacities reached a rate never met before;
- the number of worked hours has decreased markedly since 1981; and
- the relative part of the salary in the cost of manufactured goods has decreased close to current levels of the 1960s.

These positive facts from an economical point of view are marked in the last 2 years which, in addition, are characterized by an increase in temporary work and unstable working status. The number of workers in these categories was multiplied by 1.5 from June 1996 to December 1997. Then, one can just hope that our hypothesis is wrong and that economics and ULMSD are not so strongly linked and thus the disease figures will decrease in France as they do in other countries.

4. CONCLUSION

The working population aging and feminization, quality control implementation, changes in the production structure from classical industries to services and the productivity increase are all risk factors that can cause the observed development of ULMSD. Their combined action during a short period is probably a major explanation of the epidemic. The analysis of the socio-economical and demographic changes in France, as well as in other so-called developed countries, shows the tremendous challenge of an efficient prevention of musculoskeletal disorders. Indeed, if these conditions must pursuit, it is probable that the ULMSD epidemic is likely to spread out. Thus, prevention will only be effective through a strong re-assessment of work methods to keep the fit between the imposed production needs and the human producer abilities. If changes in work are not occurring, the socio-economical needs of the working population and the demographic realities will put in light the threat of work to health and even public health which depends, for a large part, on the health of the working population. The results of the present study support the idea that it is not possible to look only at economics without considering its impact on health.

At last, it could be interesting to apply at a local level, workplace or factory, the observation of the workforce and the work changes like that done here from a nationwide view. Indeed, these aspects may be the easiest to check when a product or its manufacturing is modified and can help prevention.

REFERENCES

BENSAID A. and DEJOURS C., 1994, *Psychopathologie et psychodynamique du travail. Encyclopédie Médico Chirurgicale* (Paris: Techniques).

CAISSE NATIONALE DE L'ASSURANCE MALADIE, 1996, *Statistiques nationales des accidents du travail et des maladies professionnelles* (Paris: CNAM).

DARES, 1993, *Conditions, organisation du travail et nouvelles technologies en 1991. Dossiers statistiques du travail et de l'emploi, Ministère du Travail* (Paris: Masson).

INSTITUT NATIONAL DE LA STATISTIQUE ET DES ETUDES ECONOMIQUES, 1998, *Tableaux de l'Economie Française* (Paris: INSEE).

TEIGER, C., 1989, Le vieillissement differentiel dans et par le travail. Un vieux problème dans un contexte récent. *Le Travail Humain*, 52, 21–56.

WORLD HEALTH ORGANISATION, 1993, *Study Group: Aging and Working Capacity*. Technical report 835 (Geneva: WHO).

Robot Safety Standard – R15.06

R. N. Sheat† and J. Fryman‡

†Detector Manager, Honeywell Data Instruments, and Chair, R15.06 Subcommittee on Safety

‡Manager, Standards Development, Robotic Industries Association

1. INTRODUCTION

The USA has introduced a new version of its safety standard for industrial robots, ANSI/RIA R15.06-1999. The new volume has been greatly expanded in content, and has been edited for ease of use and understanding. Goals were established and met to enhance the safety of personnel employed in robotic automation applications. Harmonization with international standards was considered during the revision process, even though the document remains uniquely a US standard. The standard introduces new criteria that will enhance the design safety of robots. Included are safety circuit performance criteria, and the introduction of a safety stop circuit separate and in addition to the required emergency stop circuits. Risk Assessment (reduction) is also presented in considerable detail. The importance of human factors in a safe work environment is inherent throughout the standard.

2. DEVELOPMENT BACKGROUND

2.1. Robotic Standards

Voluntary consensus standards used by US industry are known as American National Standards, and are developed under the auspices of the American National Standards Institute (ANSI). The Robotic Industries Association (RIA) is the accredited standards developing organization for robotic standards in the USA. Part of the ongoing evolution of standards is the required five year review. When the required periodic review of the 1992 edition of the standard became due, the RIA determined that a major revision was appropriate.

2.2. Scope of Revisions

Goals were established for the revised standard development effort. These included: enhancing robot and robot system safety; making the standard more understandable, both in organization and user friendliness; and to respond to feedback received on
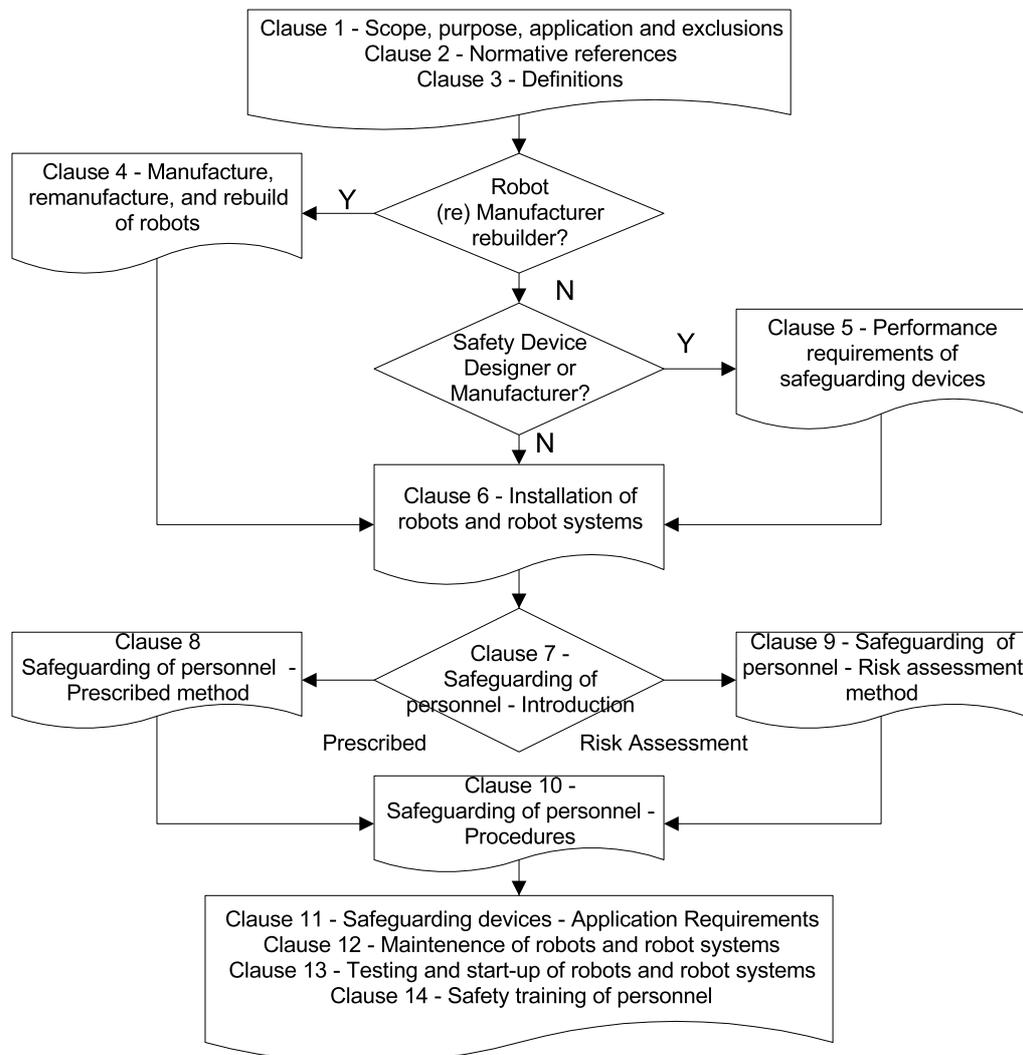
Figure 1. - Logic flow of ANSI/RIA R15.06-1999

the 1992 document regarding perceived weaknesses. Additionally, the standard needed to address realistic and achievable safeguarding requirements in industrial automation applications.

2.3. New Robot Safety Standard

The response to these challenging goals is a new 1999 edition of the robot safety standard, ANSI/RIA R15.06-1999. The new edition has a revised, open format designed to be more easily read and understood, as well as be more in harmony with international standards. The text was greatly expanded, reorganized and divided into additional sections with specific interest to specific segments of the industry. A strategy for safety was developed and set forth, which includes options for performing risk assessment (reduction) or following a more strictly prescribed method of safeguarding.

3. DOCUMENT STRUCTURE

There are 14 clauses in the new standard versus the nine clauses in the 1992 edition. The safety standard assigns responsibilities for industrial robot safety to manufacturers, integrators, installers, and the user. Figure 1 offers a graphic presentation of the flow of responsibilities through the new standard.

4. COMPLIANCE

4.1.

The newly revised standard addresses both new and existing robots, and robot systems. Some change to existing systems may be required if an existing system does not meet a minimum prescribed level of safety compliance. Ample timeframes are established to allow for compliance with the new standard. Generally, systems in compliance with the 1992 standard will probably not require any retrofit to be compliant with this edition of the standard.

4.2.

The new standard is more inclusive, but only applies to industrial robots and systems that are used in industrial automation applications. Several specific robot applications are excluded; and machinery that is not a robot is not mentioned as an exclusion.

4.3.

Normative references typically include other voluntary standards related to the topic and providing important and useful information over and above what is included in this standard. This revision has introduced normative references to regulatory standards (directives) issued by the US Department of Labor, Occupational Safety and Health Administration (OSHA). This is an important distinction, and reflects that OSHA, in turn, recognizes voluntary standards as accepted industry practices. In fact, ANSI and OSHA have a memorandum of understanding regarding expedited acceptance of ANSI approved standards by OSHA.

4.4.

The definitions section has been cleaned up and a concerted effort made to harmonize definitions with those in ISO 8373. Only terms that are used in multiple locations within the document are provided as definitions. Terms that are only used once, are described in detail within the section in which they are used. A

significant change in the new standard is the use of the word "space" to replace "envelope" in describing the volumetric areas of robot motion. Envelope was considered too much of an "engineering" term, and the committee felt the word space better conveyed the concept to a general audience.

5. MANUFACTURER RESPONSIBILITIES

Manufacturer's responsibilities have been refined. This is a performance standard, so changes are directed to how the robot (and robot system) must perform. Although this often affects the way a manufacturer designs his product, the standard contains no specific design criteria. The differences of rebuilt and remanufactured were better explained, and specific requirements applied to each.

5.1.

The requirement for axis movement without drive power was revised to better reflect its intent. There must be a safe method to release personnel caught, pinched or trapped by motion of the robot. Additionally, this method should not create or increase another hazard such as energy release due to gravity.

5.2.

Axis limiting requirements were revised to require provisions for adjustable hard stops on the primary axis. Also required is that provisions (i.e. mounting post or electrical contact) be made for installation of limiting devices on the second and third axis of the robot. These are important elements in reducing the hazard of robot motion.

5.3.

Protection from singularity is now required during teach. This protection takes the form of the robot motion being stopped prior to the robot program automatically executing a correction for the singularity condition. The teacher must acknowledge the condition, and authorize the robot motion to continue, thus allowing anyone in the safeguarded space to move clear of the robot motion.

5.4.

A new concept of a three-position-enabling device has been introduced. The new standard requires the use of a center on (full squeeze or release stops all motion) switch that must be constantly activated to enable hazardous motion. This requirement was added in response to studies that have determined that people respond differently to perceived hazards, and will often "freeze" (pull tight) rather than release an object (e.g. the motion enabling device) they are holding. Specific cautions are added regarding the importance of ergonomic considerations when installing this device.

5.5.

A more detailed list of information that is important to the robot user has been developed. This list is required to be made available by the manufacturer, and in turn, provided to the integrator, or installer of the robot and robot system, and finally to the end user. The required information varies in scope from that usually contained in a routine data sheet (sales literature) to data normally

found in installation/user manuals to specialized data that must be purchased if it is requested.

6. SAFETY CIRCUITRY

6.1.

Safety circuit performance has been introduced to require certain levels of performance for safety circuits in various applications. This concept is harmonized with, but distinctly different than the category requirements in ISO/IEC 13849-1 (EN 954-1). Definitions for four levels of performance from simple to control reliable are provided. All safeguarding devices must be integrated into the robot work cell using one or more of these performance levels.

6.2.

Closely related to circuit performance is the new requirement for separate safety stop circuits to be used by safeguarding devices that signal a stop (i.e. light curtains, gate switches). This stop is separate from the emergency stop circuit, and allows an orderly stop, and one that retains program logic for troubleshooting.

7. TEACHING

The task of programming or teaching the robot has been carefully defined as a separate task from that of verifying the robot program or path. An entire section is dedicated to the specific safeguards called out to protect the teacher. These safeguards include the use of slow speed control at all times during teach, single point of control, and clearance. For teaching, the required clearance is measured from the "operating space".

8. PROGRAM VERIFICATION

Program verification also has a section dedicated to specifying very specific safeguards required to allow a person to be in the safeguarded space when the robot is functioning at speeds above slow speed. In this mode, greater clearance requirements are needed; and therefore are measured from the 'restricted space.' A feature known as Attended Continuous Operation allowed in the 1992 edition is now prohibited.

9. RESPONSIBILITIES FOR SAFEGUARDING DEVICES

The standard has been reorganized to include a clause providing specific guidance and detailed performance requirements to the manufacturers of safeguarding devices. Another clause provides detailed guidance on the selection, installation and use of safeguarding devices. The integrator or installer of the equipment is a key element in robot and robot system safety and receives considerable instructions throughout the revised standard on proper safeguarding techniques.

10. RISK ASSESSMENT

The 1992 robot safety standard required that a risk assessment

be performed, but went on to provided very little additional information. The new 1999 standard responds to comments from the robot community by offering a choice of safeguarding methodologies. The robot user must choose between performing a risk assessment or installing all safeguards in a prescribed regimen.

10.1.

Risk assessment has become an increasingly important aspect of safety. An entire clause now discusses risk assessment with detailed performance criteria, and a suggested methodology to perform an effective risk assessment. Charts are provided to aid in determining risk and the appropriate reduction options for safeguard selection. This method pays careful attention to sources of potential hazards, and aids the effort to better understand the actual process being performed in an application.

10.2.

Alternatively to a risk assessment, a prescribed method of safeguarding is included with specific safeguards and installation requirements. This method ensures a safe environment, but does not aid in the understanding of the hazards involved in an automated application.

11. ADVANCING TECHNOLOGY

The revised standard recognizes the continuing advance in technology development. It was carefully written so as not to preclude the use of new technology when it becomes safe, feasible and recognized by other standards. Specifically noted are advances in control reliability and newly introduced software based controllers that are expected to see more widespread use in the next few years. Specific issues are addressed which will provide an assurance of safety in the application of new technology.

12. CONCLUSION

A number of other enhancements have been made in this 1999 revision of the robot safety standard. Included in these enhancements is a list of suggested hazards, expanded protection for maintenance personnel, and more detailed information on initial (first time) start-up and training programs. All of the changes made in this standard serve to ensure a safer, better designed, and more productive work environment for personnel in today's highly automated work place.

The new standard can be ordered from the Robotic Industries Association, PO Box 3724, Ann Arbor, MI 48106, USA; tel: (734) 994-6088; Internet: www.robotics.org

ANSI/RIA R15.06-1999; American National Standard for Robots and Robot Systems — Safety Requirements. ©Robotic Industries Association, PO Box 3724, Ann Arbor, MI 48106, USA.

Slaughterhouses

G. Toulouse and J.-G. Richard

Institut de Recherche en Santé et Sécurité du Travail, Montréal, Québec, Canada

1. INTRODUCTION

The activities related to the production and processing of meat (poultry, beef, pork, and others) and meat by-products range from cattle raising to consumer distribution. The slaughterhouse plays a primary role in the process. In addition to the actual slaughter of animals, it is usually the site of primary processing activities such as cutting, freezing and wrapping.

In the last two decades, the slaughterhouse industry has had to adapt to some major changes, including new consumer habits, the opening up of markets, new quality and hygiene standards, and others. The slaughterhouses are now seeking to improve their productivity and competitiveness by specializing in certain types of animals, concentrating their production by grouping their facilities, mechanizing operations, increasing production line speeds, introducing primary processing activities and diversifying their product range. As a result, the working context has changed considerably. Workers are becoming increasingly specialized, and their jobs now include a significant inspection, quality control and machine work recovery component.

During the same period, prevention organizations have observed an increase in the number of work-related musculoskeletal disorders (WMSD) associated with repetitive work. In Western countries, the meat-processing plant now rank first among the sectors affected by this problem, which has overtaken traditional risks such as cuts and animal-transmitted diseases. Accordingly, ergonomics research and interventions have been concerned mainly with the development of prevention strategies for this type of disorder. The strategies in question often focus on worker participation.

This article describes some aspects of the production system and work activities in slaughterhouses, together with the main occupational health and safety problems that arise from them. Solutions are also proposed, and possible preventive approaches are discussed.

2. THE PRODUCTION SYSTEM

In modern slaughterhouses, the production system is based on a tight flow procedure involving a large number of operations, some carried out by machines and others by human workers. The nature of the product to be processed means that a *just-in-time* supply system must be used. The quality and homogeneity of the animals often varies from one batch to the next, limiting the possibility of mechanization and increasing the workload at some workstations. Currently, the highest levels of mechanization are found in poultry slaughterhouses. In slaughterhouses that kill larger animals, many operations are still carried out manually.

Slaughtering and primary processing activities are usually carried out in four separate departments: reception and slaughtering, gutting, cutting, and wrapping. The animals and by-products are moved around on hooks hanging from a chain, or on conveyor belts. A tight flow system is implemented from reception to shipping. In the last two departments, the product must be refrigerated, which requires ambient temperatures of around 12∞C.

Production capacities have increased gradually over the years. In poultry slaughterhouses, up to 12,000 birds per hour can be slaughtered, while in pig slaughterhouses the figure is around 900 animals per hour. For poultry, work cycle times are often between 3 and 5 seconds.

3. THE RISK OF WORK–RELATED MUSCULOSKELETAL DISORDERS

In a recent review of the literature, Loppinet and Aptel (1997) make a number of observations based on epidemiological studies of work-related musculoskeletal disorders (WMSD) in the meat industry.

- WMSDs affect all joints in the upper limbs.
- They occur regardless of the type of animal being processed.
- They occur in meat cutting, grading, and wrapping operations at every stage of the production line (Kurppa *et al.* 1991). Most of these activities involve the main risk factors normally associated with WMSDs: repetition, restrictive postures, the use of force, cold, vibration, and psychosocial factors.

In poultry slaughterhouses, WMSDs are more common among women than among men. This appears to be due to the fact that women are exposed more to risk factors associated with extended periods of standing and more repetitive and faster gestures (Mergler *et al.* 1987).

There are some significant differences in the prevalence and incidence of WMSDs depending on the publications involved. These differences may be due to the diagnostic criteria used, as well as to differences in the age, sex and seniority of the populations studied (Hagberg *et al.* 1992).

4. THE CONDITIONS THAT CREATE WMSD RISK AND SOME POSSIBLE SOLUTIONS

The conditions that create WMSD risk are varied and complex, and it is therefore difficult, in most cases, to isolate a single cause. In fact, all these conditions (or risk factors) interact. For example, the risks associated with the use of a knife depend on the design of the knife, the size and layout of the workstation, the characteristics of the product, the actions involved in the task, the training received by the worker, and so on. In the following paragraphs, we will consider some of the main conditions that influence WMSD risk. Very little research has so far been done on the psychosocial risks involved, and this element will therefore not be addressed here.

4.1. Workstation Layout

In slaughterhouses, the methods used to carry out particular tasks are determined by the two main ways of transporting the products. Many operations are carried out on a product that is attached by a hook to an overhead conveyor, while in other cases the product is removed from a conveyor, processed on a nearby table and then returned to the conveyor.

The spatial constraints associated with this type of layout are as follows:

- Evacuation of products from a conveyor, or onto a conveyor located above the original transportation conveyor, often forces workers to adopt extreme postures. Currently, problems of this nature are solved by designing workstations in which evacuation takes place directly underneath the original transportation conveyor.
- The effort involved in cutting up a carcass is increased by the range of joint movement required to work on a large animal (e.g. pig or cow), and by differences in worker body size. In the former case, the problem involves the standardization of the animals slaughtered, the planning of the slaughter stations, and task distribution. In the latter case, the problem can be solved by using a platform that can be adjusted to different heights by the worker.
- Lack of space between workers also causes postural restrictions, stress, and risk of knife injury. When the product is transported by means of an aerial conveyor, the space between workers is doubly important. It must be calculated on the basis of the conveyor speed, the duration of the work cycle on the product, the possibility of incidents and the time required to clean and sharpen the knife.
- Standing postures with a high static component are frequent and generate considerable fatigue. In many cases, the fatigue can be reduced by introducing high stools or saddle chairs.

4.2. The Knife

The use of a knife is often associated with the appearance of WMSDs. The size of the handle, the grip with rubber gloves and the sharpness of the blade all affect the amount of force required to cut the product (Armstrong 1982). Researchers have proposed a number of solutions to help improve knife design and reduce cutting effort:

- The angle between the knife handle and the blade should be such that it limits wrist flexion. However, this solution is often practicable only for tasks involving a single unidirectional action.
- The length and diameter of the handle should be adapted to suit individual anatomical differences.
- The size of the handle and the blade can be optimized to suit the type of cut, so as to reduce the force required;
- Handles made of a non-slip material that promotes comfort and flexibility of use will help limit fatigue by absorbing shocks, distributing pressure and enhancing grip.

A recent study on the sharpening of knives in slaughterhouses revealed that three-quarters of the 25 workers questioned had difficulty with this operation (Chatigny and Vézina 1995). The authors, observing that the workers did not fully master this technique, developed a training session (available in video format) based on pairings of experienced sharpeners and apprentices.

Finally, knives can be replaced in some instances by circular saws, hacksaws, rotary blade knives, or power cutters that, theoretically at least, reduce the amount of force required.

4.3. Gloves

The use of gloves to protect workers against cuts and/or cold may exacerbate musculoskeletal constraints. Some authors (Armstrong 1994) have shown that the thicker the glove, the greater the force required. In slaughterhouses, workers often wear several pairs of gloves at once. Each pair serves a different function:

cotton gloves for comfort, rubber gloves for their waterproof qualities and steel mesh gloves to protect against injuries.

There is a real need to develop gloves better suited to the conditions in slaughterhouses that comply with hygiene requirements and are not too expensive.

4.4. Time-related Constraints

The time-related constraints inherent in this type of work (the pace imposed by the production line or conveyor belt) influence WMSD risks in a number of ways (Toulouse 1995):

- lack of breaks;
- effort required in awkward postures determined by the need to begin operations upstream and continue them downstream when the product is not in the best position for the worker. Increased travelling speeds and recovery of incidents exacerbate this problem;
- some related actions that are not considered in the task (sharpening the knife, cleaning and sometimes sterilizing the knife, cleaning the gloves) increase the cycle time, extending the postural range or causing the task to be performed more quickly.

Many of these problems can be reduced by planning the task so as to take into account the possibility of incidents, product variability, and related activities. When designing tasks, it is also important to leave enough space between workers (Bellemare and Richard 1992).

4.5. Worker Training

The impact of vocational training on WMSD occurrence has not been studied directly, but nevertheless appears to be an important factor. Training in slaughterhouses, often carried out 'on the job', may be deficient and may not enable the workers to develop methods that minimize the risk of WMSD. Vocational training programs should be based on data obtained from surveillance of symptoms and injuries, and on the results of ergonomic analysis.

4.6. Job Rotation

Job rotation is often proposed as a way of reducing repetition or exposure to risk factors. However, although research in this area is by no means extensive, it appears to show that job rotation does not always achieve the anticipated results. Effective job rotation is subject to a number of difficulties, such as training deficiencies, workstations designed without taking worker body size variability into account, work places that limit the ability of operators to move from one workstation to another, as well as variations in production management and human resource management.

5. ERGONOMICS AND PREVENTION STRATEGIES

A review of the preventive approaches used in Quebec's poultry slaughterhouses (Richard 1997) showed that managers and health and safety officers were often powerless when faced with the complexity of the risk factors associated with this type of operation. There is therefore a need to pay particular attention to identifying and understanding the cause-and-effect links between the risk factors and their determinants, on the one hand, and symptoms and injuries, on the other. While awareness-raising and training activities provide basic knowledge, only a symptom surveillance program combined with surveillance of high-risk

jobs can produce the information required to seek and introduce solutions. A real-time symptom reporting and monitoring procedure, and ergonomic analysis of working activities, form the basis of such a program. The program should be accompanied by a preventive assignment program following manifestations of pain.

When the risk factors specific to a working situation are understood, the factory personnel, aided by an ergonomist, are best placed to identify and design corrective measures. Recent experience has, on several occasions, illustrated the interest of participatory ergonomic approaches in this field, although their long-term effectiveness has yet to be proved. We can assume that this effectiveness will depend on how the company uses the expertise developed by participatory ergonomics projects and how it applies that expertise as part of ongoing quality improvement projects or when designing new production systems.

REFERENCES

ARMSTRONG, T., FOULKE, J. A., JOSEPH, B. S. and GOLDSTEIN, S. A., 1982, Investigation of cumulative trauma disorders in a poultry processing plant, *American Industrial Hygiene Association Journal*, 43, 2, 103–116.

ARMSTRONG, T., RADWIN, R.G., JOSEPH, B.S., and HANSEN, D.J., 1994, Repetitive trauma disorders: job evaluation and design, *Human Factors Society*, 3, 325–36.

BELLEMARE, M., and RICHARD, J.-G., 1992, Integrating ergonomics into modernization of a slaughterhouse: the implementation of an approach based on an analysis of the real work. In S. Kumar (ed.) *Advances in Industrial Ergonomics and Safety IV*, *Proceedings of the Annual International Industrial Ergonomics and Safety Conference* 92, (London: Taylor & Francis), 59–63.

CHATIGNY, C., and VÉZINA, N., 1995, Analyse du travail et apprentissage d'une tâche complexe; étude de l'affilage du couteau dans un abattoir, *Le Travail Humain*, 3, 229–252.

HAGBERG. M., MORGENSTERN, H., and KELSH. M., 1992, Impact of occupations and jobs tasks on the prevalence of carpal tunnel syndrome, *Scandinavian Journal of Environmental Health*, 18, 337–45.

KURPPA. K., ViIKARI-JUNTURA, E., KUOSMA, E., HUSKONEN, M., and KIVI, P., 1991, Incidence of tenosynovitis or peritendinitis and epicondylitis in a meat processing factory, *Scandinavian Journal of Environmental Health*, 17, 32–7.

LOPPINET, M., and APTEL, M., 1997, Les TMS dans la Filière Viande — revue de littérature, Rapport n° ISSN 0397-4529, Institut National de la Recherche Scientifique.

MERGLER, D., BRABANT, C., VÉZINA, N., and MESSING, K., 1987, The weaker sex? Men in women's working conditions report similar health symptoms, *Journal of Occupational Medicine*, 29, 417–21.

RICHARD, J.-G., 1997, Identification des outils requis pour accroître l'efficacité des stratégies de prévention des lésions attribuables au travail répétitif dans les abattoirs et usines de transformation, Rapport IRSST, A–158.

TOULOUSE, G., 1995, Étude descriptive des déterminants des facteurs de risque de LATR aux postes d'éviscération de deux abattoirs de porcs, Rapport IRSST- Études et Recherches, R–105.

Slip, Trip and Fall Accidents

R. A. Haslam

Health & Safety Ergonomics Unit, Department of Human Sciences, Loughborough University, Loughborough LE11 3TU, UK

1. INTRODUCTION

Slip, trip or fall accidents can happen on the level, on ramps, on steps or stairs, or from a height, with differing causes and consequences. Together they form a major category of accident, imposing a substantial social and economic cost. UK statistics, compiled by the Health & Safety Executive (HSE), show that fall accidents are the principal cause of major injury at work (Figure 1). Using 1993/94 figures, HSE (1996) estimated slip and trip accidents cost employers £300 million each year, with a cost to society > £800 million. Outside work, the UK Home Accident Surveillance System (HASS) results for 1996 showed falls to be 40% of all home accidents, with an annual incidence rate estimated at 1930 per 100 000 of the population (Figure 2).

Although international comparisons are impeded by differences in procedures for reporting and recording, in the USA data from the Bureau of Labor Statistics for 1996 identified that falls were involved in 11% of fatal accidents at work. Similarly, a National Safety Council analysis of "deaths due to unintentional public injuries," again for 1996, listed falls as the largest category, with a rate of 2.0 per 100 000 (excluding incidents involving motor vehicles). A similar pattern exists in other countries: slip, trip and fall accidents are a major problem worldwide.

2. SLIPS

Normal walking (locomotion) goes through a cycle of the supporting leg pushing off from the toe, the other leg swinging forward to heel strike, body weight transferring forwards onto the heel then to the toe, supporting leg pushing off from the toe, and so forth. Slipping occurs when the friction between the foot and the floor is insufficient to prevent movement between the
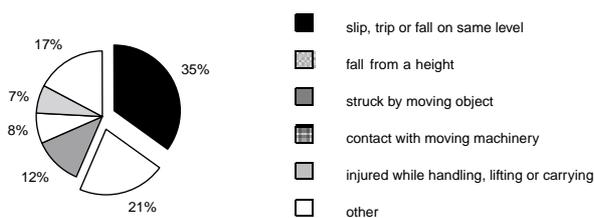
Figure 1. UK Occupational accidents 1995/96.
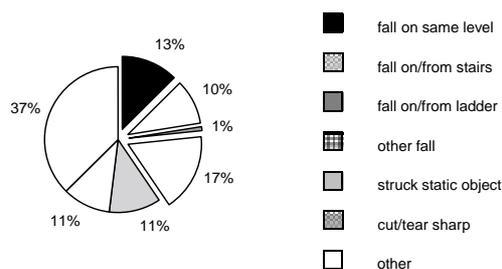
Figure 2. UK home accidents 1996.

two surfaces. Leamon (1992) described three categories of slip: a *microslip* of < 20 mm, a *slip* of between 80 and 100 mm, and a *slide*, > 100 mm. Microslips and slips are less of a problem, as even if detected by an individual, they are likely to be able to maintain balance. Slides may become unrecoverable, leading to body impact and the possibility of injury.

Slipping can happen during either toe off or heel strike phases of walking, although the latter is usually more hazardous. This is because during heel strike, the forward momentum of the body is in the same direction as the slip. Most slips occur where the properties of a walking surface change suddenly, for example due to spillage of a liquid, or when first stepping onto ice. A reduced stride length decreases the horizontal (shear) force component during both toe off and heel strike phases of walking, leading to increased stability, making it possible to maintain balance on slippery surfaces. However, this requires both detection and adjustment to the walking conditions.

2.1. Coefficient of Friction

The coefficient of friction (CoF), is a means of describing the resistance to slipping between an object and a surface, for example a particular synthetic soled shoe and a polished wooden floor. Traditionally, a distinction has been made between static and dynamic CoF. Static CoF is the ratio of the vertical force acting between an object and a surface and the horizontal force just sufficient to start the object to move. Calculating the ratio using the force necessary to keep the object in motion gives dynamic CoF.

Tribologists (those specializing in the study of friction) have devoted some considerable effort to developing techniques for measuring CoF that will provide valid and reliable results. It has been suggested that the number of different methods proposed may be approaching 100. These fall into four broad categories:

- Drag devices – a weight, with a shoe sole sample on its surface, is dragged across the floor, while the force required for slipping is measured, giving either static or dynamic CoF (e.g. the Horizontal Pull Slipmeter and Tortus instruments).
- Pendulum devices – a pendulum arm, end-fitted with a sample of shoe material, is swung across the floor surface, with deceleration of the arm allowing a dynamic CoF to be calculated.
- Articulated strut – a plate, covered with shoe material, attached to an arm, is pressed onto a surface covered with floor material; the floor material is moved at a constant speed, causing the arm to move from the vertical, until slipping occurs, giving a measure of static CoF (e.g. the James Machine).
- Walking tests – with the DIN ramp test, for example, the incline of a ramp is raised until a human subject walking on the test surface begins to slip; a CoF can then be calculated from the ramp angle.

(See Marletta 1991 and Leamon 1992 for further details of methods.) A drawback of walking tests is the need for bulky equipment, preventing measurement of floors in situ. Unfortunately, there is often poor agreement between results from the different approaches, varying according to surface roughness, whether the surface is wet or dry, and the type of shoe material used. Recent thinking has been that neither static nor dynamic CoF are adequate, as slipping occurs during the change of state

between the fully static and fully dynamic conditions. Devices are currently being evaluated and refined that deliberately set out to measure CoF during the non-steady-state period.

In practice, rougher floors tend to have better slip resistance, especially when wet. Materials range from having smooth surfaces such as polished metal and polished wood, through to ceramic and terrazzo tiling, through to rough stone and concrete. Microcellular urethane and rubber materials provide improved slip resistance over PVC and leather.

3. TRIPS

Tripping is less common than slipping and happens when the foot catches on an obstacle or object, with continuing motion of the body resulting in a stumble or fall. Tripping hazards may be fixed features of the environment, a raised paving stone or a step for example, or temporary items, such as a trailing electrical cable. Marletta (1991) suggested irregularities of as little as 7 mm may be sufficient to induce tripping. As with slipping, detecting changes in the walking surface or an object in the way allows avoidance.

4. INDIVIDUAL FACTORS

Personal characteristics involved in slip, trip and fall accidents are gait, balance, stature, strength, vision and behavior. These in turn may be moderated by health, age, fatigue, medication, alcohol, environment (e.g. lighting and floor surface) and activity (e.g. load carriage or performance of a cognitive task). Human ability to walk and balance is a complex process, involving integration of inputs from the visual, vestibular and somatosensory systems. Stature influences the location of an individual's center of gravity and inherent stability. Strength affects ability to recover from disturbances to balance, and use of supports such as handrails. Vision, in addition to its contribution to balance, also enables monitoring of the walking surface and detection of obstacles. Behavior may vary in terms of the caution individuals exercise in different circumstances.

5. ORGANIZATIONAL FACTORS

Most research dealing with slip, trip and fall accidents has been from the directions of epidemiology, biomechanics and tribology. However, researchers such as Bentley and Haslam (1998) have highlighted the significant contribution of organizational factors. These influence exposure to hazards and the control of risks. For example, within the Royal Mail postal delivery function, business targets for letter delivery, together with a "job and finish" policy, combined to provide strong informal incentives for staff to carry overweight mail pouches. The carrying of a heavy asymmetric load was identified as a significant risk factor for slip, trip and fall accidents. Considerable changes in equipment, custom and practice are needed to overcome a situation such as this. In another context, slipping accidents can be a problem in supermarkets, due to liquid spillages and loose vegetable products falling on the floor. Preventative measures include providing mats in front of vegetable counters, and good housekeeping, such as cleaning up contaminants as soon as they occur. However, the latter requires a sustained effort devoted to training and motivating staff.

6. FALLS ON THE LEVEL

Falls on the level are the most frequent type of fall, in both occupational and domestic settings, with slips occurring more frequently than trips. For example, Bentley and Haslam (1998) found slips on the level to be almost double the number of trip accidents among postal delivery staff working outdoors. Falls on the level can also happen when a surface moves as it is stepped on, such as with a loose mat, or when a moving surface stops abruptly, for example when standing on a train or moving walkway. Injuries from falls on the level are commonly sprains and strains, followed by contusions, with the wrist, arm, ankle and back being the body areas most often affected.

On ramps, the shear forces between footwear and the floor at toe off and heel strike increase with gradient. The risk of heel slip and falling backwards when descending a ramp is greater than the risk of slipping during toe off and falling forwards on ascent.

Certain manual handling tasks seem likely to increase the risk of fall accidents, although limited information on numbers and severity of incidents makes this difficult to confirm. Load carriage affects center of gravity, potentially making it more difficult to recover from imbalance. During pushing or pulling, shear forces between the feet and the floor can be very high, increasing the chances of slipping. Trolleys and trucks may also obscure vision, impeding monitoring of the walking surface.

7. FALLS ON STEPS AND STAIRS

Accidents on steps and stairs are the second major category of fall accident in terms of numbers. There is general agreement that children and the elderly are at increased risk, although the extent of this is difficult to quantify, due to an absence of data on levels of stair use by different groups. Accidents during descent result in more injuries than for ascent, where the distance traveled in a fall is likely to be less. Fractures are more frequent injury outcomes than for falls on the level, with the elderly particularly at risk. Falls on stairs may involve either a slip, trip or misstep, with the relative occurrence differing between indoor and outdoor environments.

Important features of steps and stairs from the viewpoint of safety are riser and going (tread) dimensions, design of nosing, length of flight, nature and condition of surface material, position of handrail, and lighting. Templer (1992) recommended riser dimensions should be within 117–183 mm, with goings between 279 and 356 mm. User behavior that may lead to accidents includes rushing, carrying items and leaving objects on stairs. Templer (1992) also identified dimensional irregularities between adjoining steps and content of the visual field as other possible contributory factors.

8. FALLS FROM A HEIGHT

In the UK, falls from a height > 2 m are the largest cause of fatal accidents at work (20.6% in 1995/96) with the construction industry particularly afflicted. Examples include falls from or through roofs, from scaffolding, off ladders, through windows, from machinery, or anywhere where there is a sudden change in floor level or where it is necessary to climb onto something. Because of the forces dissipated through the body on impact, accidents of this nature often lead to serious multiple injuries and death.

Human factors issues are a frequent ingredient in these accidents, either with regard to selection and use of equipment or the decision to venture into unsafe situations. Ladders are often used when scaffolding would be more appropriate. Accidents occur because scaffolding has not been erected properly or tied in sufficiently. Workers go on to fragile roofs without crawling boards. Employees climb up and stand on roller conveyer systems to reach items stored above. Many of these accidents are avoidable.

9. IMPROVING SAFETY

The most effective approach to reducing slip, trip and fall accidents is to eliminate hazards. Ideally this should be done at the design stage, for example by:

- Choosing flooring that offers good slip resistance and that retains this property with cleaning and use.
- Designing systems of work, workplaces and equipment to minimize contamination of walkways and steps with liquids, powders or dry items, such as polythene bags or paper.
- Designing systems of work, workplaces and equipment to prevent trip hazards occurring on walkways.
- Covering outside areas, to keep off rain and ice.
- Avoiding steps and stairs in buildings used by older age groups, or providing easy to access alternatives, such as lifts (elevators).
- Incorporating permanent maintenance access to buildings, machinery and equipment that does not require use of ladders.

Remedial approaches include:

- Performing risk assessments to identify slip, trip and fall risks and implementing measures for their control.
- Ensuring use of suitable footwear.

- Providing appropriate equipment for manual handling and other tasks.
- Cleaning up spillages, removing obstacles, detecting and repairing damage to flooring and equipment, at an early stage.
- Making sure there is sufficient lighting to aid detection of hazards.
- Marking trip hazards that can not be removed; using changes of floor appearance (e.g. color) to delineate steps, stairs, etc.
- Erecting barriers around temporary holes and on roof edges when work is taking place.
- Using correct materials for floor cleaning and preventing access to areas while this is in progress.
- Providing information and training.

Many of these secondary approaches require significant, ongoing commitment at the organizational level. The difficulty of achieving this should not be underestimated.

REFERENCES

BENTLEY, T.A. and HASLAM, R.A., 1998, Slip, trip and fall accidents occurring during the delivery of mail. *Ergonomics*, 41, 1859–72.

HEALTH & SAFETY EXECUTIVE, 1996, *Slips and Trips* (Sudbury: HSE Books).

LEAMON, T.B., 1992, The reduction of slip and fall injuries: Part II – The scientific basis (knowledge base) for the guide. *International Journal of Industrial Ergonomics*, 10, 29–34.

MARLETTA, W., 1991, Trip, slip and fall prevention. In Hanson, D. (ed) *The Work Environment*, vol. 1: *Occupational Health Fundamentals* (Michigan: Lewis), pp. 241–76.

TEMPLER, J., 1992, *The Staircase: Studies of Hazards, Falls and Safer Design* (Cambridge, MA: MIT Press).

Slips and Falls

W.-R. Chang

Liberty Mutual Research Center for Safety and Health, Hopkinton, MA, USA

R. Grönqvist

Finnish Institute of Occupational Health, Laajaniityntie 1, Vantaa, Finland

1. INTRODUCTION

Slips and falls are a serious problem and are among the most common accidental injuries in the world, although they have received little public awareness and concern. There were 14 900 deaths in the US due to fall accidents in 1997 (National Safety Council 1998). Risky situations occur typically during normal locomotion and during manual exertion like lifting, pushing, pulling, or load carrying.

Slips and falls are also a complicated problem. The risk of slips and falls is governed by human perception, the control of human posture and balance, and the frictional characteristics of the shoe and floor contact interface. The disciplines involved include epidemiology, biomechanics, tribology, and psychology. Epidemiology helps to identify the seriousness and causes of accidents. Psychology and biomechanics help to perceive floor condition, develop the walking strategies over different floor conditions, and understand human response during slippery conditions and during a potential loss of balance. Tribology helps to evaluate whether floors and footwear are slippery and to identify the mechanisms involved in accidents. The causality of slips and falls is not yet fully understood. Before effective prevention strategies can be put into practice, one must clarify the accident and injury mechanisms involved. The chain or network of events, constituting exposure to hazards, initiation of hazardous incidents, and final injury, should be identified using epidemiological principles and methodologies.

The primary extrinsic risk factor for falls initiated by slipping is by definition poor grip or low friction between the footwear and floor. The question of whether the risk of falling and injury initiated by a slip is more related to a constantly low friction or an unexpected, sudden change in friction still remains to be solved. Other risk factors for slips and falls are related to insufficient lighting, uneven surfaces, inappropriate design of stairs and floors, poor housekeeping, load carrying, inadequate control of posture, and aging.

2. EPIDEMIOLOGY OF FALLS

The economic and human dimensions of slip and fall injuries at work, at home, and during leisure-time activities are overwhelming. According to the data published by the National Safety Council (1998), falls represented about 15% of all unintentional-injury deaths and 21% of all hospitalized injuries in the US in 1995. The data also showed that the probability of unintentional-injury deaths due to slips and falls was approximately equal for males and females. The average unintentional-injury death rate due to slips and falls was approximately 1165 per month in the US in 1995 with a slightly lower rate in June, a slightly higher rate in December, and a coefficient of monthly variation of 4.2%. Slips and falls are also a leading cause of accidental deaths at home. There were 9200

deaths due to slips and falls at home which accounted for approximately 32.4% of the total unintentional-injury deaths at home in the US in 1997.

The unintentional-injury deaths due to slips and falls in the US were the highest in 1943 (24 701 deaths) and were the lowest in 1986 (11 444 deaths) (National Safety Council 1998). However, the unintentional-injury deaths due to slips and falls have been on the rise since 1986. There was a 27% increase from 1987 to 1997 and a 2% increase from 1996 to 1997. Deaths due to slips and falls accidents per 100 000 population were the highest in 1936 and 1943 (18.4) and the lowest in 1986 and 1987 (4.8). However, the death rate has also been on the rise since 1987. There was a 17% increase in death rate from 1987 to 1997 and a 2% increase from 1996 to 1997.

Slips and falls are also a leading cause of unintentional-injury death for the older population (National Safety Council 1998). Among the unintentional-injury deaths due to slips and falls, 79% (11 057 deaths) were over the age of 65 with a peak of 2 670 deaths within the age group between 85 and 89 in 1995.

The injuries with lost time due to slips and falls averaged in excess of $13 096 per workers' compensation claims filed in 1995 and 1996 (National Safety Council 1998). Slips and falls also accounted for 20% of non-fatal occupational injuries involving days away from work in 1996. Among nine (9) industries evaluated by Leamon and Murphy (1995), the direct costs of occupational injuries due to slips and falls were the highest for the construction, restaurant, and clerical industries. Meanwhile, slips and falls were the second highest source of losses for manufacturing, trucking, retail and wholesale stores, healthcare, food products manufacturing, and professional drivers. The incidence rates for falls and workers' compensation claims were found to be highest for young (less than 25 years) and old (over 65 years) workers. Falls accounted for 17% of all work-related injuries and 12% of worker fatalities in the US (Leamon and Murphy 1995). The common perception of fall injuries might be related to falls from elevation. Though falls from elevation usually result in a higher claim cost, falls on the same level accounted for 65% of claim cases and, consequently, 53% of claim cost in the total direct workers' compensation for the occupational injuries due to slips and falls.

3. PERCEPTION OF SLIPPERINESS

The risk of slips and falls seems to depend on a person's subjective judgement of the potential slipperiness of actual floor conditions. There have been some speculations that it is easier to adapt one's gait when a slippery condition is consistent than when rapid and unexpected changes in slipperiness occur. Experimental support for this argument can be found in a study by Strandberg (1985) in which he observed that a consistent slippery condition led to fewer falls than a borderline somewhat slippery condition. However, no accident statistics is available to confirm this argument. On the contrary, Merrild and Bak (1983) found that certain high-risk winter days can cause over tenfold increase of pedestrian injuries due to falls initiated by slipping.

Vision may be the only sensory mode which allows a person to predict the potential slipperiness of a surface before stepping onto it. However, vision may also play an important role in the activation of other postural control mechanisms. Proprioceptor (e.g. stretch reflexes) and other receptors (e.g. pressoreceptors)

require that one already has walked on a slippery surface and felt it in order to acquire the feedback to adapt one's gait properly. Visual control can lead to several avoidance and accommodation strategies of gait when challenging conditions are encountered. Vision is known to regulate step length and width, direction of gait, walking velocity, and orientation of limbs, etc. (Patla 1991), but visual control cannot be solely relied upon in unexpected perturbations such as slips and trips since it is slower than vestibular or proprioceptive control.

3. BIOMECHANICS OF SLIPS AND FALLS

3.1. Control and Adaptability of Gait

Several sensomotoric systems are used by humans to maintain upright static posture and dynamic balance during locomotion. The central nervous system rapidly and accurately processes the sensory input from vestibular organs, vision, proprioceptors, and exteroceptive pressoreceptors. When posture and balance are challenged during a sudden slip or trip, a coordinated neuromuscular motor response is needed to re-establish the balance for avoiding a fall.

Vestibular influx normally governs 65% of the body sway during sudden perturbation, while 35% is accounted for by visual and proprioceptive influx (Pyykkö *et al.* 1990). Stretch reflexes in joints and muscles and tactile cues from pressoreceptors, which operate in the early prevention of falls, are important protective responses to unexpected perturbations. Aging, which contributes to an increased risk of slipping and falling, is one of the main reasons for defective coactivation of functional stretch reflexes. The very elderly rely mostly on visual influx (latency 120–200 ms) which is slower than other mechanisms for control of balance (Pyykkö *et al.* 1990). Latencies for corrective reflex responses for healthy young men in the recovery from tripping have been reported to be much quicker (60–140 ms) indicating that balance control during locomotion involves specific movement patterns in response to a tripping perturbation (Eng *et al.* 1994).

Certain protective gait adaptations are aimed at regulating gait in hazardous situations, for instance slippery conditions. Humans can adapt to walking continuously over very slippery surfaces due to such adaptations which take place in essentially a one-step cycle after one has become aware of the slipperiness of the particular surface being approached. Llewellyn and Nevola (1992) reported the results of trials where subjects adapted their gait under low-friction conditions by adopting such a protective strategy. The strategy involved a combined effect of force and postural changes of the early stance. The subjects took shorter steps and increased their knee flexion, which, in combination, reduced the vertical acceleration and the forward velocity of the body.

3.2. Critical Gait Phases

Five (5) major motor functions are used to achieve safe and efficient propulsion of the body during the gait cycle regardless of the pace of locomotion (Winter 1991). The first three of these functions are to maintain the support of the upper body during stance, to maintain the upright posture and balance of the total body, and to control the foot trajectory in order to achieve safe ground clearance and gentle heel or toe landing. The other two major motor functions are to generate mechanical energy to maintain or increase the forward velocity, and finally to absorb the mechanical energy for shock absorption and stability or to decrease the forward velocity of the body.

Heel strike and toe-off are two critical gait phases in walking for slips and falls accidents. A forward slip on the leading foot typically occurs during the heel contact, while a backward slip on the sole forepart typically occurs during the toe-off. Since the weight on the trailing foot is being rapidly transferred to the heel of the leading foot during the toe-off phase, a slip during the push-off is considered less dangerous than a slip at heel landing. The forward slip starting at back edge of the heel during the heel landing phase would very likely result in a more dangerous fall because the body weight is being transferred to the leading foot.

The biomechanical dynamics of human heel slipping was investigated by Strandberg and Lanshammar (1981) when subjects approached a force plate covered with mixtures of water and detergent. They observed that the peak sliding velocities after the heel strike were above walking speed (1 to 2 ms^{-1}) during slips that resulted in a fall, but were usually below 0.5 ms^{-1} in slips where the subjects were able to regain their balance and avoided a fall. They concluded that a slip was likely to result in a fall if the sliding exceeded 0.1 m in distance or 0.5 ms^{-1} in velocity. The average critical slip motion during their experiments started 50 ms after the heel strike when the rear edge of the heel was in contact with the floor and the vertical load was about 60% of body weight. They also reported that a slight sliding between the footwear and floor right after the heel strike occurred often, even in dry non-slippery conditions, but that the subjects were mostly unaware of the slight movement.

Morach (1993), on the contrary, found in human slipping experiments on contaminated (oil, glycerin, and water) floors that the horizontal foot velocity in the forward direction prior to heel contact varied between 0.3 ms^{-1} and 2.75 ms^{-1}. The highest velocity after slip initiation was 2.5 ms^{-1} and occurred on a steel floor with oil as a contaminant, when there was an immediate slip following the heel strike. The average walking speed was approximately 1.5 ms^{-1} in all the above experiments. Morach's results indicate that a higher minimum sliding velocity (1.0 ms^{-1}) than that proposed by Strandberg and Lanshammar should be used when assessing the slip resistance properties of shoes and floors, particularly in the presence of oil. His findings also suggest a heel contact angle of 15 to 20 degrees for simulation of slipping at the heel and a short contact time of 30 to 60 ms.

The risk of a slip-and-fall accident is considered low when the friction coefficient (mm) is greater than the ratio of the horizontal (F_H) and vertical (F_V) force components applied to the ground, i.e.

$$F_H < F_V$$

Consequently, if the magnitudes of the horizontal and vertical force components applied to the ground during walking and the friction coefficient for the actual tribosystem are known, it would be possible to evaluate whether the system is potentially slippery or not. Strandberg and Lanshammar (1981) measured the friction use peak, F_H / F_V, approximately 0.1 s after the heel strike. The average peak was 0.17 when there was no skidding between the shoe and floor, 0.13 when the subject was unaware of the sliding motion or regained balance, and 0.07 when the skid actually resulted in a fall. Dynamic friction appeared to be more

important than static friction, because in most of their experiments the heel slid upon heel contact even without any contaminant.

4. FRICTION MECHANISMS FOR SHOE AND FLOOR INTERFACE

Friction has been widely used as an indicator of floor slipperiness. Moreover, friction at the interface between the footwear and floor is very complicated. Factors such as the footwear and floor materials, contaminants at the interface, temperature, humidity, surface roughness on footwear and floor, contact force, relative velocity at the interface, and patterns on the footwear and floor surfaces affect friction. As pointed out by Tabor (1974), the major friction mechanisms involved at a dry interface are the plowing of asperities of the harder solid surface through the softer surface and the shearing of the adhesive bonds formed at the region of real contact. When a polymer slides over a rough hard surface, some contribution to the friction will arise from the deformation of the polymer by the hard asperities. The main mechanisms for friction due to polymer deformation include tearing and hysteresis due to the viscoelastic properties of the polymer. The energy dissipation due to deformation mainly occurs at a small depth below the surface where the maximum shear stresses occur. The main sources of the adhesion are electrostatic and van der Waals' forces for the polymers. The contact area between polymers and a hard surface is greater than that predicted by the Hertz theory where no adhesion was assumed (Tabor 1974). When the polymer adheres to another surface, work must be expended in order to produce sliding since the adhesions at the interface must be broken. The formation and rupture of adhesive bonds increase the hysteresis loss of the polymer and, thus, increase the friction. An equation to predict friction coefficient based on the contribution from adhesion and hysteresis for elastomers on dry surfaces is available in the literature (Moore 1972).

The hydrodynamic lubrication and boundary lubrication are additional friction mechanisms involved when liquid contaminants are present at the interface. The surfaces of shoe and floor are completely separated by the contaminant under the hydrodynamic lubrication, while the normal load is supported by the contaminant and some direct solid to solid contact under the boundary lubrication. In the classical generalized Reynolds' equation, the generation of hydrodynamic pressure in the lubricant film is governed by wedge, stretch, and squeeze terms (Moore 1972). Hence, Reynolds' theory takes into account the vertical squeezing motion and the tangential sliding motion. The effect of the squeeze and wedge terms has been widely discussed (Moore 1972). The stretch contribution has not been discussed extensively, although it is considered important for viscoelastic materials such as rubbers and elastomers.

The squeeze-film effect governs the hydrodynamic region as indicated by Moore (1972). The formation of hydrodynamic pressure to separate two surfaces requires a sufficient relative velocity, surface smoothness, high viscosity of contaminant, and a low contact force at the interface. Therefore, an increase in surface roughness on either shoe or floor, pattern on shoe or floor to reduce the amount of contaminants at the interface, and a reduction in conformity between the shoe and floor can certainly help increase friction through a reduction in the squeeze-film region.

The independent equations for the squeeze and wedge terms

are also available in the literature (Moore 1972). For the squeeze effect, the thickness of contaminant at the interface is affected by the viscosity of contaminant, the contact area, the normal force and the descend time of the footwear. For a given contaminant, a shorter drainage time to reduce the risk of slips and falls accidents can be achieved by an increase in the normal contact force, and a reduction in the contact area and the viscosity of contaminant. For the wedge effect, the thickness of contaminant at the interface is affected by the contaminant viscosity, dimension of the contact area, the sliding velocity and the normal force. To minimize the contaminant thickness at the interface for obtaining good contact and grip between the shoe and floor surfaces, it is critical to reduce the dimension of the contact area and the viscosity of contaminant, and to increase the normal force.

5. MEASUREMENT OF SLIP RESISTANCE

Both static and dynamic (kinetic) friction coefficients have been widely used to indicate the slip resistance between the shoe and floor surfaces, but there has been continuing debate regarding how to measure it properly. Static friction has been considered important for slip initiation, while kinetic friction represents the resistance to continuation of a slip after initiation. Support for both static and dynamic friction coefficients has been reported.

Static friction is limited to the slip start before any detectable motion, while a fully developed slip condition may require the determination of steady-state dynamic friction produced only after some time delay from the slip start. The transitional period after the slip initiation is typically very short (less than 250 ms) when slipping at heel strike, and it seems to reflect the squeeze-film phenomena during early stance phase in a foot forward slip on contaminated surfaces.

Drag or towed-sled devices, pendulum strikers, articulated-strut devices, braked-wheels or skiddometers, turntables or rotating discs, and gait simulators are often referred to in the context of slip resistance measurement. The output quantity according to these different principles is typically force, torque, loss of energy, inclination angle, or rolling resistance. Among a variety of techniques currently available for measuring slip resistance of footwear and floors, Strandberg and Lanshammar (1981) found surprisingly little support in their biomechanical skidding data for the most common measurement principles.

A dynamic loading test condition, typical for the heel strike in normal walking, has received too little attention in many devices for slip resistance measurements. Only the pendulum strikers, some gait simulators, and some articulated-strut devices produce an impact at the moment when friction coefficient is measured. However, the impact forces produced are mostly not well defined and do not correspond to normal gait, where the heel touch-down is characterized by collision-type contact forces (Cappozzo, 1991).

The measurement parameters and their ranges should reflect the biomechanics and tribophysics of actual slipping incidents. As indicated by Grönqvist (1999), the condition at the floor and shoe interface should satisfy the following criteria: a normal force build-up rate of at least 10 kNs⁻¹, a normal pressure from 0.1 to 0.6 MPa, a sliding velocity less than 1 ms⁻¹, and a contact time prior to and during the friction coefficient computation between 50 and 800 ms.

Friction coefficient limit values should be correlated to the

normal variability of human gait characteristics. As summarized by Grönqvist (1999), the safety criteria for friction coefficient reported in the literature has been between 0.15 and 0.6, depending on biomechanical factors, such as walking speed, stride length, turning, pushing/pulling, stopping and anthropometric parameters, and environmental factors, such as ramps and stairs. However, friction coefficient limit values should always be regarded as rough guidelines since the safety criteria reported in the literature are very much depending on biomechanical factors as well as the normal variability of human gait. It is not yet fully understood whether preferably a high enough friction coefficient or merely a consistent friction coefficient would be beneficial for safe walking.

Human subjects can differentiate the slipperiness of floors and footwear in various dry, wet, and contaminated conditions. Purely subjective methods, such as paired comparison tests, or partly subjective and partly objective methods, such as ramp tests, and the measurement of slip distance during slipping incidents have been used to assess slip resistance as summarized by Grönqvist (1999).

6. CONCLUSIONS

Despite the complexity of the problems in slips and falls, some interventions could be quite simple such as a proper housekeeping cleaning protocol for the floor surfaces, sufficient lighting, a promotion of the awareness of the problems and periodic replacement of worn footwear. More elaborate interventions could include a proper selection of floor surface, floor treatments, a proper selection of footwear with an adequate tread pattern design and heel geometry, and employee training. However, one should bear in mind that effective interventions also depend on the type of contaminants involved. Successful interventions for one contaminant may not be successful for other types of contaminants.

Several strategies can be used to reduce slips and falls accidents. Current slip resistance measurement techniques (either objective or subjective methods) need further improvement and need to be validated by biomechanical evaluation in conditions for which they are intended to be applied. Expert systems should be used to assist designers of floorings and footwear in leading to safer products and to assist safety consultants for intervention development. Effective housekeeping interventions at workplaces and elsewhere should be implemented. Safety standards for both floorings and footwear, such as contaminant and task related standards, should be developed and adequate selection guidelines for the users of such products should be established.

REFERENCES

CAPPOZZO, A., 1991, The mechanics of human walking, in Patla, A.E. (ed) *Adaptability of Human Gait. Implications for the Control of Locomotion* (Amsterdam: Elsevier Science), pp. 167–86.

ENG, J.J, WINTER, D.D, and PATLA, A.E., 1994, Strategies for recovery from a trip in early and late swing during human walking, *Experimental Brain Research*, 102, 339–49.

GRÖNQVIST, R., 1999, Slips and falls, in S. Kumar (ed.) *Biomechanics in Ergonomics* (London : Taylor & Francis), pp. 351–75.

LEAMON, T.B. and MURPHY, P.L., 1995, Occupational slips and falls: more than a trivial problem, *Ergonomics*, 38, 487–98.

LLEWELLYN, M.G.A. and NEVOLA, V.R., 1992, Strategies for walking on low-friction surfaces, in Lotens, W.A., Havenith, G. (eds) *Proceedings of the Fifth International Conference on Environmental Ergonomics*, Maastricht, pp. 156–7.

MERRILD, U. and BAK, S., 1983, An excess of pedestrian injuries in icy conditions: a high-risk fracture group — elderly women, *Accident Analysis & Prevention*, 15, 41–8.

MOORE, D.F., 1972, The friction and lubrication of elastomers, in Raynor, G.V. (ed.) *International Series of Monographs on Material Science and Technology*, Vol. 9, 288 pp., (Oxford: Pergamon).

MORACH, B., 1993, Quantifierung des Ausgleitvorganges beim menschlichen Gang unter besonderer Berücksichtigung des Aufsetzphases des Fusses, Fachbereich Sicherheits-technik der Bergischen Universität — Gesamthochschule Wuppertal, 181 pp., Wuppertal (in German).

National Safety Council, 1998, *Accident Facts* (1998 ed.) (Itasca., IL, USA).

PATLA, A.E., 1991, Visual control of human locomotion, in Patla, A.E. (ed.) *Adaptability of Human Gait: Implications for the Control of Locomotion* (Amsterdam: Elsevier Science) pp. 55–97.

PYYKKÖ, I., JÄNTTI, P., and AALTO, H., 1990, Postural control in elderly subjects, *Age and Ageing*, 19, 215–21.

STRANDBERG, L., 1985, The effect of conditions underfoot on falling and overexertion accidents, *Ergonomics*, 28, 131–47.

STRANDBERG, L. and LANSHAMMAR, H., 1981, The dynamics of slipping accidents, *Journal of Occupational Accidents*, 3, 153–62.

TABOR, D., 1974, Friction, adhesion and boundary lubrication of polymers. In Lieng-Huang Lee (ed.) *Advances in Polymer Friction and Wear, Polymer Science and Technology* (London/New York: Plenum), Vol 5A, pp. 5–30.

WINTER, D.A., 1991, The biomechanics and motor control of human gait: normal, elderly, and pathological, Second edition, 143 pp., University of Waterloo, Canada.

The Strain Index

J.S. Moore† and A. Garg‡

†Department of Nuclear Engineering, Texas A&M University, 130B Zachry Engineering Center, College Station, TX 77843-3133, USA

‡Deaprtment of Industrial & Manufacturing Engineering, University of Wisconsin — Milwaukee, EMS E314, PO Box 784, Milwaukee, WI 53201, USA

1. INTRODUCTION

The Strain Index is a semi-quantitative job analysis methodology based upon principles of physiology, biomechanics and epidemiology (Moore and Garg 1995). Its purpose is the identification of jobs that place workers at increased risk of developing disorders in the distal upper extremity (elbow, forearm, wrist, hand). Application of the Strain Index (SI) methodology results in a numerical score (the SI score) that, based on interpretation guidelines, predicts whether a job exposes workers to increased risk of developing distal upper extremity disorders, i.e. if the job is a "problem."

Consistent with physiological, biomechanical and epidemiological principles, the SI methodology is based on multiplicative interactions among its task variables. The SI score represents the product of six multipliers that correspond to six task variables. The six task variables include intensity of exertion; duration of exertion; exertions per min; hand/wrist posture; speed of work; and duration of task per day. Intensity of exertion, hand/wrist posture and speed of work are estimated. Duration of exertion, exertions per min and duration of task per day are measured. Based on these estimated or measured data, each of the task variables is rated according to five ordinal levels using Table 1. The user finds the column heading corresponding to the appropriate task variable, moves down to the appropriate row within that column, then follows the row to the first column on the left hand side to identify the appropriate rating. The multipliers for each task variable are determined from the ratings using Table 2. The user finds the column heading corresponding to the appropriate task variable and the row corresponding to the appropriate rating, then identifies the multiplier at the intersection of the task variable column and rating row. The SI score is the product of the six multipliers.

2. STRAIN INDEX TASK VARIABLES

2.1. Intensity of Exertion

Intensity of exertion is an estimate of the force requirements of a task. It reflects the magnitude of muscular effort, in the context of percent maximum strength, required to perform the task one time. It reflects physiological and biomechanical stresses on the muscle–tendon units of the distal upper extremity related to the magnitude of tensile loading. It does not reflect stresses related to endurance or stamina.

Since tensile load cannot be measured directly *in vivo* and measurement of applied force with the hand is impractical or difficult in an industrial setting, intensity of exertion is generally estimated using verbal descriptors (Table 1). This is actually an estimate of perceived exertion similar to using the Borg CR-10 scale. To estimate the intensity of exertion, a job analyst or ergonomics team observe a worker (or workers) perform the job, then select the verbal descriptor from Table 1 that best corresponds to their perception of the intensity of exertion. Perceived exertion may also be obtained from the worker(s). Differences among estimates can be resolved by consensus if more than one job analyst is involved. The intensity of exertion rating value (1, 2, 3, 4 or 5) is then assigned using Table 1. The intensity of exertion multiplier corresponding to this intensity of exertion rating is identified using Table 2.

Table 1. Ratings for the Strain Index task variable are assigned by finding the row within the column that corresponds to the datum for each task variable, then recording the rating value listed in the first column (left hand side).

Rating	Intensity of Exertion	Duration of Exertion (% of cycle)	Efforts/ Minute	Hand/wrist Posture	Speed of Work	Duration per Day (hrs)
1	Light	<10	<4	Very Good	Very Slow	≤ 1
2	Somewhat Hard	10 - 29	4 — 8	Good	Slow	1 - 2
3	Hard	30 - 49	9 — 14	Fair	Fair	2 - 4
4	Very Hard	50 - 79	15 — 19	Bad	Fast	4 - 8
5	Near Maximal	≥ 80	≥ 20	Very Bad	Very Fast	≥ 8

Table 2. Multipliers for each Strain Index task variable are determined by finding the intersection of the appropriate rating row with the appropriate task variable column.

Rating	Intensity of Exertion	Duration of Exertion	Efforts/ Minute	Hand/Wrist Posture	Speed of Work	Duration per day
1	1	0.5	0.5	1.0	1.0	0.25
2	3	1.0	1.0	1.0	1.0	0.5
3	6	1.5	1.5	1.5	1.0	0.75
4	9	2.0	2.0	2.0	1.5	1.0
5	13	3.0[A]	3.0[A]	3.0	2.0	1.5

A If duration of exertion is 100%, then efforts/minute multiplier should be set to 3.0.

2.2. Duration of Exertion

Duration of exertion reflects the physiological and biomechanical stresses related to how long an exertion is maintained. It is characterized as the percentage of time an exertion is applied per cycle. In the SI methodology, the terms "cycle" and "cycle time" refer to exertional cycle and average exertional cycle time respectively, not production cycle or production cycle time.

To measure exertional cycle time, the job (or a representative videotape of the job) is observed for a sufficient period so that a job analyst or ergonomics team can obtain a reasonable representation of the job requirements. In general, this should be several complete job cycles (the more the better). The duration of the observation time and the duration of exertion time are measured with stopwatches. The percentage duration of exertion is calculated by dividing the exertion time by the observation time, then multiplying by 100. The calculated percentage duration of exertion is compared with the ranges in Table 1 to determine the duration of exertion rating. The duration of exertion multiplier corresponding to this duration of exertion rating is identified using Table 2.

2.3. Efforts per Minute

Efforts per min refers to the number of exertions applied by the hands per min. Efforts per min are measured by counting the number of exertions that occur during a representative observation period (measured in min), then calculated by dividing the total number of exertions by the observation time. The calculated result is compared with the ranges in Table 1 to determine the efforts per min rating. The efforts per min multiplier corresponding to the efforts per min rating is identified from Table 2.

2.4. Hand/Wrist Posture

Posture refers to the anatomical position of the wrist or hand relative to neutral position. Even though posture has not been associated with the occurrence of distal upper extremity disorders in epidemiological studies, it is likely relevant from a biomechanical perspective (Moore 1996, 1997). Specifically, shear forces at tendon sheaths are generated when loaded tendons turn corners; therefore, non-neutral hand/wrist postures may be relevant when considered in combination with the temporal pattern of tensile tendon loads.

Similar to intensity of exertion, average hand/wrist posture during exertions is estimated qualitatively by one or more observers (Table 1). If there is more than one job analyst, it is recommended that differences in ratings be resolved by consensus. The hand/wrist posture rating is assigned by comparing the verbal anchor to Table 1. The hand/wrist posture

multiplier corresponding to the hand/wrist posture rating is identified using Table 2.

2.5. Speed of Work

Speed of work is an estimate of the perceived pace of the task or job. It is included because of its modifying effects on exertions, i.e. maximal voluntary strength decreases and EMG amplitude increases with increasing speed. In addition, it is suspected that a worker's muscles do not fully relax between high speed, high frequency exertions.

A job analyst or ergonomics team subjectively estimates speed of exertion. If there are multiple job analysts, it is recommended that differences in ratings be resolved by consensus. Once a verbal anchor is selected, the speed of work rating is assigned according to Table 1. The speed of work multiplier corresponding to the speed of work rating is identified using Table 2.

2.6. Duration of Task per Day

Duration of task per day reflects the total time that a task is performed per day. It attempts to incorporate the beneficial effects of task diversity, i.e. job rotation (if the differing task is associated with reduced stresses), and the adverse effects of prolonged activity, such as overtime. Duration of task per day (expressed as hours) can be measured or recorded from authoritative sources, e.g. worker/supervisor reports of shift duration or administrative records. The duration per day rating is assigned according to Table 1. The duration per day multiplier corresponding to the duration per day rating is identified using Table 2.

3. STRAIN INDEX SCORE

The SI score is the product of all six multipliers:

SI = (intensity of exertion multiplier) × (duration of exertion multiplier) × (exertions per min multiplier) × (posture multiplier) × (speed of work multiplier) × (duration per day multiplier). (1.1)

An example demonstrating the calculation of a SI score is demonstrated in Table 3.

4. INTERPRETATION OF THE STRAIN INDEX SCORE

Initial guidance for interpretation of the SI score was based on data from a prior study (Moore and Garg 1994). In that study, a "positive" job was a job noted to have workers with reported distal upper extremity morbidity during a 20-month observation, while a "negative" job had no reported disorders during the same period. The distal upper extremity morbidity included specific

Table 3. An example to demonstrate the procedure for calculating the SI score.

	Intensity of Exertion	Duration of Exertion	Efforts/ Minute	Posture	Speed of Work	Duration per Day
Exposure Data	Somewhat Hard	60%	12	Fair	Fair	4–8
Ratings	2	4	3	3	3	4
Multipliers	3.0	2.0	1.5	1.5	1.0	1.0

SI score = 3.0 x 2.0 x 1.5 x 1.5 x 1.0 x 1.0 = 13.5

disorders (epicondylitis, DeQuervain's tenosynovitis, trigger finger, trigger thumb, CTS) and non-specific disorders. The latter category was assumed to represent strain-related disorders of the muscle–tendon units.

Among "positive" job categories, the SI scores ranged from 4.5 to 81.0 with a mean of 29.0. Among "negative" job categories, the SI scores ranged from 0.5 to 4.5 with a mean of 2.3. The difference between the two means was statistically significant ($t = 4.05$; d.f. = 23; $p < 0.01$). There was only one overlapping SI score between these two groups — 4.5. Based on these results, preliminary interpretation guidelines were that a job with a SI score < 5 probably represented a "safe" job while a job with a SI score > 5 probably represented a "problem" job. If more conservative criteria are desired, a job with a SI score < 3 is probably "safe," a job with a SI score > 7 is probably a "problem," and a job with a SI score between 3 and 7 cannot be reliably classified.

5. PREDICTIVE VALIDITY OF THE STRAIN INDEX

The primary strategy to evaluate the SI's predictive validity involves comparing the SI's dichotomous hazard prediction for a set of jobs against a dichotomous morbidity classification of the same jobs. Jobs were classified to "problem" or "safe" based on whether the SI score was greater than or less than 5 respectively. Jobs were classified as "positive" or "negative" based on whether there were distal upper extremity disorders reported by workers assigned to the job during a retrospective observation period. Using a 2×2 table, the degree of association between the SI score predictions and the observed morbidity were made. In addition, the sensitivity, specificity, positive predictive value and negative predictive value could be calculated.

The original paper reported the results of such a procedure for 25 jobs from a pork processing plant. There were 12 "positive" jobs and 13 "negative" jobs. Of the 12 "positive" jobs, 11 were predicted to be a "problem" and one "safe." Of the 13 "negative" jobs, all were predicted to be "safe." These values correspond to a sensitivity of 0.92; specificity of 1.00; positive predictive value of 1.00; and negative predictive value of 0.93. This validation was considered preliminary because the authors were not blinded to the health outcomes when exposure was evaluated with the SI and the hazard predictions made.

Since the original publication, the SI's utility as a job analysis and solution development tool was demonstrated in a project that involved the analysis of several problem jobs in a red meat packing plant by two ergonomics teams (Moore and Garg 1997). More recently, 30 jobs from two manufacturing facilities and 29 jobs from a turkey processing plant have been evaluated with the job analysts blinded to health outcomes (Moore 1999, personal communication). In the manufacturing facilities, there were six "positive" jobs and 24 "negative" jobs. Of the six "positive" jobs, all were predicted to be a "problem." Of the 24 "negative"

jobs, 21 were predicted to be "safe" and three predicted to be a "problem." These values correspond to a sensitivity of 1.00; specificity of 0.90; positive predictive value of 0.62; and negative predictive value of 1.00. In the turkey processing facility, there were 22 "positive" jobs and seven "negative" jobs. Of the 22 "positive" jobs, 20 were predicted to be a "problem" and two were predicted to be "safe." Of the seven "negative" jobs, all seven were predicted to be "safe." These values correspond to a sensitivity of 0.91; specificity of 1.00; positive predictive value of 1.00; and negative predictive value of 0.78. When all 84 jobs are combined into one 2×2 table, sensitivity was 0.93; specificity was 0.93; positive predictive value was 0.93; and negative predictive value was 0.93.

Based on this validation strategy, the SI is an effective method for identifying jobs likely to be have workers with reported distal upper extremity disorders.

6. FUTURE WORK

Other aspects of validation, such as inter-rater variability and test–re-test repeatability, are subjects of current research projects sponsored by the National Science Foundation Industry/ University Cooperative Research Center in Ergonomics. Strategies for quantifying intensity of exertion and hand/wrist posture are being explored with an objective of possibly defining explicit mathematical relationships between task variable measurements and multipliers. A theoretical approach for applying the SI to jobs that involve multiple levels of a similar exertion (called complex tasks) and jobs that involve multiple tasks is being developed. In addition, the ability to predict specific disorders is being explored by comparing unique patterns of exertions for specific musculo-tendinous structures with proposed pathogenetic models for specific disorders affecting those structures (Moore 1996, 1997)

REFERENCES

MOORE, J.S., 1996, Proposed pathogenetic models for specific distal upper extremity disorders. In *Proceedings of the International Conference on Occupational Disorders of the Upper Extremities*, 24 October, Ann Arbor, Michigan.

MOORE, J.S., 1997, DeQuervain's tenosynovitis. Stenosing tenosynovitis of the first dorsal compartment. *Journal of Occupational and Environmental Medicine*, 39, 990–1002.

MOORE, J.S. and GARG, A., 1994, Upper extremity disorders in a pork processing plant: Relationships between job risk factors and morbidity. *American Industrial Hygiene Association Journal*, 55, 703–15.

MOORE, J.S. and GARG, A., 1995, The Strain Index: a proposed method to analyze jobs for risk of distal upper extremity disorders. *American Industrial Hygiene Association Journal*, 56, 443–58.

MOORE, J.S. and GARG, A., 1997, Participatory ergonomics in red meat packing plant. Part 2: Case studies. *American Industrial Hygiene Association Journal*, 58, 498–508.

Surveillance for Work-related Musculoskeletal Disorders

K. A. Grant

US Air Force, Institute for Environment, Safety and Occupational Health Risk Analysis, Ergonomics and Hearing Conservation Branch, Brooks AFB, TX 78235–5123, USA

1. INTRODUCTION

Surveillance is classically defined as, "the ongoing systematic collection, analysis and interpretation of health and exposure data in the process of describing and monitoring a health event" (CDC 1988). Perhaps more simply, surveillance seeks to identify patterns of health and disease among groups of workers and the risk factors that influence these trends. In recent years, establishing surveillance procedures has been recognized as a necessary first step in the establishment of an industrial ergonomics program (OSHA 1990). There is growing recognition that poorly designed workplaces are a major contributor to many occupational injuries and illnesses, including work-related musculoskeletal disorders. Surveillance determines the need for occupational safety and health action, and provides the data necessary to direct prevention efforts and assess the efficacy of ergonomic interventions.

2. FUNDAMENTALS OF SURVEILLANCE

Surveillance activities can be divided into two broad categories: (1) health or medical surveillance and (2) hazard surveillance. Health or medical surveillance activities seek to identify established or suspected cases of occupational injury or illness for follow-up. This approach can be viewed as reactive, i.e. the occurrence of illness or injury is the driver for further action. Hazard surveillance monitors and characterizes exposures to physical hazards associated with injury and illness in the workplace. This approach is more proactive, and can lead to the identification and elimination of hazards before illnesses occur. In reality, health and hazard surveillance activities are complementary, and the best musculoskeletal disorder prevention programs usually incorporate both.

2.1. Monitoring Injury and Illness Trends

Owing to Occupational Safety and Health Administration (OSHA) requirements, almost all workplaces maintain at least some information about work-related injuries and illnesses. Because of their ready availability, almost all musculoskeletal surveillance efforts begin with a review of existing injury and illness records to identify jobs that may have resulted in work-related musculoskeletal disorders in the past. Potential data sources include the OSHA *Log and Summary of Occupational Injuries and Illnesses*, on- and off-site medical records, workers' compensation records, accident reports and insurance claim data. Initial review of these records usually provides a count of the number of musculoskeletal disorders associated with a specific job or work area within a given time frame.

2.1.1. Calculating rates and percentages

To permit meaningful comparisons, e.g. across jobs, departments or work sites that employ different numbers of workers, or across periods of fluctuating employment, injury/illness occurrences are usually expressed as a rate.

The incidence rate is the rate of work-related musculoskeletal disorders that appear for the first time during a given period (usually a year). The value is commonly expressed as the number of illnesses per 100 full-time workers per year and is calculated as follows:

(Number of new cases in the past 12 months x 200 000 h (2.1)/ number of work hours in the past 12 months).　　(1)

Information on work hours can usually be obtained from personnel or payroll records. A common assumption is that each employee works 2000 h per year (8 h a day, 5 days a week, 50 weeks a year). Therefore, an estimate of the value of the denominator can be obtained by multiplying the number of full-time equivalent workers employed in the job, department, or plant by 2000 h.

The prevalence rate is the percentage of all workers who are experiencing musculoskeletal symptoms at a specific instance in time, regardless of when the problem first appeared. It is calculated as follows:

(Total number of cases at a given point in time)/number of workers at the same point in time). (2)

The severity index is similar to the incidence rate, but substitutes the number of lost or restricted work days due to illness for the number of cases in the numerator. The severity index is calculated as follows:

(Total number of lost or restricted workdays in the past 12 months x 200 000 h)/number of work hours in the past 12 months). (3)

2.1.2. Interpreting incidence, prevalence and severity

Although closely related, incidence and prevalence rates provide somewhat different information. Prevalence depends both on the incidence and the duration of the illness from its onset to its resolution. For example, if the incidence of neck pain is low, but workers take a long time to recover, the prevalence will be high relative to the incidence. On the other hand, even if the incidence of musculoskeletal disorders among workers is high, the prevalence may be relatively low if workers recover quickly, or if they are forced to leave the workforce because of their illness.

The severity index provides information about the seriousness and costs of reported musculoskeletal disorders. Because medical treatment practices, the health benefits available to employees, and the opportunity for light duty influence the severity index, it is usually a good indicator of the effectiveness of the worksite's medical management program. Note that the severity index can be skewed if a few employees are away from their jobs for an unusually long period of time.

Incidence and prevalence rates provide valuable information for directing intervention efforts to truly hazardous workplaces. Unfortunately, incidence and prevalence rates are sometimes used as performance indicators in management reviews of worksite occupational safety and health programs. This practice can inadvertently penalize employers who have recently established an

ergonomics program. Many organizations find that incidence and prevalence rates initially increase after the introduction of an ergonomics program. This is usually due to heightened awareness and reporting of musculoskeletal disorder risk factors and symptoms, rather than a true increase in the disorders. In these situations, the severity index may provide a more appropriate yardstick, since the seriousness and recovery period associated with musculoskeletal disorders are usually diminished with earlier reporting.

2.1.3. Limitations

The quality and utility of existing record systems for surveillance purposes often varies; as a result, different data sources may offer strikingly different conclusions about the incidence of work-related musculoskeletal disorders associated with a specific job. One unfortunate problem is the lack of consistency or standardization in the way work-related musculoskeletal disorders are defined. "Musculoskeletal disorders" encompass a broad spectrum of illnesses that can be caused or aggravated by a wide variety of work activities. Many of these conditions can result from either acute trauma or chronic exposure to adverse working conditions; therefore, distinguishing work-related musculoskeletal disorders from acute sprains and strains is difficult unless a code for etiology is provided in the database. Many databases do not have fields for such codes, and even if they do, the codes are often applied by administrative or safety personnel who are more likely to make erroneous determinations of cause than medical professionals. For example, at most Air Force installations, the injury and illness log is maintained by a junior enlisted member who is usually instructed to code all back disorders as "injuries," without regard to their actual cause.

Another problem that limits the utility of record-based surveillance efforts is that a lag often exists between the appearance of a hazard and the onset of injury. As a result, records may not accurately reflect the current situation.

Finally, linking injury or disease data with exposure to a hazard can be especially challenging. Job titles are often poor indicators of exposure to risk factors for musculoskeletal disorders, and most records do not provide a detailed description of the worker's job or the disorder. Additional data needed to link musculoskeletal disorders to specific tasks or job processes may not exist.

2.2. Soliciting Information from Workers

Because of limitations in record-based surveillance, many employers conduct additional activities specifically to solicit information about adverse health outcomes from their employees. These activities may involve distributing questionnaires or other surveys, or administering health interviews and physical exams to workers employed in physically demanding jobs.

2.2.1. Worker surveys

Worker surveys can take several forms. They can be lengthy or quite short; they can be oral (i.e. administered by an interviewer) or written. Common features of surveys include:

- Use of charts and pictures, where workers can indicate the location of pain or other musculoskeletal symptoms.
- Questions about the onset and duration of musculoskeletal symptoms.

- Use of numerical rating scales to indicate the severity of pain, fatigue or discomfort.
- Questions about the duration of employment and the nature of job activities.

Worker surveys have several advantages over other methods of collecting surveillance data. First, the investigator has a great deal of control over the data. Once the investigator has decided what questions need to be answered, s/he can include survey items that provide the necessary information. Second, surveys are cheap and relatively easy to administer – in most cases, workers can complete the surveys at their convenience, and responses can be kept anonymous. These features can encourage high participation rates and candid responses, although the opportunities for individual follow-up are more limited. These features also allow worker surveys to be re-administered periodically to aid in early recognition of emerging problem areas.

Worker surveys also have some inherent limitations. The time and resources required to develop and administer surveys may make them more costly than surveillance activities that rely solely on existing records. The survey information may also be unreliable if there is animosity or a lack of trust between management and workers. Finally, the design of the questionnaire can have a dramatic impact on the quality of the data and the number of workers who complete the survey. The length of the questionnaire, the phrasing of the questions, and method of administration must be carefully considered. In a diverse or multilanguage workforce, translating the questionnaires into several languages may be necessary.

2.2.2. Physical exams

The purpose of physical exams is to identify workers at risk of developing more serious musculoskeletal conditions, so that they may be referred for early treatment. It must be noted that most musculoskeletal conditions are not amenable to detection in a preclinical state, and surveillance for these conditions constitutes secondary, and often tertiary, prevention of progression and disability rather than primary prevention of disease. Nonetheless, OSHA (1990) has published recommendations for conducting baseline and periodic medical examinations of workers assigned to jobs with significant musculoskeletal demands. These examinations should include a medical and occupational history, and inspection, palpation, range of motion, and other pertinent maneuvers of the upper extremities and back.

2.3. Conducting Hazard Surveys

Ideally, actions to prevent work-related musculoskeletal disorders should proceed before injuries/illnesses occur. Screening jobs for risk factors that lead to the development of work-related musculoskeletal disorders can highlight jobs for additional, more detailed surveys. Even without clear medical evidence that musculoskeletal problems exist, hazard surveillance activities can provide the data needed to begin an effective primary prevention program.

Hazard surveillance efforts depend heavily on walk-through surveys. The purpose of the walk-through is to identify risk factors that might otherwise go unnoticed and provide additional basis for prioritizing jobs for further evaluation. Typically, investigators observe job activities, speak with workers and supervisors to obtain information not apparent from observation, and use

checklists to score job features against a listing of risk factors. The walk-through survey is distinguished from more formalized job analysis efforts by the amount of detailed information collected.

Although most ergonomic checklists are designed for use by non-experts, some minimal level of training is usually needed to use checklists properly. Hazard checklists can vary in length and in scope. At one extreme, there are "generic" checklists that are applicable to nearly all jobs in all industries. These checklists can be contrasted with more focussed checklists, that are tailored to evaluate conditions typical of a specific job or industry. While some checklists are intended to serve primarily as mnemonics (i.e. to remind users to evaluate a particular job characteristic), other checklists incorporate a scoring system for assessing the degree of hazard associated with a job or process.

Ideally, hazard surveys should be administered (1) whenever a job, task or process is changed substantially, (2) when new jobs are introduced and (3) periodically (especially after new cases of musculoskeletal disorders are reported) to detect whether trends exist across jobs that use similar equipment, tools or processes. Hazard surveys can also be incorporated into regularly scheduled safety and health inspections, expanding the scope of these inspections to include identification of musculoskeletal disorder risk factors.

3. ESTABLISHING INTERVENTION PRIORITIES

Ultimately, surveillance activities obligate the employer to intervene to alter the workplace factors that produce adverse health events and hazards. There have been several attempts to recommend various thresholds or "triggers" when intervention in the workplace becomes necessary. The ergonomics standard adopted by the California Occupational Safety and Health Standards Board on 17 April 1997 requires employers in that state to take certain actions when a single WMSD is reported. This requirement is based on the recognition that formal surveillance activities usually detect only a small proportion of the musculoskeletal problems and that one reported case may lead to several times as many unreported cases.

As stated previously, health and hazard surveillance data are highly complementary, and considered together, they do suggest a rational means of establishing intervention priorities. Stumpp and Sparks (1997) suggest that areas where the incidence of musculoskeletal disorders is high and risk factors have been identified should get first attention, followed by areas where only musculoskeletal disorders have been found, and then by areas where only risk factors have been found. If additional prioritization is required in the top tier (e.g. because intervention resources are limited), giving preference to jobs that employ many people, or jobs where major changes are already planned can be a sensible and cost-effective approach.

4. CONCLUSIONS

Surveillance is essential to the prevention/control of musculoskeletal disorders in the workplace. In general, no single data source provides enough information to direct a program for preventing work-related musculoskeletal disorders. Therefore, effective surveillance programs make use of multiple data sources to identify problem areas and determine intervention priorities. Even in the absence of health data, hazard surveys conducted in workplaces where significant or well-defined hazards exist can provide the data needed to mount an effective primary prevention program for work-related musculoskeletal disorders.

Once established, surveillance should become an on-going process. As corrective actions are taken, surveillance data can provide the information needed to show the beneficial effects of these efforts. By integrating surveillance efforts with existing quality assurance and cost containment programs, their utilization and success will be maximized.

REFERENCES

BAKER, E.L. and MATTE, T.P., 1994, Surveillance for occupational hazards and disease. In Rosenstock, L. and Cullen, R. (eds) *Textbook of Clinical Occupational and Environmental Medicine* (Philadelphia: W.B. Saunders), pp. 61–7.

CALIFORNIA OCCUPATIONAL SAFETY AND HEALTH STANDARDS BOARD, 1997, *Ergonomics*, General Industry Safety Orders, Section 5110.

CENTERS FOR DISEASE CONTROL, 1988, *Guidelines for Evaluating Surveillance Systems*. MMWR CDC Surveillance Summaries, (Atlanta: CDC) 37(S-5), pp. 1–18.

OCCUPATIONAL SAFETY AND HEALTH ADMINISTRATION, 1990, *Ergonomics Program Management Guidelines for Meatpacking Plants*. No. 3123 (Washington, DC: USDOL, OSHA).

STUMPP, D. and SPARKS, P., 1997, Screening and medical surveillance for cumulative trauma disorders. In Erdil, M. and Dickerson, O. B. (eds) *Cumulative Trauma Disorders, Prevention, Evaluation and Treatment* (New York: van Nostrand Reinhold), pp. 203–20.

System Safety Engineering and Risk Assessment

N. J. Bahr

Hernandez Engineering, Inc., 3509 Kent Street, Kensington, MD 20895, USA

1. INTRODUCTION

The desire to be free of accidents and injuries has been a part of mankind since the beginning of time, with one of the earliest written references to safety from the Code of Hammurabi *c*.1750 BC, stating that the builder of a house would be put to death if poor construction caused death of the owner. System safety engineering and risk assessment has changed much since then; however, many still confuse the terms, and frequently use them interchangeably.

System safety engineering is the application of engineering and management tools and practices to assure that a system is safe for operators and the public (i.e. free from injury), environment and other equipment. In this definition it is critical to understand that a *system* is the combination or interaction between people, hardware, software and operating environment.

Risk assessment is the formal process of determining the risk of a particular event to the system. It can be used to determine the safety of a system or it can be used for non-safety trade-off assessments of the probability that a particular event and its resulting consequences occur. One such example is the risk of late product launching or financial risk of a particular operational strategy.

2. FUNDAMENTALS OF SYSTEM SAFETY

2.1 What is an Accident?

The purpose of system safety engineering is to *design into a system* the prevention of an accident or mitigation of its consequences. Accidents now are no longer confined within national borders but many times are transnational. The Chernobyl nuclear accident in Ukraine in 1986 not only immediately killed 31 people but also sent radioactive particles through wind currents to over 20 countries. Society's rapid technological development many times has created very complex human–machine interfaces. Poor design and operations of these interfaces can also lead to undesired mishaps.

An *accident* is an unplanned series of actions that causes undesired loss of life, injury or unwanted damage to equipment or the environment. However, they do not just occur, they follow a discernible process and series of steps before a catastrophic event occurs. If one can prevent or control one or more of these steps in the process many times a mishap can be prevented altogether or its consequences diminished.

The *initiating or trigger event* is the action, either machine or human that forces the mishap to occur. For example, an electrical short near a fuel line causes a fire or explosion. Likewise, the omission of critical steps in an operator's fill procedures of an anhydrous ammonia storage tank over pressurizes the pressure vessel.

Intermediate events tend to exacerbate or propagate the initiating event or alleviate the results. The use of a fire detection
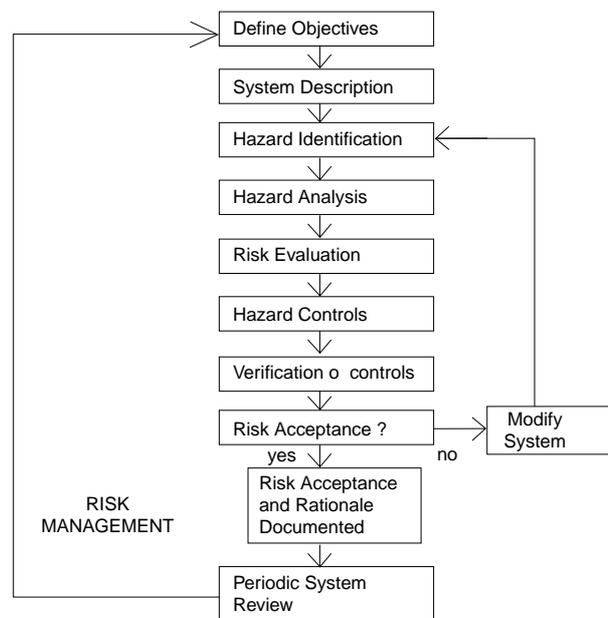
Figure 1. The system safety process (Bahr 1997).

and suppression system can mitigate the damage from a fire caused by an electrical short in a fuel line. However, increased fuel flow or storage of oxidizers in the vicinity of that fuel line can significantly increase accident damage. If one can understand which events (personnel, equipment, or both) create a hazardous condition then it is possible to control or ameliorate the consequences of those events.

2.2 System Safety Process

The system safety process (Figure 1) is a closed-loop system that helps one to identify hazards, evaluate their risk to the system, develop controls to prevent the hazard or mitigate its effects, and then periodically to review the process. This process is a comprehensive systematic combination of engineering analyses and management oversight.

The first step is to define the objectives and system under consideration. Next the entire system (people, equipment, software, operating environment, etc.) is reviewed. Hazards are identified and analyzed through various techniques. Each hazard is evaluated for its corresponding effects and the overall risk to the system. If the evaluation determines that the risks are unacceptable then controls are developed and implemented to mitigate those risks. It is critical that the controls are verified to be adequate to control the hazard and that they are in place.

Management must make a formal risk acceptance decision. If the risk is still unacceptable then the system is modified. This can be implemented through design changes, procedural controls or changes in the work environment. If the risk is now sufficiently controlled then the entire process is documented (this is typically done in accordance with ISO-9000 and ISO-14000 requirements).

2.3 Hazard Reduction Precedence

Once a hazard has been identified there is an infinite number of ways to control it. However, there is a hierarchy of methods that are followed in most industries.

The first step is to *design the hazard out of the system*—make it physically impossible for the hazard to exist and, therefore, an accident sequence to occur. For example, if fire is the hazard, then removing any one of the following four items out of the system—combustible material, oxygen, chemical reaction or ignition source—absolutely prevents a fire from occurring.

Many times it is not feasible due to costs or other constraints to remove the hazard. The next step is to use *fail-safe* devices. This is a component that automatically returns the system to a safe state. Pressure relief valves and fuses are two such examples. But a fail-safe device alone usually is not sufficient to prevent a catastrophic event. In this instance *caution and warning* devices are warranted. Smoke detectors or oxygen monitors in a process plant are frequently used.

Other times, fail-safe devices and caution and warning systems still are inadequate. One such example is with an ammonia leak. If it is impossible to design the hazard out (e.g. ammonia is the working fluid in the system and cannot be substituted) then a fail-safe device can help avoid further system damage. Caution and warning systems will alert personnel to the hazard, but further action is still required. *Special procedures and training* of operators and other personnel can help contain the consequences and mitigate the effects. Use of personnel protection equipment and special emergency response actions can significantly affect how serious the resulting accident will be. But because people tend to commit more errors when under stress, this is the least preferred method of hazard abatement.

2.4. Safety Management

An effective and pro-active safety organization with upper management commitment is essential to preventing accidents. Most companies and large governmental research and development organizations have a corporate or executive safety office that reports to the organization director.

A good safety program consists of two major sections: safety management and safety requirements. Table 1 shows the primary elements of both.

2.5. Safety Analysis Techniques

There are numerous safety analysis tools in existence today. Most are very similar, using the concept of hazard reduction precedence, but different industries tend to apply their own techniques.

The manufacturing industry traditionally has used standards-based compliance methods. *Safety checklists* verify that the plant

Table 1.

Sec. 1 Safety Management System	Sec. 2 Saftey Requirements and Norms
Policy	Hazardous Operations Procedures
Safety Organization	Emergency Response Procedures
Safety Assessment	Mishap Reporting and Tracking Procedures
Safety Training and Awareness	Safety Design Requirements
Safety Compliance System	National and International safety requirements
Subcontractor Safety Control	

follows national norms required for personnel safety. Many industrial countries have national rule-based occupational health and safety standards. However, this is beginning to change. For example, recently the US Occupational Health and Safety Administration (OSHA) enacted the Voluntary Protection Program (VPP). Because it is impossible and undesirable to audit every plant in the USA, the government has proposed that companies implement a comprehensive safety program. Both plant unions and management must be part of the program. The advantage of a VPP is that companies can avoid the frequent OSHA inspections. Like the similar Malcolm Baldridge Award for quality programs, costs are high; however, after the initial investment, significant reduction in accident rates and enhanced plant efficiencies have more than paid for the effort.

The chemical process industry applies a plethora of safety tools, the most prevalent *HAZOP* (hazard operability analysis). This tool was first developed in the UK by Imperial Chemical Industries in the early 1960s and later adopted in the USA in 1974. A team of five-to-seven experienced engineers review process drawings and through brainstorming sessions assesses deviations to the process and their effects on safety. A complicated system is divided into nodes; the team analyzes each using predetermined guidewords (i.e. no flow, less flow, low temperature, etc.) to determine the effects of potential changes to the intended operation. A corollary benefit to the safety weaknesses identified is determination of process inefficiencies. Compliance-based safety checklists are still used in the industry, though to a lesser extent.

The 1990s has seen an increase in the application of risk assessments—especially *probabilistic risk assessments* (PRA). Many European countries use the term *quantitative risk assessment* or *quantitative (or probabilistic) safety assessment*. PRA, first developed by the commercial nuclear power industry, uses quantitative equipment failure rates and human reliability error rates to determine the likelihood of an accident and its consequence. Widespread application across all industries has still not occurred because of the high cost and labor-intensive nature of the technique. Also, effective application requires a good database of equipment failure rates and the still controversial quantitative human error rates. Process industry PRA have also proven very useful as inputs to environmental risk and impact assessments.

After the 1979 Three Mile Island commercial nuclear power plant accident, the US Atomic Energy Commission documented PRA in their "WASH 1400, The Reactor Safety Study." The commercial nuclear industry has spent considerable time and money developing the needed equipment failure rates and human error probability calculations and databases. Most commercial nuclear power plants (and nuclear powered spacecraft) use this technique for licensing.

The aerospace and military industries have been using numerous system safety techniques since the late 1950s. Frequent missile and aircraft failures forced various military forces to develop some of the most commonly known system safety tools used around the world today. Some of these are *hazard analysis, operations and support hazard analysis (O&SHA)*, and *fault tree analysis*.

The hazard analysis is a methodical and systematic approach to assessing hazards and controls to a system. Through understanding the design and operation of a system during all

phases of the life cycle, hazards and their causal factors are identified. Hazard scenarios are developed and assessed for accident severity and likelihood. This severity and likelihood classification allows one to compare hazards and determine a resolution hierarchy through a ranking method. The likelihood assessment typically is a qualitative determination, though probabilistic methods are now becoming more common. The advantage of performing a qualitative frequency determination is that one can still assess the likelihood of an accident if quantitative data are not available, reliable or cost-effective. Hazard analysis is frequently used in almost all industries even unlikely ones such as food safety. The O&SHA is similar except it focuses on human operators and how they interact with their environment and equipment.

Fault tree analysis is a graphical safety tool used both in system safety and reliability engineering. This deductive methodology postulates a top event or fault, and then develops functional branches or events that must occur for the top event to occur. Through logic gates and standard Boolean algebra one can determine which events or combination of events must occur for the top event to exist. Fault trees can be either quantitative or qualitative and are particularly useful in merging equipment failure data with human error probabilities. Fault trees were very helpful for the National Aeronautics and Space Administration in determining the causal factors to the *Challenger* accident. Fault trees are now found commonly used across all industries such as oil drilling platforms and automobile safety assessments.

The safety tools common to the mass transit industry are hazard analysis, fault tree analysis and regulatory compliance.

Most of the safety techniques described above are primarily safety assessment techniques and are not strictly human factors safety tools. Additional human performance assessment tools that typically are used in conjunction with a safety analysis or part of one are THERP, HEART, SLIM-MAUD, expert estimation and human cognitive reliability models.

3. FUNDAMENTALS OF RISK ASSESSMENT

3.1. What is Risk?

Risk is defined as the triplet: event scenario, probability of occurrence and consequence. The event scenario is the description of a particular event, along with its likelihood of occurrence and the severity of the consequences of said event. There is no zero risk—everything has at least some risk (albeit infinitesimal in some cases). One of the disadvantages of safety engineering is that many times worst-case scenarios that are very unlikely to occur are identified. Using risk assessments one can better optimize the safety process. Also, as stated earlier, risk assessments do not have to focus only on the safety of a system.

3.2. Risk Assessment

The risk assessment process is very similar to the system safety process described earlier, and uses many of the system safety analysis techniques described above. The risks are then quantified and evaluated.

The assessment objectives and system description should be defined first. The next step is to develop initiating event scenarios. The safety analysis techniques described earlier can be used separately or in combination. For example, one may wish to perform a hazard analysis of the facility to identify the primary hazard concerns or initiating events.

The following step is to develop event trees. An event tree is a graphical representation that describes the relationship between a particular hazard and the control system used to mitigate that hazard. The event tree is particularly useful to elucidate if one's safety controls are capable adequately of controlling the hazard. Next, the scenarios are quantified using the calculated failure probability derived from the fault tree input.

Then the consequences are determined through analysis of the potential damage states created if the hazard continues to the end result. Damage states are determined both quantitatively and qualitatively and typically use a cost–benefit analysis. One could say that the resultant damage-state is system incapacitation and US$20 million of damage. The next step is then to evaluate that risk.

An event tree can illustrate the probabilities of a particular event scenario along with its consequences – say, different damage to equipment. Damage states are shown as qualitative (critical equipment damaged) and quantitatively (equipment damage value). The corresponding risk expectation value is the multiplication of the two numbers. This is very useful because one can see immediately the relationship between a particular safety barrier and its corresponding risk of accident.

3.3. Risk Profiles

Even with a risk ranking system or with event trees it is not always easy to understand which risks are most important. For example, it could be difficult to determine which risk is more important inside a set of various risks all ranked equally or near-equally. In this instance, risk profiles are helpful. As Figure 2 illustrates, it is very easy to determine that scenarios one and three are most important.

4. RECOMMENDATIONS

- Develop a system safety program early in the system life cycle.
- Use one or more of the accepted system safety techniques with national and international regulatory compliance standards.
- Perform the system safety analysis techniques before hardware and operational procedures are developed.
- Use the system safety process as part of the organization's safety training and awareness program.
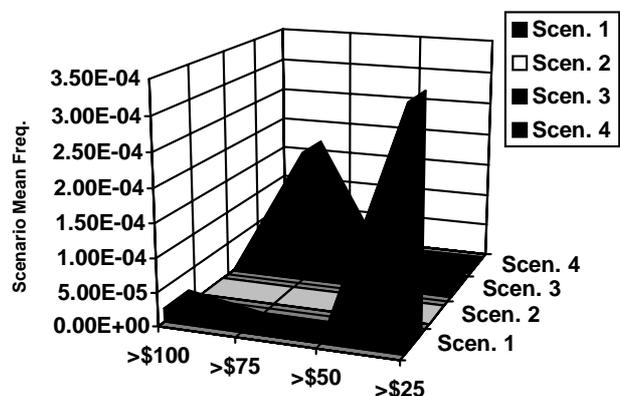
Figure 2. Risk expectation profile.

- Study near-miss accidents for clues to unidentified hazards or accident causal factors.
- Every time the system has had a significant change, review and assure that hazard controls are still valid.

REFERENCES

BAHR, N.J., 1997, *System Safety Engineering and Risk Assessment: A Practical Approach* (London: Taylor & Francis).

CENTER FOR CHEMICAL PROCESS SAFETY, 1992, *Guidelines for Hazard Evaluation Procedures* (New York: American Institute for Chemical Engineers).

HENLEY, E.J. and KUMAMOTO, H., 1992, *Probabilistic Risk Assessment, Reliability Engineering, Design, and Analysis* (New York: IEEE Press).

ROLAND, H.E. and MORIARTY, B., 1985, *System Safety Engineering and Management* (New York: John Wiley).

SWAIN, A.D. and GUTTMAN, H.E., 1983. *Handbook of Human Reliability Analysis with Emphasis on Nuclear Power Plant Applications.* NUREG/CR–1278 (Washington, DC: USA Nuclear Regulatory Commission).

Work Organizations: Health and Productivity Issues

K. Lindström

Finnish Institute of Occupational Health, Department of Psychology
Topeliuksenkatu 41 a A, FIN-00250 Helsinki, Finland

1. INTRODUCTION

A healthy work organization can be seen as an organizational strategy aiming to optimize effectiveness and the well-being of the personnel. Effective organization needs healthy and competent people and through its organizational practices and culture promotes the learning and well-being of people. The organization should be effective and profitable in order to justify its existence. The functioning of workplaces depends strongly on the global business market, national laws and regulations, and of course on the partners and customers. The focus should therefore be broad when defining a healthy organization in order to meet the interests of individual employees, teams, stakeholders, and the whole society (Jaffe 1996). In the long run, their interests are easier to fit together, but short-term financial profitability can be in conflict with the ideals of the healthy work organization.

The definitions of a healthy work organization include many aspects which are also relevant in todays turbulent business environment where downsizing, mergers, and other restructuring processes are common. The organization should be effective in various kinds of environment and have the readiness to foresee and react to the external changes in its market environment.

One basic human characteristic of organizational healthiness is the congruence between its written values and overt practices (Cox and Leiter 1992). The written organizational values of the company should to be shared and realized in everyday practice, especially in critical situations where no ready practical solutions are available. This chapter has three aims:

- To describe the main job and organizational characteristics of a healthy organization.
- To present ways to assess the healthiness of an organization.
- To present some practical intervention approaches which guide the organization towards healthiness.

A healthy and productive work organization doesn't mean at the practical level anything radically new, but it integrates aspects from existing theories of individual well-being and organizational effectiveness into a value-based strategy. At the organizational development level, the approach should also be multilevel, combining the individual and organization-oriented interventions.

2. CHARACTERISTICS OF HEALTHY ORGANIZATIONS

The definitions of a healthy organization have covered issues related to the basic existence of organizations and their relations with the external environment. Some theoretical and pragmatic models which have been related to the healthy organization are job stress models, job redesign and organizational development models, schools of psychodynamics, and unconsciousness of organizations, plus applied corporate strategies for health promotion emphasizing certain types of policies and culture (Cooper and Cartwright 1994, Jaffe 1996). From the viewpoint of organizational psychology, its inner psychological and social functioning is important, as well as the perceived task characteristics and organizational practices and climate. The traditional work safety issues and physical work demands should be seen as basic requirements, especially in manufacturing and in other physically heavy work.

Here I describe those characteristics which have been shown to promote the employees well-being and competence, and company effectiveness and profitability. The characteristics describing a healthy organization usually have both individual and organizational focus. The quality of social relations can be viewed from the individual perspective and as an indicator of the group and organizational climate. Absenteeism is an individual health indicator, but a high sickness absenteeism rate indicates health at the organizational level. The following characteristics are typical of a healthy organization.

2.1. Appreciation of People

Respect and appreciation of people and their work has been regarded as a basic value of a healthy organization. At the workplace level, it includes mutual trust, fair treatment, and the consideration of ethical issues in decision making. Besides help and support from supervisors, older workers perceived respect as one of the main factors in maintaining their working capacity (Tuomi *et al.* 1997).

2.2. Job Demands and Control

Job stress research has shown that high job demands and low control at work can lead to elevated stress and contribute to say cardiovascular diseases. Low social support has been added as a third dimension to this model (Karasek and Theorell 1990). The regulation and control of the risks at work, especially job stress, is a central issue in a healthy organization.

2.3. Leadership and Management Practices

Management and leadership practices are important contributors to the well- functioning of the organization and the well-being of the personnel. Teamwork and empowerment mean a new division of power and decision making between the management and the personnel. Managerial practices can in this way facilitate the development of competence and the health of employees. A participative management style with low task orientation and empowering have been associated with lower stress and ill health. Challenging tasks together with good leadership practices can contribute to job satisfaction and job involvement.

2.4. Change Management

Change management practices are crucial for the future of the company and its people. It is usually a question of combining the interests of individuals, company stakeholders, and customers. Change management should be a part of human resource management practices at company level. The basis is good information about changes at all levels in the organization. The management of recent organizational change processes, based on outsourcing or a lean organization, has often failed in both human and financial respects, because the changes have been based only on economic calculations; the participation of people

and the utilization of their ideas have been neglected. The Finnish downsizing examples from the municipal sector showed that the anticipated saving of money was lost when the sickness absenteeism rate increased because of elevated and prolonged job pressures (Vahtera *et al.* 1997).

2.5. Continuous Improvement

Adopted continuous improvement practices associated with an innovative climate promote the idea of lifelong learning and the development of the personnel's competence. Awareness of the future of the organization, and customer needs are necessary when implementing the individual and group-level learning efforts. Learning at work and worksite training opportunities are especially important for aging workers, otherwise they are vulnerable to marginalization when changes occur in the production, services, or technology.

2.6. Work and Private Life Interface

The two main individual-centered characteristics of a healthy organization to be considered are the balancing of work and private life, and the managing of diversity issues. Both are concerned with basic human values and rights and are usually included in company values. When the flexibility of working hours and work arrangements is increased both quantitatively and qualitatively, it brings more degrees of freedom but also an impossible situation to be solved by individuals. Flexibility at work can increase insecurity or lead to conflicting demands at the individual level, e.g., part-time employment, outsourcing, multiskilling. On the competitive labor market, those who have more alternatives to make choices are better off. Flexibility of working hours and work arrangements has either facilitated or complicated the balancing of private life and work life. Longer weekly working hours have increased job pressures and left less time for recovery and for the other sectors of life.

2.7. Diversity Issues

Awareness of the richness of human diversity is a central topic in a healthy organization and in the society as a whole. Neglecting it will easily lead to marginalization and discrimination of older persons, women, and people who are not in the majority. This means that the experience and knowledge based on the diversity of age, gender and ethinicity, and also equality issues are violated because they are not necessarily seen as beneficial by the traditional groups having the power and the best advancement opportunities in the organization.

It is often difficult to show a direct relationship between organizational effectiveness and well-being along with social and psychological factors. Job satisfaction coupled with good supervisory support and continuous improvement practices contributed to productivity and profitability of small and medium-sized enterprises (SMEs) (Lindström *et al.* 1999).

3. ASSESSING ORGANIZATIONAL HEALTH
3.1. Continuous Monitoring

Assessing the state of the work organization requires continuous information on its critical characteristics and outcomes, indicating organizational effectiveness and individual competence and well-being. Usually the customers' needs and satisfaction should also be surveyed in order to update the situation on the market

and to revise the future vision and strategic goals of an organization. A combination of feedback from the personnel on work and organizational practice, and from clients about their needs and satisfaction, gives valuable development ideas for continuous improvement practices or quality management. These should also be related to so-called objective data on organizational effectiveness and personnel productivity. This information could be summarized in so-called human resource accounting and reporting of the enterprise, which gives summarized data for the management and the stakeholders about the human resources.

3.2. Personnel Surveys

Repeated personnel surveys should at least include measures about the task characteristics, like quantitative workload and the challenges and learning possibilities, as well as leadership practices and other organizational practices relevant for a healthy organization. The readiness to change and a sensitivity to external changes are important when striving towards continuous improvement practices. The quality of organizational practices and climate reveals the possible risks for individual ill health and organizational ill health, like social conflicts, dissatisfaction, burnout, and sickness absenteeism.

Longitudinal measures within the company and benchmarking possibilities with other organizations from the same field are valuable. Various survey methods for assessing these job and organizational characteristics are available, e.g., organizational health questionnaire (Cox and Leiter 1992), occupational stress indicator (Cooper and Cartwright 1994), and healthy work organization questionaire (Lindström 1997). These methods mainly measure job and organizational characteristics and can be completed by measures of well-being, like the Mashlach burnout inventory (MBI), and the general health questionnaire (GHQ). However, the most common practice is to tailor the methods for a particular company, but taking into account the availability of reference data and the scientific basis of the method. Repeated measures with the same method allow the effects of natural and planned changes to be monitored.

Organizational surveys are useful, not only for monitoring the current state but also for use by managements, teams, and employees when planning interventions. The survey feedback method followed by some interventions is the most traditional way to start an organizational intervention.

4. PROMOTING ORGANIZATIONAL HEALTH
4.1. Multilevel Approach

The main actors in an intervention should be the management in a participatory way; but human resource experts, occupational safety and health personnel, and training experts have a special role in facilitating the development. The characteristics of a healthy organization include organizational practices which can be seen in the form of developmental methods, like change management and continuous improvement practices. These organizational practices regulate job pressure and improve the functioning of the company. Access to occupational health services for the personnel, as well as implementing health promotion programs related to the job and the organization, can help to correct these problems.

Increasing ones own control at work and employee participation are fundemental issues in preventing harmful stress effects

and creating personal involvement in the job and commitment to the organization (Cooper and Cartwright 1994, Karasek and Theorell 1990). Especially in SMEs where no such institutional support and help is readily available, the management is more responsible for these aspects. The competence and material resources can nevertheless be problematic (Lindström *et al.* 1999).

Although the healthy organization is a multilevel model, the intervention actions should also be multilevel, and they should be implemented at the organizational and individual level, and preferably also at group level. Business organizations operate in a societal and global context where they are exposed to various challenges and threats. The organization's awareness of the market changes is the way to cope with them and maintain the inner healthiness and competence of the organization. The decline in profitability easily leads to negative consequences at company level, and the principles of the healthy organization are forgotten. The pure market economy approach at the organizational level with shortsighted profit making, this is a source of people's ill health.

4.2. Organizational Interventions

When thinking of recent ways to develop organizations, it is crucial to consider the values of the management and human resource management functions. Here it is advantageous to have value-based interventions investing in the long-term development of human resources. Generally it is possible to distinguish three different organizational change processes in which the principles of the healthy organization are relevant:

- Improve and promote existing ways to function, e.g., promote the health and work capacity of employees, the quality of services and products, or effectiveness in general.
- Shift towards a learning-oriented and knowledge-based organization, and implement or facilitate teamwork and networking
- Restructure the organization, e.g., downsize, outsource, and reengineer the business process.

Both the learning organization and knowledge management intervention are beneficial for a healthy organization, because they are usually based on long-term development of human resources. Structural changes like downsizing, business processes re-engineering, and outsourcing are today's practice and often unavoidable; the way in which these kinds of changes are implemented is crucial. They lead easily to adverse effects on business itself, loss of competence, and individual suffering. Here the participatory approach would soften the possible negative effects, and even innovative solutions may be a by-product.

Common organizational interventions usually aim to improve the present way of functioning (Lindström 1996). It is the easiest approach because the power structure of various interest groups is not greatly affected, and all efforts can be focused on the development itself. But without the management's commitment, this kind of implementation model can lead to ineffective or negative side effects in the long run, because when the participative approach is used in interventions, it commits people to the present way of functioning.

When structural changes based on business are implemented soon after these kinds of interventions, the results of the intervention are lost, and people are dissappointed; their commitment and belief in joint development decrease in the future. The

vision and the new strategic goals should therefore be the starting point in order to facilitate real changes. Practical experience has shown that the management as well as the personnel should be committed to the goals and process of the intervention, and both business and well-being goals should be combined. This is especially important in a rapidly changing, turbulent environment undergoing unpredictable changes. Joint participation of management and personnel gives more sustainable results and corresponds to the definition of a healthy organization.

Interventions can also be focused on traditional job redesign, which has proved successful in large companies and in SMEs. Starting the developmental process with working seminars based on democratic dialog strengthens commitment at all levels in the organization. All the developmental ideas of the participants are then based on that vision and the ways of implementing it. Opening channels for ideas within the organization and supporting their growth, both require an atmosphere of trust. Creating trust is a demanding task which takes time and requires positive feedback, so that individuals get enough courage to express their own ideas without fear of being ignored or punished. This kind of innovative climate has been very successful in developing SMEs (Lindström *et al.* 1999). Innovative interventions also require time and financial resources.

4.3. Individual interventions

At the individual level, a healthy organization means that individual job redesign solutions are undertaken and social support is available when needed. Both aspects require an infrastructure to be created at the workplace and there should be access to consultative help. Work-related risks and their harmful effects, such as job stress and burnout, are two particular situations requiring individual support. The ideal would be access to occupational health care services outside the work. However, the development of individual competence is the best assurance for the individual when considering the continuity of his or her life-long work career. At the individual level, the crucial point in a healthy organization is respect of people and investments in long-term individual growth and learning. Therefore the idea of the learning organization should also include development of the personnel's professional and organizational competences. In summary, the key elements in the healthy and productive organization approach are respect of people, simultaneous development of business and personnel, and an open-minded approach to the society and the business partners and clients.

REFERENCES

COOPER, C.L. and CARTWRIGHT, S., 1994, Healthy mind; healthy organizationæa proactive approach to occupational stress. *Human Relations*, 47, 455–71.

COX, T. and LEITER, M., 1992, The health of health care organizations. *Work and Stress*, 6, 219–27.

JAFFE, D.T., 1996, The healthy company: research paradigms for personal and organizational health. In Sauter and Murphy (eds) *Organizational Risk Factors for Job Stress* (Washington DC: American Psychological Association), pp. 13–39.

KARASEK, R. and THEORELL, T., 1990, *Healthy Work* (New York: Basic Books).

LINDSTRÖM, K., 1996, Finnish research in organizational development and job redesign. In Murphy, Hurrell Jr, Sauter, and Keita (eds) *Job*

Stress Interventions (Washington DC: American Psychological Association), pp. 283–93.

LINDSTRÖM, K., 1997, Assessing and promoting healthy work organization. *From Experience to Innovation, IEA '97*, Vol. 1, *Organizational Design and Management*. (Helsinki: Finnish Institute of Occupational Health), pp. 504–6.

LINDSTRÖM, K., SCHREY, K., AHONEN, G., and KALEVA, S., 1999, The effects of promoting organizational health on personnel well-being and company effectiveness in small and medium-sized enterprises. In Murphy and Cooper (eds) *Healthy and Productive Work: An International Perspective* (Louisville KY: Taylor & Francis), forthcoming.

TUOMI, K., ILMARINEN, J., SEITSAMO, J. *et al.*, 1997, Summary of the Finnish research project (1981-1992) to promote the health and work ability of aging workers. *Scandinavian Journal of Work and Environmental Health*, 23, 66–71.

VAHTERA, J., KIVIMÄKI, M., and PENTTI, J., 1997, Effect of organizational downsizing on health of employees. *Lancet*, 350, 1124–28.

Work Organization and Psychosocial Work Factors: Definitions

P. Carayon†‡, M.L J. Smith‡ and M. C. Haims‡

†Ecole des Mines de Nancy, Parc de Saurupt, F-54042 Nancy Cedex, France
‡Department of Industrial Engineering, University of Wisconsin-Madison, Madison, WI 53706, USA

1. INTRODUCTION

Work-related musculoskeletal disorders (WRMD) are increasingly a concern to occupational health and safety professionals, ergonomists, industrial engineers, employers, unions and workers. The US Bureau of Labor Statistics (BLS) has shown that all types of WRMD have increased as a percent of total occupational illnesses. In 1995, WRMD associated with repeated trauma accounted for 308 000 cases or 62% of new illness cases in private industry (BLS 1997). WRMD include tendon disorders (e.g. shoulder tendinitis), nerve disorders (e.g. carpal tunnel syndrome, thoracic outlet syndrome), muscle disorders (e.g. tension neck syndrome), joint disorders (e.g. osteoarthrosis), vascular disorders (e.g. hypothenar hammer syndrome), bursa disorders (e.g. knee bursitis) and unspecified musculoskeletal symptoms or multiple-tissue disorders (Hagberg *et al.* 1995). Recent studies indicate a potential link between work organization, job stress and WRMD (NIOSH 1992, Smith *et al.* 1992), and theories on these links have been proposed (Sauter and Swanson 1996, Smith and Carayon 1996). Here are reviewed theories and empirical studies on the relationship between work organization, psychosocial work factors, stress and WRMD.

2. CONCEPTUAL FRAMEWORK

Work organization has been defined as the way in which work is organized, supervised and carried out (Hagberg *et al.* 1995). It can contribute to WRMD problems by specifying the nature of the work activities (variety or repetition), the extent of loads, the exposure to loads, the number and duration of actions, workstation design, tool and equipment design, and environmental features. The policies and procedures of a company can affect WRMD risk through the design of jobs and the defining of work-rest cycles, work pace (i.e. work pressure), the psychological climate regarding socialization, career and job security, the level of employee training, availability of assistance and supervisory relations. All of these factors interact as a system to produce an overall stress load on the person that can lead to WRMD (Smith and Carayon-Sainfort 1989). Stress may logically have a role in WRMD because such disorders are much more prevalent in occupations with features that produce job stress, i.e. short-task cycles, monotony, low control, low content and high work pace.

There are theoretical reasons to believe that work organization factors can play a role in the report and development of WRMD (Sauter and Swanson 1996, Smith and Carayon 1996). The model presented in figure 1 suggests the potential relationship between work organization and WRMD. Mediating this relationship are (1) stress reactions and (2) individual factors. Organizational, ergonomic and psychosocial factors can produce stress reactions on the individual: these stress reactions can have physiological, psychological and behavioral aspects, such as biomechanical loading of muscles or joints, increased levels of catecholamine release, or adverse psychological mood states (Smith and Carayon-Sainfort 1989, Hagberg *et al.* 1995). Stress reactions can lead to various types of strain, including WRMD. These various concepts
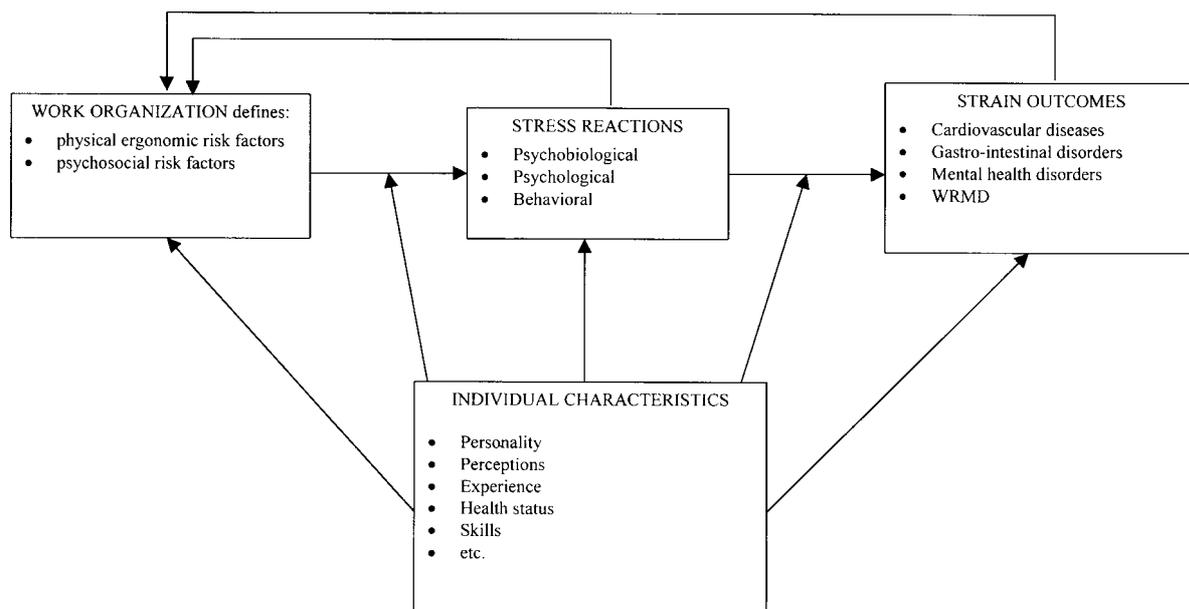
Figure 1. Model of the relationship between work organization and WRMD.

of work organization, ergonomic and psychosocial work factors, stress reactions, and strain are hypothesized to describe various stages of the chain of events that links work organization to WRMD (figure 1).

Work organization determines the psychosocial work factors and ergonomic risk factors to which employees are exposed. For instance, an organization based on teamwork may provide for high levels of worker participation, therefore fostering a positive psychosocial work environment. It may also allow workers to rotate between tasks, therefore reducing exposure to high physical loads of certain tasks. According to Smith and Carayon-Sainfort (1989), a work system is composed of five elements: (1) individual, (2) tasks, (3) physical and social environment, (4) tools and technologies and (5) organizational conditions. The way work is organized, structured, supervised and carried out (i.e. the design of the work system) will define the various elements of the work system, as well as their interactions. These elements and their interactions can have positive and/or negative aspects that represent various psychosocial work factors and ergonomic risk factors. For instance, an assembly-line organization is characterized by short-cycled, fast-paced tasks. These tasks not only have negative psychosocial work factors, such as little variety and job control, but also negative ergonomic risk factors, such as high repetitiveness.

3. MECHANISMS

There are psychobiological mechanisms that make a connection between job stress and WRMD plausible and likely (Smith and Carayon 1996). Psychological stress can lead to an increased physiological susceptibility to WRMD by affecting hormonal, circulatory and respiratory responses that exacerbate the influences of the traditional ergonomic risk factors. In addition, psychological stress can affect employee attitude, motivation and behavior that can lead to risky actions which increase WRMD risk. Here the importance of work organization and job stress in the development, reporting and experience of WRMD is emphasized. Two different mechanisms for the relationship between WRMD and job stress are proposed: (1) a psychobiological mechanism and (2) psychological and behavioral reactions to stress. Highlighted are the importance of work organization in the development, experience and reporting of WRMD.

3.1. Psychobiological Mechanism

When an individual undergoes the psychological, physiological and behavioral effects of job stress, there are changes in body chemistry that may increase the risk of WRMD. Changes in the body include: increased blood pressure, increased corticosteroids, an increase in peripheral neurotransmitters, an increase in muscle tension, less effective immune system response and hyperventilation or overbreathing (Selye 1956, Levi 1972, Theorell and Karasek 1996, Westgaard 1996). While it has long been known that these stress reactions can contribute to increased cardiovascular and psychological strain (Cooper and Marshall 1976, Smith 1987), it is believed that they also can increase the risk of WRMD.

Several studies have shown a link between blood pressure and job stress. In particular, high workload and work pressure, and lack of job control have been related to an increase in blood

pressure (Matthews *et al.* 1987, Schnall *et al.* 1990). The increased blood pressure may reduce blood flow to the extremities, which could accelerate or exacerbate tissue damage during high workload, and may precipitate the occurrence of WRMD. An increase in corticosteroids is another physiological reaction to stress (Frankenhaeuser 1986, Daleva 1987). Increased levels of corticosteroids, in particular cortisol, can lead to increased fluid retention in body tissues. This could be an important risk factor for carpal tunnel syndrome (CTS). Corticosteroid-induced fluid retention may be similar to the fluid retention in the extremities of the pregnant woman, which is believed to increase their risk of CTS. Excess fluid and tissue swelling can place pressure on and pinching of the nerve(s) that can cause the parasthesia and pain associated with peripheral neuropathy. Another consideration related to physiological stress is the amount of tension in the muscles. With increased levels of norepinephrine, the tension in muscles has the potential of being greater, as does the extent of recruitment of the muscle fibers in performing an activity (Westgaard 1996). This heightened muscle tension may be increased by adverse psychological moods such as anxiety or anger, and can lead to excessive muscular force when working. Some studies have suggested that increased muscle tension may be a mediating variable between negative psychosocial work factors and psychological stress, on one hand, and musculoskeletal disorders on the other hand (Theorell *et al.* 1991, Waersted and Westgaard 1991, Westgaard 1996).

Schleifer and Ley (1996) assert the importance of breathing as a psychophysiological pathway through which work organization and psychological stress factors contribute to WRMD, in particular among computer users. Under stressful conditions, a chronic hyperventilation/overbreathing response occurs, which is characterized by reductions in the percent of CO_2 in exhaled air, as a psychological stress effect. Physiological effects of overbreathing include heightened muscle tension and paresthesia, which may lead to WRMD. An experimental study of computer workers under stressful conditions provides some empirical support for this model (Schleifer *et al.* 1996). At some point in the stress process, the organism is unable to continue responding normally and exhaustion occurs (Selye 1956). During this stress-induced exhaustion, the immune system cannot function normally and thus cannot provide the typical resources for repairing damaged tissues. Studies by Vaernes *et al.* (1991) and Endresen *et al.* (1991) have shown that job stress, anxiety and depression are correlated with changes in the immune system. Chronic exposure to ergonomic risk factors while the organism is undergoing psychological stress may create micro damage that cannot be fully repaired, and which over time can lead to permanent damage.

3.2. Psychological and Behavioral Reactions to Stress

A second major way in which stress can influence the occurrence of WRMD is through its effects on a person's psychological and behavioral reactions. Thus, stress can affect psychological moods, work behavior, coping style and actions, motivation to report injury and motivation to seek treatment for a WRMD injury, or symptoms of impending injury.

WRMD involve significant pain. Many times, diagnosis of a disorder is based on the nature and extent of pain reported by

the person. Stress may serve to increase the frequency of reporting of pain because of a general increase in personal sensitivity to pain brought on by negative psychological moods. Increased pain or greater severity of pain has been related to psychological stress among patients with low back pain (Atkinson *et al.* 1988) and among large samples of adults (Mechanic and Angel 1987, Korff *et al.* 1988).

A related issue is a social psychological aspect of illness behavior. It is possible that a person under psychological stress could develop specific physical symptoms (such as sore wrists) that would "legitimate" their general psychological discomfort and pain. Having pain in the wrists and fingers is an acceptable disorder, while feeling depressed is not considered as acceptable. Thus, the effects of psychological disturbances may be reflected in physical disorders of the musculoskeletal system. This is similar to mass psychogenic illness (Colligan and Murphy 1979) and psychosomatic disorders (Wolf 1986) where psychologically induced disturbances lead to physical impairment.

The occurrence of WRMD can itself act as a source of stress. Experiencing WRMD can trigger a stress reaction. This relationship has been found between back pain and stress (Feuerstein *et al.* 1987). Job Stress can also affect the behavior of a person in dealing with the work environment. For instance, a person who is stressed may become angry, and this could lead to using improper work methods and/or forceful work techniques (e.g. gripping a tool too tightly). Persons under stress often develop poor attitudes and motivation about the job and about their personal health and well-being (Caplan *et al.* 1975, Kahn 1981). Maladaptive coping behaviors could make people more susceptible to injury or disease and lead to a diminished capacity to work, both conditions increasing the potential for WRMD.

3.3. Work Organization and WRMD

Work organization can influence job stress which, as shown earlier, can affect the risk of WRMD. Work organization can also determine or influence ergonomic risk factors of WRMD. Studies have shown a link between work organization, psychosocial work factors and musculoskeletal disorders, such as symptoms in the back, neck and shoulders (Bongers *et al.* 1993, Moon and Sauter 1996). Some empirical studies have also examined the relationship between work organization and upper extremity WRMD (NIOSH 1992, Smith *et al.* 1992, Ferreira *et al.* 1997). For instance, studies by Smith *et al.* (1992) and Schleifer *et al.* (1996) have shown a link between electronic performance monitoring and musculoskeletal discomfort.

NIOSH (1992) conducted a cross-sectional study of 533 telecommunications workers in five different jobs. Several measures of work organization were related to upper extremity musculoskeletal disorders and symptoms. For instance, fear of being replaced by computers was related to increased neck and elbow symptoms, while high information processing demands were related to increased neck and hands/wrists symptoms. Other work organization factors, such as surges in workload, lack of decision-making opportunities, high task variety and the lack of production standards, were also related to upper extremity WRMD (Hales *et al.* 1994). A more recent cross-sectional survey study of 114 teleservice representatives also conducted by researchers at NIOSH shows further evidence of a link between work organization, psychosocial factors and WRMD (Hoekstra

et al. 1996). Those employees who reported a lack of control over their job tended to have a higher risk of back WRMD, and perceived workload variability was associated with neck WRMD.

Some studies have examined the effect of both psychosocial and ergonomic risk factors on WRMD. Kerr *et al.* (1997) show that psychosocial work factors, such as decision latitude and co-worker support, are important predictors of low back pain, even when adjusting for biomechanical factors. Similar results are reported by Skov *et al.* (1996) in a group of salespeople and by Wahlstedt *et al.* (1997) in a group of 655 postal workers. Wahlsted *et al.* (1997) found that high psychological work demands was associated with symptoms in the lumbar region and that low social support at work was associated with symptoms in the neck–shoulders–thoracic region. Another study of computer users shows that occupational stress can even have a determinant effect on WRMD that outweighs the effect of workstation design (Patterson 1997).

Work organization can define or influence ergonomic risk factors of WRMD, such as repetition, force and posture (for a review of ergonomic risk factors of WRMD; Putz-Anderson 1988). Work organization can define the nature of, strength of and exposure time to these ergonomic risk factors by specifying how a job is to be carried out, establishing product levels and defining pay structure. Work organization may define, for instance, the degree of repetitiveness of the job. In a highly fractionalized job, the worker tends to do the same tasks over and over, which produces repetition and boredom. The work organization also establishes cycle times, through the design of tasks. Short-cycle times and task repetitiveness define the repetition of motions. In this example, the work organization defines that the worker will be exposed to high repetition, which is an ergonomic risk factor for WRMD (Silverstein *et al.* 1987). Work organization also defines the strength of the ergonomic risk factors. A work organization that designs job tasks that do not encourage movement, and do not allow workers to take mini-breaks when needed, may induce static awkward postures. For instance, machine-paced work is a work organization system where workers have little freedom for influencing the pace or standard operation of their work. Such a work system does not allow for any variation in work, and usually does not give workers time to take mini-breaks when needed (Smith 1987). Work organization can also define the exposure time to ergonomic risk factors. By setting work standards and pay schemes, management sets the pace at which a worker is supposed to work. This will then define the exposure time to certain risk factors. For example, if the worker is supposed to produce a certain number of products per time period, then this work standard will in turn define the duration of exposure to certain forces and postures. In addition, management defines the number of hours of work. Overtime, for instance, is a work organization factor that increases the duration of exposure to ergonomic risk factors. Such exposure may be particularly risky, due to increased worker fatigue.

4. CONCLUSION

To control, reduce, eliminate and prevent work-related musculoskeletal disorders, it is important to consider the range of factors which can directly and/or indirectly influence them. Here we argue that from an ergonomics point of view work organization is the driving factor that determines the exposure

to physical ergonomic risk factors and psychosocial work factors. We have defined two types mechanisms of the potential role of stress in the relationship between work organization and WRMD: (1) psychobiological mechanisms (e.g. increased blood pressure, muscle tension and cortisol) and (2) psychological and behavioral reactions to stress (e.g. sensitivity to pain and reporting of symptoms). Several empirical studies have demonstrated a link between work organization and WRMD. Other articles in this volume (e.g. Work Organization Interventions) propose models and methods for reducing WRMD with emphasis on work organization.

REFERENCES

ATKINSON, J.H., SLATER, M.A., GRANT, I., PATTERSON, T.L. and GARFIN, S.R., 1988, Depressed mood in chronic low back pain: relationship with stressful life events. *Pain*, 35, 47–55.

BONGERS, P.M., de WINTER, C.R., KOMPIER, M.A. and HILDEBRANDT, B.H., 1993, Psychosocial factors at work and musculoskeletal disease: a review of the literature. *Scandinavian Journal of Work, Environment, and Health*, 19, 297–312.

Bureau of Labor Statistics, 1997, *Safety and Health Statistics: Workplace Injuries and Illnesses in 1995* (Washington, DC: Bureau of Labor Statistics, US Department of Labor).

CAPLAN, R.D., COBB, S., FRENCH, J.R.P., HARRISON, R.V. and PINNEAU, S.R., 1975, *Job Demands and Worker Health* (Washington DC: US Government Printing Office).

COLLIGAN, M. and MURPHY, L., 1979, Mass psychogenic illness in organizational settings. An overview. *Journal of Occupational Psychology*, 52, 77–90.

COOPER, C.L. and MARSHALL, J., 1976, Occupational sources of stress: a review of the literature relating to coronary heart disease and mental ill health. *Journal of Occupational Psychology*, 49, 11–25.

DALEVA, M., 1987., Metabolic and neurohormonal reactions to occupational stress. In Kalimo, R., El-Batawi, M. A. and Cooper, C. L. (eds) *Psychosocial Factors at Work and their Relation to Health* (Geneva: WHO), pp. 48–63.

ENDRESEN, I.M., ELLERTSEN, B., ENDRESEN, C., HJELMEN, A.M., MATRE, R. and URSIN, H., 1991, Stress at work and psychological and immunological parameters in a group of Norwegian female bank employees. *Work and Stress*, 5, 217–27.

FERREIRA, M.J., DE SOUZA CONCEICAO, G.M. and NASCIMENTO SALDIVA, P.H., 1997, Work organization is significantly associated with upper extremities musculoskeletal disorders among employees engaged in interactive computer–telephone tasks of an international bank subsidiary in Sao Paulo, Brazil. *American Journal of Industrial Medicine*, 31, 468–73.

FEUERSTEIN, M., CARTER, R.L. and PAPCIAK, A.S., 1987, A prospective analysis of stress and fatigue in recurrent low back pain. *Pain*, 31, 333–44.

FRANKENHAEUSER, M., 1986, A psychobiological framework for research on human stress and coping. In Appley, M.H. and Trumbull, R. (eds) *Dynamics of Stress – Physiological, Psychological and Social Perspectives* (New York: Plenum), pp. 48–63.

FRANKENHAEUSER, M. and GARDELL, B., 1976, Underload and overload in working life. A multidisciplinary approach. *Journal of Human Stress*, 2, 35–46.

HAGBERG, M., SILVERSTEIN, B., WELLS, R., SMITH, M.J., HENDRICK, H.W., CARAYON, P. and PERUSSE, M., 1995, In Kuorinka, I. and Forcier, L. (eds) *Work-related Musculoskeletal Disorders (WMSDs): A Reference Book for Prevention* (London: Taylor & Francis).

HALES, T.R., SAUTER, S.L., PETERSON, M.R., FINE, L.J., PUTZ-

ANDERSON, V., SCHLEIFER, L.R., OCHS, T.T. and BERNARD, B.P., 1994, Musculoskeletal disorders among visual display terminal users in a telecommunications company. *Ergonomics*, 37, 1603–21.

HOEKSTRA, E.J., HURRELL, J., SWANSON, N.G. and TEPPER, A., 1996, Ergonomic, job task, and psychosocial risk factors for work-related musculoskeletal disorders among teleservice center representatives. *International Journal of Human–Computer Interaction*, 8, 421–31.

KAHN, R., 1981, *Work and Health* (New York: Wiley).

KARASEK, R.A., RUSSELL, R.S. and THEORELL, T., 1982, Physiology of stress and regeneration in job-related cardiovascular illness. *Journal of Human Stress*, 8, 29–42.

KERR, M.S., SHANNON, H.S., FRANK, J.W., NORMAN, R.W.K., WELLS, R.P., NEUMANN, P. and GROUP, T.I.S.W., 1997, The relative importance of biomechanical and psychosocial risk factors in a case-control study of occupational low-back pain. In Seppala, P., Luoparjarvi, T., Nygard, C.-H. and Mattila, M. (eds) *13th Triennial Congress of the International Ergonomics Association* (Tampere: Finnish Institute of Occupational Health), pp. 64–5.

KORFF, M.V., DWORKIN, S.F., LE RESCHE, L. and KRUGER, A., 1988, An epidemiologic comparison of pain complaints. *Pain*, 32, 173–83.

LEVI, L., 1972, Stress and distress in response to psychosocial stimuli. *Acta Medica Scandinavia*, 191.

MATTHEWS, K.A., COTTINGTOM, E.M., TALBOTT, E., KULLER, L.H. and SIEGEL, J.M., 1987, Stressful work conditions and diastolic blood pressure among blue collar factory workers. *American Journal of Epidemiology*, 126, 280–91.

MECHANIC, D. and ANGEL, R.J., 1987, Some factors associated with the report and evaluation of back pain. *Journal of Health and Social Behavior*, 28, 131–39.

MOON, S.D. and SAUTER, S.L. (eds), 1996, *Beyond Biomechanics – Psychosocial Aspects of Musculoskeletal Disorders in Office Work* (London: Taylor & Francis).

NIOSH, 1992, *Health Hazard Evaluation. US West Communications* (Washington, DC: US Department of Health and Human Services, PHS, CDC, NIOSH).

PATTERSON, K., 1997, Typing skill, exposure, and occupational stress as risk factors for neck and shoulder injury in non-secretarial computer users. In Seppala, P., Luoparjarvi, T., Nygard, C.-H. and Mattila, M. (eds) *13th Triennial Congress of the International Ergonomics Association* (Tampere: Finnish Institute of Occupational Health), pp. 96–8.

PUTZ-ANDERSON, V. (ed), 1988, *Cumulative Trauma Disorders – A Manual for Musculoskeletal Disease of the Upper Limbs* (London: Taylor & Francis).

SAUTER, S.L. and SWANSON, N.G., 1996, An ecological model of musculoskeletal disorders in office work. In Moon, S. D. and Sauter, S.L. (eds) *Beyond Biomechanics: Psychosocial Aspects of Musculoskeletal Disorders in Office Work* (London: Taylor & Francis), pp. 3–21.

SCHLEIFER, L.M. and LEY, R., 1996, Macroergonomics, breathing, and musculoskeletal problems in computer work. In Brown, O. and Hendrick, H.W. (eds) *Human Factors in Organizational Design and Management – V* (Amsterdam: Elsevier), pp. 261–6.

SCHLEIFER, L.M., GALINSKI, T.L. and PAN, C.S., 1996, Mood disturbances and musculoskeletal discomfort: effects of electronic performance monitoring under different levels of VDT data-entry performance. *International Journal of Human–Computer Interaction*, 8, 369–84.

SCHNALL, P.L., PIEPER, C., SCHWARTZ, J.E., KARASEK, R.A. and SCHLUSSEL, Y., 1990, The relationship between job-strain,

workplace diastolic blood pressure and left ventricular mass index. *Journal of the American Medical Association*, 263, 1929–35.

SELYE, H., 1956, *The Stress of Life* (New York: McGraw Hill).

SILVERSTEIN, B.A., FINE, L.J. and ARMSTRONG, T.J., 1987, Occupational factors and carpal tunnel syndrome. *American Journal of Industrial Medicine*, 11, 343–58.

SKOV, T., BORG, V. and ORHEDE, E., 1996, Psychosocial and physical risk factors for musculoskeletal disorders of the neck, shoulders, and lower back in salespeople. *Occupational and Environmental Medicine*, 53, 351–6.

SMITH, M.J., 1987, Occupational stress. In Salvendy, G. (ed) *Handbook of Human Factors and Ergonomics* (New York: Wiley), pp. 844–60.

SMITH, M.J. and CARAYON, P., 1996, Work organization, stress, and cumulative trauma disorders. In Moon, S.D. and Sauter, S.L. (eds) *Beyond Biomechanics – Psychosocial Aspects of Musculoskeletal Disorders in Office Work* (London: Taylor & Francis), pp. 23–41.

SMITH, M.J., CARAYON, P., SANDERS, K.J., LIM, S.Y. and LEGRANDE, D., 1992, Employee stress and health complaints in jobs with and without electronic performance monitoring. *Applied Ergonomics*, 23, 17–27.

SMITH, M.J. and CARAYON-SAINFORT, P., 1989, A balance theory of job design for stress reduction. *International Journal of Industrial Ergonomics*, 4, 67–79.

THEORELL, T., RINGDAHL-HARMS, K., AHLBERG-HULTEN, G. and WESTIN, B., 1991, Psychosocial job factors and symptoms from the locomotor system – a multicausal analysis. *Scandinavian Journal of Rehabilitation Medicine*, 23, 165–73.

VAERNES, R.J., MYHRE, G., AAS, H., HOMNES, T., HANSEN, I. and TONDER, O., 1991, Relationships between stress, psychological factors, health and immune levels among military aviators. *Work and Stress*, 5, 5–16.

WAERSTED, M. and WESTGAARD, R.H., 1991, Working hours as a risk factor in the development of musculoskeletal complaints. *Ergonomics*, 34, 265–76.

WAHLSTEDT, K., BJORKSTEN, M.G. and EDLING, C., 1997, Psychosocial work situation and musculoskeletal symptoms among Postal Workers. In Seppala, P., Luoparjarvi, T., Nygard, C.-H. and Mattila, M. (eds) *13th Triennial Congress of the International Ergonomics Association* (Tampere: Finnish Institute of Occupational Health), pp. 412–15.

WESTGAARD, R.H., 1996, Effects of psychological demand and stress on neuromuscular function. In Moon, S.D. and Sauter, S.L. (eds) *Beyond Biomechanics – Psychosocial Aspects of Musculoskeletal Disorders in Office Work* (London: Taylor & Francis), pp. 75–89.

WOLF, S., 1986, Common disorders and grave disorders identified with occupational stress. In Wolf, S. and Finestone, A.J. (eds) *Occupational Stress – Health and Performance* (Littleton: PSG), pp. 47–53.

Work-related Musculoskeletal Disorders of Upper Limb and Back: Review of Guidelines for their Prevention

F. S. Violante†, L. Isolani‡ and G. B. Raffi‡§

†Occupational Safety Service and ‡Postgraduate School of Occupational Medicine, Servizio di Medicina del Lavoro, and §Occupational Health and Safety Service, Servizio di Sicurezza, Igiene e Medicina del Lavoro, Policlinico Sant'Orsola-Malpighi, Bologna, Italy

1. INTRODUCTION

Although the first definitions of the concept of Ergonomics go back to the previous century, ergonomic guidelines have been developed mostly during the past 15–20 years thanks to the increasing interest in biomechanical factors (force, posture, frequency, vibrations) and work tasks (manual handling of loads, repetitive movements) related to musculoskeletal disorders.

Ergonomic standards appeared late compared with the guidelines of other work-related disorders, such as the those caused by chemical and physical agents.

The development of guidelines in fact reflects the prevalence of the interest in a certain subject at a certain time. Thus, during this century, in the industrialized countries heavy industry prevailed first and chemical and petrochemical later, with the occupational risks related to chemical (metal, dust) and physical (noise, radiation) agents. Starting from the 1970s, the tertiary sector rapidly expanded and at the same time informative systems spread in all the working places. At the same time, great industrial hygiene problems in the work environments (noise, dust, fumes, gas and vapors in the air) were solved or drastically reduced and the prevalence of work-related pathologies changed as a consequence of the prevalence of other risk factors: biomechanical and relational factors. Today, upper limb and back musculoskeletal disorders prevail in terms of health and costs for the worker, industry and society.

This led to the need to develop ergonomic standards for the prevention of work-related musculoskeletal disorders (WRMD) of the upper limbs and back. The development of this type of standards presents many difficulties since the WRMD is a complex group of disorders in terms of definition and content. These difficulties will be clarified if, for example, the pathways to develop guidelines for chemical agents versus ergonomic standards are compared.

In Industrial Toxicology guidelines for prevention of disorders due to the exposure to chemical substances hinge upon the possibility of establishing threshold limit values. The limit values of exposure to chemical substances are fixed by experimental studies on animal models: increasing doses of toxic substances are inoculated in guinea pigs to establish the no-adverse effect level (NOAEL). At the same time data from epidemiological studies are evaluated, in which adverse effects are related to levels of exposure. Therefore, threshold limit values of exposure guaranteeing reasonable protection can be defined.

To define ergonomic standards preventing the WRMD is not easy because:

- musculoskeletal disorders are multifactorial and occur also in the absence of exposure (work environment);
- causal factors such as force, repetition, posture, vibrations can be conceptually difficult to evaluate. For instance, is repetition a primary causal agent or a different way of expressing another causal factor such as for example strength? In Industrial Toxicology this problem is easier to solve: in fact the causal agent is the chemical substance and the level of exposure can be measured quantitatively;
- dose–effect curves are difficult to define due to intrinsic difficulties (absence of animal models for experimental research/studies) and extrinsic (WRMD are also caused by factors other than working activities). Consequentially it is difficult to establish a cause–effect relationship between biomechanical factors and WRMD.

As Bernard (1997) reported in a literature review, epidemiological studies identified a number of mechanical agents strongly associated with WRMD when exposure was intense, prolonged and most of all if a combination of factors occurred.

Despite the difficulties underlined above, in the past two decades many ergonomic standards to prevent WRMD have been developed. They can be divided as follows:

1. Standards with limit values of exposure.
2. Organizational-preventive standards.

Prescription type standards define an acceptable limit of exposure to biomechanical factors related to WRMD. These limits were developed from the evaluation of different biomechanical parameters. These standards usually involve both employer and employees. The employer is required accurately to estimate the risk of WRMD, while employees need to be given an adequate ergonomic training so that the working task can be carried out avoiding or minimizing the risk factors.

Organizational standards assume that you cannot define reliable numerical values to be considered as acceptable limits of exposure. The ergonomic approach is less strict. These standards indeed define general ergonomic measures, delimiting a field of action in which the employer has large liberty of movement. In other terms, the employer is free to create the ergonomic program according to his business.

These standards are directed mostly towards employers who are advised to have a flexible organization able to:

1. survey the working environment and the employees;
2. seek the presence of problems; and
3. develop strategies to decrease or eliminate the problems.

Employee training is always regarded as an essential component because it is a valid measure in order to limit risks.

There are various guidelines for preventing WRMD. Some are laws, some are just technical suggestions: the difference is substantial mostly for the results obtained. Ergonomic standards with the value of law come from countries with a political awareness of the significance of ergonomics concerning health and safety prevention and economical aspects.

Compulsory application of ergonomic parameters in the short-term has a purely educational meaning, whereas in the long-

term it aims to decrease the frequency of WRMD, occupational accidents, amount of sick leave and related health costs.

Among the best-known ergonomic standards there are standards indicating limit values such as the revised NIOSH equation for the design and evaluation of manual lifting tasks (NIOSH), ISO 11228-1 and organizational standards such as the National Standard for Manual Handling (Australia), Ergonomics for the Prevention of Musculoskeletal Disorders (Sweden), Occupational Health and Safety Regulation of Worker's Compensation Board of British Columbia (Canada), OSHA Draft Ergonomics Program (USA) and ANSI Z-365 (USA).

2. STANDARD WITH LIMIT VALUES

2.1. USA: Revised NIOSH Equation for the Design and Evaluations of Manual Lifting Tasks (NIOSH)

In 1981 the NIOSH published *Work Practices Guide for Manual Lifting*. This standard presented analytic procedure including a lifting equation for calculating a recommended weight related to symmetric lifting tasks with both hands and suggestions for controlling the risks of occupational accident to the back due to manual lifting.

In 1993 the revised lifting equation was published. This reflects the result of newer and wider research, but was developed on criteria (biomechanical, psychophysical, physiological) used also in the previous standard and coming from the review of the literature and advice of experts in the field. In the revised equation differences are concerned with: standard position of lifting (increased), limits of weight (lower), multiplicational factors (increased from four to six).

The new standard covers a wide range of tasks. This has a double advantage: on one hand it also allows the evaluation of asymmetric lifting tasks and of the lifting of objects with difficult handles; on the other hand it involves an increased number of protected workers and, since recommended limit values for weight are stricter, also more protection/safety for workers.

The revised lifting equation has application limits. It applies only if the load being lifted:

- is carried with both hands;
- is carried standing;
- excludes other manual activities (pushing, pulling, carrying). These can be if necessary very small (< 10% of the worker's total activities);
- has an adequate friction between floor and feet;
- is carried out smoothly;
- concerns loads not too hot, cold or contaminated with unsteady content; and
- occurs in good microclimatic conditions.

2.2. ISO 11228-1

The International Organization for Standardization (ISO) in 1998 proposed the international standard ISO 11228 (*Ergonomics-manual handling*). ISO 11228 is divided into three parts: (1) lifting and carrying; (2) pushing pulling and holding; and (3) handling of low loads at high frequency.

The first part (ISO 11228-1) is a typical prescriptive standard. ISO 11228-1 has an ergonomic approach similar to that of NIOSH (revised NIOSH equation for the design and evaluation of manual lifting tasks, 1993) for the type of analysis and the evaluation of

manual handling of loads, and for the ergonomic recommendations. ISO 1128-1 does not apply if the object to be handled weighs < 3 kg. Evaluation of the risk of manual handling is made according a step-by-step approach, considering the weight of the object, its initial and final position, the frequency, and duration of the task. This step-by-step approach ceases only when ergonomic conditions of the task are adequate or re-established.

Risk evaluation implies identification, estimation and reduction of risk. The tasks of just lifting and of lifting and carrying are analyzed comparing the weight of the object handled, its initial position, frequency and duration of handling with the suggested related limits. Quality of handling is also considered. Each suggested limit is expressed in equations summarized in tables that show all the other parameters at the same time, supplying in this way an accurate ergonomic evaluation. Limits suggested by ISO hinge on integration of data from epidemiological, biomechanical, physiological and psychophysical studies.

ISO 11228-1 is directed towards designers, employers, employees and anyone involved in the ergonomic programs. In fact, given the complexity of ISO1128-1 (contemporary evaluation of different parameters, comparison with limit values) its application is limited mostly to those in the know. ISO 11228-1 is described as applicable to both the working and domestic population doing manual handling. It is clear that the involvement of the non-working population implies diffusion of ergonomic practices and the consequent familiarity with them: this being just a future goal at the moment.

ISO 11228-1 includes two annexes. Their aim is to simplify and make more understandable what has been previously illustrated in the standard. The first attached sheet gives instructions about factors you must take into account in manual lifting and carrying. It is thus useful to quantify the manual handling risk and define the related ergonomic intervention. The second attached sheet gives examples of evaluation and ergonomic approach to manual lifting tasks. Thus, it makes it possible to apply equations and lifting limits previously determined, and puts into practice the necessary ergonomic intervention.

3. ORGANIZATIONAL STANDARDS

3.1. Australia: National Occupational Health and Safety Committee

In Australia, in 1990, the National Occupational Health and Safety Committee (NOHSC) approved the National Standard for Manual Handling with the ambitious goal of developing, facilitating and increasing a national strategic program for health and safety in the working environment. This standard has legal value and proposes a multifactorial ergonomic approach (risk identification, evaluation, control) and is believed to be better than the simple use of the suggested limit values. Its aim is to prevent accidents due to manual handling or at least to decrease their severity. It includes precise duties for employers and for employees. Employers have duties about design, risk assessment and risk control.

The National Standard for Manual Handling comes together with the National Code of Practice for Manual Handling which supplies explanations about the former, by suggesting the right tools for identification, evaluation and control of manual handling risk. The National Code presents many illustrations, checklists,

questionnaires, tables and summaries of the standard to simplify for employer and employees its application and understanding.

3.2. Sweden: Ergonomics for the Prevention of Musculoskeletal Disorders

In Sweden AFS 1998:1 (*Ergonomics for Prevention of Musculoskeletal Disorders*) is a standard developed by the Swedish National Board of Occupational Safety and Health. It consists of two parts: (1) Provisions; (2) General Recommendations on the implementation of the Provisions. Provisions have a legal status. They define responsibilities in the prevention of WRMD of the employer and other professional figures such as producer, importer and supplier.

In accordance with the traditional attention of the Northern European Countries to psychosocial factors, among the employer's duties it is compulsory, for example, to minimize monotony and repetition. All these relational factors may increase the risk of accident or illness due to manual material handling. Furthermore, the employer deals with his employee's training and they are reprimanded when they do not follow the instructions.

Unlike the other ergonomic standards, AFS 1998-1 gives responsibilities also to the producer, importer, supplier of weights which are dangerous for the health of the people handling them. This illustrates how Sweden has created an ergonomic standard which makes everyone (importer, producer, trader) comply with ergonomic requirements. Special dispensation to this law is represented by products from the European Union in order to avoid hindering free commerce among the different countries. This shows how far European Countries are from Sweden concerning ergonomic programs for prevention and how Europe still lacks unitary ergonomic perspective.

The National Board of Occupational Safety and Health promulgated also General Recommendations on the implementations of the Provisions. Unlike Provisions, Recommendations do not have mandatory value. They rather consist of a wide collection of useful explanations to clarify the meaning of Provisions through images, examples of cases, information and references. Thus they explain and comment on the law that might be difficult to understand as it is basic and concise.

3.3. Canada, British Columbia: Occupational Health and Safety Regulation (Workers' Compensation Board)

In Canada, British Columbia, the Occupational Health and Safety Regulations (1998) was developed by the Workers' Compensation Board. Occupational Health and Safety Regulations consists of three main parts: (1) core requirements; (2) general hazard requirements; and (3) specific industry requirements. Ergonomic requirements (general conditions) are in the first part. They aim to eliminate, or at least minimize the risk of musculoskeletal accidents. Ergonomic requirements are:

1. Definition of risk of musculoskeletal accident.
2. Detection, evaluation and control of the above. Risk factors to be taken into account are related to physical factors, layout and workstation, characteristics of the weight handled, microclimate and work organization.

3. Preparation and training of workers including identification of signs and symptoms of WRMD.
4. Employer and employees' evaluation of efficacy of ergonomic measures selected through a program based on periodic monitoring.
5. Consultation of employees and the Health and Safety Committee by the employers on all the above.

3.4. USA: ASC Z-365 (ANSI)

In 1999 the American Standard Institute (ANSI) was developing an ergonomic standard (ASC Z-365) for the control of cumulative trauma disorders (CTD). ASC Z-365 describes processes and principles for the control and management of CTD and it is directed towards people who have responsibility tasks in health surveillance and security programs, or people who design working environments and procedures. ASC Z-365 clearly requires professional expertise and trained personnel in order to be applied as a general way of controlling specific working situations. The control process of CTD involves mostly employers who are responsible for the health and safety of their employees: this responsibility of course includes CTD.

Compared with other ergonomic standards, ASC Z-365 introduces a novelty: it provides periodic training not only for employees, but also for managers. It aims at a continuous confrontation and the creation of a common ergonomic language about periodic surveillance, work analysis, working place planning and medical care. Involving employees it is relevant and it implies active attendance to ergonomic programs.

In ASC Z-365 surveillance is the main aspect from which stem two branches: on one hand evaluation and management of CTD, on the other hand working tasks analysis. A suspected CTD implies the intervention of a Health Care Provider (HCP) for therapy and follow-up. The aim is an effective secondary prevention. Early diagnosis permits the immediate start of therapy, therapeutic efficacy increases and prognosis improves. At the same time, days missed for sickness decrease and health costs consequently decrease.

Whenever surveillance identifies risk factors for CTD o HCP assesses CTD work-related, work analysis is carried out. The aim is to detect risk factors for CTD. It is not necessary if the ergonomic solution of the problem detected is obvious and simple. Once work analysis is performed, the ergonomic intervention program follows to eliminate risk factors and possible cases of CTD. Subsequently the health periodic surveillance will verify whether the problem is under control.

3.5. USA: Draft Ergonomics Program Standard (OSHA)

In 1999 the Occupational Safety and Health Administration (OSHA) developed a Draft Ergonomics Program Standard. This standard is intentionally simplified to be easily understood. It has been developed from the answers to the questions most often asked by the employer when he faces an ergonomic standard. It clarifies if, how and when it needs to be applied; what the duties are; how and when detection, analysis and control of risk for WRMD need to be performed; how to have training, health surveillance, evaluation of the ergonomic program.

It is a simple and clear draft: it consists also of tables

illustrating different steps of the ergonomic program, duties and people who have them, and periodicity of the health surveillance.

REFERENCES

ARMSTRONG, T.J., 1998, Voluntary industrial ergonomics standard. In International Conference on Occupational Disorders of the Upper Extremities, San Francisco.

BERNARD, B.P. 1997, *Musculoskeletal Disorders and Workplace Factors*. Publication no. 97-141 (Cincinnati: Department of Health and Human Services, Public Health Service, Centers for Disease Control and Prevention, National Institute for Occupational Safety and Health).

ISO, 1998, *Ergonomic-manual Handling. Part 1: Lifting and Carrying*. Draft International Standard ISO/DIS 11228-1 (International Standardization for Organization).

NATIONAL OCCUPATIONAL HEALTH AND SAFETY COMMISSION, 1990, *National Standard for Manual Handling and National Code of Practice for Manual Handling* (National Occupational Health and Safety Commission of Australia).

OCCUPATIONAL HEALTH AND SAFETY REGULATION, 1998, *Ergonomics Requirements* (Canada: Workers' Compensation Board, British Columbia).

OSHA, 1999, *Draft Ergonomics Program Standard* (Occupational Safety and Health Administration).

SWEDISH NATIONAL BOARD OF OCCUPATIONAL SAFETY AND HEALTH, 1998, *Ergonomics for the Prevention of Musculoskeletal Disorders*. AFS 1998:1 (Swedish National Board of Occupational Safety and Health).

WATERS, T.R., PUTZ-ANDERSON, V., GARG, A. and FINE, L.J., 1993, Revised NIOSH equation for the design and evaluation of manual lifting tasks. *Ergonomics*, 36, 749–76.

Work-related Musculoskeletal Disorders: General Issues

B. Silverstein

Safety & Health Assessment and Research for Prevention (SHARP),
Washington State Department of Labor and Industries, Olympia, WN
98504-4330, USA

1. INTRODUCTION

Musculoskeletal disorders (MSD) are injuries and illnesses of the muscles, tendons, ligaments, joints, nerves, vessels and supporting structures that are involved in locomotion. They are usually manifested by pain, numbness, tingling, swelling or loss of function, and are primarily located in the upper limb, back and, to a lesser extent, the lower limbs. The term "MSD" is not a diagnosis but a category of specific and non-specific diagnoses related to the above tissues that have some common features. MSD related to work occur when there is an imbalance in the work system that overwhelms the individual either suddenly (acute onset) or over weeks, months or years (gradual onset). Examples of these disorders most often found related to workplace activities and conditions are shown in table 1. There are at least three major areas of confusion regarding MSD and work.

First, workplace factors are not the *sole causes* of MSD. There can be a number of personal factors that can contribute to MSD either with or without a contribution of work factors, including systemic diseases, recreational activities, age and gender. It is only when workplace factors are significant contributors to the cause or exacerbation of these MSD that they become "work-related" or WMSD. Although the diagnosis of a specific order (e.g. rotator cuff tendinitis) may be made in the same way irrespective of what "caused" it, the importance of distinguishing that factors are primary has more to do with prevention possibilities in the workplace. Personal factors such as age and gender are not as easily modifiable as changing the location or type of tool used in a workplace.

The second area of confusion has to do with *onset* of MSD. These disorders can have either a seemingly sudden onset or gradual onset over weeks, months or years. For example, while

Table 1. Examples of work-related musculoskeletal disorders.

Muscle	trapezius myalgia
Tendon-related	tenosynovitis, tendinitis, peritendinitis epicondylitis, deQuervains disease rotator cuff tendinitis, bicipital tensosynovitis
Nerve	carpal tunnel syndrome, pronator teres syndrome Guyon canal syndrome, cubital tunnel syndrome radial tunnel syndrome sciatica
Vessel/nerve	hand–arm vibration syndrome
Ligament	ulnar collateral ligament laxity of the thumb (gamekeeper's thumb)
Joint	cox arthrosis of the hip

cox arthrosis of the hip can take years of heavy physical load to manifest itself in farmers, low back problems can occur suddenly from a fall or gradually with over exertion due to repeated lifting with an awkward trunk posture over weeks, months or years depending on the exposure and individual capacity. To add to the confusion, how the "onset" of the back trouble is reported often depends on medical payment systems in different locations. For example, low back problems identified as being caused by a sudden or unusual event are more likely to be accepted by many workers compensation systems whereas the more gradual onset disorders are not.

The third area of confusion has to do with "how much is too much" exposure in the workplace. This confusion is largely focused on the "gradual onset" MSD, although the lines of demarcation are not always clear. Designers and engineers want specific "safe" limit values or at least ranges depending on the population for whom they are designing jobs, equipment and processes. The answer "well, it depends" is an area of enormous frustration for practitioners, employers and workers.

The multifactorial nature of gradual onset work-related MSD is not dissimilar from other areas of occupational health such as occupational hearing loss or occupational asthma in which there is an interplay between individual and workplace factors occurring over time that eventually manifest in loss of function or frank disease or injury. There is much that remains to be learned about causation and prevention of WMSD but as Hill aptly wrote: "All scientific work is incomplete. … All scientific work is liable to be upset or modified by advancing knowledge. That does not confer upon us a freedom to ignore the knowledge we already have, or to postpone the action that it appears to demand at a given time."

While other chapters present different ways to evaluate system imbalance, the focus here is to provide an overview of the scope and magnitude of the problem, what is known about the workplace risk factors for MSD and to recommend general prevention strategies based on what can be modified. The discussion concentrates on gradual onset WMSD although much of this discussion is also applicable to sudden onset disorders.

2. CONCEPTUAL MODEL OF GRADUAL ONSET WORK–RELATED MUSCULOSKELETAL DISORDER

Armstrong *et al.* (1993) proposed a cascading conceptual model for gradual onset WMSD where the external environment included exposure (work requirements) that lead to an internal dose (e.g. tissue loads) which sets off a series of internal cascading responses which are in turn modified by individual capacity (figure 1).

This cascading process can lead to training effects or, if overwhelmed without sufficient recovery, to MSD. This model allowed researchers to assess where their own research fits into the process. The model was expanded in 1998 at a US National Academy of Sciences workshop to look at the interplay of a variety of "external" factors that contributed to the development and consequences of prolonged overloading of the internal system (figure 2). This extended model looks at the basic interplay between load on and tissue tolerance in the individual. The social context in which it takes place, organizational and physical work conditions, as well as individual physical and psychological factors

Figure 1. Cascade model of gradual onset work-related musculoskeletal disorders.

Figure 2. National Academy of Sciences Steering Committee model of work-related musculoskeletal injuries.

all have impacts on this relationship. This load-tolerance interplay alone and in the context of the external factors results in physiological responses and perhaps symptoms, which in turn may be reported and eventually result in disability. The many feedback loops in this system illustrate the complexity of causation and suggest multifactorial solutions.

3. MAGNITUDE OF GRADUAL ONSET WORK–RELATED MUSCULOSKELETAL DISORDER

Figure 3 illustrates the difficulty in estimating the true magnitude of the problem of MSD. Case definitions, populations surveyed and the severity of the reported "cases" differ between studies and workers compensation laws differ. Usually studies of workers

Figure 3. Estimation of magnitude depends on how musculoskeletal disorders are identified.

with symptoms report two-to-three times the prevalence as those based on symptoms and clinical findings. While there is some indication of under-reporting of these disorders through official registers, this data usually covers large populations. Focus on disability has a large impact on direct monetary costs but limits prevention opportunities in at risk work groups.

Workers compensation data from Washington State (table 2) can be used to estimate the magnitude of the problem in most industrialized countries for gradual onset upper limb and back disorders over all and for some specific diagnostic categories (Silverstein and Kalat 1998). The more specific the diagnosis, the longer the lost workdays and the higher the cost. Women tend to be disproportionately represented in hand/wrist/carpal tunnel syndrome, whereas males are in the back/sciatica categories. Compared with all industry, those industries with at least a 5-fold increased rate of upper limb claims included temporary help in assembly and machine operations, shake mills, seafood canneries, meat products manufacturing, wallboard installation and roofing. For gradual onset back disorders, the industries with at least a 5-fold increased rate included shake mills, wallboard installation, roofing, pipe manufacturing, temporary help in assembly and moving companies.

4. WORKPLACE RISK FACTORS

Workplace risk factors have been identified in a number of epidemiological studies, population surveys, clinical studies, human and animal experimental studies. Bernard et al. (1997) conducted an exhaustive review of the epidemiological literature.

Table 2. Washington State workers compensation average claims incidence rate (CIR) and cost for gradual onset upper limb and back musculoskeletal disorders.

	CI per 10 000/year	% 4+ lost workdays	No. of lost workdays	Cost ($)	% female	Median age (years)
Upper extremity	90.6	35.4	171	5986	43.2	34
Shoulder	30.6	38.5	202	7692	35.6	34
Rotator cuff sydrome	20.9	64.3	250	15 370	33.2	38
Elbow	12.3	36.2	229	7428	36.5	38
Epicondylitis	12.1	46.0	201	6536	40.0	39
Hand/wrist	44.2	41.3	201	7401	51.0	33
Carpal tunnel syndrome	28.2	68.6	219	12 821	55.9	36
Back	161.9	35.4	171	5986	28.4	32
Sciatica	5.0	80.9	442	39 797	31.7	37

Table 3. Epidemiological evidence for causal relationship between physical work factors and gradual onset musculoskeletal disorders (Bernard *et al.*).

	Repetition	Force	Posture	Vibration	Combination	Other
Neck	**	**	***	+/-		
Shoulder	**	+/-	**	+/-		
Elbow	+/-	**	+/-		***	
Carpal tunnel syndrome	**	**	+/-	**	***	
Hand/wrist tendinitis	**	**	**		***	
Hand–arm vibration syndrome				***		
Back		lifting ***	**	whole body ***		heavy work**
						static postures +/-

+/- Insufficient evidence; **evidence; ***strong evidence.

Based only on epidemiological studies of working populations, they concluded that there was sufficient evidence for causal relationships between a number of risk factors including high repetitiveness, high force, awkward postures, and vibration and disorders of different body areas (table 3). The evidence was strong particularly when risk factors were combined. The NRC Report (1998) came to similar conclusions based on a wider array of evidence. Psychosocial aspects of work, such as low autonomy and high demand with little supervisory support have also been identified as at least exacerbaters of physical load and WMSD. The problem with many of the epidemiological studies is the lack of specificity in exposure assessment that makes quantitative dose–response relationships difficult to determine. On the other hand, laboratory studies can quantify exposure but the outcome measure is not illness or injury but rather some measure of localized fatigue, discomfort, capacity or acceptability. The relationship between fatigue and illness/injury requires further research. Nonetheless, the quantitative estimates are more useful for designers and practitioners who are responsible for ensuring a safe work design. The mere existence of a "hazard," such as a 50 kg box, does not put a worker "at risk" until there is some interaction between the worker and the hazard that can be described in terms of the temporal variation, duration and intensity of the contact. Although a similar critical review of the laboratory data has not been published, different combinations of risk factors have provided quantitative estimates based on reduced capacity or acceptability ratings. The psychophysical

tables for material handling "acceptable limits" of Snook and Cirello (1991) and 1991 Revised NIOSH Lifting Equation (Waters *et al.* 1993) have provided valuable guidance for designers, despite some of their limitations. Similar work is taking place for the upper extremity, particularly in looking at "repetitive work."

There are basically three methods for conducting exposure assessment: ask (interviews, questionnaires, diaries), look (checklists, videotapes) and direct measurement (surface electromyography, electrogoniometry, force transducers, accelerometers). There are tradeoffs in simplicity, precision, interpretation and utility between these methods. An important goal for ergonomists is developing a common taxonomy for describing the duration, intensity and temporal patterns of exposure to different risk factors, irrespective of method. This would allow for better comparisons of findings between studies. The University of Michigan's effort to provide written descriptions of different upper limb risk factors intensity levels and a catalog of videotapes of jobs that match those level descriptions, may provide the basis for being able to combine the "ask" and "look" domains of exposure assessment.

The Swedish regulation's guidance on an approach to assessing risk in repetitive work (1998) (table 4) can assist employers in evaluating a number of workplace factors that contribute to upper limb disorders. This guide looks at short work cycle, lack of opportunity to change postures and movements, lack of freedom of action, single-task job content and lack of learning possibilities as putting an individual at

Table 4. Assessing physically monotonous, repetitive work (adapted from Swedish National Board of Occupational Health 1998).

	Red	Yellow	Green
Work cycle	Repeated several times/minute for at least half a shift	Repeated several times/min for at least 1 h or many times an hour for at least half a shift	Repeated a few times every hour
Working postures and movements	Constrained or uncomfortable	Limited scope for changing work postures and movements	Good possibilities of varying working postures and movements
Job decision latitude	Work is completely controlled external to the worker	Work partly controlled external to worker. Limited possibility to influence way task is performed	Good possibilities to adapt work to one's ability. Influence on planning, work arrangement
Job content Training/competence	Does isolated task in production process. Short training phase	Does several tasks in production process. Training for several tasks	Does several tasks or whole process including planning, instruction. Continuous competence development

extreme risk. This allows the employer and workers to identify why these factors are present and what to improve to provide a healthier work environment. All four of these dimensions of work could be obtained by asking or looking, while more precision could be obtained through direct measurement of work cycle time, and postures and movements. These dimensions could also be used to design new jobs. Similarly, Kilbom (1994) provided quantitative guidance to practitioners on when jobs with repetitive motions of different upper limb areas should be evaluated more closely when combined with > 1 h duration and high physical loads or "force." While all of these guidelines are provided with various caveats about not necessarily being safe for all individuals, they try to provide sufficient specificity for designers and practitioners to use in assessing jobs. Recognizing that these quantitative estimates provide no guarantes, they should be used, tested and the results reported in the literature for future refinement.

A recent employer survey of ~5000 employers from different sizes and industry sectors indicated that ~10% of employers had no employees exposed to musculoskeletal risk factors associated with manual handling or repetitive work (Foley and Silverstein 1999). In multivariate modeling of employer reported numbers of employees exposed to different durations of 14 risk factors, there were significant increases in workers compensation claims rates with increases in total available employee work hours exposed to most manual handling risk factors (lifting heavy loads, frequent lifting, lifting with awkward postures) and upper limb risk factors (working with hands above shoulder height, repetitive arm movements and vibrating tool use). This suggests that employers have the capacity to identify, at least crudely, significant risk factors. The same survey indicated that for those employers taking prevention steps, they not only had reductions in injury frequency and severity (> 50%) but also in collateral benefits of reduced absenteeism and turnover and increases in productivity, quality and employee morale (25–35%).

5. PREVENTION STRATEGIES

The critical next step after performing risk assessment is in identifying why unacceptable risk may be present in any particular job or process. A systems approach provides the best opportunities for process changes that can both improve health and performance. The opportunities vary depending on the job life cycle. The greatest opportunity for engineering controls is in the product design and then the process development phases. Engineering controls are still possible in the production implementation and personnel use phases but administrative improvements also play a role at this point and tend to be less effective.

Basic principals of risk reduction efforts in individual job design at the process, production and personnel phases include:
- Provide workstations where workers can reach, see and fit (avoid extremes in range of motion).
- Provide variability in tasks that can lead to variability in postures and motions (reduce risk from fixed postures and highly repetitive movements).
- Keep loads close to reduce biomechanical disadvantage.
- Provide opportunities for adjustment in tools, equipment and workstations especially where there are multiple users.

- Provide preventive maintenance for tools and equipment to reduce exertion requirements.
- Provide opportunities for workers to control the way they perform the job and to learn new things.
- Use a collaborative problem-solving process to identify the solutions that solve the most problems simultaneously without introducing new ones.

6. CONCLUSIONS

Most practitioners would suggest incorporating ergonomics into a regular part of doing business. Implementing ergonomics related processes and practices requires ongoing top management commitment, supervisor and employee knowledge and involvement and ongoing evaluation of the effectiveness of the process in reducing MSD. In most countries, voluntary efforts in this direction are necessary but have rarely been demonstrated to be sufficient to prevent WMSD in all workplaces. While there is much to learn about quantifying relevant exposures and their interactions, enough is known to use general principles to reduce hazardous exposures now.

REFERENCES

ARMSTRONG, T.J. *et al.*, 1993, A conceptual model for work-related neck and upper-limb musculoskeletal disorders. *Scandinavian Journal of Work and Environmental Health*, 19, 73–84.

BERNARD, B.P. *et al.*, 1997, *Musculoskeletal Disorders and Workplace Factors: A Critical Review of Epidemiologic Evidence for Work-related Musculoskeletal Disorders of the Neck, Upper Extremity and Low Back*. CDC/NIOSH 97-141 (Cincinnati, OH:NIOSH).

FOLEY, M. and SILVERSTEIN, B., 1999, *Washington State Employer Survey on Musculoskeletal Disorders, Risk Factors and Prevention Steps*. SHARP 53-01-1998 (Olympia: Washington State Department of Labor & Industries).

HILL, A.B., 1965, *Proceedings of the Royal Society of Medicine*, 56, 295–300.

KILBOM, A., 1994, Repetitive work of the upper extremity. Part I – Guidelines for the practitioner; Part II – The scientific basis for the guide. *International Journal of Industrial Ergonomics*, 14, 51–86.

SILVERSTEIN, B. and KALAT, J., 1998, *Work-related disorders of the back and upper extremity in Washington State, 1989–1996*. Technical Report 40-1-1997 (Olympia: Washington State Department of Labor and Industries).

SNOOK, S.H. and CIRELLO, V.M., 1991, The design of manual handling tasks: revised tables of acceptable weights and forces. *Ergonomics*, 34, 1197–214.

STEERING COMMITTEE FOR THE WORKSHOP ON WORK-RELATED MUSCULOSKELETAL INJURIES, COMMITTEE ON HUMAN FACTORS, NATIONAL RESEARCH COUNCIL, 1998, *Work-related Musculoskeletal Disorders: A Review of the Evidence* (Washington, DC: National Academy Press).

SWEDISH NATIONAL BOARD OF OCCUPATIONAL HEALTH, 1998, *Ergonomics for the Prevention of Musculoskeletal Disorders*. AFS 1998:1 (Stockholm, Sweden: SNBOH).

WATERS, T.R., PUTZ-ANDERSON, V., GARG, A. and FINE, L.J., 1993, Revised NIOSH equation for the design and evaluation of manual lifting tasks. *Ergonomics*, 36, 749–76.

Work-related Musculoskeletal Disorders: Overview

L. Forcier† and I. Kuorinka‡

†Institut de recherche en santé et sécurité du travail, 505 Blvd Maisonneuve O., Montréal Québec Canada H3A 3C2

‡F-46 160 Gréalou, France

1. INTRODUCTION AND DEFINITIONS: SORTING THROUGH THE CONFUSION

Work-related musculoskeletal disorders (WMSD) are known by a variety of terms across the world. In the USA, "cumulative trauma disorders" (CTD) is sometimes used to refer to this group of disorders. In Japan, they have been known as "occupational cervicobrachial disorders" (OCD), in Canada and Australia as "repetitive strain injuries" (RSI), and more recently in Australia as "occupational overuse syndrome" (OOS). Lately, WMSD has gained popularity, worldwide, as the preferred designation for this group of disorders.

Although the exact description of what constitute WMSD may vary among countries, states or provinces, and indeed among researchers, generally speaking they are chronic disorders (as opposed to due to an accident) of the upper and lower limbs and are, by definition, related to work. These work-related chronic disorders of the limbs may involve tendons, muscles, bones and cartilage, bursa, or peripheral nerves and also include some selected vascular disorders (Table 1) (Kuorinka and Forcier 1995). Of course, these same disorders may occur in non-working populations or their development may be associated with factors not related to work; when such is the case, these disorders are not considered *work-related musculoskeletal disorders*.

Theoretically speaking, musculoskeletal back disorders can also be work-related and may share quite a few similarities with WMSD of the limbs; however, historically, research on back disorders has been separate and often the term WMSD excludes the back (in this chapter the back is *excluded*). There is far more information on WMSD of the upper limbs than of the lower limbs; however, evidence is mounting which shows that lower limbs, their tendons, bursa, etc., are also at risk of developing WMSD.

It is important to make here another small a part on the terminology involved in WMSD. Dorland's *Illustrated Medical Dictionary* (1974: 465) defines *disorder* as: "A derangement or abnormality of function; a morbid physical or mental state." Disorders as in WMSD, therefore, encompasses some clinically well defined *diseases* ("a definite morbid process having a characteristic train of symptoms … its etiology, pathology, and prognosis may be known or unknown"; Dorland's 1974: 453), for example, de Quervain's disease, and some well defined *syndromes* ("a set of symptoms which occur together; the sum of signs of any morbid state"; Dorland's 1974: 1575), for example the carpal tunnel syndrome. WMSD also include some clinically less defined conditions, for example tension neck syndrome, and some non-specific groupings of musculoskeletal symptoms (Table 1).

Finally the term "work-related musculoskeletal *problems*" may also be used, especially in research projects. This term may only refers to symptoms (subjective evidence, e.g. pain, ache, fatigue, discomfort) and/or signs (objective evidence, e.g. limited movements) of musculoskeletal dysfunction without a full pledge diagnosis, although sometimes the term "problems" may be used to embrace both the disorders *and* various non specific groupings of symptoms and/or signs. When the latter occurs, of course, work-related musculoskeletal *problems* and work-related musculoskeletal disorders mean the same thing.

Definition of the WMSD is partly problematic because the underlying phenomenon is a medical issue, described in medical

Table 1. Examples of work related musculoskeletal disorders (based on Kuorinka and Forcier 1995).

Anatomical structures involved	Examples of possible WMSD[1]
Tendons	•Tendinitis/peritendinitis/tenosynovitis/ insertion tendinitis (enthesopathy)/synovitis of most joints, in particular shoulder, elbow and hand-wrist •Epicondylitis •de Quervain's disease (stenosing tenosynovitis) •Trigger finger
Nerves	•Median nerve entrapment: Carpal tunnel syndrome (CTS) (entrapment at the wrist) and Pronator teres syndrome (entrapment at the elbow) •Ulnar nerve entrapment: Cubital tunnel syndrome (entrapment at the elbow) anf Guyon canal syndrome (entrapment at the Guyon canal) •Radial tunnel syndrome (radial nerve entrapment at the elbow) •Thoracic outlet syndrome (TOS) (neurogenic TOS = entrapment of the brachial plexus at different locations) •Cervical syndrome (radiculopathy) (compression of nerve roots)
Circulatory/Vascular structures	•Hand-arm vibration syndrome (involves vascular and nerve damage) •Hypothennar hammer syndrome •Raynaud's syndrome
Joints (cartillage and bone)	•Osteoarthritis of most joints/degenerative joint disease
Muscles	•Tension neck syndrome
Bursa	•Bursitis of most joints
	•NON SPECIFIC GROUPINGS OF SYMPTOMS[1]

[1] It should be noted that sometimes the terms WMSD, CTD, RSI, etc have been used as umbrella terms to group everything shown in this table and at other times they have been used to soley refer to the non-specific groupings of symptoms.

terms and organized using a medical taxonomy. A problem arises when the medical issue is associated with determining "work relatedness," which makes the whole concept fuzzy and imprecise rather than clarifying it. For an ergonomist, it might be most useful to adopt a pragmatic approach and to accept the WMSD concept in its largest sense. Such attitude is the basis for preventive actions, actions in which the ergonomist has a far more useful role than in arguing the work relatedness of WMSD cases.

2. SCOPE OF WMSD: WHICH COUNTRY, INDUSTRY, AND JOB? A LOOK AT STATISTICS

In the USA, the incidence of disorders due to repeated trauma, as they are known by the Bureau of Labour Statistics (BLS), has risen. In private industry, from 1984 to 1997 it increased from 5.1 to 32 cases per 10 000 full time worker (Hales and Bernard 1996, BLS 1999). Some of the explanations *postulated* for this increase include (1) increased awareness that these disorders may be work-related leading to improved recording, improved detection and increased reporting, and expanded workers compensation law, and (2) increased number of cases due to (a) increased productivity (b) increased number of women on the work force (c) shift to service industry jobs and (d) increased VDT use (Brogmus *et al.* 1996, Hales and Bernard 1996).

WMSD have been observed worldwide, in developed, developing and underdeveloped countries. They are also found in many industries and jobs; indeed they can be found wherever there is the presence at work of certain risk factors thought to be associated with the development or aggravation of these disorders. They have been noted in agriculture, forestry, fishing, mining, construction, manufacturing, transport, wholesale trade, retail trade, and service industries and public administration, in blue and white collar workers, trades and professional jobs. However they can be more prevalent in some industries and jobs (Table 2). Among the occupations or industries more at risk, many are

good-producing related. Indeed, in an analysis of the 1993 CTD of the upper extremities claims from the Liberty Mutual Group (largest writer of worker's compensation insurance in the USA since 1936), 30 of the 39 highest risk jobs for CTD claims were manufacturing (although manufacturing jobs account for only 21% of US employment), and 11% of claims were associated with jobs with heavy VDT use (Brogmus *et al.* 1996).

There is no perfect agreement on the rate of WMSD between geographical areas, or on which industries or jobs are more at risk (Table 2). This is not surprising since compensation statistics represent only what is legally acceptable as a compensation claim at one point in time, in any given country, state or province.

In fact, workers compensation statistics are likely to show only part of the scope of WMSD, not only because their "case definition" (i.e. acceptable claim) is restricted, but because other mechanisms may also be operational and have an impact on the number of disorders reported. Known since 1885 (Fox and Collier 1976), the *healthy worker effect* may be one such mechanism. Although this effect is comprised of many element (Choi 1993, Last 1995), one element is particularly pertinent to compensation statistics (and OHS research). It is known that workers who have developed health problems in their work have a tendency to quit their work, which may result in fewer WMSD being reported and claimed. Also noteworthy, in terms of reducing the number of claims for WMSD, workers may not report a WMSD but may use days of absence from work to deal with a WMSD.

When researchers study a workplace with known risk factors they do tend to find that the *size* of the problem *is greater than that shown by compensation rates*. Of course, the case definition (i.e. what will be considered as a WMSD) used by researchers is very different than that used in the compensation system and in fact varies from one research project to the next. The results of a study by Chiang *et al.* (1993) for 207 workers in the fish processing industry will be briefly mentioned here to illustrate

Table 2. Examples of workers compensation statistics for WMSD (top five industries).

USA (Private industry–disorders associated with repeated trauma in 1997 (Incidence rate per 1,000 full–time workers[1])	CANADA (Manitoba Workers Compensation Board, 1991 RSI data for the upper limbs) (Rate per 1,000 full–time workers[2])	AUSTRALIA (RSI for the upper limits for 1985–1986 for South Australia) Industry incidence (cases per 1,000 person–years[3])	
		Men	Women
•Meat packing plants (119.2)	•Manufacturing: Packing houses, manufacturing meat products, poultry processing, stockyards (23.5)	•Textiles, clothing footwear manufacturing (4.7)	•Basic metal manufacturing (16.7)
•Motor vehicles and car bodies manufacture (74.1)	•Manufacturing of aeroplanes (9.1)	•Food and beverage manufacturing (4.5)	•Transport equipment manufacturing (16.2)
•Poultry slaughtering and processing (52.3)	•Manufacturing: Bakeries, food, and dairy products (3.4)		•Electricity, gas, water (13.4)
•Knit underwear mills (49.4)		•Miscellaneous manufacturing (4.4)	•Miscellaneous manufacturing (13.4)
•Men and boys' trousers and slacks	•Manufacturing: Planing mills, tanneries, sash and door factories, monument dealers, manufacture of furniture and concrete blocks (3.3)	•Public administration (3.9)	•Fabricated metal manufacturing (9.0)
	Manufacturing: Foundries, rolling mills, machine shops, scrap metal dealers, agricultural machinery, manufacturing of vehicles (3.0)	•Agriculture (3.1)	

[1] In 1997 the incidence rate for all US private industry combined was 3.2 per 1,000.
[2] Manitoba overall RSI claim rate was 1.0 per 1,000.
[3] South Australia overall RSI claim rate was 1.9 per 1,000.

the type of rates that can be found in a research study. Besides having been conducted in a high risk industry, this study was selected here because it met the 4 quality criteria mentioned in the a National Institute of Occupational Safety and Health (NIOSH) review of the literature (1997) (70% participation rate or more; used a physical exam to assess WMSD; used observation or measurements for risk factor analysis; and investigators were blinded to cases and exposure status). For physician observed disorders of shoulder girdle pain, Chiang *et al.* found that 10% of workers were a case in their *low* risk factor exposure group, 37% in the *mid*-exposure group and 50% in the *high* exposure group (for 1000 workers this means a rate of 100 in the low, 370 in the mid-, and 500 in the high). For the carpal tunnel syndrome the case rates per 1000 workers were, respectively for each group: 80 for low, 150 for mid-, and 290 for high.

Statistics based on different case definitions will, therefore, define a greater or lesser population of workers with the WMSD. The road between the beginning of a work-related musculoskeletal problem to the declared and compensated claim is long and not understood; consequently to correctly establish OHS priorities, ergonomic interventions and prevention programs, it would be preferable to have the broadest picture possible, using WMSD data collected in the workplace itself and not just the company's WMSD compensation data.

3. RISK FACTORS AND EVIDENCE OF THE WORK RELATEDNESS OF VARIOUS DISORDERS

Factors which are known, on the basis of epidemiological evidence, to be associated with the development or aggravation of a health outcome (musculoskeletal disorders for our purposes here) are considered risk factors (Last 1995). These may be factors to do with inborn or inherited characteristics (grouped here under the term "personal factors") or lifestyle and environmental exposures (divided here into "*work* risk factors" and "outside work risk factors"). Although presented as separate categories of risk factors, there are links between factors (a certain personality or physical stature may be predisposed to choose or be given a certain type of job), and interactions between factors (cumulative exposure to repetitive use of a limb at work and in sports), irrespective of category. This section of the article will concentrate on work risk factors, with only a mention of the other risk factors. Table 3 shows a list of possible work risk factors; usually workers"

exposure is not limited to one work risk factor but combinations of risk factors are present.

It should be noted that the terms risk factors and work factors are often used loosely and interchangeably to describe any factor at work which could possibly be associated with WMSD even in the absence of statistical/epidemiological evidence or even if far remove in the chain of events (e.g. poor lighting is called a risk factor when the underlying risk factor is the poor posture that may result).

Research investigating the possibility of a link between *physical* work factors and these musculoskeletal disorders has been reviewed (Kuorinka and Forcier 1995, Hales and Bernard 1996, NIOSH 1997, Viikari-Juntura 1995, 1998). These reviews *included all studies with a determined level of quality, irrespective* of whether the results *did not support* a link with work or *did indicate* the possibility of an association. Generally speaking, there is agreement between these *independent* reviews: there are, indeed, links between various risk factors at work and musculoskeletal disorders. Table 4 represents a fusion the NIOSH (1997) and Viikari-Juntura (1998) findings after examination of the association between some selected disorders and some physical work exposures. There is no systematic examination of the work-relatedness of lower limbs in either of these reviews, however Kuorinka and Forcier (1995) did review knee bursitis and found that there was evidence for a relationship between posture (kneeling) and the disorder. Finally, among physical risk factors at work, exposure to cold or working in a cold environment has been postulated as associated with musculoskeletal disorders. Although none of these reviews systematically assessed the relationship with cold, there is a section in the Kuorinka and Forcier (1995) which discusses briefly some existing studies.

Currently there is more information available on physical risk factors than on psychosocial risk factors. Generally speaking, studies have shown associations between WMSD and psychosocial factors related to job and task demands (especially heavy workload, time and work pressure, limited job control, and monotonous work) (Hales and Bernard 1996, NIOSH 1997). There are also studies that have found links between lack of job clarity and WMSD (Hales and Bernard 1996, NIOSH 1997). Finally, poor social support at work (from supervisor and/or co-workers) has also been linked with WMSD. Mechanisms postulated to explain such associations are based on the possibility

Table 3. Possible risk factors for WMSD at work (inspired from Huorinka and Forcier 1995, Hales and Bernard 1996, and NIOSH 1997).

Physical risk factors at work	
•Force	Forceful exertions, and/or shock, impact and impulsive loading, and/or local mechanical stresses
•Repetition	Cyclical work activities, repeated tasks or work that involve frequent and/or prolonged cyclical movement or repetitive movement of selected body segments/joints
•Posture	Awkward, adverse or extreme postures and static postures leading to loading of anatomical structures
•Cold	Working in a cold environment, drafts and other local/segmental exposure to cold (e.g. gripping cold objects)

Psychosocial risk factors at work[1]	
Factors related to job or task demands	e.g. heavy workload, monotonous tasks, poor work content, time pressure, increasing work pressure, lack of autonomy, limited job control . . .
Factors related to the organisation as a whole	e.g. job insecurity and lack of job clarity, employee–supervisor relationship and poor social support, shiftwork . . .

[1] Factors between and within categories overlap.

Table 4. Epidemiological evidence for association between physical risk factors at work and selected WMSDs (adapted from the NIOSH review, 1987, with notes from the Viikari-Juntura, 1998, review where there are differences).

Physical risk factor at work	Neck (including neck/shoulder studies which were likely to be specific to neck)[1]	Shoulder tendinitis[2]	Lateral Epicondylitis[3]	Carpal Tunnel Syndrome[4]	Hand/wrist tendinitis[5]	Hand–Arm vibration
Force	Evidence	Insufficient evidence	Evidence	Evidence	Evidence	
Repetition	Evidence	Evidence	Insufficient evidence	Evidence	Evidence	
Posture	Strong evidence	Evidence	Insufficient evidence	Insufficient evidence[6]	Evidence	
Vibration	Insufficient evidence	Insufficient evidence		Evidence[7]		Strong evidence
Combination of the above factors	Insufficient evidence	See table note[8]	Strong evidence[9]	Strong evidence	Strong evidence	

[1] Includes some studies on tension neck syndrome and some studies of neck and/or neck and shoulder problems based on symptoms alone or based on symptoms and physical examination.

[2] Some studies included a diagnosis of shoulder tendinitis while others used a combination of symptoms and physical examination findings or symptoms alone but do not necessarily diagnose them as tendinitis.

[3] Includes some studies with physical examination for epicondylitis and some studies using symptoms only as case definition.

[4] Includes a few studies where CTS was assessed by symptoms alone, and studies where CTS was assessed by symptoms and physical findings and/or electrophysiological tests of nerve function.

[5] Includes studies on de Quervain's disease, trigger finger, tenosynovitis, peritendinitis, tendinitis. All studies used at least a physical examination to determine cases.

[6] Viikari-Juntura denotes the results of research as showing "Evidence".

[7] Viikari-Juntura denotes the results of research as showing "Strong evidence".

[8] NIOSH, in its summary table, does not comment on the possibility of a link between combination of these risk factors and neck disorders, however, Viikari-Juntura, in her summary table, indicates that there is "Evidence" for an association between combination of these factors and Rotator cuff tendinitis.

[9] Viikari-Juntura denotes the results of research as showing "Evidence".

of a direct relationship (i.e. these factors increase stress which in turn increases muscle strain and subsequently the response to physical risk factors) and/or indirect relationship (i.e. these factors mediate the employees" perception and response to the physical risk factors). Job dissatisfaction has also received much attention and has shown some links with WMSD (NIOSH 1997); however here we did not consider it as a work risk factor since it is likely to made up of very many components and appears to lie more at the junction of work/personal/outside work risk factors.

Some personal factors (e.g. gender, age, weight, some anatomical differences, etc.) and outside work risk factors (e.g. tasks at home) have also been linked with the development of musculoskeletal disorders. Many studies have reported that women have a higher WMSD rate than men (for example, see South Australian incidence rates in Table 2). Depending on culture, some women may report and seek treatment more readily than men; as well, they may have added personal or outside work exposure. However, generally speaking, women also occupy very different jobs than men, usually jobs of a more repetitive nature. Further, workplace design may have traditionally been thought of in terms of men. In some studies where occupational exposure factors are controlled there is no longer an association with gender (e.g. Stevens 1988). Beside gender, Kuorinka and Forcier (1995) and Hales and Bernard (1996) have examined some other personal factors such as age, weight and a few others.

Obviously, if only personal factors and/or outside work factors contributed to the development of a particular disorder, for that particular case in that particular individual, we are no longer dealing with a *work-related* musculoskeletal disorder. However, in cases declared at work, some could be associated with work risk factors alone, while others may involve a mixture of work and personal/outside work factors. Currently, unless in the presence of very compelling evidence concerning one type of factor versus another, it is difficult to apportion the various contributions. Rather than deal with a case by case scenario, it is by far preferable to look at the work environment of employees with problems and their close co-workers, and do population statistics and risk factor detection to see if risk does exist at work, irrespective of each individual's personal and out of work exposure profile. After all, in work environments where there are extensive WMSD risk factors identified, the role of personal and outside work risk factors is likely to be far less important (Kuorinka and Forcier 1995). In the Chiang study (1993) cited in the previous section, the low exposure group showed low prevalence of WMSD. In a recent article, Battevi *et al*. (1998) observed workers *not exposed to* tasks considered at risk of WMSD and found low prevalence of WMSD (1.9% for males and 5.2% for females for a total of 3.9%) as well as low prevalence of pain, irrespective or duration and frequency (13.5 % for men and 9.6% for women for a total of 11.2%). The authors conclude that "the prevalence were on average quite low . . . hence the authors recommend that even minimal prevalence detected in particular work environment should not be underestimated" (see the Waldermar surveillance article in this volume).

4. A NOTE ON THE LINK WITH WORK, WMSD ETIOLOGY AND PATHOPHYSIOLOGY OF THE DISORDERS

Partly, what renders understanding the mechanisms behind the links between work risk factor and the development or aggravation of the disorders difficult is that, in many cases, the pathophysiology processes involved in the varied disorders are not understood. Each type of disorder is likely to have its own process, although some may share similarities. It is also hard to know if the development or the aggravation of WMSD occurs suddenly (sudden onset) or over days, weeks, months or years of exposure (gradual onset) and in a cumulative fashion even in the presence of small exposures and/or in one or a series of overdose. The safe "value" for exposure to each risk factor is unknown. The resulting exposure "value" of the interaction between various work risk factors (additive impact?, synergism?, some cancellation?) is also unknown. Further, the role of remedial factors (e.g. time of recuperation allowed between each exposure) or possible protective factors (e.g. individual physical and mental fitness) is not fully understood. And, again, all of this is likely to vary according to the actual disorder.

5. EXPLANATORY MODELS

WMSD are a heterogeneous group of pathological conditions. Many efforts have been made to explain the natural history and the chain of events leading to WMSD, however, no generally accepted model is available (e.g. Johansson and Sojka 1991, Armstrong *et al.* 1993). The reason may be that it can be hypothesized that each WMSD condition may have different causes and the same cause may produce different outcomes. However, a relative consensus seems to exist that the final link in the chain of events leading to a WMSD pathology is the mechanical stress on the musculoskeletal system.

6. IDENTIFICATION OF FACTORS AT WORK

Whether talking of physical work risk factors (mechanical exposures) or psychosocial work risk factor, there is no perfect method for identifying and measuring risk factors. Each method has its advantages and its drawback; the choice of which method to use will depend on the context, the level of depth required of the results, the costs and resources available, and the factor(s) to be measured. This section will briefly discuss detection of physical risk factors at work with a passing mention of psychosocial factors.

Table 5 provides a brief overview of various types of methods available to identify and measure physical risk factors in order to be able to estimate the mechanical exposure which result from being exposed to these risk factors. As one travels across this table from exploratory "measures" to direct measurements, the precision/quality of the results increases, however so are the costs and efforts involved in collecting and analyzing the data.

Some of the same types of methods which are used to measure physical risk factors (Table 5) can also be used to measure other work factors. We define here "other work factors" as work variables (e.g. workstation design, design of work areas and other

Table 5. Overview of various types of methods available for identifying and measuring physical risk factors at work (inspired from Kuorinka and Forcier 1995 and van der Beek and Frangs-Dresen 1998).

	Exploratory	Subjective measurements	Systematic observations	Direct measurements	Interpretation of results
Types of methods	•Informal interviews •Self-reports	•Self-reports •Questionnaire techniques •Expert judgement	•Observations at workplace •Video observations	•Direct measurement at workplace •Measurements during simulation in laboratory	•Experts opinion •Panels •Reference groups
Examples	•Interviews •Discussions •Collecting opinions	•Questionnaires •Diaries •Quick checklists	•Systematic samples of tasks and work of individual workers	•Physiological (e.g. EMG) •Biomechanical (e.g. tendon buckle transducers to measure achilles tendon forces) •Activity analysis	•Group discussions •Expert groups •Autoconfrontation

architectural features, tools available, objects lifted, their weight, shape, etc.) which are part of the chain leading to the presence (or absence) of risk factors. In the prevention of WMSD, actions taken are often aimed at these work factors thought to be associated with the risk factors (e.g. re-designing a tool so that the force exerted is lessened).

Where to start? If a company has an active risk factor surveillance program, some basic data on risk factors is already available. If not, it is best to start with the less costly methods (e.g. a quick checklist) to first examine if there is the presence of risk factors and what seems to be the problem. Then, depending on what was found, one can move on to more precise methods (and more expensive) to measure risk factors and work determinants, as required. Van der Beek and Frings-Dresen (1998) have compared the usefulness of the types of methods for measuring physical risk factors and some other work factors. Irrespective of the method(s) chosen, once having measured and analyzed the work factors (risk factors or otherwise), there will be a need for some form of discussion and consultation with appropriate individuals to interpret the results and decide on the best approach possible to deal with the findings.

For most psychosocial factors there are few direct measuring methods. Although with variables such as workload there is the possibility to measure objectively, for many psychosocial factors the measurements are of the subjective type.

It should be noted that it is usually very difficult to collect in depth information on all risk factors at once (all physical factors or all psychosocial factors or both); choices usually need to be made. These choices will inevitably direct the possibilities for intervention and prevention; one can only deal with issues one sees. Further, even if it were possible to collect at once information on all factors, it could be difficult to later establish which are likely to be the more important factors, that is to establish those factors on which to intervene in an attempt to reduce exposure. Besides all that is unknown about safe values and interactions between factors, there is also the problem that not all factors were measured with the same degree of accuracy (the same type of measures). If a factor seems more important it could be that its measure was more precise. Still, even with our current limits, much can be done to detect and exert control over risk factors at work and thus prevent WMSD.

7. POSSIBLE ERGONOMIC APPROACH TOWARDS PREVENTION OF WMSD

In this context prevention of the WMSD is presented mainly from the point of view of primary prevention (prevention at the source). The authors are convinced that ergonomics is an important, maybe the most important, mean of realizing prevention of the WMSD in the working environment.

7.1. Prevention Strategies

From an operational point of view it is necessary to analyze different levels of preventive actions. Depending on the social context where the prevention is practiced, different actors are engaged and different means may be employed – in more generic terms different strategies may be adopted.

The *most general level is societal* which operates through legislation, norms and indirectly through social insurance schemes. Most industrialized countries have legislated at least some aspects (recognition, surveillance, compensation, etc.) of WMSD. In the international scene, the International Labour Organisation (ILO) is the authoritative agent in that respect. Recently important initiatives for creating new legislation and to establish new norms concerning the WMSD have been undertaken for example in Europe, Canada and in the United States. Viikari-Juntura (1997) provides an overview of some recent standards and specific research based guidelines.

One step down, but still at a wide level, various programs and campaigns, often combining awareness raising and some forms of activation, have been run by diverse organizations (e.g. the state or province's compensation body, unions, etc.) and by various public funded voluntary/lobby groups. These programs generally aim to motivate the target population into action to deal with WMSD. Programs may also contain an educational aspect, for example to help the targeted individuals analyze their own work places with respect to WMSD risk. The results of campaigns and programs are rarely followed up, thus we do not know much of their efficacy.

From an ergonomics point of view, *strategies employed in industry* are of special importance. Ergonomists may have on this level direct influence on the content and on the accomplishment of such strategies.

The industrial strategies are as varied as the industries

themselves. The comparative analysis of the efficiency of individual strategies is usually not possible due to lack of published data. Examples of possible strategies follow.

Strategies using normal existing OHS channels. In cases where occupational safety and health (OSH) is firmly rooted in to the practices and customs of an organization, it is only natural to channel prevention through it. A definite advantage of such situation is the already acquired high competence of OHS officers (and ergonomists) on the subject matter. However, in such a strategy, depending on how existing OHS is structured (e.g. separated from the operational side of the company and falling under the resource/personnel management aspect), prevention may be understood as a specialist isolated function. This could reduce the interest and involvement of the production folks, that are, really, in a central position to realize the prevention.

Strategies using existing safety audits. Some North-American industries have adopted systematic safety audit practices. Incorporating the necessary elements for the control of WMSD in the safety audit has shown to be useful in prevention of the WMSD at least in some cases known to authors. Lack of data prevents a more detailed analysis.

Strategies using existing company standards. Major industries may have their own standards on widely varied issues. This makes it at least theoretically possible to incorporate preventive elements for WMSD to such a standard base. The internal standards (which may in some cases be huge bases of knowledge) may concern principles of technical design, including work places, tools, production, and working methods, etc. They may also be guidelines for acquisition of production means, tools, furniture and others. Such issues, if they may be influenced, are most useful in prevention of the WMSD.

Strategies using the creation of special programs. Prevention strategy may take the form of special programs. For example, some companies support physical activity programs which may concern WMSD either as part of general schemes for improving physical fitness or sometimes specifically oriented towards WMSD (most often these programs have been geared towards the prevention of low back pain). The usefulness of such programs remains to be demonstrated.

Finally, irrespective of the chosen strategy, *it should be noted that commitment is the first building block*. Raising awareness about the WMSD is proposed as means to uncover the supposedly hidden problems and permit the organization to take adequate decisions to prevent it. "Let the sleeping dogs lie" has been often preferred to this approach and it seems that the organizational culture must be strongly committed to prevent WMSD to be able to take the risk of waking up the dogs.

7.2. Preventive Actions

Whatever strategy is adopted, it has to converge towards a certain number of actions. Theoretically, the number of preventive actions is large but the feasibility and economic factors limit the number to a fairly small collection of identifiable issues.

The preventive action may be aimed at a collectivity of workers, a company as a whole, a professional group or other, or it may have the individual considered at risk (or already experiencing symptoms) as a target. The preventive action may be aimed at generic risk factors themselves or it may be directed towards specific work factors which are thought to be in the chain of events leading to WMSD.

The following is not exhaustive list but shows the types of actions reported in the literature:

- Preparing or acquiring informative material on WMSD.
- Exploring and employing strategies to convince management of the usefulness of preventing WMSD.
- Mounting awareness raising programs aimed at individual workers. Such a program should be completed with means that permit individuals to react and act adequately to improve his/her situation.
- Designing and implementing tools with technical designers, supervisors and work-study engineers.
- Finding ways to incorporate the prevention into the company culture.
- Working with production engineers, architects, etc. to design work environments with reduced WMSD risk.

7.3. Ergonomics as Means of Preventing WMSD

Ergonomics is by definition an important tool and mean of preventing WMSD in an industrial and other occupational context. Ergonomists are usually capable of analyzing risk factors in a specific context and of understanding how to correct the situation. The following list describes elements often found in an ergonomic intervention.

- Identification of risk factors.
- Analysis of relevant aspects of the production process and organization which may result in these risk factors existing. This phase is necessary to be able to proceed to the next step, identification of risk situations.
- Identification of risky situations

An ergonomist's competence should be – and usually is – sufficient to be able to identify in a given operational situation the generic risk factors of WMSD. An ergonomist must be able to find out when, where and how these risk factors occur, how they are linked to other relevant factors and to the production process in general. One of the problems is that WSMD research has described poorly the interrelations between risk factors themselves and with other relevant factors. Further knowing an epidemiological correlation between factors is not sufficient.

- Identification of potential solutions and preventive approaches.

In a given situation, many different solutions and preventive approaches may be present. Each of them may have specific merits and seem to promise results in the preventive sense. A closer look may show, however, that certain approaches may not be realizable for some theoretical, practical, economic or other reason. This leads to an iterative priority setting process, involving not just the ergonomist but employees and management at various levels of the company, where proposed approaches are scrutinized over and over again until a balanced approach is identified which seems to be most feasible.

- Choosing feasible preventive approaches.

After identifying the most appropriate approach a *formal* decision should be taken which should enable all concerned parties to agree upon and to commit themselves to work towards a common goal.

- Implementing of prevention.
- Securing the follow up and evaluating the impact of the chosen approach(es).

8. CONCLUSIONS

WMSD are a heterogeneous group of conditions. Many aspects of WMSD are not clearly defined and understood. Still, statistics report that these are an important problem in industry and cannot be ignored. Literature reviews also show that there are, indeed, links between work and these musculoskeletal disorders, and there is information available on risk factors at work. Although the knowledge base is incomplete, ergonomics has an important role to play towards prevention of WMSD.

REFERENCES

ARMSTRONG, T.J., BUCKLE, FINE, L.J., HAGBERG, M., JONSSON, B., KILBOM, Å., KUORINKA, I.A.A., SILVERSTIEN, B.A., SJØGAARD, G. and VIIKARI-JUNTURE, E.R.A., 1993, A conceptual model for work related neck and upper limb musculoskeletal disorders. *Scandinavian Journal of Work and Environmental Health*, 19, 73–84.

BATTEVI, N., MENONI, O. and VIMERCATI,C., 1998, The occurrence of musculoskeletal alterations in worker populations not exposed to repetitive tasks of the upper limbs. *Ergonomics*, 41, 1340–6.

BLS, January 1999 [http://stats.bls.gov/os/ostb0631.pdf]

BROGMUS, G. E., SOROCK, G. S. and WEBSTER, B. S., 1996, Recent trends in work-related cumulative trauma disorders of the upper extremities in the United States: an evaluation of possible reasons. *Journal of Occupational and Environmental Medicine* 38, 401–11.

CHIANG, H.-C., KO, Y.-C., CHEN, S.-S., YU, H.-S., WU, T.-N. and CHANG, P.-Y., 1993, prevalence of shoulder and upper-limb disorders among workers in the fish-processing industry. *Scandinavian Journal of Environmental Health*, 19, 126–31.

CHOI, B.C.K., 1993, Definitions, sources, magnitude, effect modifiers and strategies of reduction of the healthy worker effect. *Journal of Occupational Medicine*, 35, 890–2.

DORLAND'S ILLUSTRATED MEDICAL DICTIONARY (25th edition) (Philadelphia: W.B. Saunders, 1974).

FOX, A. J. and COLLIER, P. F., 1976, Low mortality rates in industrial cohort studies due to selection for work and survival in the industry. *British Journal of Preventative and Social Medicine*, 30, 225–30.

GUN, R.T., 1990, The incidence and distribution of RSI in South Australia 1980–81 to 1986–87. *Medical Journal of Australia*, 153, 376–80.

HALES, T. R. and BERNARD, B. P., 1996, Epidemiology of work-related musculoskeletal disorders. *Orthopedic Clinics of North America*, 27, 679–709.

JOHANSSON, H. and SOJKA,P., 1991, Pathophysiological mechanisms involved in genesis and spread of muscular tension in occupational muscle pain and chronic musculoskeletal pain syndromes: a hypothesis. *Medical Hypotheses*, 35, 196–203.

KUORINKA, I., FORCIER, L. (eds) *et al.*, 1995, *Work-related Musculoskeletal Disorders (WMSDs): A Reference Book for Prevention* (London: Taylor & Francis).

LAST, J. M., 1995, *A Dictionary of Epidemiology* (New York: Oxford University Press).

NIOSH, 1997, *Musculoskeletal Disorders and Workplace Factors: A Critical Review of Epidemiologic Evidence for Work-related Musculoskeletal Disorders of the Neck, Upper Extremity, and Low Back*. Publication No. 97-141 (Cincinnati: DHHS NIOSH).

VAN DER BEEK, A.J. and FRINGS-DRESEN, M.H.W., 1998, Assessment of mechanical exposure in ergonomic epidemiology. *Occupational and Environmental Medicine*, 55, 291–9.

VIIKARI-JUNTURE, E., 1995, The role of physical stressors in the development of hand/wrist, and elbow disorders. In Gordon, Blair, Armstrong *et al.* (eds) *Pathophysiology and Clinical Factors in Upper Extremity Repetitive Motion Syndrome* (Bethesda: American Academy of Orthopedic Surgeons), 7–30.

VIIKARI-JUNTURA, E.R.A., 1997, The scientific basis for making guidelines and standards to prevent work-related musculoskeletal disorders. *Ergonomics*, 40, 1097–1117.

VIIKARI-JUNTURA, E., 1998, Risk factors for upper limb disorders; implications for prevention and treatment. *Clinical Orthopaedics and Related Research*, 351, 39–43.

YASSI, A., SPROUT, J. and TATE, R., 1996, Upper limb repetitive strain injuries in Manitoba. *American journal of Industrial Medicine*, 30, 461–72.

Work-related Musculoskeletal Disorders in Dental Care Workers

D. C. Murphy

Occupational Health Specialist, New York University College of Dentistry, USA

1. OBJECTIVES

After reading this article, the reader will be able to:

- describe historical developments in both the practice of dentistry and federal regulations which impact on the current situation
- list three or more risk factors associated with the development of WMSDs in dentistry
- identify three strategies for preventing/reducing WMSDs in DCWs

2. INTRODUCTION

This article offers a comprehensive overview of the current state of ergonomics, specifically, work-related musculoskeletal disorders (WMSDs), as they relate to dentistry and dental care workers (DCWs). Apart from some historical background information and statistics which assist the reader in appreciating the scope of the problem, the focus is on identified risk factors and effective preventive strategies.

An abbreviated summary of the most relevant literature offers multiple examples of research findings which demonstrate a positive association between dental tasks/procedures and adverse effects on various parts of the DCWs' musculoskeletal system — namely, cervical spine/neck area, shoulder, elbow, wrist, hand, fingers, and lower back. Methodologies employed by the researchers include, but are not limited to, worker surveys, videotapes, electromyography, visual analog scale, and records review. Preventive strategies are also enumerated from two different viewpoints: (a) recommendations from independent researchers and (b) interventions based upon a "hierarchy of controls" developed by OSHA (Occupational Safety & Health Administration); however, there are many noticeable similarities. Application of the various preventive interventions to dentistry are given with examples. There are statements made that refer to the forthcoming legislation and its supporters and opponents, and a plea for the integration of official policies with people, practice, and product development.

3. HISTORICAL OVERVIEW TO THE PRESENT

Ergonomics is defined in many ways according to the discipline providing the definition; however, a simple straightforward description is "the science of fitting the job to the worker and when there is a 'mismatch' between the physical requirements of the job and the physical capacity of the worker, work-related musculoskeletal disorders — WMDSs — can result" (OSHA 1999). Historically, in dentistry, fitting the job to the dental care worker, was not a common practice, and in fact some of the same attitudes continue to this day.

3.1. Working Position

Historically, the dental chair was designed to treat the patient in the upright position — hence the need for the dentist to stand, resulting in the following problems: (1) muscle fatigue (2) need to lean awkwardly over the chair in a twisted position, and (3) need for one foot to always be seeking the foot lever to operate the chair and equipment, thereby putting most of the weight on one leg rather than evenly distributing it. As years in dental practice increased (i.e. "duration of exposure" increased), these fixed contorted positions resulted in a high incidence of circulatory and musculoskeletal conditions among dentists: scoliosis, kyphosis, lordosis, varicosities, and other leg and foot problems were common.

In an effort to reduce the physical fatigue of standing and the resultant adverse effects on health, dental stools began to be used in the 1950s and thus the term "*sit down*" dentistry was coined. However, the initial stools were single pedestaled, and often tilted and rotated, presenting a safety problem. While the design of the stool has been greatly improved since that time, current researchers have identified potential ergonomic conditions related to sitting. In a sample of 3316 dental care providers, lower back pain was associated with sitting for four or more hours per day.

3.2. Tools and Equipment

Traditionally the old "drills" used in dentistry were run by stepping on a pedal that rotated a belt. This slower, labor-intensive method has been replaced by compressed air-driven high-speed handpieces. While the increased speed of the new instruments improves the quality of care and reduces the time required to complete the dental procedure, it has also resulted in vibration-related hand and wrist conditions.

3.3. Personnel

A third significant development in dentistry involves the move from a single practitioner, i.e. the dentist, to "four-handed dentistry". This switch now incorporates a professionally trained "helper", usually a dental hygienist, into the daily practice. Like most changes there have been benefits such as a reduction in the length of the patient visit due to increased efficiency — and there have been liabilities — such as the development of WMSDs in a whole new cadre of workers.

3.4. Current Situation

As early as the 1950s, researchers recognized the fact that the practice of dentistry is associated with several musculoskeletal disorders that may be attributable to risk factors found in the typical dental work environment. Since that time, we have seen the passage of the Williams-Steiger Act (Occupational Safety and Health Act of 1970), which led to the creation of the Occupational Safety and Health Administration (OSHA) as well as the National Institute for Occupational Safety and Health (NIOSH), whose respective missions, while somewhat different, converge in a shared concern and responsibility for the health of American workers.

During the past decade, both federal agencies (NIOSH and OSHA) have recognized a disturbing increase in the incidence of musculoskeletal disorders among various worker populations, including (but not limited to) office workers, telephone operators, post office employees and supermarket baggers. Despite national

and international studies documenting the existence of this type of condition in dental care workers, limited attention has been paid to this worker population by federal agencies. Additionally, private organizations including the American Industrial Hygiene Association (AIHA) and the American National Standards Institute (ANSI) have supported the position associating workplace risk factors in the dental care environment with the development of musculoskeletal injuries. In fact, a comprehensive report recently issued by the National Academy of Science (NAS) concluded that "ergonomic hazards at work cause musculoskeletal disorders, and intervention measures to prevent these disorders currently exist" (Work-related MSDs: a Review of the Evidence, National Academy of Science, 1998).

4. SCOPE OF THE PROBLEM

Since the early 1980s, NIOSH has been waving a red flag alerting the nation to the ever-increasing number of cases of WMSDs reported in a variety of workplaces.

Although death is not a likely outcome due to the presence of the more common deficiencies that would be addressed by an OSHA Ergonomics Standard, disability is common. In 1997, there were 276 600 cumulative trauma disorders (CTDs) in the United States, representing 64 percent of all workplace illnesses, according to the Bureau of Labor Statistics (BLS) in their BLS Annual Survey of Workplace Injuries and Illnesses for 1997. There were more than 1.25 million work-related back injuries, with an average cost of $24 000 for each claim filed. According to the National Council on Compensation Insurance Inc (NCCI) in its Annual Statistical Bulletin 1998, it is estimated that these disorders cost American industry more than $20 billion annually and the cost is escalating. Unfortunately BLS statistics on dental care workers are questionable due to limited or underreporting of conditions as well as a "clumping together" of all illnesses and injuries among this worker population in the Bureau's current system.

In March 1995, OSHA issued a Draft Proposed Ergonomic Protection Standard. It was a comprehensive, prescriptive document several hundred pages in length including non-mandatory appendices. There were two criteria for assessing ergonomic risk: signal risk factors, and the presence of work-related musculoskeletal disorders.

Comment was solicited and a firestorm of response was received. In June 1995, Congress prohibited the use of OSHA funds to issue a proposed or final Ergonomics Standard or guidelines. With this prohibition no longer in effect, the agency director is determined to have an approved standard in place by the year 2000.

Although the American Dental Association (ADA) directed interest toward this emerging health issue as early as 1963, when it published a lengthy article in its national journal entitled "Body Mechanics Applied to the Practice of Dentistry" (ADA 1963), its current stance is guarded at best. Some of the concerns voiced by the ADA and its members include the financial costs to the dentist (for example, to purchase new instruments or retrofit equipment) as well as lack of conviction regarding "a scientific documented linkage between cause and effect".

5. REVIEW OF THE LITERATURE

The inclusion of dental care workers in the forthcoming OSHA

Ergonomic Standard as a worker population at risk of developing MSDS is still unresolved, but a review of the relevant literature helps answer this question. To summarize: (1) much of the literature discussing musculoskeletal disorders and dental care workers is found in non-dental journals — primarily those whose audience consists of physical therapists, industrial hygienists, occupational therapists, orthopedic surgeons, and ergonomists; (2) the earliest international article found was the publication in 1955 of a research study of dentists done in the United Kingdom, while the first American article was a 1963 entry in the Journal of the American Dental Association; (3) there are multiple studies of various design — using different research methodologies — that demonstrate an association between dental tasks and musculoskeletal disorders; (4) the prevalence of WMSDs is an international occurrence, not just national; and (5) more than one anatomical site is commonly affected.

- *Data collection* has been on-going by independent researchers as well as various organizations including, but not limited to, the Occupational Safety and Health Administration, the National Institute for Occupational Safety and Health, the American Industrial Hygiene Association, the American National Standards Institute, the National Academy of Science and the American Dental Association.
- *Study design* varied, and included case-control, cohort, cross-sectional, and retrospective approaches while the populations studied were dental hygienists, dentists, dental assistants, and control groups such as pharmacists who were matched for significant variables with the study (DCW) population.
- *Tools and methods* included worker questionnaires, surveys, videotaping, motion analysis, biological monitoring (heart rate, nerve conduction, x-rays) and review of pertinent records, i.e. worker compensation data, medical charts, etc.

6. RISK FACTORS

The question is: How do we assess the risk of developing a work-related musculoskeletal disorder in a particular worker population (i.e. dental care workers)?

The suggested answer is to survey the workplace of concern for the presence of specified risk factors. Independent researchers have identified six risk factors for the development of WMSDs and they are: repetition, awkward postures, force, vibration, direct pressure, and insufficient rest. In addition, they state that a

Figure 1. Deviated upper-body positions during a dental task.

Table 1. Risk factors: comparison

General risk factor	Signal risk factor
Repetition	Performance of same motion/motion pattern every few seconds for 2–4 hours at a time
Awkward postures	fixed or awkward postures for more than 2–4 hours
Vibration	use of vibrating or impact tools/equipment more than a total of 3–4 hours per work shift
Force	using forceful hand exertions for more than a total of 2–4 hours at a time

combination — i.e. the presence of more than one risk factor — markedly increases the potential for the development of CTDs.

An observation of dental care providers in their workplace clearly demonstrates the presence of more than one of these risk factors during the performance of most dental procedures (see figure 1).

The Occupational Safety and Health Administration has cited two criteria for assessing risks, including the presence of "signal risk factors" and the presence of work-related musculoskeletal disorders. The "signal risk factors"— similar to the previously stated list of six risk factors — are more specific (see table 1).

7. ANATOMICAL SITES INVOLVED

While work-related musculoskeletal disorders have been known to affect most parts of the body, disorders of the back and upper extremities are the most frequent in the general population. Conditions that have been associated with dentistry specifically include disorders of virtually every sector of the musculoskeletal system, including (but not limited to) upper and lower back, wrist and hand, arms, lower extremities, neck, shoulders (including osteoarthritis), and shoulders and neck.

7.1. Neck and Shoulder

Various study methodologies have been used including videotaping, questionnaires, time studies, Electromagnetic Graphs (EMG), and direct observations made on DCW groups since the 1960s. There was sufficient evidence to identify prolonged static neck flexion, repetitive neck movements, shoulder abduction or flexion, lack of upper extremity support and insufficient work breaks, as risk factors for developing WMSDs in the neck and shoulder regions (see figure 2).

7.2. Wrist and Hand

Various study methodologies have been used including vibrometry. Reported and observed signs and symptoms such as

Figure 3 Awkward posture of elbow and wrist with fingers in pincer grip.

numbness, pain, and degenerative changes in distal interphalangeal joints were significantly greater in the study group than those in selected control groups. Demonstrated risk factors which result in wrist and hand WMSDs include highly repetitive work, forceful work, wrist/hand vibrations, extended work periods, repetitive wrist flexion and extension, and high precision grip (see figure 3).

7.3. Low Back Disorders (LBD)

Various study methodologies were used including electromyography, X-rays, and self-reported lumber pain and neuralgia among practicing dentists. Evidence was found supporting an association between LBDs and the following risk factors: heavy physical work, forceful movement, work-related awkward postures, whole body vibration, prolonged sitting and static work in the sitting position requiring spinal flexion and rotation (see figure 4).

8. MANAGEMENT AND CONTROL STRATEGIES

The development of an effective management and control program requires the inclusion of several components, including:

   (a) prevention — the most important component
   (b) diagnosis — early and accurate
   (c) treatment — to relieve signs and symptoms and to minimize disabilities
   (d) possibly a change of careers if all else fails

While the focus of this section is on prevention, a few words about early diagnosis and treatment are warranted. It is imperative that any DCW experiencing signs and symptoms of WMSDs be evaluated as early as possible and by a qualified clinician. There are many board-certified occupational physicians who are oriented to an inclusion of workplace conditions as a potential

Figure 2 Awkward posture involving neck and shoulders.

Figure 4 Twisting of lower back during dental procedure.

cause and/or contributor to any illness or injury reported by the patient. This physician, after assessing the role of the workplace on the patient's condition, will usually recommend a conservative treatment approach which may include but is not limited to: exercise, physical therapy, a work-hardening or work-conditioning program, medication, modification of work process, work layout, work schedule or equipment/tools, or alternative therapies such as herbs, massage or acupuncture. Surgical intervention is usually only considered after the aforementioned conservative modalities are tried unsuccessfully. As a last resort, the worker may be advised to consider a change of careers.

9. PREVENTION: AN OVERALL APPROACH AND SPECIFIC STRATEGIES

The area of "prevention" also comprises multiple components. First, one must make an accurate assessment of one's workplace and this can be done by the use of various tools such as employee survey, videotaping of actual tasks and postures, or reviewing relevant records — for example, workers' compensation claims or medical records. Using these investigative techniques will then allow the identification of high(er)-risk tasks and procedures as well as higher-risk personnel. (There is some documentation in the literature associating non-work-related personal characteristics such as arthritis, diabetes, pregnancy, and some activities (knitting) with the development of musculoskeletal disorders.) These findings can then be utilized to identify and apply effective intervention measures. As with all diseases, the emphasis needs to be on prevention or limitation of the adverse event (i.e. work-related musculoskeletal disorders); similarly, as with other disease (illnesses and injuries), there are proven strategies for effectively preventing or reducing the incidence of WMSDs. There are two sources of information for preventing WMSDs in dental care workers: (a) the reported results of various independent researchers, and (b) recommendations made in the most recent draft of the Ergonomic Standard proposed by the Occupational Safety and Health Administration. While varying in subject categories, their recommendations are similar.

9.1. Prevention Based on Independent Research Findings

As mentioned previously, there have been a relatively large number of domestic and international studies conducted on both the predictors of risk and prevalence of musculoskeletal disorders in the dental care worker population, and as a result, researchers have been able to identify, implement, and recommend effective interventions. Specific preventive strategies as categorized by anatomical site are outlined below.

9.1.1 Neck and Shoulders
Support for upper extremities
Elimination of static neck flexion
Elimination of shoulder abduction or flexion
Adequate work breaks
Avoiding work with elbows higher than shoulder
Exercise to strengthen musculature and improve flexibility

9.1.2 Wrist/Hand

Avoid forceful exertions
Reduce repetition of task
Reduce length of time doing a repetitive task
Reduce/eliminate repetitive wrist flexion and extension
Exercises to strengthen musculature and improve flexibility

9.1.3 Low Back
Alternate between standing and sitting
Avoid prolonged sitting
Use back support
Avoid awkward postures
Avoid spine flexion and rotation/twisting while sitting or standing
Exercises to strengthen musculature and improve flexibility

9.1.4 Elbow
Avoid repetitive rotation of wrist
Avoid repetition
Avoid reaching beyond 15 degrees
Exercises to strengthen musculature and improve flexibility

The following case example, which applies these preventive strategies, demonstrates the feasibility of using a safe approach in dentistry.

9.1.5 Case example
There is currently an American dentist (Dr Daryl Beach) who has been living and practicing dentistry in Japan for the past 35 years. In the early 1970s, he developed a unique approach to practice for the dental care worker. The original intention of this system was to intuitively — or "proprioceptively" — derive the most natural, non-stressful body position for performing dental procedures (see table 2). This would provide the operator with the optimal hand and finger control needed for fine motor tasks, which, in turn, would result in the operator's acquisition of superior technical skill and accuracy in the performance of dental practice.

There have been many articles published pertaining to this unique practice system (the Beach System). However, there is only one research article (Kaneko 1994) which focuses on the ergonomic aspects of the system rather than the measurement of skill, accuracy, and productivity. While there are limitations to Kaneko's study design and methodology, his findings — based on interview and videotaping of Japanese dentists — were as follows: (a) a head shift of >15 cm from the neutral centralized position results in high stress in that area; in dentists using conventional practices, for 60% of their workday their head shift was > 15 cm compared to the Beach group which employed a head shift of > 15 cm in only 1.3 % of their work day; (b) reported

Table 2. Sensory receptors.

Exteroceptor	receptor for external stimuli, for example, heat
Interoceptor	receptor: internal organs, for example, thirst
Proprioceptor	internal receptor that: allows one to sense body positions, relationships, movements and forces; is located in the joints and muscles; in the most important type of receptor for fine and precise finger control; 'pd'

Table 3. Reported musculoskeletal pain (Kaneko 1994).

Site of pain	Group A (conventional)	Group B (beach)
Back	51%	8%
Shoulder	44%	0%
Neck	36%	0%
Arms	40%	0%

Group A = 121 participants
Group B = 12 participants

musculoskeletal pain in several body sites was significantly greater in the group using conventional practice methods (see table 3)

Beach introduced and implemented a system of practice that incidentally resulted in eliminating many of the researcher-identified risk factors for WMSDs, including flexed or bent cervical neck positions, abducted elbows and shoulders, twisting and turning of the lower back, and wrist and hand deviation from the neutral plane (see figures 5 and 6).

9.2. Preventive Strategies Based on OSHA's Hierarchy of Controls:

The Occupational Safety and Health Administration has devised its own systematic approach to work-related health and safety issues, namely its 'Hierarchy of Controls' which in order of importance are as follows: engineering, administrative, work practice and personal protective equipment. Prior to deciding which type of control is needed, it is imperative that an analysis of the workplace is conducted (see table 4) to determine both the risks and, eventually, appropriate control measures.

As mentioned earlier in this article, determination of risk is made by the presence of signs and symptoms of WMSDs (see table 5) and the presence of "signal risk factors" (see table 1). If it is determined that both signs and symptoms and signal risk factors exist, application of one of the control strategies is warranted.

10. ENGINEERING CONTROLS:

Engineering controls are physical changes to jobs that control exposure to WMSD hazards; they act on the source of the hazard. Examples applied to dentistry include:

- *Processes*: use safe work practices
- *Equipment*: for example, the stool or seat for the dental
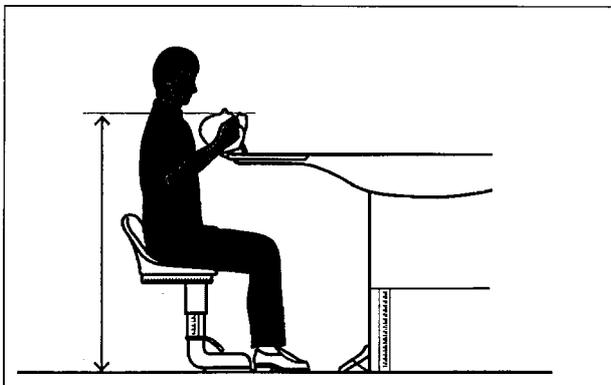
Figure 5a The balanced operator with patient in horizontal position.
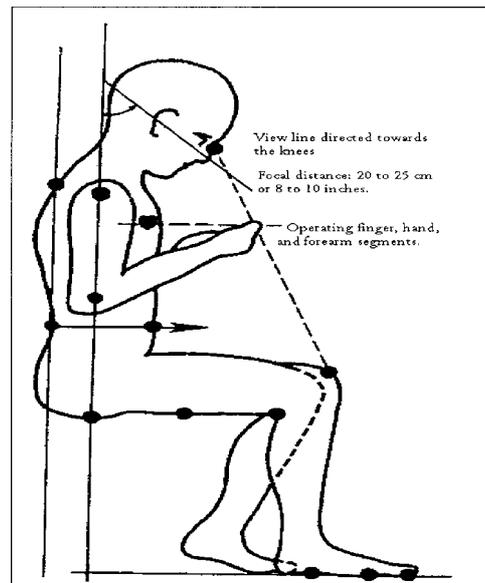
Figure 5b The balanced operator.

operator should be adjustable in height, have a lumbar support, and possibly a frontal support
- *Tools*: for example, dental instruments should be located within 15 inches reach, should be of high quality and sharp, should be lightweight to reduce fatigue, have increased surface holding area to minimize the force needed and should be angled so that the wrist may be maintained in a neutral position, i.e. the instrument bends rather than the wrist.

11. ADMINISTRATIVE CONTROLS

Administrative controls are procedures and methods instituted by the employer that significantly reduce daily exposure to WMSD hazards by altering the way in which work is performed. Examples in dentistry include rotation of employees for tasks such as scaling, adjusting the work pace (slowing it down), alternation of tasks done by individual workers to reduce repetitions, provision of adequate rest breaks in between procedures, reduction in patient load, and shortened recall periods when possible to reduce the amount of work needed on patients per visit.

12. WORK PRACTICE CONTROLS

Work practice controls are controls that decrease the likelihood of exposure to WMSD hazards through alteration of the manner in which a job/work activity is performed; it acts on the source of

Figure 6 Correct alignment of operator's shoulders, elbows and wrists.

Table 4. Job hazard analysis.

Job factors	Components
Physical demands	Force, repetition, work postures, duration
Workstation layout and space	Work reaches, work heights, seating and floor surfaces
Equipment used/objects handled	Size and shape, weight, vibration, handles and grasps
Enviromental conditions	Cold, heat, glare (as related to awkward postures)
Work organization	Work rate, work recovery cycles, task variability

the hazard and examples include: safe work practices such as the use of automation rather than hand-scaling, use of proper body positioning, i.e. avoid bending neck and/or twisting lower back, and maintenance of neutral positions — wrist, neck, torso, etc. — whenever possible. Conditioning for newly hired or newly returned employees is also recommended and may involve simple brief exercises that can be done at the work site or extensive work-hardening/work-conditioning programs of longer duration conducted under the guidance of a trained professional such as a physical therapist, ergonomist or occupational physician.

Randolph conducted a research study on a group of workers in a meat processing plant whose risk factors were comparable to those of dental care workers — namely, low force with high repetitions. She taught the workers a series of exercises for flexors and extensors of the upper extremities to be done daily over a period of six weeks. At the end of the study period, using employee surveys and a visual analog scale (VAS), Randolph reported 78% of the participants said their upper extremities felt better, 69% said the exercises made their job easier, 69% reported less fatigue at the end of the work shift and 65% said they had increased strength (Randoph 1999).

13. PERSONAL PROTECTIVE EQUIPMENT (PPE)

Personal protective equipment or PPE are interim control devices worn/used while working to protect the employee from exposure to WMSD hazards. An example of PPE used in dentistry are gloves. From an ergonomic perspective, recommendations pertaining to gloves include the use of those designed specifically for the left and right hands rather than ambidextrous gloves, and selection of the correct size for the individual since gloves that are too small can increase pressure to the hand and wrist, while gloves that are too large increase the grip force needed.

14. SUMMARY

The documented cases of WMSDs in the USA by the Bureau of Labor Statistics, coupled with strong evidence in the research

Table 5. Signs and symptoms of WMSDs.

Signs of WMSDs	objective, physical findings examples are decreased range of motion, decreased grip strength, loss of function, swelling, redness
Symptoms of WMSDs	subjective, physical indications of a developing WMSD vary in severity and can appear gradually examples are numbness, tingling, burning, aching, pain, stiffness

literature, support the need for official intervention — namely, in the form of the proposed Ergonomic Standard from OSHA. There has been a great deal of opposition to the standard and some of the reasons stated are: economic costs to the employer (purchase of new equipment, etc.), the complexity of identifying hazards, and the sometimes debatable linkage between cause and effect (for example, hobbies such as knitting have frequently been cited as non-workplace activities that contribute to some wrist and hand disorders). Despite the resistance, it seems quite certain that a standard will be in place by the year 2000. This author would like to recommend an integration of various related segments in order to achieve optimum ergonomic working conditions for all dental care workers; these segments — referred to as "4 Ps"— are Policy (the OSHA Standard as well as formal supportive policies from bodies such as the American Dental Association), People (DCWs themselves need to become involved and take responsibility for their musculoskeletal well-being through awareness, education, and continued research), Products (manufacturers must continue to improve upon instruments and equipment used in dentistry), and Practice (utilize rest breaks, task rotation, and exercise programs).

15. RELEVANT TERMINOLOGY AND ABBREVIATIONS

WMSDs:work-related musculoskeletal disorders
CTDs:cumulative trauma disorders
RMIs:repetitive motion injuries
OSHA:Occupational Safety and Health Administration
NIOSH:National Institute for Occupational Safety and Health
ADA:American Dental Association
DCWs:dental care workers (includes dentists, dental hygienists and dental assistants)

REFERENCES

BRAMSOM, J.B. *et al*,, 1998, Evaluating dental office ergonomic risk factors and hazards. *JADA*, 129.

BEACH, D., 1990, A logic for specifying human centered settings and technology. Paper presented at the 3rd International Symposium on Human Factors in Organizational Design and Management, Kyoto, Japan.

GREEN, E.J. and BROWN, M.E., 1963, Body mechanics applied to the practice of dentistry, *JADA*, 67.

KANEKO, H., 1994, The effect of dental equipment on human behavior, stress and pain. Paper presented at the FDI World Congress, October.

MURPHY, D.C., 1998, *Ergonomics and the Dental Care Worker* (Washington, DC: APHA).

National Academy of Sciences, 1998, *Work-related Musculoskeletal Disorders: a review of the evidence*, October, 1.

RANDOLPH, J., 1999, A comparison of flexion and extension exercises to reduce localized discomfort and increase strength in workers at risk for developing cumulative trauma disease. Paper presented at the American Occupational Heath Conference, New Orleans, April.

US Department of Labor, 1970, *Occupational Safety and Health Act*, Public Law No. 91-596.

US Department of Labor, 1999, *Occupational Safety and Health Administration*, Draft of the Proposed Ergonomics Program Standard, March, 4

Part 11

Social and Economic Impact of the System

Analysis of Worker's Compensation Data

F. T. Conway

State of Wisconsin Department of Workforce Development, Worker's
Compensation Division, Madison, WI 53703, USA

1. INTRODUCTION

Worker's compensation data are a rich source of information that
can be used to identify and subsequently to control occupational
injuries. Worker's compensation law exists at the state, federal
and national levels of jurisdiction. The purpose of these laws is
to provide some financial support for injuries and diseases that
arise out of the and in the course of employment. This legal
requirement involves various parties including employers, the
employee, insurance companies, medical providers and the state,
federal or national entity that serves to distribute such compen-
sation and adjudicate the disagreements resulting from such
claims. This chapter will identify critical information that is
currently collected through current claim forms and associated
documentation related to claims, using a form filed for compen-
sable injuries in the State of Wisconsin as a template. It will pro-
pose some opportunities for utilizing the claim information to
study and control occupational injuries. It will identify some of
the shortcomings of such information and discuss the implications
of such research to control occupational injuries.

2. CLAIM INFORMATION RELEVANT TO OCCUPATIONAL INJURIES

The information collected on a compensable claim from the
Employer's First Report of Injury or Disease (Wisconsin WKC-
12 [Revised 2/98]) includes salient aspects of the occupational
injury. The injury description field in the form is interpreted
through a coding structure into categories of the nature of the
injury, part(s) of body affected, type of accident and source of
the injury. The nature of the injury categorizes the injury
description noted on the form into groupings describing the injury
in medical terms like amputation, contusion or strain. The part
of body field categorizes the area(s) impacted by the injury like
the head, finger or wrist. The type of the accident is a categoriza-
tion of how the accident occurred, such as caught between, struck
by or fall. The source of the injury is a categorization of what
caused the injury like equipment, hand tools or co-worker.
Temporal aspects of the injury are collected on the form in the
date of injury, time of day and the time the injury occurred in the
shift being worked fields. Employer and insurer information is
included on the form in the name, mailing address and Fein
number fields. Employee information is collected regarding the
name, age, gender and employee length of tenure with the cur-
rent employer. The days of healing period for the injury is calcu-
lated by subtracting the date of return to work following the
injury from the last day worked. This computation provides an
indication of the extent of the injury. Additionally, the perma-
nency of the injury is noted in the medical evaluation of the
injury as well as the disabling nature of the injury as total or
partial. As a claim is paid, information regarding the indemnity
cost of that claim is associated with the claim and tracked to
ensure that claims agreed to be compensable or deemed by liti-
gation to be compensable are paid appropriately. Litigated claims
collect information regarding the hearing scheduling and settle-
ment time periods associated with the legal administration of
each claim as well the legal outcome.

3. OPPORTUNITIES FOR THE ANALYSIS OF CLAIMS FOR OCCUPATIONAL INJURY CONTROL

3.1. Macro-analysis of Compensable Claims

One area of opportunity for the analysis of claims is the surveil-
lance of injury trends at a macro-level that collects aggregates of
compensable claims and looks at injuries of significant volume
across all types of business. This information is often examined
yearly and provides totals of the different classifications men-
tioned above. The analysis can identify in broad terms the types
of injuries occurring frequently and some of the associated costs
for the injuries (for example, Berkowitz and Burton Jr 1987, Leigh
et al. 1997, Feuerstein *et al.* 1998). The most dangerous indus-
tries can be identified. Specific factors that led to injuries can be
determined. Injury control can benefit from this analysis by draw-
ing attention to the types of injuries occurring most often as well
as the frequency of occupationally related diseases like asbes-
tosis and silicosis. Such a frequency can be compared across time
to get a perspective on the changing nature of that injury or ill-
ness. A comparison of occupational injuries across industries can
provide some context for additional research into reducing injuries
and illnesses in target industries as is often done by the US
Occupational Safety and Health Administration.

3.2. Micro-analysis of Compensable Claims

A different focus on occupational injury control can be addressed
by examining the lost-time experience of a specific employer or
a group of employers within the same industry. This can provide
a comparison of the nature, frequency and severity of injuries in
a particular industry that can be used to identify employers with
excellent safety records regarding compensable claims as well as
those that may benefit from safety education or training. One
such comparison by the State of Wisconsin was part of an educa-
tional program titled "Safety Works." This program targeted
employers with no lost-time injuries over a time frame and
recognized their success. Additionally, educational materials were
sent to those employers with an incident rate of claims many
times higher than similar industries. It is expected that this pro-
gram will lead to collaboration opportunities for employers with
labor to work to reduce hazards in the workplace. Another
analysis of compensable claims has been undertaken as part of a
study of the medical care provider practices of some State of
Wisconsin employers. Five industry groups have been compared
in terms of their incident claims rate, indemnity cost per claim,
percentage of claims litigated and median healing period for
the time away from work. These were compared with others
with compensable claims that are in the same industry. This

comparison will shed some light on how the medical care provider practices can result in an improved safety climate and a reduction in compensable injuries and associated costs.

At the individual worker level, worker compensation data provides an excellent opportunity to track an individual employee's health over time. An individual with a repetitive strain or cumulative trauma injury can be tracked over time to examine resulting health effects of the work assignments. This would evaluate the appropriateness of the specific work tasks assigned to an individual with such a disorder. Aggregating this analysis for employees with such injuries would provide an employer the opportunity to establish low-risk tasks for employees with injuries, and to maintain productivity and reduce lost work time. The analysis of compensable claims for a particular company could combine the lost-time injuries information for a period with specific safety practices and equipment in a work setting to determine the injury costs and highlight areas of improvement for injury reduction. This would provide the detail of the cost-saving benefits of job re-design strategies like rest breaks or safety precautions like machine guarding. The continual collection of compensable claims also provides the opportunity for the longitudinal analysis of injuries for a workforce. Compensable claims also provide information regarding the permanency of an injury as well as fatal injuries. Analysis of this type of information can provide context regarding the most hazardous industries and accidents for a particular business.

At the company financial level, analysis of worker compensation data provides critical information regarding lost time and indemnity payments to injured employees to help manage the business. The direct demonstration of a financial relationship between work injuries and financial expenditures can provide additional incentive for increased efforts for occupational safety and health at work. Additionally, the lost time for the employee is recorded on a worker compensation system and indicates a measure of the amount of lost work time for that employee, and in aggregate, for the company. Such lost time has to be compensated for with additional hiring and training of temporary staff to make up for the absent employee. The importance of return to work for injured employees has prompted studies examining the factors related to successful return to work. One such study examines which factors were related to return to work when the employee had a low back pain claim (Gallagher *et al.* 1995).

At the company safety and health level, there is additional value in the evaluation of compensated injuries as they are coded in terms of part of body, nature of injury, description of injury and cause of injury. All of these fields can provide insight into the details regarding how the nature of the work or the environment in which the work is performed may cause injuries and illnesses. This detail can point out areas of work that may be hazardous or may be associated with a disproportionate number of injuries, which can provide a focal point for efforts to control occupational injuries. Worker compensation data can also provide a valuable comparison of an employer to other similar employers, like those in similar industries. This comparison can provide a competitive advantage for that company relative to other companies performing similar work. Shift work can also be studied in compensable claims. A direct comparison of night- and dayshift injury rates and severities within a given company will indicate the influence of shiftwork schedules on employee health and the associated loss of work time from injuries.

Another strength of worker compensation data are that it can be linked to the adjudication process for the compensable claims to provide an indication of the potential for legal action for certain injuries, as well as the settlement rates for such litigated claims. A comparison of employers in a similar industry in terms of the percentage of claims litigated can provide an interesting insight into which injuries are contested often. One study suggested some evidence of an increased likelihood for a poor health outcome when a worker's compensation claim was litigated (Vaccaro *et al.* 1997).

4. LIMITATIONS AND IMPLICATIONS OF THE ANALYSIS OF COMPENSABLE CLAIMS

A shortcoming of the information available regarding compensable claim data is that some claims lack some necessary information to make the comparisons suggested above. The number of compensable injuries needs to be associated with an employment base to have a reasonable ratio of compensable injuries for that workforce. A total number of employees for that company is not collected on claim forms and must be attained elsewhere to compare rates of compensable injuries across employers. A field that omitted on some compensable claim forms is the field of job title. This omission weakens the ability to identify hazardous occupations and link exposures to hazardous substances or working conditions directly to specific occupations. Further, an indication of how often the individual performs that task is also rarely included in the claim. This weakens the potential to infer the worker knowledge or expertise in the job and the familiarity of the employee with the risks associated with that job. A weakness of the system of sending in compensable claims is that there is no presumption of safety knowledge for the individual filing the claim, which can result in a loss of information relevant to the claim that may be important in determining the causal nature of the injury. Similarly, accidents often result from a complex series of events that cannot be adequately described with a single classification system (Sanders and McCormick 1993). As with any reporting system, there is the potential for a bias in the reporting of injuries, and some research has indicated that accidents may be underreported by as much as 15-20% (Beaumont 1980). A serious limitation to the analysis of compensable claims is that a compensable claim carries with it a minimum number of days of lost work time necessary for the claim to be filed. Thus, there are a lot of injuries and accidents that do not appear in the data as they fail to rise to that level of lost work time. This results in a bias of compensable claim data in the direction of more serious injuries and only represents a portion of the occupational injuries and illnesses experienced at work.

5. CONCLUSIONS

Worker's compensation data provides many opportunities to perform analysis that results in occupational injury control. The data can identify hazardous conditions and exposures and the associated health consequences. Combining the data with specific company safety measures can evaluate the effectiveness of interventions like machine guarding. Comparison of compensation data between employers in a similar industry can identify the competitive advantage of a safety and health program. Analysis

of worker compensation data are more powerful when combined with other sources of information to provide an accurate context of all of the factors related and a work-related injury or illness.

REFERENCES

BEAUMONT, P., 1980, Analysis of the problem of industrial accidents in Britain. *International Journal of Manpower*, 1, 28–32.

BERKOWITZ, M. and BURTON, Jr, J.F., 1987, *Permanent Disability Benefits in Workers' Compensation* (Michigan: W.E. Upjohn Institute for Employment Research).

FEUERSTEIN, M., MILLER, V.L., BURRELL, L.M. and BERGER, R., 1998, Occupational upper extremity disorders in the federal workforce: prevalence, health care expenditures, and patterns of work disability. *Journal of Occupational and Environmental Medicine*, 40, 546–55.

GALLAGHER, R.M., WILLIAMS, R.A., SKELLY, J., HAUGH, L.D., RAUH, V., MILHOUS, R. and Frymoyer, J., 1995, Workers' compensation and return-to-work in low back pain. *Pain*, 61, 299–307.

LEIGH, J.P., MARKOWITZ, S.B., FAHS, M., SHIN, C. and LANDRIGAN, P.J., 1997, Occupational injury and illness in the United States: estimates of costs, morbidity, and mortality. *Archives of Internal Medicine*, 157, 1557–68.

SANDERS, M.S. and MCCORMICK, E.J., 1993, Human error, accidents, and safety. In *Human Factors in Engineering and Design* (7th ed.) (New York: McGraw-Hill).

VACCARO, A.R., RING, D., SCUDERI, G., COHEN, D.S. and GARFIN, S.R., 1997, Predictors of outcome in patients with chronic back pain and low-grade spondylolisthesis. *Spine*, 22, 2030–5.

WISCONSIN WORKER'S COMPENSATION FORM WKC-12. Revised 02/98.

Collaborative Learning Environment in Higher Education: Implementing Job and Organizational Learning Theories in Academia

K. Sanders and C. Carlson-Dakes
University of Wisconsin-Madison, Madison, WI 53706, USA

1. INTRODUCTION

As Human Factors Engineers specializing in Sociotechnical Systems, we have found it intriguing and greatly rewarding to apply our understandings of job and organizational design to the world around us: higher education. Specifically, we have spent the last five years working to help faculties redesign their jobs, beginning with teaching, so that they find the experience more rewarding, satisfying, stimulating, and less isolated. We have done this by designing faculty development programs that are completely voluntary, intense, and centered on collaboration, helping faculty learn how to collaborate effectively in order to change the institutional policies and procedures so that they and their students learn and grow.

To those around us, specifically those from educational theory and/or faculty development backgrounds, we have been surprisingly successful. Our program grew from an Industrial Engineering dissertation (Sanders 1993) to a college-wide pilot program, then to a three-year, federally funded campus research program, and most recently to be integrated into the institution as a center, "Creating a Collaborative Academic Environment," which is now part of the Provost's Office at UW-Madison.

It was clear from the start of the dissertation study in 1993 that many people doubted that faculties from a large, decentralized research institution would volunteer to spend 1.5 hours each week learning about learning in small groups of diverse colleagues. We were told that faculties would not be interested in teaching or collaboration, and certainly would not have time to devote to professional development in teaching. In fact, we were told that unless we did paid them, there would be no hope of participation.

We are glad to report that those concerns did not turn out to be our reality. We have found that it is possible to translate the classic theories from Herzberg, House and Kahn, Lawler, and Trist into a meaningful and motivating learning experience for faculties. In fact, our particular interpretations share much in common with research findings from adult education. The effects of CCLE have been shown to be lasting in over 135 faculties, and the resultant changes in their attitudes and behaviors have resulted in changes in organizational policy and procedure at the department, college, and divisional levels. Additionally, with our integration into the central administration, we now have an opportunity to bring together groups that previously have not communicated or collaborated. We feel that we are at the very beginning of wider scale organizational change, all resulting from a small, grassroots program designed from classic literature in the meaningfulness of work and the reduction of stress.

In this article, we offer to you the principles on which we base our work. We have used these principles to design our original faculty development program for learning and teaching, Creating a Collaborative Learning Environment (CCLE), as well as the foundation for our newer projects, the Peer Review of Teaching Project, the Teaching/Learning Colloquium Series, and our parts of collaborative projects in the NSF Foundation Coalition and the Leadership Institute. Our future projects will include sister programs for faculty development in research and service, the design of which will also be grounded in these few principles. The specific program structure and effects of CCLE are reported elsewhere (Sanders *et al.* 1993), and will not be included here.

2. KEY PRINCIPLES

2.1. Program Structure: Collaboration, Connectedness, Diversity, Safety, Responsiveness

Our faculty development approach centers on our core beliefs about how learning happens, and the atmosphere necessary for personal and professional growth. Essentially, we create structures where people can work together in a safe, respectful manner, so that they all feel their contributions are not only heard, but valued and an integral part of the group's progress.

2.1.1. Collaboration

We view learning as a complex, social process of making meaning. As we know from classic job enrichment literature, learning which is meaningful to an individual, can be motivational. That is, people will want to apply their new understandings to their work environment if they view their new learning as relevant, and have the social support and resources to experiment. Given our assumptions about the social nature of learning, collaborative learning is a logical choice. Additionally, collaboration skills are required for organizational movement on a larger scale. We suggest that program structures and activities be designed to teach collaboration explicitly, in addition to any specific program content.

2.1.2. Connectedness

We also consider learning to be an ongoing process of making connections and interpreting relationships between previous experiences and new information and experiences. This is key in designing faculty development experiences, because many workshops and short-term interactions do not take into account that continuity and time engaged in meaning-making are key to deep learning. We ask that participants continue to engage in deep discussions every week over the course of eight months. This intense approach works because we all know that people are more likely to remember what they have experienced, lived, or discussed at length. When you seek deep change, in our opinion, you must create a learning experience that creates strong connections between individual learners and their peers, their teachers, and the content matter. We create a community of learners by structuring the program so that all participants recognize that a group of colleagues knows them, their concerns, and their contributions.

Diversity is also a central focus in terms of team construction. We place people in deep conversation with colleagues they might never meet in another setting. We have found that the groups with the most diversity (for example, a team might consist of faculties from African Languages and Literature, Genetics, Mechanical Engineering, Physiology, Children's Literature, Rural Sociology, Human Ecology).

2.1.3. Safety

In addition to enhanced creativity, the diverse teams also help to ensure that people are not placed in groups with colleagues they might know from other settings. For example, we do not recommend placing people from the same department in the same team because they carry with them certain expectations of each other and personal histories, perhaps not positive, that inhibit fresh insights and inquiries. In order to be safe and respectful, groups must be formed in such a way that individuals do not feel as if they can predict the perspectives and positions of others. Instead, we wish people to feel that they are exploring new territories, new ideas and vantage points, that they would not be able to create as individuals, through the social construction of knowledge.

We structure working groups by scheduling weekly meetings around participants' availability, so that in as much as possible, the programs accommodate their schedules. Faculties meet each week for 1.5 hours over an entire academic year, in small groups of seven to eight with a program facilitator. It is important this facilitator is not a colleague, but instead someone with no vested outcome in the conclusions reached by the group.

2.1.4. Responsiveness

Formative program assessment must be built into the program structure from the very beginning. That means that there must be a way for the developers to gather and assess information about what the teams are learning, and what actions they are taking with that new knowledge. In CCLE we interview faculty participants at the end of each semester of participation to hear what they have learned, and how the program is or is not meeting their needs. The interviews are then transcribed and reviewed for patterns across groups and over time with individuals. We are an action research program, however, even when we run purely applied programs we always have the groups discuss their progress with each other in structured discussions, and we ask them to write down as individuals, what they are learning, what is working well for them, and what is not working so well. We actively solicit their input to help us redesign and refine the program structure and activities as needed.

2.2. Program Activities: Challenge, Complexity, Creativity, Interpretation, Reflection

The exact program activities necessary for any particular professional development program are highly dependent on the topic area under exploration. However, we do have a number of underlying principles that we consistently use when designing new programs, which are also interwoven with our assumptions about the nature of learning described in the previous section.

2.2.1. Challenge

Challenge people to reconsider their underlying, unconscious beliefs about their work and their colleagues. We find that many

people assume that they are alone, isolated in their desires and concerns. They assume no one understands, that they are the only people who feel something is missing in their work, that things are the way they are because someone put thought into creating them that way. Bringing people together and structuring conversations in a safe and nurturing environment brings to light the fact that indeed, we are in this together, no one of us is alone. The current system is just that — the current system. We do not have to continue in the same pattern. To do this, we have participants explicitly reconsider their core ideas and assumptions about what they are trying to accomplish as individuals and as members of the university, at work and in their lives as a whole. We ask them to read literature that is likely to be outside their areas of expertise, not for the purpose of believing everything, but instead for the purpose of awakening to new possibilities, new disciplines, new avenues of thought that might take them to different destinations. We consciously attempt to avoid feeding the idea that there is 'one answer' somewhere in the world, and that they need only find that answer. Instead, we challenge them to create their own answers, their own questions, their own action plans. Our role is to stimulate their questioning, open doors of opportunity for exploration, and to provide social support and feedback for their journeys.

2.2.2. Complexity

Do not simplify what is not simple. We believe that one of the reasons we have been successful in engaging faculty at a pre-imminent research university in extended discussions about learning and teaching is because we do not reduce the complexity of the topic or the system it exists in. That is, we hold the complexity central by drawing it into the activities explicitly. We have faculty describe and map out how institutional policies and procedures affect the outcomes they are interested in — in this case, student learning. They make connections between the tenure system, grading, mass education, and teaching approaches. If we were to simplify this into the "ten priorities for institutional change" we will have defeated our purpose, which is to have these people harness their creativity, skills, and motivation to change their own institution. Additionally, we feel safe in the knowledge that there is no one solution, or question, or direction. There are many possibilities, and we offer a forum for them to discuss the complexity, instead of denying it.

2.2.3. Creativity

Foster creativity by using diversity of worldviews and perspectives in an inclusive way. Use multiple types of activities in addition to reading and discussion, such as drawing, role playing, concept diagramming, game playing, reflective writing, reaction papers, journal writing, in order to keep people actively engaged in seeing the content from new perspectives. Have participants read outside their disciplines or fields, and ensure they listen to and understand the multiple perspectives that others in their diverse groups create. Create activities where the group must create a way to represent multiple views of the same system, and in this way, every diverse view holds value in that it must become part of the whole. Additionally, in order to become part of the group's project, each idea must be understood by the others in the group, or else it could not be represented effectively.

2.2.4. Interpretation

Hold central the act of interpretation and making meaning. It is important for people to realize that the truths they hold on to so tightly have been constructed over years, and often are no longer valid or useful. We make certain that participants understand *we* are not the experts in higher education reform; instead, *they* are the experts. They are the people who will make changes in themselves, their departments, their colleges, and the university. We exist to help them discover options and interpret their possibilities. We do not hold out the "answer" or the "best way." Those are contextual truths at best.

2.2.5. Require reflection

A gift that can be given to faculty or any other employee who feels overwhelmed, rushed, and pressured is a place to go each week where they can think, breathe, and reflect. Just as students do not automatically see the value in reflecting on their learning, so too you might find faculty surprised when you ask them to write out and present what they have learned. Moreover, the very act of writing or speaking what they have learned, and what they plan to do with that new understanding, will help them make new connections. We find this part of learning — reflecting on the issues, how we came to be here, and the possibilities for movement — can make all the difference. It gives space for creativity and insight to bloom.

3. WHAT TO AVOID

It is typical for program developers to be tempted or pressured to do the following:

> grow too rapidly
> preach "the answer"
> pay participants
> problem-solve for the administration

Resist!

3.1. Growing Too Rapidly

When a program becomes successful, word spreads to faculty and administrators (or management). It becomes tempting to "offer" it to more and more people, with the potential for destroying the very characteristics that create success. With collaborative programs centered on work in small groups, an explosion of participants can mean one of two things, neither of which will be advantageous. First, it may mean that the number of small groups grows to a scale where staffing cannot accommodate them. That is, you may find that from six groups of seven to eight people, you now have twelve groups. That will mean that the formative program assessment must change, the number of trained facilitators must increase, and the facilitators will have more difficulty in communicating and organizing themselves. Program coherence will be lost or challenged.

A second possibility is that the structure of the program changes to accommodate larger groups. This is also dangerous in that the nature of the experience in a group of seven or eight is quite different than in a group of twelve. The facilitator will not know all the individuals to the same depth; neither will they know each other. There will be difficulty in having all participate during the same time period, and it will be easier for participants to feel that it is all right if they skip meetings because the group is not depending on them (i.e. there will be enough other people

there, and so it will not matter too much to miss a meeting). We caution that before a program expands, the developers feel secure that they understand what it is about the experience that makes it rewarding for participants. These aspects must be preserved at all costs, even if that means remaining relatively small.

3.2. Preaching 'the Answer'

In programs where it is hoped that people will take what they learn and apply it to situations which are complex and unique, it is vitally important to refrain from reducing complex issues into simple models. People are naturally attracted to things that look simple — like the ten best ways to teach, or the seven sure-fire tips to success. However, everyone is smart enough to look at those lists and critique why they will not or cannot work. We suggest that you should not make the list, but instead create activities where people make and critique their own lists — continually. Our goal is to have people understand what they are intending to achieve, options for exploration, ways to find out if they have been successful, and colleagues to help them plan the next iteration. In this way, we feel individuals and organizations learn and grow. No one can do the learning and experimenting for another.

3.3. Paying Participants

We have been lucky in many ways in that we never had the resources or the belief system that would set up monetary compensation for participation in CCLE or any of our programs. Because we work from job enrichment theories, we believe that the nature of the experience is its own reward. People take what they have learned and use it in their everyday work lives, in teaching, research, and in administration. It changes the way they look at themselves, their roles, their colleagues and their students. This opening up, these new possibilities, would be affected if people were paid for their time with us. We feel that different people might volunteer, and that they would view these discussions as something outside their real job — something extra, added on, not part of. Instead, we desire to infuse a new experience into the "regular" work week. We see ourselves not as offering more work to be done, but as an option to consider for personal and professional growth.

3.4. Problem Solving for the Administration

Finally, it is also common for the administration to see a successful faculty development program as an opportunity to impart their own wisdom to faculty, or to serve out an agenda of problems to be solved. We see that this would be a logical connection for any administrator, and have been very fortunate that pressures applied in this area were momentary and rather light. When a program becomes driven by the administration, it cannot but help buy into the current power structure, the current system of work. If this happens, it not only becomes disingenuous to ask people to creatively re-imagine their work, down to their most basic assumptions, but it probably also becomes near impossible. The more the program becomes bogged down in specific issues to 'solve' beyond those which emerge naturally from the participants, the more it becomes ineffectual. This danger also connects to the issue of safety mentioned early on. If people do not feel safe, respected, and valued, if they know that the agenda is being set from a position of power elsewhere in the organization, they are unlikely to trust, to open up, and to learn.

4. CONCLUSION

In summary, we have found creative implementation of classic job design, organizational learning, and occupational stress literature immensely valuable in helping employees redesign their own work, as well as the policies and procedures at a large state research university. Our implementation of the theoretical frameworks shows that people do respond to meaningful learning and social support in a very positive way. They change their teaching, their research, their attitudes, and their very outlooks. They tell us they are more stimulated and engaged by their work, and are striving in directions they may have never considered previously. The basis in Sociotechnical Systems Engineering has marked overlap with research in adult education, and we feel that large-scale innovation in the cultures and organizational structure of higher education is increasingly likely as practitioners from diverse fields come together to help faculty redesign higher education.

REFERENCES

HERZBERG, F., 1987, One more time: How do you motivate employees? *Harvard Business Review*, September–October, 109–20.

HOUSE, J.S., and KAHN, R.L., 1985, Measures and concepts of social support. In S. Cohen and L. Syme (eds), *Social Support and Health* (New York: Academic Press), pp. 83–108.

LAWLER, E. E., 1986, *High Involvement Management* (San Francisco: Jossey-Bass).

SANDERS, K., 1993, The effects of a participatory process to improve college teaching. Unpublished doctoral dissertation, University of Wisconsin-Madison.

SANDERS, K., CARLSON-DAKES, C., Dettinger, K., Hajnal, C., Laedtke, M. and Squire, L., 1997, A new starting point for faculty development in higher education: creating a collaborative learning environment. In DeZure (ed.) *To Improve the Academy*, 16, 117–50 (Stillwater, OK: New Forums Press).

TRIST, E., 1981, *The Evolution of Sociotechnical Systems: A Conceptual Framework and Action Research Program* (Toronto: Ontario Ministry of Labour).

Communication Processes in Small Groups

B. S. Caldwell

Department of Industrial Engineering, University of Wisconsin-Madison, Madison, WI 53706 USA

1. INTRODUCTION

Group performance represents a significant and distinct element of human activity and forms of work processes. Small groups are often created, or evolve, to coordinate, communicate, and coincide information, task allocations, and strategic alignment to complete assignments that are too complex, difficult, or physically distributed for one person to perform effectively alone.

Human-system interfaces for group communication and task performance have a critical role in the design of technologies to support effective and coordinated distributed complex systems. These technology interfaces must facilitate styles of information flow at distinct levels of aggregation, and domains of human activity, which are not required for individual interface designs. For an individual human—machine interface, the coordination of separate tasks and communication between display, actuation, and other feedback systems occurs within the human controller. As task functions are dispersed between group members, technology interfaces must be developed to replace information flow channels which occur in the human nervous, endocrine, or vestibular system. Therefore, an adequate discussion of technology support for group performance must begin with a recognition of the range of group tasks and styles of performance.

2. DOMAINS AND STYLES OF GROUP PERFORMANCE

McGrath (1984) has developed a taxonomy of group tasks that distinguishes conceptual (or cognitive) tasks from behavioral (or conative) tasks, and cooperative from conflict-oriented tasks. Using these distinctions, plus the differences between planning and execution, McGrath created a circumplex of eight distinct types of tasks:

- Planning: creating plans for action
- Creativity: generating ideas for later use
- Intellective: solving problems with an identifiable correct answer
- Decision-making: solving problems where there is no objectively correct answer
- Cognitive conflict: resolving conflicts of viewpoint or perspective
- Mixed-motive: resolving conflicts of interest, relative value, or intention
- Contests: competitive performances against one or more other groups
- Performances: executing specific tasks towards identifiable goals

Group performance and task type distinctions are also described by Sundstrom and colleagues (Sundstrom, DeMeuse, and Futrell 1990), who emphasize demands of coordination based on shared expertise, and task role differentiation or integration between group members. Of course, timescales differ greatly between purely psychomotor synchronization required for coordinated physical performance, and channels that permit the flow of strategic information and exchanges of opinion. Both task type and group composition therefore represent critical dimensions of group task performance that must be defined in order to provide appropriate technology facilitation for group interactions and appropriate information flow.

Timescales and group members' shared expertise or coordinated knowledge place significant constraints on the bandwidth of feasible information exchange in group performance. In addition, the degradation of coordination depends strongly on the demands for synchrony in the information flow, which in turn is based on expectations by group members of appropriate ranges of acceptable delay for the style of communication and type of task performance. The impact of delayed information flow differs on whether communications are based on a synchronous communications model (based on real-time feedback error correction) or an asynchronous model (based on sequential actions or "move and wait" strategies) (Caldwell 1997).

2.1. Dynamics of Group Networks and Information Flow

Shaw (1981) points out that, as the number of group members increases, the number of potential information flow patterns grows. The issue of the network flow patterns between group members is therefore an important structural concern in technology support for group task performance. (This multiplicity of communication possibilities is the primary justification for group dynamics researchers who insist on a minimal group size of three members. Since the number of possible flows varies as $N*(N-1)/2$, the dyadic case (N=2) provides no variation in possible information flow channels.) Coordination and synchronization is therefore limited to network architectures supported by either technology design or group interaction patterns.

Highly centralized networks, where a small fraction of group members must process the majority of messages in the group, have advantages in speed for simple information exchange situations. The multiple linkages of central members make it possible for information to flow quickly to all members of the group with a minimum of delays and distortions. However, the bandwidth capacity of the network as a whole is limited to the processing capability of the more central members. Overloads of information (either due to technology bandwidth constraints or human information processing limits) to a single central position in the group network can cause catastrophic and progressive deterioration of all communications which do not flow exclusively through peripheral network channels.

Decentralized networks are effective in overcoming degradation due to overload or incapacitation of a single group member, and are therefore desirable in many complex systems where multiple task performance options are required. Because there is no single point of coordination in a decentralized network, however, propagation of delayed, distorted, or misdirected information is more likely to cause disorganization or breakdowns

in group task coordination. Group coordination or network centralization must therefore be sensitive to the technology limitations of the human-system interfaces used to sustain task performance, as well as the social interaction limitations of group members' levels of individual expertise, shared experience, and ability to cope with changes in task characteristics or group performance resources relative to task demands.

3. TOOLS TO FACILITATE DOMAINS OF GROUP PERFORMANCE

3.1. Generation, Negotiation, and Choice Tasks

A full census of the population of group tasks which could be supported by information and communication technologies is neither feasible nor within the scope of this article. However, McGrath and Hollingshead (1994) note that the development and evaluation of technologies designed to support group performance does not adequately represent the distribution of task domains described in Section 2 above. They distinguish four types of group information and communication support tools:

- Group Internal Communication Support Systems (GCSS)
- Group Information Support Systems (GISS)
- Group External Communication Support Systems (GXSS)
- Group Performance Support Systems (GPSS)

It should be noted that "performance" in this sense is not equivalent to the "performance" task type of the McGrath (1984) task circumplex. GPSS systems, in McGrath and Hollingshead's description, also include decision support and other electronic tools for supporting meetings in organizations. (Relatively few organizational meeting tasks can be described as "performances" in the domain of sports or surgical team based execution tasks, but may be any of a variety of conflict, generation, or negotiation tasks.)

Both GCSS and GXSS tools are intended to enhance the information processing capability of group members for coordination, both internally and externally, and recognition of temporal factors on information flow. GISS tools are also commonly described as information repositories, databases, or intranets. These tools represent collections of information resources that can be accessed by group members to support shared knowledge and capture the additional experiences of group members as tasks are performed.

3.2. Execution Tasks

Relatively few tools have been developed to support distributed groups conducting coordinated psychomotor performances or execution tasks. Unlike the communication support tools described above, coordinated physical behaviors require information synchronization at the precision of the human nervous system (approximately 200–400 ms for feedback-based psychomotor or sensory tasks). This level of precision functionally restricts the capability of computer technologies to support group-level coordination and performance.

In fact, the trend towards increasing fidelity and display resolution of individual human–computer interfaces increases the difficulty of designing effective distributed synchronous interfaces. Two paths towards developing group-level interfaces have been followed. One path emphasizes "parallel" presentation of the same interface elements to all members of the group, with

simple separations of task allocations according to a sectioning of the single task among group members. This approach corresponds to a more undifferentiated or non-expert model of team performance (Sundstrom *et al.* 1990). Distinct, but coordinated, interface displays pay a much greater performance penalty for increasing interface resolution with limited bandwidth channels. Frequently used methods for increasing interface validity in individual supervisory control or navigation tasks (such as shape rendering, interactive hints, or other computationally intensive display algorithms) can degrade, rather than enhance, performance in distributed group execution tasks requiring psychomotor and sensory performance coordination (Caldwell and Everhart 1998).

4. IMPACT OF NON–TECHNOLOGY FACTORS ON COMMUNICATION PROCESSES

The technological capabilities of group communications systems place distinct bounds on what information flow is possible between group members. It is, however, critical to recognize that non-technology factors, particularly situation constraints and group behavior norms, can more significantly determine the overall effectiveness of a particular communication system implementation. Time available as a task resource or constraint is clearly one element affecting communication processes in groups. Even for cognitive and strategic tasks, communications effectiveness is severely limited if the time required for information flow between group members is long compared to the time available for task performance. (This issue is most clear in feedback-controlled physical execution tasks, where the time available for coordinated display and control information flow is frequently under one second in order to maintain task performance standards.)

Task urgency or the time progression of the group task situation is a factor that may make the same communication technology effective in one task setting and useless in another. Both the lags (elapsed time between initiation of information flow and the start of its receipt) and flow restrictions (the amount of information which can be received and processed through available channels in a unit time) of information exchange between group members can significantly affect task coordination depending on the momentary changes in task demands. Therefore, communication process adequacy must be examined as a dynamic variable in group communication, and not simply on average or aggregate.

Technology appropriateness to support effective group coordination may also be a function of the alternatives available to group members. Several theories have been put forth describing the characteristics which most influence communication media use. No one theory has managed to fully explain differences in medium usage, suggesting that choices are made for a variety of convenience, consistency, or other non-technology factors. When a specific communications channel does not possess sufficient capability to support certain types of information flow, and adequate alternatives are not available, group members will adapt communications patterns to enable coordination and task performance within the constraints of the technology. (The use of ASCII characters as "emoticons" in electronic mail, and the adoption of specialized vocabularies to minimize communication degradation in radio exchanges, are examples of these processes.)

However, group communication is not arbitrarily flexible or elastic to adapt to a technology which is fundamentally incapable of exchanging task critical information in the time and task constraints that define minimal standards for acceptable group communication and task coordination.

5. TOWARDS QUANTITATIVE MEASURES OF GROUP COMMUNICATION VARIABLES

Small group communication and task performance shares many characteristics with other complex engineering systems. Mathematical tools developed to advance the description, design, and prediction of engineering systems have been shown to be effective across a range of physical representations, from fluid and mechanical energy flows through thermodynamic heat transfers and electromagnetic component systems. Although most researchers of group communication and performance would hesitate to suggest that social and organizational processes can be described at the same level of mathematical rigor as engineering technology systems, some progress has been made to describe *patterns* of communication processes in group activity.

Social network analysis techniques have begun to focus on the influence of the organization of links between group members on the response of the group to task demands and communications breakdowns. One measure of link organization is *warp*, which refers to the degree to which communications links are centralized to flow through a small number of critical nodes. The higher the warp measure, the more centralized is the network. As described above, such networks can provide greater efficiency and minimize delays or degradation due to multiple link paths. On the other hand, high warp networks are susceptible to massive breakdown if one of the most central nodes is unable to support the information traffic through it (either due to situational demands or technology or other breakdowns).

Information theory and entropy measures have long been applied to the study of information flow and communication. The original conceptualization of information bits and sender–receiver channels is limited, however, to the flow of data, not the social and experiential human factors which allow a group member to translate data into information. Thus, bits are not the appropriate level of analysis to continue the development of quantitative analyses of information flow and group communication in complex task environments. A more general tool for analyzing the quality and effectiveness of information sharing is *coupling*, which is used in distributed engineering systems analysis to examine the capability of transferring energy from one component of the system to another. As coupling increases, components begin to interact with less loss of energy and less uncertainty of the effect of one component on the other. (Zero coupling indicates that components are functionally independent, and have no direct influence on each other.) Coupling is affected by both the *inertia*, or resistance to change, of the affected components, and the *damping*, or opposition to energy flow, of the medium in which the components operate.

In electromagnetic or fluid flow environments, coupling, inertia, and damping are well-defined and consistently quantified measured variables. As of this writing, strong mathematical analogues to these variables have not been developed or agreed upon in the research literature. Some previous allusions to these types of variables have been made in the psychological literature, including the theories of Lewin (1938) and Gibson (1979). Prescriptive definitions of "the right communication network" are no more likely than accurate and general pronouncements of "the perfect electronic component" or "the universally applicable flow description". In the latter cases, the questions of critical design parameters or optimization criteria are acknowledged as an important element of the engineering design process. Nonetheless, the same mathematical tools are known to apply effectively to those, and other, systems engineering projects. The next major advance in group communication and performance technology design rests in the identification of quantitative measures which can apply effectively to the structure, function, and effectiveness of information sharing and coordinated task performance.

REFERENCES

CALDWELL, B. S., 1997, Components of information flow to support coordinated task performance. *International Journal of Cognitive Ergonomics* 1(1): 25–41.

CALDWELL, B. S. and EVERHART, N.C.,1998, Information flow and development of coordination in distributed supervisory control teams. *International Journal of Human–Computer Interaction* 10(1): 51–70.

GIBSON, J. J., 1979, *The Ecological Approach to Visual Perception*. Boston: Houghton Mifflin.

LEWIN, K., 1938, *The Conceptual Representation and the Measurement of Psychological Forces*. Durham, NC: Duke University Press.

MCGRATH, J. E., 1984, *Groups: Interaction and Performance*. Englewood Cliffs, NJ: Prentice-Hall.

MCGRATH, J. E., and HOLLINGSHEAD, A. B., 1994, *Groups Interacting With Technology*, *Sage Library of Social Research #194*. Newbury Park, CA: Sage.

SHAW, M. E., 1981, *Group Dynamics: The Psychology of Small Group Behavior* (3rd ed.) New York: McGraw Hill.

SUNDSTROM, E., DEMEUSE, K. P. and FUTRELL, D., 1990, Work teams: applications and effectiveness. *American Psychologist* 45(2): 120–33.

Community Ergonomics Theory: Applications to International Corporations

A. D. Bayeh† and M. J. Smith‡

†Purdue University, Department of Industrial Engineering

‡University of Wisconsin, Department of Industrial Engineering

1. INTRODUCTION

Community ergonomics (CE) was developed from two parallel directions. One came from existing theories and principles in human factors/ergonomics and behavioral cybernetics, which have been applied to assess situations and find solutions at a macro-ergonomic (Hendrick 1986) level. The second direction emerged from the careful and thorough documentation of specific case studies of communities, which in turn led to additional principles and theories with more contextual relevance.

Thus far, CE has been applied at the community level, some of these applications have been documented within the last few years (Cohen and Smith 1994, Smith and Smith 1994, Smith *et al.* 1999). This chapter will examine the applicability of CE at the international level, specifically for corporations engaged in international development and trade. This will be done in two steps. First, the transferability and applicability of CE theories, principles and tools from the local community level to the more global international level of analysis will be examined. Then, CE principles (Smith *et al.* 1999) will be restated so that they can be applied to the international level of analysis.

2. BACKGROUND

Organizations are complex systems which have been defined in terms of their characteristics such as structure (formal and informal), culture, purpose, products and services, climate, workforce composition, leadership, management style, processes, tasks, technology, tools, image, reputation, information flow networks and mechanisms, labor-management relations, geographical context or scope (global, national, regional, local), societal context, and size. Organizations have technical and personnel subsystems that interact to produce outputs. The coordination, integration and combination of these aspects of organizations need to be properly managed to provide effective and efficient operations. This is particularly true for international corporations.

The world has become increasingly more open, taking organizational design principles and management practices to a new level of acceptance and application in the last half of the twentieth century. Simultaneously, this time period has seen struggles for fairness, equality, freedom, and justice for all peoples of every nation. In addition, a more complex global economy has emerged throughout the world. Accordingly, organizations have had to operate within a complex world economy characterized by the following issues, among others: continuous change, a heterogeneous (often international) workforce at all job levels, increased spatial dispersion of physical and human assets, decreased standardization, increased diversification of products and markets, decreased centralized control of resources

and information, uneven distribution of resources, variable performance within and between locales, increased operational and safety standards, and increased susceptibility to global changes in the economic, social, political, and legal conditions. Many organizations have responded to the opportunity presented by the unique openness of the world economy by expanding production and trade worldwide.

In this climate of increased international trade, important issues of social and cultural values need to be carefully examined. Culture, in and of itself, is hard to define in a way in which all its facets and dimensions are appropriately addressed. However, there exists a common understanding and acceptance of what specific culture and cultural values mean. In terms of a global value, there seems to be a universal agreement on the fact that mutual respect and fairness are at the heart of healthy multicultural dynamics. The right to be respected requires showing respect for others and not violating this universal right. The same can be said for fairness, and other considerations such as commitment and responsibility. There is a requirement to be mutually tied so that the welfare of one party is important to other parties' interests.

There is little agreement about what makes effective multicultural organizations. This has led to the emergence of more refined international management practices in response to numerous obstacles encountered when organizations expand abroad or when they transfer processes abroad. Difficulties encountered in growth, expansion, internationalization and globalization are mainly due to the following factors:

- The lack of a process for effective transnational transfers.
- The lack of knowledge of operational requirements and specifications in newly entered markets.
- Ignorance of cultural norms and values in different countries.
- The lack of adaptability mechanisms.
- Low tolerance to uncertainty, ambiguity, change and diversity, etc.

Even though much progress has been made to address some of these issues, there are critical problems that still remain at the organizational level. One of the most important problems in this regard is the lack of a "genuine" corporate commitment to social responsibility with the same intensity as striving for profits. In fact, some managers may believe that social responsibility may conflict with corporate financial goals and tactics. We believe that long-term corporate stability and success (profits) follows from community support for the enterprise and its products. Social responsibility nurtures the mutual benefits for the enterprise and for the community that leads to stability.

We propose that multinational corporations must accept a corporate social responsibility that recognizes the universal rights of respect and fairness for all employees, neighbors, purchasers and communities. Whether these rights are profitable in the short-term, or difficult to attain, such corporate social responsibility is a requirement if global ventures are going to be successful in the long-term. The survivability, acceptability and long-term success of the corporation will depend not only on profits, but also on social responsibility. For this will lead to greater accessibility to worldwide markets. Organizational structure must go beyond the micro-ergonomic and macro-ergonomic levels (Hendrick 1986) to also include the societal (or cultural) level. This will enhance the "fit" between the organization and the economic

markets within which it operates. They must address multicultural design, a comfortable corporate culture, and principles of respect and fairness for employees, customers, and neighbors. This will improve the "fit" with the international "cultures" in which the corporation operates.

3. TRANSFERABILITY CRITERIA

The transferability and applicability of theories and principles are often not universal. This is especially so when an *ergonomic* concept of improving the "fit" is applied, since ergonomic approaches focus on the "unique" characteristics of the environment (culture) and of the person (or the organization) to get the best match. There are key factors that may be considered transferability criteria that determine whether a set of situations is comparable. These comparability criteria are given below.

3.1. Content

This defines what the internal situation is. For instance community characteristics such as learned helplessness, dependency, cumulative "social" trauma (CST), poverty, deficiency in access to and appropriate use of resources affect the potential fit. Often, fundamental improvements must be made in combination before fit can be addressed. \h Table 1 points out some symptom characteristic of settings where CE has been shown to be useful at the community level, and we have expanded this to show how these same symptoms are present at the international level.

3.2. Context

These are conditions external to the situation itself, which are affecting or being affected by the situation. The context captures the environmental, political, social, and economic climate. The context is sometimes highly influenced by religion, social class, political, ethnic, age and gender issues which may dominate the particular place where the situation takes place.

3.3. Process

These are mechanisms describing how the personal, corporate or community situation operates. The process is defined by the mechanisms linking people or corporations to their particular environment, leading to predictable outcomes. Understanding the process is essential in pinpointing crucial points of intervention to eliminate societal hazards or to minimize the negative impact of such hazards.

These transferability criteria may be used to expand CE from the local or community level to the global or international level of analysis and application as shown in Table 1.

4. COMMUNITY ERGONOMICS PRINCIPLES IN AN INTERNATIONAL AND MULTICULTURAL CONTEXT

The principles focus on a goal of social responsibility, fairness and social justice, and do not threaten the prosperity of a company or organization. They are built on the premise that the society and communities in which a corporation operates and from which the corporation generates profits should benefit from the presence of the company. Thus, the corporation should contribute to the development, growth and progress of the local communities. These principles propose a symbiotic relationship between the hosting community or society and the "guest" corporation.

4.1. The FIT Principle

Owing to the diversity of the workforce in multinational ventures

Table 1. Hazards at the community and international levels of Community Ergonomics

Societal Symptom	Definition	Local or Community Level of Analysis	International Level of Analysis
CST - Cumulative Social Trauma	This is the outcome of long term exposure to detrimental societal conditions leading to a cycle of dependency and ineffective coping.	Inner cities with low quality of life, high unemployment rates, low quality of educational, health, security, and social services.	Developing countries with low quality of life, high unemployment rates, low quality of educational, health, security, and social services.
Adverse Societal Conditions	These are the specific detrimental societal conditions in the environment within which people have to function. The severity of the adverse condition determines the likelihood of survival or social proper development.	Communities with high unemployment rates, high crime records, high drug trafficking occurrence, high teenage pregnancies, low education level, low security, health care low income, low skill level, high corruption.	Countries with high unemployment rates, high crime records, high drug trafficking occurrence, high teenage pregnancies, low education level, low security, health care, low income, low skill level, high corruption, political and economic instability.
Societal Isolation	This is the phenomenon that results from segregating, outcasting and alienating groups of people, and treating them with inequity and unfairness.	Inner cities, ghettos, barrios.	Developing countries with little economic power and weak political leadership.
Economic Dependency	Requiring assistance from others. This is the result of needing to survive and seeking help, with no skills or mechanisms to ensure long-term improvement and growth.	Welfare systems, federal funding, unemployment programs, etc. which provide immediate assistance but no survival skills. Once down it;s hard to get up.	International Monetary Fund, World Bank international institutions lending money with high interests and restricted stipulations but no financial management or administrative leadership or knowledge
Learned Helplessness	This is a long-term self-perception of recurrent social isolation and economic dependency that disables people from improving their lives.	Multiple generations of welfare recipients, conformity with the surrounding and observed standard of living as one's highest potential goal.	Countries that are constant recipients of loans and support leading to a perception of disability and helplessness.

and the ongoing trends of today's world economy, accommodations need to be made by companies for a diverse workforce. That is, companies need to design for cultural diversity. Corporations must understand and incorporate the norms, customs, beliefs, traditions and modes of behavior of the local community into their everyday operations. In some communities there will be multiple cultures that need to be included in this process. Often there is a need to strike a balance among the various cultures as they are not always compatible. The corporation's organizational structure and operational style need to be flexible to bridge the gap between the corporate and local cultures. This cultural accommodation will produce commitment to the corporation from employees, customers and neighbors.

4.2. The BALANCE Principle

This is taken from a concept proposed by Smith and Sainfort (1989) which defines the need to find the proper relationship among components of a larger system. Based on this concept we believe that there is a need for balancing corporate financial goals and objectives with societal rights and corporate social responsibility. Companies are a part of a larger community and through their employment, purchasing, civic and charity activities they can influence community development and prosperity. As an integral part of this larger system companies have a responsibility to promote positive balance for the benefit of the community and the corporation. Positive balance often cannot be achieved by optimizing any single or subset of components of an operation such as profits or market share. Rather, the best overall system outcome is achieved by trading short-term optimization for long-term stability by accommodation of all elements of the system including the community.

4.3. The SHARING Principle

Traditionally, a corporation's success has been measured in terms of its financial growth. There are many factors that may lead to this financial growth including the value of the company (stock price), market share, sales, profits and assets. However, there are other factors that are not usually taken into account but will become more critical as social awareness becomes more prominent. For instance, customer loyalty, community support and acceptability of products will be critically related to the corporation's long-term financial success.

For example, if a corporation chooses to invest some of its profits back into the community in ways that are significant to that community, then the corporation may be viewed not only as a business but also as a community partner. Such contributions may take the form of building schools, providing educational scholarships, providing technical training and skill development, giving assistance to small or family owned businesses, or developing community utilities. In giving something back to the community the corporation is developing loyalty to its products and protecting its long-term profitability.

4.4. The RECIPROCITY Principle

This deals with the mutual commitment, loyalty, respect and gain between producers and consumers. It results from the BALANCE and SHARING principles. A bond results from the corporation giving something back to the community, which builds loyalty from the consumers to the company, and eventually leads to a

genuine sense of loyalty from the organization back to the community. In this respect, what might have started as responsibility, will over time become mutual loyalty and commitment. Within the corporate organization, the same phenomenon takes place when the organization shows responsibility towards its employees (producers), who in turn become loyal and committed partners with the corporation.

4.5. The SELF-REGULATION Principle

Corporations should be viewed as catalysts of self-regulation and socio-economic development in host communities. Communities and countries in disadvantaged economic conditions (Table 1) typically show symptoms of learned helplessness, dependency, isolation and cumulative social trauma. Instead of perpetuating conditions that weaken people and institutions, an effort should be made to help people to self regulate, grow, flourish and become productive. In this effort, corporations are very important because they provide employment, training and professional development opportunities that give people the tools to help themselves. Corporations can also invest in the community infrastructure, such as schools, clinics and hospitals, which leads to stronger, healthier and more independent communities in the future.

Through community ergonomics, corporations can help disadvantaged people understand that they have the right to a better, healthier, more productive life. The self-regulation principle asserts that it is in the self-interest of those who have economic privileges and resources to share them with those who do not. For such sharing should produce prosperity and self-regulation, and less need for future sharing. It brings about stability.

4.6. The SOCIAL TRACKING Principle

Awareness of the environment, institutional processes, and social interaction is necessary for people and corporations to navigate through their daily lives, and for communities to fit in the broader world. Clear awareness helps to control the external world and leads to more robust, flexible, open system design. It is important for community members, employees and corporations to be aware of their surroundings to be able to predict potential outcomes of actions taken. Similarly, it is important for corporations to develop a certain level of awareness regarding the workforce, the community, and the social, economic and political environment within which they operate. This includes the cultural values of the people affected by the corporation's presence in a particular community. Considering that change is normal in the global economic arena, it is critical for effective decision making to track this change, and to understand the implications of diversity.

4.7. The HUMAN-RIGHTS Principle

This principle underscores the belief that every individual has the right to respect, a reasonable quality of life, fair treatment, a safe environment, cultural identity, respect and dignity. There is no reason for anyone not to be able to breathe fresh air, preserve their natural resources, achieve a comfortable standard of living, feel safe and dignified while working, and be productive. The term "human rights" implies that people are entitled to these considerations. Communities and corporations should be consciously fair and righteous in achieving their economic goals and objectives. People should not be assigned a difference in

worth based on class, gender, race, nationality, or age. The workplace is a good starting point to bring about fairness and justice in societies where these do not exist as a norm. It is left to the corporations' discretion to use their economic and political power to bring about social change where human rights are neglected.

4.8. The PARTNERSHIP Principle

This proposes a partnership among the key players in the system in order to achieve the best possible solution: corporation, community, government, employees and international links. By doing this, balance may be achieved between the interests of all parties involved, and everyone is treated fairly. In addition, partnership assures commitment to common objectives and goals.

5. CONCLUSION

Operating a business beyond traditional boundaries and entering other cultures and societies should be undertaken with a sensitivity for the needs of the new communities. The essence of internationalization, globalization and multiculturalism is in the culture and climate that the corporation develops to be sensitive to the community and the diversity of the workforce. Just having a diverse workforce in a corporation does not lead to equality or fairness, and it does not ensure that the workplace is receptive or appreciative of diversity or multiculturalism. This includes respect, partnership, reciprocity, and social corporate responsibility towards employees, community and society as a whole. It requires seeking a balance between the corporate culture and that of the community where the business operates and the cultures brought into the company by the employees. In the past corporations have entered new markets all over the world, profiting from cheap labor, and operating freely with little or no safety or environmental liability. However, the level of social awareness has increased all over the world, exposing sweat shops, inhumane work conditions, labor exploitation and environmental violations across the board. The focus as we enter the 21st century will be in doing business with a "social" conscience. By doing this, corporations will become welcome in any part of the world they wish to enter. This is how their products will be sought and accepted, and that is how consumers, employees, communities and societies will come to support businesses so they will remain successful into the future. The key to international business success is not only in just being financially shrewd, but also in balancing economics with social awareness and social corporate responsibility for the community's benefit.

REFERENCES

COHEN, W.J. and SMITH, J.H., 1994, Community ergonomics: past attempts and future prospects toward America's urban crisis. In *Proceedings of Human Factors and Ergonomics Society 38th Annual Meeting*, Santa Monica, Human Factors and Ergonomics Society, 734–8.

HENDRICK, H.W., 1986, Macroergonomics: a conceptual model for integrating human factors with organizational design. In O. Brown and H. Hendrick (eds) *Human Factors in Organizational Design and Management* (Amsterdam: North Holland), 467–77.

SAINFORT, F. and SMITH, M.J., 1996, Community quality. In J.O. Brown and H.W. Hendrick (eds) *Human Factors in Organizational Design and Management — V* (Breckenridge: Elsevier Science), 541–6.

SMITH, J.H., COHEN, W.J., CONWAY, F.T., CARAYON, P., DERJANI BAYEH, A. and SMITH, M.J., 1999, Community ergonomics. In H. Hendrick and B. Kleiner (eds) *Macroergonomics* (Mahwah: Lawrence Erlbaum).

SMITH, J.H., COHEN, W., CONWAY, F. and SMITH, M.J., 1996b, Human centered community ergonomic design. In J.O. Brown and H.W. Hendrick (eds) *Human Factors in Organizational Design and Management — V* (Amsterdam: Elsevier Science), 529–34.

SMITH, J.H., CONWAY, F.T. and SMITH, M.J., 1996a, Methods for the conduct of community ergonomic research: findings from a community setting. In O. Brown and H.W. Hendrick (eds) *Human Factors in Organizational Design and Management — V* (Amsterdam: Elsevier Science), 535–40.

SMITH, M.J. and SAINFORT, P.C., 1989, A balance theory of job design for stress reduction. *International Journal of Industrial Ergonomics*, 4, 67–79.

SMITH, J.H. and SMITH, M.J., 1994, Community ergonomics: an emerging theory and engineering practice. In *Proceedings of Human Factors and Ergonomics Society, 38th Annual Meeting*, Santa Monica, Human Factors and Ergonomics Society, 729–33.

Community Ergonomics: Planning and Design Solutions for Urban Poverty

W. J. Cohen† and J. H. Smith‡

†Massey University, College of Business, Department of Management Systems, Private Bag 11–222, Palmerston North, New Zealand

‡University of Wisconsin-Madison, Department of Industrial Engineering, 1513 University Avenue, Madison, WI 53706, USA

1. INTRODUCTION

The socio-technical complexity of work organization in US society and the pace of technological change and global competition demand new patterns of political, social, economic, community, and industrial organization. While the US Government is faced with the responsibility of designing democratic foundations for economic and work opportunities, it is the human factors and ergonomics professions that must define, design and improve systems of work and social organizations. Failure to view work as the primary determinant of the human condition and as defining the adaptive status of individuals and groups to guide the overall structure of society, may be the fundamental basis of environmental disorder and chaos confronting the US City (Smith 1962). Community ergonomics as a scientific design practice is field-tested and grounded in practical research to redress matters of striking urban poverty. The problems of low self regulatory control, lack of community–environment fit, societal isolation and dependency on government social programs have created a segment of community members who are in extremely dire circumstances.

The USA, as a complex socio-technical society, has experienced a significant decline in the urban neighborhood environment. The decline is reflected in crime, unemployment, deterioration in schools, change in the family unit as an economic social structure, ethnic conflict and expanded demand for State and Federal resources for prisons, mental healthcare, housing and other forms of welfare to correct so called "human defects."

Pioneers in human factors and ergonomics envisioned a broad application of multidisciplinary sciences to benefit society (McCormick 1970). Today, ergonomists are re-evaluating the more recent conceptualizations of ergonomics as narrowly focused to office or industrial work settings (Nadler 1992, Sheridan 1991). The goal of CE is to improve the fit between people and society to achieve higher levels of self-regulatory control and individual effectiveness. A principle outcome is that institutions designed to help the poor are better able to achieve their intended purposes.

Community ergonomics is more than prescriptive; it is always based on the need to achieve implemented solutions through a purposeful problem solving methodology. The process is designed as a decision and action system, with the benefit that the steps necessary for success can be detailed, measured, and adjusted during the activities. The community ergonomist is a trained professional who can manage complex systems in highly turbulent environments. CE as an engineering practice has emerged to confront these new systems realities through the design and organization of work as a social and economic necessity and to improve the compatibility between people, technology, and environments within which people function.

2. THE NEED FOR COMMUNITY ERGONOMICS

The intended missions of institutions such as welfare are no longer justified (Cohen *et al.* 1994, Smith and Smith 1994). Socio-technical reliability is the probability that a given system will perform its intended function for a range of users. The delivery of services and products do not match the needs, desires, or capabilities of those they claim to serve. In effect, people cannot track the dynamics of the missions and objectives of these institutions, and neither can institutions understand the behavior of people effectively to reduce their concerns. The inability of people and societal components to track each other results in severely impacted systems performance and damage for both.

When community ergonomics is applied to the worst urban blight, the assumption is that socio-technical reliability is low and cannot provided for the needs, abilities, and limitations of residents. Defectively designed (ill-fitting) environments trigger behavior that appears contrary to effective and purposeful lifestyles in systems designed for people with stable jobs and safe residential arrangements.

The continued exposure to interactions of individuals with institutions exhibiting low socio-technical systems reliability over an extended period of time, and the resulting human behavior, lead to the development of a definable and observable symptomology. CST is a term to describe, "individuals, groups, or communities [who] have limited control over the environment, are forced into life habits, coerced into interactions in which they are not motivated, and custodial routines which limit self-regulation" (Smith *et al.* 1996). Smith and Smith (1994) describe CST as daily life routines that over gradual time, "produce wear and tear on motivation, skill, and emotional habits leading to emotional, behavioral, and psychosomatic disorders, trauma, and stresses."

An important aspect of the CST concept is that there is nothing inherent about the condition, thus dispelling the notion of poverty based on hereditary (or any other group characteristic). Long term exposure to extreme social hazards is disabling to the point that people loose skills, motivation or resources to self regulate their lives or functionally to thrive in society. These impairments become barriers that limit access to education, employment income, motivation and positive life-effecting choices. One of the major roadblocks impeding people from leading functional and fulfilling lives in isolated urban environments is the lack of access to resources available and services provided.

3. THE PROCESS OF COMMUNITY ERGONOMICS

Community ergonomics is concerned with the improvement of very poor central city environments neglected, but still firmly present in the consciousness of US politics. However, government efforts to improve very poor central city environments have never been well coordinated, systematic or durable.

The process is structured in stages to prevent participants for becoming overwhelmed during the process. Every meeting and activity has defined outcomes that are necessary to achieve

in order to move the process forward. Each stage is comprised of one or more activities, which are action based so that participants do not simply attend meetings passively. The input, output and transformations that are required at incremental stages successfully to implement solutions are specified by using concepts of social tracking and feedback control. Through this action, there is perturbation among the participants that allows for and promotes growth of social tracking relationships. The need to form bonds among the participants requires that members present their ideas as a function of the others in the group. In the end, it is expected that each member can sense their unique contribution to the effort reflected in the overall design of the interface.

The process results in a conceptual understanding of how to guide people in planning and design efforts and helps to explain human behavior in dynamic group situations. For the first time, behavioral cybernetic principles (Smith 1962) are applied in a complex societal setting coupled with an intensive stage dependent approach to problem solving to create a planning, design, and implementation process. Thus, certain transformations occur at certain stages of the planning process in the form of milestones achieved, social tracking bonds established and feedback perceived. Without this understanding and learning, success is limited at the implementation stage. Overall the CE approach is based on purpose satisfaction and the achievement of goals.

4. ACTIVITIES, PURPOSES AND EXPECTED OUTCOMES

The following sections introduce the seven activities constituting the CE process and are adapted from Nadler's Breakthrough Thinking™ (1994). The detailed CE process specifies the decision steps and criteria necessary to achieve the purposes for each activity (Cohen 1997). The process of stages and activities is shown in Figure 1.

4.1. Activity One

The first activity of the process identifies the people necessary to the effort in its current state and defines their roles and responsibilities in the solution finding structure. For the purpose of this activity, it is more important to identify people with skills needed to ensure the successful completion of the solution finding process than simply choosing names of people typically involved in community planning efforts. The expected outcome is to define specific roles and responsibilities of community members and participants. This, in turn, leads to the assignment of appointments to the planning and design team.

4.2. Activity Two

The second activity specifies the issues of concern, aspirations, needs, and goals of the initial planning team. The purpose of this stage is to define perspectives of each member regarding the sub-
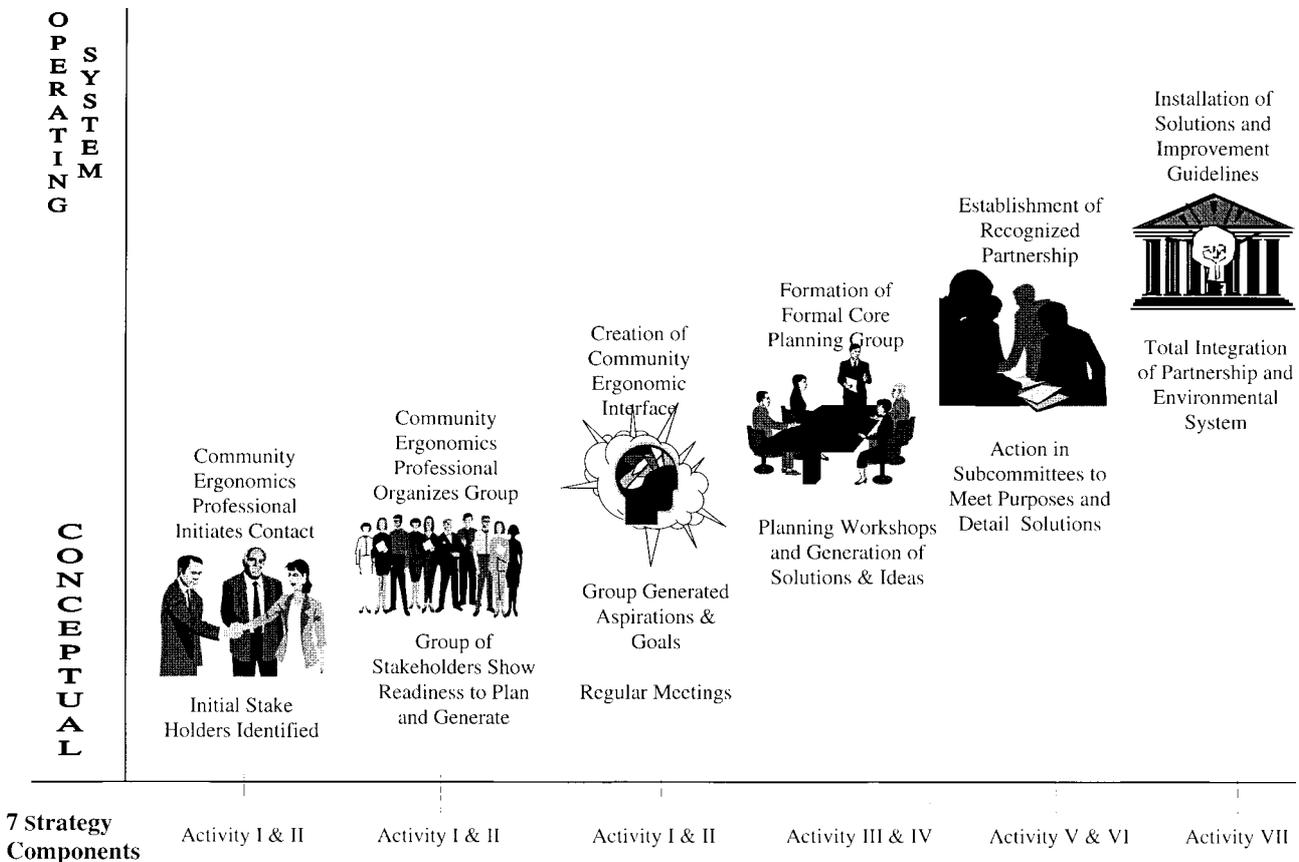
Figure 1. Stages of group formation, control and solution accomplishment in the CE process.

stantive issues to be confronted as well as to identify the values and motivations that make these issues troubling to the participants. The expected outcome is to describe the issues, specific problems, and the values and measures associated with each problem. This first planning document functions as the initial guidelines and project history for prospective recruits. It outlines the entire strategy to be pursued, the types of skills and individuals needed, the issues of concern, and aspirations of the initial planning team. It does not specify solutions or detailed ideas. It is generally a conglomeration of concerns and value statements with the intention to coalesce a group of people into a recognized "effort."

4.3. Activity Three

The third activity establishes the purposes for the community activity by establishing the reason for the effort and the things to accomplish. This stage usually occurs some time after the first two stages are well developed and the planning team is firmly established. It is expected that the specifications of what is to be accomplished and the function of the changes sought are determined. However, the methods and solutions to achieve these purposes are not yet specified. The primary purpose of this stage is to determine the scope and level at which the effort will progress. Project participants decide how the project should progress based on the self-regulatory skills they acquire concerning their abilities and the motivations of others.

4.4. Activity Four

The fourth activity establishes the interface solutions and ideas for the selected purpose level. The goal is to organize ideas and suggestions that will eliminate the need to work on the selected purpose. In other words, ideas are sought which lessen the need to work on the purpose because the conditions that require the need are alleviated. The expected outcome is to target a particular solution or set of solutions for the community ergonomics project. The ideas are grouped into major alternatives with specific components that support the central purpose. The selected idea (major alternative) becomes the target interface for the remainder of the effort.

4.5. Activity Five

The fifth activity allows for the development of sufficient details to ensure the initial and continued feasibility of the selected and approved solutions. Using a systems design matrix from Nadler and Hibino (1994), the complex and varied components necessary for solution achievement are detailed. The expected outcome is to specify the solution in this format of elements, dimensions, properties, attributes, and their interrelations.

4.6. Activity Six

The sixth activity defines the implementation and installation strategy for the target system. The purpose is to identify individual and group attributes, change behavior, conditions of readiness and organization attitudes that influence the acceptance and success of the solution details. The expected outcome is the specification of modifications for change in the form of behavior and actions. This stage occurs in tandem with the sixth activity and also concerns the monitoring and adjustments that are inevitably needed with new systems.

4.7. Activity Seven

The seventh activity establishes a system for continuous review of the community ergonomics planning team effort. The purpose is to describe and specify how change and improvement to the community environment interface building will occur. The expected outcome is an interwoven set of future specifications that allow for improvements to the detailed system and a continued search for improvements to meet changing needs. This stage is most concerned with predicting future needs and aspirations.

5. CONCLUSIONS

The operational requirements of decision and action systems resulting from the CE process are specified, produced and measured during the planning and design activities. The community ergonomist, leading the process, is a trained interface professional, capable of engineering and managing complex systems of people, information, organizations and environmental dynamics required to establish social tracking bonds and feedback control within the planning group structure.

The approach provides new specifications of how to achieve solution implementation in extremely difficult and complex planning environments and how it can be measured, explained, taught, and replicated in future efforts. It is not enough to bring people to the same room to investigate problems. The research details how and why positive results can be expected and obtained; it also explains how to predict success or failure in community project situations.

If urban recovery is a real possibility then the first step for community leaders, residents, thinkers, and policy makers is to play a believing game. In order for positive change to occur, everyone involved has to believe that bringing about change is worth the extraordinary effort it requires. However, few effective design solutions exist for the numerous interface problems prevalent in these inner city communities.

The emphasis of CE is not on research, but on planning and design, structuring, creating, and arranging for new facilities to improve the compatibility between the inner city and the urban areas of concern. Within this context, the process is ergonomically managed by using knowledge about human abilities, human limitations and other human characteristics to design community systems, tasks, products, technology and environments.

The CE approach improves the understanding of effective design procedures for confronting the characteristics of systems, products, institutions and human interactions as interdependent phenomena that may contribute to cumulative social trauma and socio-technical reliability problems. The work will contribute significantly to the new realities of human factors and ergonomics by equipping the ergonomics profession to help with immense socio-technical problems emerging in US society and by developing a broad body of research and application.

REFERENCES

COHEN, W.J., 1997, Community ergonomics: design practice and operational requirements. Doctoral dissertation, University of Wisconsin-Madison.

COHEN, W.J., SMITH, J.H. and SMITH, M.J., 1994, Community ergonomics: past attempts and future prospects toward America's urban crisis. In *Human Factors and Ergonomics Society, 38th Annual Meeting* (Nashville: Human Factors and Ergonomics Society).

McCORMICK, E.J., 1970, *Human Factors Engineering* (3rd ed.) (New York: McGraw-Hill).

NADLER, G., 1992, The role and scope of industrial engineering. In G. Salvendy (ed.) *Handbook of Industrial Engineering* (New York: Wiley), 3–27.

NADLER, G. and HIBINO, S., 1994, *Breakthrough Thinking: The Seven Principles of Creative Problem Solving* (2nd ed.) (Rocklin: Prima).

SHERIDAN, T.B., 1991, New realities of human factors. *Human Factors Society Bulletin*, 34, 1–3.

SMITH, K.U., 1962, *Behavior, Organization and Work* (Madison: University of Wisconsin-Madison).

SMITH, J.H., BLOCK, R.T. and SMITH, M.J., 1996, Cumulative social trauma and human stress disorders in the age of technology. In A. Derjani (ed.) *The American Society for Ergonomic Systems Engineering Conference* (Madison: ASESE).

SMITH, J.H. and SMITH, M.J., 1994, Community ergonomics: an emerging theory and engineering practice. In *Human Factors and Ergonomics Society, 38th Annual Meeting* (Nashville: Human Factors and Ergonomics Society).

Economic Models for Ergonomists

M. S. Oxenburgh and P. S. Marlow

77 Ryan Street, Lilyfield, NSW 2040, Australia

1. INTRODUCTION

Economic models can be used to illustrate the benefit of an ergonomics intervention or can be used to assist implementing ergonomics in a workplace.

To introduce better working conditions management will usually require a financial return. This is no different to the engineer wanting new equipment — he/she must show its return in monetary terms. The major difference between the engineer and ergonomist is that the economic models used by the engineer have been long accepted whereas the ergonomist is yet to reach that point. Here we will introduce the concept of economic models and show how they can assist in introducing methods for workplace improvement.

In all economic models certain assumptions are made. These assumptions are usually not susceptible to scientific analysis or proof but are based on experience, common sense and knowledge of what is acceptable to the particular enterprise. Economic assumptions must be clear from the start and the validity agreed to by those involved, at least for that particular workplace.

The differences in culture, organization, financial flow, technology level, capital and intellectual investment will dictate different assumptions between enterprises, even those in the same industrial field. Clearly wide differences in assumptions will exist between governments and private enterprise, between primary industry and manufacture and so on. No one set of assumptions will suit all enterprises.

2. CHOOSING A MODEL

There are several types of economic and accounting models that can be used to analyze the costs and benefits of introducing ergonomics. *Economic* cost–benefit modelling seeks to capture the total cost and benefit of the ergonomic intervention whereas *accounting* models have a more limited scope. For example, traditional accounting systems do not consider the skills of workers to be of monetary value to the company. Therefore spending money to train workers is seen as a cost that cannot be offset.

Accounting seeks to capture only those factors that are seen to directly influence the firm's profit and loss.

A typical accounting model is termed "Return on Investment" (ROI) and is calculated in monetary terms over a defined period:

ROI = income stream produced by investment/cost of
　　　 investment.

The more complex economic cost–benefit models place a dollar value on all aspects (direct and indirect) of cost and benefit. However, in ergonomics it is often impracticable to quantify all the indirect benefits expected from its introduction and thus assumptions must be made. This is particularly the case for intangible factors such as:

- improvements in worker morale;
- reductions in pain and suffering due to fewer or less severe injuries; and
- reductions in usage of health services.

Without capturing the indirect benefits of the ergonomic intervention, the cost–benefit analysis will underestimate total benefits. Other economic models (e.g. costs saved model) do not attempt to quantify these indirect benefits, deeming the risk of over or under estimation to be too great. For instance, where it is anticipated that the ergonomic intervention will save lives, a monetary value for a human life is rarely agreed to by those involved.

Accounting models and economic models both seek to capture costs and benefits over the lifetime of the intervention. This prediction into the future must, of necessity, carry with it a number of assumptions. This can be seen by examining the costs saved model.

The *costs saved* model is one of the simplest cost–benefit models. It is often used where it is not feasible to estimate hidden costs or anticipate fully the benefits flowing from an intervention. The model just examines changes in cost items including costs saved. For a single year this can be calculated as:

$$TC_1 - TC_2,$$

where

TC_1 = current costs
TC_2 = predicted costs after implementation of ergonomics
　　　　 intervention.

A project is considered worthwhile where $TC_1 - TC_2 = 0$.

Once the costs or benefits occur over a period of several years the calculation becomes:

$$\hat{A}\ (e_1 - c_1) + (e_2 - c_2) + (e_3 - c_3) \ldots + (e_n - c_n),$$

where

e_1 = 　present value of the total costs for the ergonomic
　　　　 intervention in the first year
c_1 = 　present value of the total costs for the current system in
　　　　 the first year
n = 　　number of years it will take to pay for the intervention.

The further into the future the project extends, the more uncertainty will be the cost structure and assumptions will need to be made about the *most likely* outcome. Because the value of money changes over time (due to inflation rate, interest rates) it is necessary to make further assumptions about the ways its value will change and build them into the model. This is termed determining the present value of future costs. In a cost–benefit model this is achieved by applying a discount rate to future events.

One of the key assumptions of rational economic theory is that firms will prefer to have a return on their investment today rather than to wait some time for uncertain future benefits; obtaining a return on investment this year is considered to be of greater value than obtaining a return on the investment next year. To reflect this preference for the present, returns on investment received in the future may be discounted by a greater percentage than the anticipated inflation rate. (The percentage chosen is referred to as the *discount rate*).

Thus, the discount rate chosen will reflect not only anticipated

inflation rates but also the firm's culture and values, including their concern about risk. Discount rates and can have a significant impact on the final calculation of cost–benefit. For example, if a firm spends $100 this year to get a return of $110 (110%) this year then the project is clearly worthwhile; if the return of $110 did not occur until the next year then a discount rate would be applied. If the firm used a discount rate of 4% then the $110 return would effectively be worth ($110 – 4%) $105.60; this is still greater than the original investment so the intervention would be considered worthwhile. If the firm used a discount rate of 11% then the $110 return would effectively only be worth ($110 – 11%) $97.90; as this is less than the original investment the intervention would not be considered worthwhile.

As the cost–benefit model becomes increasingly complex it also becomes less and less precise.

A firm may find it easy to calculate the direct costs of a production process, but the indirect costs such as overheads, wages and on-costs may be more difficult to calculate and may need to be estimated. For example, future labor costs require assumptions, which include:

- changes in the labor turnover rate;
- payment to replace staff;
- wages increases; and
- changes in the number of injuries and/or the severity of injuries?

It is *critical* that the assumptions chosen for the cost–benefit model appear reasonable to the management and specifically those persons who will make a decision on funding. In this way there will be agreement that the model gives the *best estimate* for the likely return on investment.

Cost–benefit modeling over multiple year periods are needed where very large and complex interventions are proposed and the intervention takes a considerable time to implement. However, for the majority of ergonomic interventions it is practical to assume that costs are all incurred over a 1-year period and to offset costs against one year of benefits. This simplifies the modeling with many advantages; fewer assumptions are required, the mathematics is more transparent and, with less explanation required, the focus stays on the ergonomics intervention itself.

Unfortunately there is a dearth of models designed specifically to address working conditions. Economic models exist to describe engineering, fiscal and other functions but not for workplace people. Here the Productivity Model will be used to show the general principles of economic models which describe working conditions. The basic assumptions for this model cover:

- *employee working conditions* — namely working hours and productive working hours (paid time at work, excluding paid time for vacation, illness, etc.), wages/salary and on-costs (supervision and administration, workers compensation insurance and taxes, training, recruitment, etc.), overtime and over-employment.
- *productivity factors* — namely quality and quantity of output, error and warranty costs.

3. DATA REQUIRED

From a pool of assumptions, those applicable to the particular ergonomics intervention can be agreed upon and data collected.

Table 1. Cost indicators and how an ergonomics intervention may affect these indicators.

Cost factor	Loss from a calculated 100% efficiency (present situation)	Future value (forecast range for a reduced loss after the ergonomic intervention)	
		minimum achievement	maximum achievement
Example 1.			
Unsuitable equipment	90%	95%	98%

If the loss of 10% in present efficiency is due to unsuitable equipment the gain due to the ergonomics intervention will be in labor costs. If labor costs are $20 per working hour then a gain of 5% in efficiency (the minimum forecast) is $1 per hour or about $1600 per worker per year. If the intervention costs less than this, then the pay-back period will be less than one year.

| **Example 2.** | | | |
| Injury absenteeism | 5 days per employee per year | 2 days per employee per year | 4 days per employee per year |

The productive time is the net working time for which the employee is paid. By reducing avoidable absences, for which wages/salary are still paid, the productive time is increased leading to cost savings. In this example, the ergonomics intervention aims to reduce the injury absenteeism from 5 days per employee per year to between 2 or 3 days per employee per year.

This is a straightforward calculation based on the cost of daily employment multiplied by the number of people involved; this value is compared with the cost of the ergonomics intervention required to achieve this reduction in absenteeism.

| **Example 3.** | | | |
| Reduction of waste | 10% waste of raw materials | 7% waste of raw materials | 5% waste of raw materials |

Waste is the loss of materials during production. The ergonomics intervention seeks to reduce the waste from 10% to between 7% and 5%. The intervention will be financially successful if the savings in waste is greater than the cost of the intervention. In this example there may be extra gains to be made in less energy use, smaller stocks required and less handling of materials.

| **Example 4.** | | | |
| High warranty costs due to poor quality and/or errors | 5% value of produce | 3% value of produce | 2% value of produce |

The ergonomics intervention seeks to improve labor and management skills to overcome the quality deficits and the cost of repairing returned goods. Gains will be in a reduction in labor costs, reduced waste and reduced energy use.

Some of the data required are straightforward and easy to obtain. These include:

- labor items (productive time, wages, on-costs, training, absenteeism, labor turnover, etc.);
- equipment and material costs (equipment failures, waste, errors, etc.); and
- organizational costs (supervisory structure, manning levels, etc.).

Other data are less easy to obtain but, in the experience of the authors, give the greatest return from ergonomics investment. These include:

- productivity (gross output and quality); and
- error and warranty costs.

Where does one get the data? Some can be obtained from the accountant (wages and similar items), some from management (production, waste, errors and warranty costs) and some from the directly involved workers (workplace layout, unsuitable or worn-out equipment).

To assign costs, one can either use the raw figures (wages, absenteeism, cost of labor, value of production, etc.) or one can use a difference value. As the model is a difference model (the difference between the present and an estimated future date) difference values can be based on, for example, labor costs. For example, the ergonomic intervention may be aimed at reducing the time taken to do a particular task; the change would be the difference in time taken between the two tasks.

The present costs are added together and the envisaged reduced costs, due to the ergonomics intervention, subtracted; this value is the "estimated benefit". The total cost of the intervention is calculated (new machinery and installation costs, ergonomics consultant, extra management time, etc.) and is the "cost for improvements."

If one makes the additional assumption that these cost savings occur in one year and that the intervention occurs over one year then the payback period (in years) is:

cost for improvements/estimated benefits.

Table 1 illustrates examples of cost indicators and a forecast of how an ergonomics intervention may affect these indicators.

DEFINITIONS

- *Direct costs* — wages/salary paid to employees, obligatory and other costs (tax, insurance, retirement fund, etc.).
- *Indirect costs* — costs involved in the production of goods or delivery of services excluding the direct costs. It can include supervision, administration and similar personnel costs.
- *Hidden costs* — injury and production/service losses built into the enterprise structure but which are not separately identified.
- *Direct benefit* — identified savings in direct costs and improvements in productivity and/or quality.
- *Productive time* — paid work time used for the production of goods or delivery of a service. It excludes paid time that is not used productively e.g. illness and injury absence, vacations.

REFERENCES

OXENBURGH, M.S., 1991, *Increasing Productivity and Profit through Health & Safety* (Sydney: CCH international).

OXENBURGH, M.S., 1997, Cost–benefit analysis of ergonomics programs. *American Industrial Hygiene Association Journal*, 58, 150–6.

OXENBURGH, M.S. and FENESTRA SOFTWARE, 1993, The Productivity Model (computer software) (Sydney: author); e-mail: maurice_oxenburgh@compuserve.com

Education: The Teaching of Ergonomics

A. Woodcock†, M. D. Galer Flyte‡ and H.G. Denton§

†Design Research Centre, Colour and Imaging Institute, University of Derby, Derby DE22 3HL, UK

‡Department of Human Sciences and §Department of Design and Technology, Loughborough University, Loughborough LE11 3TU, UK

1. INTRODUCTION

The manner and level to which ergonomics (or human factors) is taught is important for the progression of the discipline itself, its furtherance among other professions, and the influence it is likely to have on globalization and technological change. Here we briefly indicate the benefits of the inclusion of ergonomics in design practice, review some of the traditional barriers to communication of ergonomics concepts, and lastly consider the manner in which ergonomics may be taught in tertiary, secondary and primary education.

There are two distinct areas of ergonomics that need to be considered:

- Education in the discipline itself, i.e. the *mastery of skills* such as usability assessment, workplace evaluation, the measurement and recommendation of ways in which job related (mental and/or physical) stress may be alleviated. Education to this level will enable one to *practice* ergonomics as a profession, and is traditionally taught at undergraduate and postgraduate level (see below).
- Education for *awareness* of ergonomic factors. Engineers and designers, for example, need to be aware of the function of ergonomics in relation to their own disciplines. In addition we need to consider general education as members of an increasingly complex, large and varied social environment. This is growing in importance as individuals from widely disparate backgrounds can now work together, sometimes without meeting face to face, using ever more rapid, effective and efficient means of communication. Effective working in such an environment requires individuals to understand and be tolerant of others, and to know how to work as a team. Ergonomics, with its emphasis on the importance of human factors and individual differences, can provide a means of unifying diverse curricular activities that might actually help children and adults work together in larger systems.

2. ERGONOMICS IN PRACTICE

To understand the importance of education in relation to ergonomics, it is necessary to appreciate the practice of ergonomics in relation to the disciplines it seeks to inform, namely engineering and design. Well-designed products must be safe, efficient, comfortable and convenient to use, durable, realistically priced, have a pleasing appearance and be pleasurable to use. All of these are human factors issues.

Rapid technological development and the movement towards concurrency in design require that a systematic account of human factors is incorporated as early as possible in the design life cycle i.e. to predict in advance of build who the users are, what their reactions are going to be, and where usability issues will arise.

The complexity of many systems means that it is often impractical or excessively costly to make changes late in the design process. Incorporating ergonomics early into the product design life-cycle has proven benefits in reducing problems with usage, increasing the acceptance and functionality of the finished product and reducing the need for highly expensive re-design.

Attempts to enhance the human factors input into design and engineering have focused on both the product (i.e. ergonomics information) and the process (i.e. the way in which ergonomics can be incorporated into the design process). "Product initiatives" include improving the accessibility of ergonomics information by presenting it in a more designer friendly manner, e.g. in web pages, CD-ROMs, guidelines and research papers which include practical recommendations. "Process initiatives" include the development of tools and techniques tailored to specific design tasks, such as different methods for capturing and communicating user requirements, e.g. through storyboarding, simulation and scenario building and looking at the manner in which this information can be fed into the design process e.g. through the development of ergonomics brainstorming decision support systems (such as the HUFIT Planning Analysis and Specification Toolset and ADECT). Anecdotal evidence would suggest that these initiatives are having an effect.

However, to improve the effectiveness of the discipline one still requires education to:

- train future ergonomics experts in industry and academia; the *practitioners*;
- provide those who are not "ergonomics experts" with an *awareness* of the human issues they should be considering; and
- provide a working environment favorable to the uptake of ergonomics issues.

Below, are the ways in which this is taking place.

3. POSTGRADUATE VOCATIONAL TRAINING

As indicated above, designers and engineers need to consider the human element in their designs at the start of the design process. However, most will have been trained on degree programs that did not consider ergonomics. To some extent an engineer or designer can develop an awareness of the value of ergonomics to their work through the normal process of professional reading. In competitive, niche markets this approach is now often too hazardous. This shortfall in expertise is increasingly being met by postgraduate vocational ergonomics courses, which may form part of a larger post-vocational qualification such as an MSc. Many organizations, especially those in engineering, invest in vocational training to upgrade their work force and to keep their knowledge and skills relevant.

One example of a postgraduate vocational ergonomics course is offered by Loughborough University. Short courses on vehicle ergonomics have been run for the automotive industry since 1981. These are offered either as in-house courses tailored to the specific needs of the company or as open courses attended by people from a variety of companies. Within the automotive industry there is a need for ergonomics training applied to both the manufacturing process and to the design and development of the product.

An example of a postgraduate module at Loughborough is

Vehicle Ergonomics, part of an MSc aimed at engineers. This 5-day residential module includes an introductory lecture on ergonomics and its application to vehicle design and the integration of ergonomics into the design process from concept to customer. Methods used in vehicle ergonomics include the identification, specification and implementation of user requirements, evaluation, the role of simulation, packaging trials, prototyping tools and road trials. Anthropometry in vehicle design is taught as a theoretical and a practical topic including the importance of anthropometry and biomechanics to vehicle design and the application of anthropometric and biomechanics data. It is considered important to give the students an understanding of the methods used to reach ergonomics solutions for themselves rather than being given only ready made solutions. The ergonomics of control and display design covers selection and location of controls and displays, methods of communication, interface and dialogue design and new technology applications. Seat design and the driving package indicate an application of anthropometry and biomechanics and the importance of integrating the driver's workplace into the whole vehicle. A major component of the module is a practical project involving the design and conduct of a user trial with "real" users/customers to tackle a specific aspect of vehicle design such as access and egress or seat comfort.

By the end of the module two things have happened. First, students report that they never before realized how pervasive a topic ergonomics is. Second is a dramatic change in attitude after the project work that involves direct interaction with members of the driving public (their customers and the people they design the vehicles for). The students become very aware that their customers are not like them. They are not all fit healthy young men (only a few students have been women) who are interested in and know about cars. The post-module evaluation often reflects a realization that ergonomics is much more relevant to their work than they originally expected. This is not an isolated example, other organizations also recognize the need for ergonomics training for their work force and courses are provided either by educational establishments or by in-house ergonomists.

It should be noted, however, that concern ergonomists once had that such training would make their role redundant as the designers and engineers would be able to "do the ergonomics" for themselves has disappeared. There is strong evidence that the demand for specialist ergonomics input to design and engineering activities within a company escalates as the designers and engineers recognize the human factors issues for themselves, the extent (or limitations) of their own knowledge, the specialist input the ergonomist can then provide, and the ultimate effect on the end product and customer satisfaction.

4. ERGONOMICS WITHIN TERTIARY EDUCATION

Here we look at ergonomics as taught in UK universities as either a first degree, postgraduate degree, or as part of a degree. Before going into detail we should recognize two areas that have caused difficulty: Firstly there has been a barrier between ergonomics and engineering due to the limited common "language" and the philosophical stances of the disciplines. Secondly, specialist ergonomics degrees tend to recruit students from both a science or design background. This means that courses need to be sensitive to the different backgrounds of the students. This has

not always been the case; e.g. Lombaers (1990) criticized courses that failed to produce well-rounded team players. Students with a science background tended to have little knowledge of product development and had difficulty with the compromising, integrative nature of the design process, whilst those looking at ergonomics from a design background tend to have little knowledge of how to conduct experiments. Recommendations have included:

- Widening the scope of engineering education to include modules in ergonomics, in which the relevance of ergonomics to engineering is outlined and ergonomics information and processes are provided.
- Ergonomics courses offered to all business and technology (and engineering) students with a focus on enabling them to conduct their own human factors investigations.
- Postgraduate level modules in ergonomics (or vocational courses) for those destined for senior management. It is graduates at this level who will be making key decisions.
- Joint degree courses. Besides acting as a scientist, the ergonomist should play a role as a contributing specialized designer with an understanding of the whole process that is being supported. This requires knowledge not only of ergonomics, but ergonomics as it is practiced in a particular industry, for example, product, industrial or automotive design. This can possibly be achieved through modularization and joint degree courses. This has the added benefit of not only having ergonomics applied to the problems faced by a sector of the workplace, but in reducing interdisciplinary communication problems, which all too frequently inhibit co-operative working, and in helping to set joint goals and foster environments in which ergonomics can be used more effectively, especially during the crucial early design stages.
- Training researchers and future teachers as a means of guaranteeing the status of ergonomics among other sciences and humanities and the manner in which it is applied by practitioners.

Some of these are issues that the ergonomics community has taken on board, but some are of a more general nature which can only be addressed through sweeping reforms to the school based approach to tertiary education. Below we look at the manner in which ergonomics is taught at degree level and also how it may be taught as a series of modules in, for example, a design degree. It has been argued (Woodcock and Galer Flyte 1998) that primary and secondary education also need to be considered as possible avenues for making potential students receptive to ergonomics issues. This is reviewed below.

4.1. As an Under- or Postgraduate Degree

Ergonomics, as taught in the UK, may be taken at diploma, degree or masters levels, either as degree in its own right, in conjunction with a related discipline, or as specific modules (options). Here we provide examples of each of these types of courses. For more details of these and courses offered in other countries, see, in the first instance, the Ergonomics Society [http://www.ergonomics.org.uk] or the International Association of Ergonomics [http://ergonomics-iea.org].

The courses, outlined below, have different areas of specialization but meet the requirements of their professional

bodies. These require that students demonstrate knowledge of the principles of ergonomics, and of the relevant human characteristics in the areas of anatomy, physiology, psychology and social organization. It also requires knowledge of how the physical environment affects people and of at least three specialist areas of application selected by the candidate. It requires also evidence that the candidate is able to use statistics, experimental designs, equipment and methods to investigate, modify or design situations an equipment for ergonomic benefits (Centre for Registration of European Ergonomists (CREE) information leaflet). A similar, and much fuller description of the skills and knowledge bases which need to be addressed can be found in the Board of Certification in Professional Ergonomics Handbook (1999).

It should be noted that the professional bodies have different categories of membership. For example, to use the designation, European Ergonomist (EurErg), requires a demonstration of full responsibility in the use and application of ergonomics knowledge and methods in practical situations for at least two years following training.

4.1.1 Masters (MSc) level courses

These are offered by five UK universities. Three offer the course as an MSc: Ergonomics, one offers Work Design and Ergonomics and another a more specialist course in Health Ergonomics. The aims of these programs are to equip students with the requisite ergonomics knowledge needed for them to become professionals recognized by the appropriate professional body for ergonomists (see above), and the necessary expertise to become effective practitioners.

The courses typically run over 12 months (full time) and 24 months (part-time). Assessment is through coursework (normally including a three month research project) and examinations. All courses are strongly supported by industries at local, national and international level, which provide guest visits and research projects for students.

Although each of the courses has its own particular emphasis they all cover basic training needed by a practicing ergonomist. For example, one course offered as a joint program is suited for those who wish to work in an industrial environment, another specializes in the application of ergonomics to problems that occur in health systems such as accidents, ill health and job satisfaction, a third emphasizes the practical applications of ergonomics in for example manufacturing, the health service, computer systems and North Sea oil and gas industries.

As these are masters level courses, the students are likely to be practitioners in a specific industry (e.g. health service, automotive engineering), or have a relevant prior degree perhaps in psychology, physiology or sports science. They will therefore be adding ergonomics competencies to their core professional skills.

4.1.2. Undergraduate degree (BSc) in ergonomics

In the UK only Loughborough University offers a BSc in Ergonomics. This can be taken as a three year undergraduate degree. If this is supplemented by a year working in industry the student will also be awarded a Diploma in Professional Practice. The program fulfills the requirements for full membership of the Ergonomics Society and meets the technical requirements for accreditation as a European Ergonomist.

In terms of the structure of the degree, the first year covers "introductory modules" such as ergonomics and design, anatomy and physiology and basic cognitive psychology; the second year consists of more specialized "degree level" modules such as occupational psychology, thermal environment, organizational behavior and professional practice. For those opting for a four year courses (to gain the additional Diploma in Professional Practice) the third year is a placement one, where the student is in paid employment in a field relevant to ergonomics. This helps to relate academic studies to practical situations and help students to gain an insight into the responsibilities associated with the profession. A close liaison with the university is maintained with the students having both a professional and academic supervisor. Students opting to take this additional year often perform better in the final year and obtain employment more readily after graduation. The final year consists of core subjects such as project work, systems ergonomics and psychology, and optional, more specialized modules such as applied vision, transport safety and human computer interaction.

Discrete areas of ergonomics may also be offered by universities as part of standard joint honors degrees such as Psychology with Occupational Psychology or as modular options. The modularization of degree courses enables students to tailor (within reason) their degree to suit their interests and requirements. For example, modules might occupational, cognitive and environmental ergonomics. These can then be used as an area of specialization when accompanied with core modules, for example in engineering or design. The modular approach to teaching ergonomics is outlined below.

4.2. As a Module in Other Courses

In this section we take as an example the way in which ergonomics can be related to product design degrees, similar approaches may be applied to engineering oriented courses.

As with the incorporation of ergonomics into the product life cycle so it is with incorporating ergonomics into design education: the earlier it is done the more effective and acceptable the outcome. As part of their design education, would-be designers (especially those specializing in industrial design) receive some introduction to ergonomics. Usually this is in the form of a series of lectures or a complete module in ergonomics. At Loughborough University, for example, industrial design students attend a complete module on Ergonomics and Design.

The module syllabus includes both lectures and practical work in which students apply ergonomics principles to the design of everyday objects. The lectures cover topics such as sources of ergonomics information and data and the appropriate application of that data design of displays and controls and their application in a control panel design exercise console and workplace design, tool design, anthropometry and other methods appropriate for workstation design and evaluation, seating, the effect of poor design and musculoskeletal complaints, and methods and tools which can be used to answer new design queries. Other related optional modules are also available such as ergonomics and product design, design of advanced technology systems, and ergonomics and transport.

Ergonomics may be taught by people who are not formally trained in ergonomics but have taken an interest in the subject or by qualified ergonomists either through "service" teaching

where the specialist is not a member of the design department, or by the growing number of ergonomists who are actually part of the design department's complement of staff or by guest lectures from practicing designers or ergonomists. Guest lectures provide a valuable addition to the students' experience in terms of case studies of good ergonomics in practice or examples of how an attention to ergonomics earlier in the design process could have avoided problems such as re-tooling or customer complaints later on. They are also valuable in putting ergonomics firmly into the context of design in real world practice.

The nature of the ergonomics material taught to design students is critically dependent on the approach taken by the person responsible for the subject as there is no agreed content. In the past, and still to some extent at present, ergonomics has primarily been seen and taught as a combination of anthropometry and "knobs and dials." In this approach, which is certainly better than no ergonomics at all, the student designer is made aware of existing sources of ergonomics information (or data) such as Pheasant, *Bodyspace: Anthropometry, Ergonomics and Design* (1986, 1996); Woodson, *Human Factors Design Handbook* (1981) or Sanders and McCormick, *Human Factors in Engineering and Design* (1987) and is urged to refer to these during their design. This encourages the student designer to take account of known ergonomics information particularly physical characteristics such as shape, size, strength and reach.

However, this existing information is not always exactly the information the designer requires for the particular task in hand. For example, reach data may not provide the specific dimension needed for the innovative product under development, or the product may be for a market for which there is only limited ergonomics information on the characteristics and lifestyles of the potential product user population. There is, therefore, a growing trend for student designers to be provided with tools to enable them to gather information for themselves. Design textbooks, such as Baxter, *Practical Methods for the Systematic Development of New Products* (1995), include sections on task analysis and other ergonomics methods. Garner (1991) has produced a book on ergonomics, *Human Factors*, specifically aimed at "A" level and undergraduate design students.

Where ergonomists are part of the design education department their experience in communicating ergonomics to student designers tends to be one of gradual change in emphasis over time. Initially the students simply want to be provided with ergonomics information but gradually, through encouragement, become enthusiastic about a more user-centered design approach. In user-centered design, the potential user(s) of the product is the central focus of attention. This includes using ergonomics methods to more readily understand actual user requirements as well as their physical characteristics and limitations as it not only includes the shapes and sizes, reach and strengths of the range of users, but also requires consideration of the task(s) the product is to support and the environment in which the product is to be used. The environment could be the physical one, for example, the lighting available for seeing the product to use it, the noise environment, for example, to be able to hear a radio in an HGV; it could also be the psychological or organizational environment in which the product is used. What, for example, are the user's expectations for the product based on their previous experience of similar products? Does the organizational environment in

which the product is to be used demand that it has a 'privacy' function for instances where confidential work could be overlooked? It is no surprise, therefore, that all this information is not readily available, yet the failure to provide this specific information has been one of the criticisms continually leveled at the discipline.

Once this latter approach has taken hold on the designer's imagination and practice their own requirements of ergonomics change from simply requiring look up tables, to knowing what questions to ask and how to relate user, task and environmental characteristics to product requirements. This is the basis of user-centered design. Information on the "how to" aspects of using ergonomics methods and tools is still not readily available to design students. Wilson and Corlett's *Evaluation of Human Work: A Practical Guide to Ergonomics Methodology* (1990) covers a great deal more than methods for use in product design, but is primarily aimed at ergonomists who have the expertise to identify the most appropriate method(s) for their particular application. This is where access to an ergonomist in the design education department is of considerable benefit.

Teaching methods employed include lectures on ergonomics principles and information, and individual or group assignments. Typically assignments require evaluation and comparisons e.g. "to evaluate a domestic product such as a kettle for ease of use by the elderly" or "to compare the ease of assembly of two brands of kitchen unit." In completing such assignments students are encouraged to take a user-centered design approach, considering the characteristics of the user(s), the tasks the product is to support and the conditions of use. They are also encouraged to identify design deficiencies as well as the good aspects of the design that should be preserved, and suggest improvements to enhance the product's usability. This encourages a positive as well as a critical approach as it is recognized that ergonomics does have a reputation (now happily diminishing) of only providing designers with negative feedback after product evaluations. To undertake projects of this nature, students require specific assistance with the methods used, such as the design of the evaluation procedure or of a questionnaire.

These assignments are often carried out in groups with the interchange of ideas among the students being seen as a valuable and crucial learning experience, given the nature of the employment they will be seeking after graduation. As the students progress through their design course they undertake a number of design projects, increasing in size and complexity. These design projects differ from the purely ergonomics related assignments described above in that they are primarily focused on product design. In these, considerations of ergonomics have to take place alongside other factors that will influence the final product, such as materials, production process and costs. This is seen as more typical of actual design practice. In these projects, the lecturer tends to act as facilitator, usually on a one-to-one basis with the student. Their role is ensuring that the student takes adequate account of the product's user requirements, refers to existing data or information (so no wheels are re-invented), collects additional user information as needed and undertakes user trials with product prototypes to establish how well potential design solutions meet user requirements. Most importantly, they enable the student to deal adequately with the compromises that will inevitably need to be made in the product development process.

It is only by having a clear understanding of the real requirements of the product users that satisfactory compromises can be made in this process.

The ergonomics input to design education can be seen, therefore, to take two principal forms, the introduction of students to sources of data and information, and the support for user-centered design as part of the whole design process. It is the combination of the provision of ergonomics information and the methods and tools to enable the design student to discover or generate appropriate information for their particular design activity that is the most powerful in design education. It is often only when the designer finds him or herself in a competitive industry that is awake to the market edge that good ergonomics can bring to a product that they see the real relevance of this subject in their design education.

5. ERGONOMICS IN SECONDARY EDUCATION (11–18 YEAR OLDS)

Education is a key factor in the development of any society and should be seen as a continuum, stretching from nursery school, through further and higher education, into the work place and retirement. We see two basic roles for the teaching of ergonomics at school level: general awareness and pre-vocational. General awareness of ergonomic factors is important to us all, if only as users and selectors of products, systems and environments. By understanding something of ergonomics we can become more discriminating consumers and we can set up our working and leisure spaces to greater effect. Going further, an awareness of ergonomics at the critical 16–18 age range can be important in relation to students selecting degree programs in areas where ergonomics are important or possibly choosing to focus on ergonomics as a specific area of study at degree level.

In this section we look at:
- the role ergonomics can play in a curriculum which will prepare learners to take their place in a global society; and
- the extent to which ergonomics is taught as a discipline in its own right.

5.1. Potential for Ergonomics in the Curriculum

Although the introduction of topics in the school curriculum is too slow an agent for change in a modern society, the school phase is critical. In the UK, the National Curriculum (NC) was designed to ensure that all children receive a broad basic education in which they firstly develop an awareness of, and subsequently a capability in specific areas of learning. The two are not, of course, exclusive and are normally seen as part of a continuum where both grow to levels appropriate for the program involved. Here we will review three areas which we believe to be pertinent to education and ergonomics, in terms of curriculum development; namely, the role of information technology, learning about individual differences and team building.

5.1.1. Information Technology (IT)

A review of the current National Curriculum for the UK has acknowledged the vital role of IT awareness. Pupils are required to be ready to respond to the new opportunities and problems that they are going to encounter. These may be summarized as the expansion of communication technologies, changing modes of employment, and new work and leisure patterns resulting from economic migration and the continued globalization of the economy and society.

This may be taken as evidence that general education is becoming inextricably associated with advanced communication technologies such as the web. Children, classes and schools are producing web pages, publishing poems, art work and class experiments, with a view to sharing experiences with their peers in other countries. The only limits to this growth are teachers' ability to guide and integrate these ventures to make them worthwhile learning experiences, and the high cost of the technology involved.

The use of IT in this manner, does indeed open up new vistas for schoolchildren, and should, if fully integrated into the curriculum provide them with greater insights into environmental and cultural differences and similarities. Whereas this has been possible through the use of television and multimedia images, the Internet provides children with the potential to initiate their own interactions. Ergonomics could be used to provide a framework to guide these interactions.

5.1.2. Individual differences

More critically, the proposed changes to the National Curriculum are required to promote equal opportunities and enable pupils to challenge discrimination and stereotyping. This raises a number of issues. First, we can see *general educational* issues associated with using advanced communication technologies such as the web, the globalization of work / leisure and the need to educate children to be sensitive to differences whether they be due to race, gender, religion or ability. Secondly we see *pre-vocational* issues in terms of preparing children for the next phase of training.

The teacher can exploit this diversity as a rich learning environment. In this way ergonomics could be seen to provide uniting glue to disparate projects, in a variety of subjects, as children are taught tolerance and understanding of individual differences at earlier stages in the curriculum.

For example, in the Fostering a Community of Learners approach, Brown and Campione (1996) the teacher acts as a facilitator to children's learning. Staff draw upon the expertise of the students (thereby actively recognizing and valuing diversity) and the wider community. The three main aspects of this are that the children research, share and perform on topic work, so mutual learning, sharing and a division of labor may develop in each group. This implicitly recognizes that individuals possess different skills, that individuals can learn from each other and that these skills can be employed to maximum in effective teamworking

5.1.3. Team working

The ability to work independently and in teams is recognized as a general educational objective from the early teaching of 7–11 year olds, through to graduate programs, for example in engineering. Bertodo (1994) emphasized the personal qualities working in such an environment would require — a breadth of knowledge and understanding, a team player, good communicator and problem solver. These abilities he felt, rank above the more traditional engineering competencies of inventiveness, creativity, rationality and thoroughness. However there are indications that not all school or university staff understand how to organize team-based learning and, particularly

the ergonomics of team working, whether in one team base or remotely using computer equipment.

Developing a team culture and having the ability to rapidly understand and adapt to other organizational structures must be a high priority in research and education, as failure to do so creates barriers to effective international projects. Again, the ability to work and organize team activity is laid down in school, where children work with individuals and groups in their class, across classes for larger joint initiatives, and through the school, e.g. for school plays, and mentoring initiatives, where older children help younger ones with reading. This approach needs to be maintained into tertiary education and through to the workplace.

5.2. Teaching of the Discipline of Ergonomics in Secondary Education

We continue the "reverse engineering" approach to the teaching and learning of ergonomics by looking firstly at the 16–18 age range (mainly "A" level, but also a range of pre-vocational courses such as GNVQs), the majority of whom will go onto tertiary education. We then look at the 11–16 age range that precedes it and, below the primary (5–11) age range.

5.2.1. Advanced ("A") level

The national curriculum in the UK only applies up to age 16. This means that post-16 courses do not have to meet formal requirements in the way they do at 11–16. In the UK, at the time of writing the norm is still very much the traditional three or four "A" levels route despite the fact that it has been heavily criticized in relation to broader approaches such as the International Baccaluriat or Scottish "Highers." Typically students intending to study engineering at degree level in the UK will take "A" levels in Maths, Physics and a third subject such as Further Maths, Geography, etc.

Only in Design and Technology "A" levels is any study of ergonomics made a requirement. However, little detail is given to guide staff by the syllabii produced by examination boards. Currently the Qualifications and Curriculum Authority (QCA) is conducting a major review of national curriculum requirements and qualifications such as "A" level. Their guidelines for "A" level subjects [http://www.qca.org.uk/] indicate that all new "A" level Design and Technology syllabi should also include a section on "planning and evaluating." This will include the use of databases, drawing and publishing and design software with a view to interpreting design data in terms of the properties of materials, ergonomics and nutritional information. Also when the student is developing and communicating ideas, they are required to take into account functionality, aesthetics, ergonomics, maintainability, quality and user preferences.

"A" level science options may occasionally include aspects of human factors. For example "medical physics" options in which the student might look at forces within the spine when a person picks up weights in various postures. Similarly a mathematics teacher may use ergonomics as a basis for project work in statistics at "A" level. The key point is, however, that there is no requirement for this.

A recent survey by Denton and Woodcock (1999) showed that in a sample of over 300 undergraduates on engineering and design programs in a UK university, a large proportion had no experience of ergonomics in schools as 16–18 year olds, unless taking Design and Technology at "A" level. So for many their understanding of ergonomics was very limited.

The overall position showed by the survey was that a student with "A" levels in mathematics or a science may have looked at ergonomics, but it is more likely that they will not. A student with "A" level Design and Technology should have learned something about ergonomics and should have applied this knowledge to project work. However, the older teachers of Design and Technology will not have learned any ergonomics as part of their initial training. Students are dependent on these staff having read up on the subject when it entered these syllabi in the 1970s; not all can be relied upon to have done this. Individually, students may have covered aspects of ergonomics in independent study or where ergonomic data are used to illustrate a certain point in class. It was encouraging to note that a small amount of ergonomics was being taught (and remembered) in other disciplines.

5.2.2. General Certificate of Secondary Education ("GCSE") level

All children are required by the national curriculum to take Design and Technology (in UK) up to the age of 16 and most take a GCSE examination on a subject within that area. The national curriculum (1995 edn) does not make the study of ergonomics explicit in any subject. It does, however, require teachers of Design and Technology to help students to consider "the needs and values of intended users" (*National Curriculum, Design and Technology*, p. 6, para. 3d) or "ensure that the quality of their products is suitable for intended users" (p. 11, para. 3i).

GCSE syllabi tend to be a little more explicit in relation to ergonomics teaching. An example requires candidates to identify and collect data relevant to the product(s) and its users: dimensions, ergonomic and anthropometric data. The survey by Denton and Woodcock (1999) confirmed that up to the age of 16, the majority should be receiving some education on basic ergonomics if design and technology teachers were interpreting the national curriculum and GCSE requirements correctly. There was also evidence that some teachers in science, PE, dance, mathematics and humanities (notably geography) also raised ergonomics in their teaching as well as in the more design related courses. However, this was not the norm, and comprised a very small proportion of the teaching in those disciplines.

Work at "GCSE" level typically involves an understanding of ergonomic principles, and the ability to assess appropriate data (at a very simple level). Children become aware of the differences in people and that products can be designed to suit certain portions of a population and the contribution of environmental factors (such as noise and light) to the product gestalt. Practical exercises can contribute to learning ergonomic principles and developing some capability in applying these principles. In terms of general education (up to 16) we can see potential benefits if children are taught to be aware of aspects of ergonomics. This could increase children's sensitivity to the differences in people, their needs and abilities possibly improving understanding, empathy and relationships between peoples. This awareness could also enable them to become more discriminating purchasers of products or systems in relation to their actual use. This might also act as a market-place lever, forcing manufacturers to design for greater inclusion of users in the first instance.

Crucial to supporting the use of ergonomics in the later Key Stages is the need to develop resource material to support teachers. Such support material could, for instance be in the form of:

- Design projects; where children are encouraged to deconstruct the problem, starting with a consideration of ergonomic problems which could lead to the development of different design solutions, e.g. for forms of seating. These could then be evaluated and compared with the original solutions, and displayed on the Internet for other children to comment on.
- The development of curriculum material on ergonomics which matches the needs, requirements and interests of its users (both staff and students).
- The application of ergonomics tools and techniques (e.g. focus groups, interviews, anthropometric measurements) which take into account the knowledge and skills which need to be acquired at each Key Stage.

6. ERGONOMICS IN PRIMARY EDUCATION (5–11 YEAR OLDS)

Baynes (1992) considers that young children start school already able to design and make, through their everyday interactions with the world. Part of this ability is a concern for users in their designs (whether they be LEGO "people", real or pretend playmates). This might be relatively simplistic, e.g. making sure that the LEGO "person" fits into the house, that the layout of an emergency hospital is satisfactory for both patient and doctor. From a very early are children are aware of what it is like to be excluded from parts of the world; interesting items are placed out of their reach, jars are designed specifically to prevent them from accessing the contents.

Kimbell *et al.* (1996) observed that at Key Stage 1 (age 5–7) children handle user issues "more than at any other Key Stage. They easily empathize with the users of their designs." They are indeed "natural born ergonomists." However, with the progression from primary to secondary school he found a marked change in the way in which user issues are tackled, with users being far more important at Key Stage 2 (age 7–11) than Key Stage 3 (age 11–14). This change in emphasis towards the need to be able to actuate one's creations, rather than take into account and reflect upon user issues in the design, appears to be reinforced through the educational system, and may be one of the reasons of the failure of the discipline to make an impact in design and engineering. A structured understanding of the design activity adopted by young children is necessary for educators and resource developers to understand, support and progress the activities in which children are engaged. This is important if ergonomics and users centered design are to be adequately supported throughout the Key Stages.

7. CONCLUSIONS

Education must be seen as a continuum to be effective. In this chapter we have "reverse-engineered" the general approach to the teaching and learning of ergonomics, though the professional autonomy of the teacher in the UK does allow a wide range of approaches. We have identified the need to see, firstly, ergonomic awareness as an important contribution to general education for all children and students, whatever their intended roles in life. This awareness will assist the individual to be a better discriminating purchaser and a better arranger of working, living and leisure spaces together with an enhanced capacity to work cooperatively with others. Secondly we have identified the need for a "practitioner" strand of ergonomics education growing from that of awareness. This is important to enable engineers and designers to work more effectively and be better able to design products for both general and specific user fields. In addition the engineer or designer will be better able to work with specialist ergonomists when appropriate. Finally the specialist ergonomist will be better able to relate to other specialists and push our understanding of this important filed of human understanding further. Education, at all levels, and as a continuum, is the key.

REFERENCES

BAYNES, K., 1992, *Children Designing* (Loughborough: Loughborough University of Technology).

BCPE (BOARD OF CERTIFICATION IN PROFESSIONAL ERGONOMICS), 1999, *Candidate Handbook: Certification Policies, Practices and Procedures*, 4th edn, at [http://www.bcpe.org]

BERTODO, E., 1994, Engineering design: academic dilemma, *International Journal of Vehicle Design*, 15, 545–62.

BROWN, A.L. and CAMPIONE, J.C., 1996, Psychological theory and the design of innovative learning environments: on procedures, principles and systems. In Schauble and Glaser (eds), *Innovations in Learning: New Environments for Education* (Mahwah: Lawrence Erlbaum), 289–325.

DENTON, H.G. and WOODCOCK, A.,1999, Investigation of the teaching of ergonomics in secondary schools — preliminary results. In N.P. Juster (ed.), *The Continuum of Design Education: Proceedings of the 21st SEED Annual Design Conference and 6th National Conference on Product Design Education*, 7–8 September, Glasgow (London: Professional Engineering), 129–38.

KIMBELL, R., STABLES, K. and GREEN, R., 1996, *Understanding Practice in Design and Technology* (Milton Keynes: Open University Press).

LOMBAERS, J.H.M., 1990, Supermarketing ergonomics, *Ergonomics*, 35, 541–5.

SHAPIRO, R.G., 1995, How can human factors education meet industry needs, *Ergonomics in Design*, 32.

WOODCOCK, A. and GALER FLYTE, M.D., 1998, Ergonomics: it's never too soon to start. In *Product Design Education Conference*, 6–7 July, University of Glamorgan.

Enhancing Industrial Performance

H. Kragt† and R. Janssen‡

†University of Technology Eindhoven, The Netherlands
‡Philips Centre for Manufacturing Technology (CFT) Eindhoven, The Netherlands

1. INTRODUCTION

In *Enhancing Industrial Performance: Experiences of Integrating the Human Factor* (Kragt 1995) the realities you can capture by reading the book are:

- Ergonomics/Human Factors should be involved right from the beginning of any design study.
- Human Factors audits in the prototyping phase of any system can "save" the company.
- User participation in all design phases is not a luxury, but a requirement.
- In the organization of large-scale projects the attention to a human-oriented design is of vital importance.
- Last, but not least, realising a good fit between human resources and technical investment is not only a challenge, but vital in many ways.

Corporations worldwide confronting new market demands and intensified competition are responding with infrastructural change and a customer-centred approach. Such initiatives aim for business and manufacturing excellence, and involve enhancing the quality of performance in all areas of client service. Rapid technical change offers new tools to meet market challenges, and multi-skilled employees operating flexible computer-integrated manufacturing systems are central in the struggle for the competitive edge. These multi-skilled employees need also tools to evaluate their workplace(s) from a human factors point of view. Here a self-assessment tool, called Self-assessment Tool Ergonomics and Human Factors (SATEH) is dealt with and related to manufacturing excellence.

2. HUMAN FACTORS IN MANUFACTURING EXCELLENCE

In a world that is continually changing, it is difficult to remain competitive. Companies do everything they can to satisfy changes in customer wishes, in applied technologies and in the strategies which are used to compete with each other. Some companies react by moving the production activities to low-wage countries as quickly as possible, leaving a core of development and other strategic activities behind in the mother country. Others feel they are forced to think about the production activities that have to lead to best class practices.

Thinking about shop-floor activities has attracted an enormous amount of attention at Philips. Van Breukelen (1996) describes the paths that can lead to manufacturing excellence. Suzaki (1993) describes an approach that should make the organization ripe for continuous improvement. Wouters and Hendrickx (1997) plead for a transition from *manu-* to *mento*facturing, by which they mean that there will be no difference in the new organization between those who think and

those who do (the production workers). The workers on the shop-floor must be empowered not only to perform their work, but also to control it independently. The giving of the power to steer themselves to people on the shop-floor is called "empowerment" in the literature (e.g. Suzaki 1993). In the new organization, one makes effective use of the collective wisdom of people. Under the motto "nobody is perfect, but a team could be," self-steering teams have been introduced on the shop-floor in many companies, and with success (Peeters and van der Geest 1996).

This chapter explains how ergonomic design approaches can contribute to the design of production systems on the basis of the first results from a practical case in which SATEH was used.

3. HUMAN FACTORS AND MANUFACTURING EXCELLENCE

The questions arises: "What actually makes a company an excellent enterprise?" On the basis of 20 years experience in the auditing of industrial organizations, van Breukelen (1996) concludes that successful factories have a number of similarities over the complete breadth of the company management. With regard to the flexibility of the production function, excellent factories strive for such things as: short throughput times, small production batches, minimum transport activities and controlled processes.

However, it is not only the technological developments, but also the workers who make a company successful. To be able to use the possibilities of new technologies effectively, people and man–machine interfaces must be regarded as key elements. A company like Philips (Kragt 1995) also recognizes this.

Ergonomic approaches play an important role in the struggle of factories to deliver world-class performances. From the field of ergonomics, a contribution is provided to these excellent performances by placing the task, the person executing task and the task environment of the production system central. On the basis of ergonomic design methods, one tries to (re)-design production systems systematically so that optimum use is made of the possibilities of the human being and his limitations are properly estimated. The models of Döring, Pikaar and Rouse can be mentioned as an example of such methods (Janssen 1997a).

To stimulate the integral vision in production system development from the field of ergonomics as well, a research project has been started at the Philips Centre for Manufacturing Technology, in which the contribution from the human factors engineering discipline has an important position. This project is intended to make the currently used design model more ergonomic. On the basis of a number of practical research projects at Philips factories in Heerlen, Aachen and Eindhoven, we especially looked at the possibilities of integrating the SATEH in the present development approach. Not only is the usability of the tool determined, but also there is an approach developed for SATEH that ties up closely with the thought that the workers on the shop-floor are the people who make a company successful. The results of the research and the mentioned approach are explained below.

Looking back at the term "manufacturing excellence," it has to be mentioned that the struggle for excellent performances demands that management is aware of the situation on the shop-floor (attitude). Here, it is especially important that the production system to be (re)-designed is looked at not only from a technical

vision, but also from logistic, economic, socio-technical and ergonomic perspectives, etc. This demands a holistic/ integral view from both production system designers and users.

4. SELF–ASSESSMENT TOOL ERGONOMICS/ HUMAN FACTORS (SATEH)

4.1. Introduction, Purpose and Background

At the time, Kragt and van der Avort (1994) developed SATEH at the request of the Corporate Industrial Policy department of Philips. Stoppels (1995) improved SATEH and applied it in the Philips factories of Bruges and Nijmegen. In 1996, the present version became available (Kragt *et al.* 1996). This version is being tested at the moment, both in the framework of manufacturing excellence projects at Philips and in various teaching projects of the University of Technology, Eindhoven.

The purpose of SATEH is to enable industrial organizations independently to determine to what extent the (design of a) production system is fitted to the user of the system, the task to be executed and the environmental factors. It is also a tool that,

through its use, can start up an effective discussion, both between members of the organization themselves and with experts. By applying the tool, one obtains:

- An insight in, and knowledge of, the state in which the production system is located in the field of ergonomics/ human factors.
- An indication and prioritization of the areas that need improvement.
- An overview of possible solutions.

The tool supports the drive for manufacturing excellence. It is a further elaboration of the ergonomic/human factors aspect, which is globally specified in the Manufacturing Excellence Self-assessment Tool (Bolwijn *et al.* 1993) by the term "Quality of Working Life." Both tools form an essential component of the policy of Philips in struggle for perfection and fit in the philosophy of continuous quality improvement.

The tool consists of six tables, i.e.:
- Table 1. Design process of work places; given in detail here.
- Table 2. Task contents (see table 2 for details).

Table 1. Design process of work places.

	1	2	3	4	5	Score
1. Orientation	In the design, only the technique is looked at and not the people		In the design, mainly the technique is looked at and hardly the people		Design is oriented on people and socio-technical aspects*	
	Problems are solved by the employees adapting themselves		Problems are solved by changes in design after installation of equipment		Problems are avoided by changes during the design process (e.g. the use of prototypes, simulations, scale models)	
2. Approach	Ergonomic factors are not looked at		Only ergonomic factors are looked at		All ergonomic factors are looked at and included in the design	
3. Design team	One or only a limited number of functions/professions is represented		Several functions/professions represented, little discussion between the numbers		Several functions/professions represented, intensive consultation between the members (integral approach	
4. Participation by the end-users	Absolutely no input from the producions workers		Participation by production workers after important decisions have already been made by the design team		An early and continuous attention on nad by the produciton workers	
	Designer has a different idea of the operation or the use of machines of production systems to the production workers (designer has a different mental model), there is of a misfit				Designer has the same idea of the operation of the use of machines of production systems as the production workers (designer has the same mental model)	
5. Analysis	With the wet finger, common sense		There is only use of a specialist		Process, task and need for information are mapped. Existing faults and wishes of all parties involved are mapped	
	Lack of functional specifications		Functional specifications, made by the designer		Functional specifications are formulated by the designers and users	
6. Introduction	It is unclear who is responsible for the introduction of a (re)-design		It is clear for everyone who has the final responsibility for the introduction of a (re)-design		Everybody involved is aware of his individual responsibility in the introdiction of a (re)-design	
7. Learning ability	Hardly any documentation of earlier experiences		Little documentation of earlier experiences		Complete documentation of earlier experiences. Documentation is used.	
	Nothing has been learned from earlier experiences		Poorly learning organisation. Only parties who are directly involved learn form experiences		Optimally learning organisation	
8. Evaulation	No evaluation		Evaluation with little or no documentation/reporting		Evaluation of both the design process and the results	

Table 2. Graphic profile of SATEH.

Aspects	Criteria	Score 1	2	3	4	5	
1. Design process of work places	1.1 Orientation						1.1
	1.2 Approach						1.2
	1.3 Design team						1.3
	1.4 Participation by the end users						1.4
	1.5 Analysis						1.5
	1.6 Introduction						1.6
	1.7 Learning ability						1.7
	1.8 Evaluation						1.8
2. Task content	2.1 Variety of tasks						2.1
	2.2 Varirty of skills						2.2
	2.3 Completeness of the task						2.3
	2.4 Working speed						2.4
	2.5 Working sequence						2.5
	2.6 Independence						2.6
	2.7 Feedback of skills						2.7
3. Physical activity	3.1 Static load						3.1
	3.2 Dynamic load						3.2
	3.3 Work attitude						3.3
	3.4 Variety of work posture						3.4
	3.5 Lifting techniques						3.5
	3.6 Working position						3.6
	3.7 Energy required						3.7
	3.8 Work place						3.8
	3.9 Tools						3.9
	3.10 Shift work						3.10
4. Mental activity	4.1 Number of mental operations						4.1
	4.2 Concentration						4.2
	4.3 Pressure of time						4.3
	4.4 Stress						4.4
	4.5 Variation between mental and physical activity						4.5
	4.6 Vigilance						4.6
5. Environmental factors	5.1 Lighting						5.1
	5.2 Noise						5.2
	5.3 Vibrations						5.3
	5.4 Climate						5.4
	5.5 Reflections						5.5
	5.6 Cleanliness						5.6
	5.7 Housekeeping						5.7
	5.8 Total comfort						5.8
6. Communication	6.1 Provision of information						6.1
	6.2 Amount of information						6.2
	6.3 Operating equipment						6.3
	6.4 Result of operation						6.4
	6.5 Meters and screens						6.5
	6.6 Visual contact between employees						6.6
	6.7 Oral communication						6.7

- Table 3. Physical activity (see table 2 for details).
- Table 4. Mental activity (see table 2 for details).
- Table 5. Environmental factors (see table 2 for details).
- Table 6. Communication (see table 2 for details).

Table 1 gives the first table from SATEH as an example. It has eight aspects, which are each included as a row. The three columns include descriptions that should be applicable to the relevant aspect. The column on the right, under the figure 5, corresponds with the situation of manufacturing excellence.

The two other columns are under the figures 1 and 3. The person who assesses a work place can also use the score 2 or 4 for those cases where there is a "boundary case." The scores that are thus filled in are graphically represented per aspect in the overview table (see the profile in table 2). By means of this profile, one can quickly get an overview. After all, points that should be candidates for improvement, for example with a score < 3, are immediately noticed.

As will be seen below, it is the intention that several people will make an assessment of the relevant work place. The profiles thus obtained form an important input for the discussion about the work place and the possible improvements to be made.

4.2. A Recipe for SATEH

The use of SATEH gives an insight in the ergonomic strengths and weaknesses of the assessed work place. At least as important as the results is the manner in which the tool is used and how the results are used. After all, ergonomic design means that one takes a critical look at both the design result and the design process, as emphasized by the presence of table 1 in the tool. In this framework, the question arises: "What is (an) ergonomic design?"

On the basis of a literature research into different ergonomic design models by Janssen (1997a), the following definition was made for ergonomic design: "The *systematic* development of a product or system, whereby account is taken of the *requirements, wishes, limitations and possibilities* of the *user* in an *early stage* of the design process and by letting him participate *actively and passively* in the design process ...." Especially the active participation of operators and the application of a clear system in the assessment is of great importance in getting a directed discussion going. These elements are, therefore, given extra attention within Philips in the manufacturing excellence programme.

Taking account of the mentioned elements, the following approach arose in a pragmatic manner:

- Awareness – the ergonomist becomes acquainted with the operators, designers, ARBO (Factories Acts) experts and other interested parties of the factory where the research is being done. The importance of ergonomic research is explained on the basis of the criteria comfort, health, safety and efficiency. This creates the first support for the (ergonomic) changes and improvements.

- Analysis – next, a first impression of the work place is gained by looking at the task and the environment and by talking to the people executing the task. On the basis of this data, the solutions to be chosen can later placed in the light of the task, the performance of the task and the task environment.

- Assessment – by operators, ARBO (Factories Acts) experts, designers and experts, etc., the work place can next be assessed with the aid of SATEH. The scores are visualized in graphs and will later form the first starting-point for the discussion about the work place.

- Interpretation – after the filling in of the tool, the scores can be collected and a first interpretation of the results follows on the basis of a specially developed table. This table gives an insight in the strengths and weaknesses of the work place. It can also be used to determine which factors the fillers-in of the tool do not agree about (lack of consensus), so that special attention can be given to these factors in the discussion.

- Discussion I – in a first discussion, the results are discussed with the fillers-in. Special attention is paid to the factors in SATEH on which there is strong disagreement. The discussion finally results in an accepted overview of the strong and weak points of the work place. This gives the participants a grip on the ergonomy of the work place.

- Re-design – from his knowledge and practical experience, the ergonomist gives the initial start to improvement possibilities with regard to the weak points in the present work place design.

- Discussion II – on the basis of the global improvement possibilities, a discussion is held with the participants about improvement possibilities and solutions regarding the ascertained bottle-necks. Proposals for improvement plans are formulated, in which the responsibilities of the participants are clearly defined.

- Implementation – next, the proposed improvement plans can be implemented. Here, the responsibility is mainly with the operators, designers and ARBO (Factories Acts) experts, etc. The ergonomist mainly fulfils a sounding board function.

- Evaluation – finally, implemented improvements must be evaluated. SATEH can be used again for this, so that one can make a good comparison between the previous and the new situations.

The approach has been successfully applied in the redesign of work places at Philips factories, including Heerlen (Janssen 1997b), Aachen and Eindhoven. As an illustration, the following section takes a close look at the results as have been realised in a Dutch Philips factory. For reasons of privacy, the name of this factory has been omitted.

4.3. SATEH in Practice

The mentioned approach has been followed at a factory of the largest Product Division but one within Philips. The relevant factory produces mainly discrete products for customers in Asia and The Netherlands. It also supplies part of these products to its own development department for tests. The factory also functions as a knowledge centre, whereby knowledge development and transfer are important activities with regard to the product, processes and production tools.

Rising organizational costs, the fall in the dividend on invested capital and a reduction of the profit within the business group to which the factory belongs made the Board of Directors decide to carry out a radical organizational change to be able to safeguard continuing growth and a stable future. In the framework of the mentioned terms of manufacturing excellence, glass wall management (Suzaki 1993) and self-steering teams (van Eijnatten 1996), this led to the introduction of "mini companies" on the shop-floor.

During the introduction of the mini-company project, bottlenecks within the factory organization were discussed with the employees on the production floor. From this, it appeared that one had to fight with several ergonomic problems in part of the production. This resulted in physical complaints from operators and a higher time lost through illness than elsewhere in the factory. A first research project was started from this realization.

On the basis of manufacturing documents, interviews with operators and designers and observations during production hours, the task, performance of the task and the task environment were analysed. Also, the SATEH tool was filled in by three operators, an instructor, a technical assistant, an ARBO (Factories Acts) employee and the ergonomic expert. The scores were then taken together, after which the mathematical average was determined. This gave a first picture of possible bottle-necks for the assessed work place.

The individual scores from which the mathematical average of various factors was composed differed. With some factors from the tool, the lack of consensus appeared to be about the score to be allocated. (For two reasons, one cannot simply express the lack of consensus in the standard deviation: (1) the measurement level of the scale is at best ordinal and (2) the number of observations is small. A solution for this was found in the use of the basis for the spread: differences between scores.) Next, it was ascertained that the calculation of sums of squares about all given scores should give more insight in the factors about which the fillers-in had different opinions (Dijkstra 1997, personal communication). This led to the following table with four cells, in which, on the one hand, on the basis of the average, it can be determined whether the relevant ergonomic factor is strong or

Table 3.

Enhancing Industrial Performance

Table 3. Interpretation table of SATEH sccores.

	There is consensus (sum of squares < $\underline{X}$)	There is no consensus (sum of squares >= $\underline{X}$)
Work place scores well (average ≥ $\underline{Y}$)	**Strength** This cell contains the <u>strong points</u> that one can include in a following Design.	**???** This cell contains points, of which it is <u>not clear</u> whether they are strong or weak. Discussion is necessary.
Work place scores badly (average < $\underline{Y}$)	**Weakness** This cell contains the <u>weak points</u> that one must improve in a following design.	**???** This cell contains points, of which it is <u>not clear</u> whether they are strong or weak. Discussion is necessary.

weak and, on the other hand, whether there is consensus on this point between the fillers-in. With a lack of consensus, the factor is placed in the right of the table and discussion is required.

On the one hand, the strength or weakness of a certain ergonomic factor is determined by the calculation of the average for this factor. This average is compared with a so called threshold value (Y). Y states the minimal required average for a factor. In this way, a principal can determine the desired threshold value of the average himself, fitting in the ARBO (Factories Acts) policy of the enterprise. In the case of manufacturing excellence, a minimum of 3.0 (Y) must be scored on the ergonomic factor. This value was, therefore, kept for the Philips factory mentioned above.

On the other hand, the degree to which there is consensus between the fillers-in about the allocated scores translated to a sum of squares. (Take the following example, in which seven people filled in SATEH: *Individual scores for a certain factor:* 2, 3, 3, 3, 3, 3, 4 – average = 3 and KS = 14 (calculated from the differences of *all* possible combinations of scores!). Here, therefore, the factor scores well on average and one agrees about it (there is consensus about the average). Discussion is not interesting. *Individual scores for a certain factor:* 1, 1, 1, 3, 5, 5, 5 – average = 3 and KS = 168. Here, the factor scores well on average, but one does not agree about it; there is insufficient consensus about the average. Discussion is necessary to be able to explain the difference in opinion and get a balanced assessment.) The higher this sum, the more lack of consensus there is in the individual answers. With the aid of a simple computer program, it can be determined which sums of squares can occur for a given number of operators. In practice, it has appeared that, on the basis of this, a limit for the sum of squares can be chosen (X), whereby sufficient discussion arises without each factor having to be examined separately. With seven fillers-in, this limit was 68 (based on earlier research) and it was, therefore, chosen.

After interpretation of the data by placing the factors in the correct cell, a first discussion was started with the fillers-in of SATEH. Especially the points about which there was a lack of consensus were discussed and, if possible, also assessed, after which an accepted set of good and bad points remained. The results were placed on the appropriate "glass wall." A second session was used to think about solutions; some that could be implemented immediately and with limited impact, others that were suitable for implementation in the long-term. A simple qualitative model was used to give a first picture of the possible

yields of ergonomic interventions (Alexander 1998). As a result of the participation of the fillers-in of the tool in this phase, greater support was created for the improvements. Also, an initial start was given for the formulation of improvement plans to give the actions direction. These plans will again be hung on the "glass wall."

In the near future, the work place will be assessed again, with the aid of SATEH, so that an insight will be gained in the success of the improvements. In that phase, the ergonomist will fulfil a sounding board function.

5. CONCLUSIONS AND RECOMMENDATIONS

The struggle for excellent performances means that, among other things, management is aware of the situation on the shop-floor (attitude). It will have to create conditions whereby such matters as empowerment (see Section 1) of the production worker is guaranteed.

One of the possibilities for realising empowerment is the stimulation of the filling-in of SATEH (see Section 4.1) and presenting the results on the "glass walls" (see Section 4.2). After all, the production floor is the place where the different disciplines meet and where an integral vision about the design and implementation of the production system can be developed. SATEH is powerful because it gives operators the opportunity systematically to evaluate their own workplace and to come up with improvement ideas.

The approach described in Section 4.2 specifies the actual situation of a production system with regard to the ergonomic and human factors. To get to the desired situation, people such as ergonomists must cooperate in the formulation of the programme of requirements for the new situation to be designed. Here, making the extensive literature in this field accessible is a first requirement. We have made a start with this by drawing flow charts, which clearly show the large amount of ergonomic design information in schematic and visual form.

The strength of SATEH is that individual workers start to think about their own work situation and the possible improvements in it. To obtain interpretable group scores from individual scores (shown in clear profiles in table 2), the sum of squares (Section 4.3) has been introduced and successfully applied in the current and other cases. The consensus matrix (table 3) indicates on which aspects there is consensus, and on which aspects discussion is still required. The discussions finally contribute to an accepted idea of the ergonomics of the work place and create stronger support for possible improvements. In practice, the feedback on the shop-floor organization also appears to work well. Profiles, as given in table 2, must be made visible on the production floor next to other performance indicators to continuously stimulate *the improvement* (glass wall management). This also facilitates communication about possible improvements between the people in the factory and, if possible, with experts.

From the practical study, it appeared that ergonomic analyses and solutions must fit in the strategy of the enterprise. Subsequent to "The Philips Way," which includes "The motivation of our people is decisive for the success of the enterprise," the following of a participative approach must be regarded as almost obvious. In the execution of the case in the Philips factory, attention must also be paid to the exemplary function that the relevant factory has with regard to its sister factory in Asia. In this framework,

the knowledge gained of the ergonomic research must be transferred, so that one can learn from each other.

Determining the profitability of ergonomic solutions contributes to its acceptance by management. In the first instance, the economic consequences of ergonomic redesign can be qualitatively mapped. However, it is not always a case of "earning money on ergonomics"; there can be a case of reducing running expenses. Here, it is recommended to look at the ergonomic solutions from a different viewpoint: "Not participating in ergonomics costs money."

The Organization must be "ready" for ergonomic research. This is achieved at Philips by such actions as the education of the management with the aid of three courses: Shop-floor Management, Middle Management and Top Management. In this way, everyone within the enterprise will speak the same language.

REFERENCES

ALEXANDER, D.C., 1998, Strategies for cost justifying ergonomic improvements. *IIE Solutions*, 30.

BOLWIJN, P. *et al.*, 1993, *Het Manufacturing Excellence Self-assessment Tool* (Eindhoven: Philips Corporate Staff).

BREUKELEN, Q.H. VAN, 1996, Facts on factories: in search of manufacturing excellence. Thesis, Erasmus University, Rotterdam.

EIJNATTEN, F.M. Van, 1996, *Sociotechnisch Ontwerpen* [*Sociotechnical Design*] (Utrecht: Lemma).

JANSSEN, R., 1997a, *Mensgericht Ontwerpem: van curatief naar preventief gedrag* [*Human Oriented Design*]. Special assignment (Eindhoven: University of Technology, Technology Management faculty).

JANSSEN, R., 1997b, *Mensgericht ontwerpen van productiesystemen: (G)een kwestie van gezond verstand?* [*Human Oriented Design of Production Systems*]. Graduation report (Eindhoven: University of Technology, Technology Management Faculty).

KRAGT, H., 1995, *Enhancing Industrial Performance: Experiences of Integrating the Human Factor* (2nd ed.) (London: Taylor & Francis).

KRAGT, H. and AVORT, P. VAN DER, 1994, *Self-assessment Tool Ergonomics & Human Factors* (Eindhoven: KPMG Lighthouse).

KRAGT, H., KESSEL, W. VAN and CULLEN, R., 1996, *Self-assessment Tool Ergonomie/Human Factors voor fabricage-omgevingen* (Eindhoven: CFT Production Systems).

PEETERS, M. and GEEST, L. VAN DER, 1996, *Zelfsturende teams; de praktijk aan het woord* [*Self-directed Teams*] (Deventer: Kluwer).

SITTER, U. De, 1989, Moderne sociotechniek. In *Gedrag en Organisatie 4/5*, October (themanummer sociotechniek), pp 222–52.

STOPPELS, J., 1995, *SATEH beproefd; de ontwikkeling van een vragenlijst om te komen tot mensvriendelijke productiesystemen* [*SATEH tested*]. Graduation report (Eindhoven: University of Technology, Technology Management Faculty).

SUZAKI, K., 1993, *The New Shop Floor Management; Empowering People for Continuous Improvement* (New York: Free Press).

WOUTERS, F. and HENDRICKX, W., 1997, Philips Industrial Policy: the vital role of manufacturing managers in Philips' future. Contribution to the MMP (Manufacturing Management Programme) block 1, 22–25 April (Eindhoven: Philips Corporate Training).

Ergonomics and Quality of Life

F. Baptista and P. Moro
Department of Industrial Engineering, University of Wisconsin-Madison, Madison, WI53706, USA

1. INTRODUCTION

Quality of life is an important aspect of one's life, and is composed of two main parts – quality of work life and quality of non-work life. The former refers to an individual's work domain, while the latter refers to an individual's out-of-work domain. If there is quality in both of these components, it is right to say that an individual has quality of life. Lack of quality in one or both components mean that an individual lacks quality of life or has a limited quality of life. One way of increasing someone's quality of life is by increasing each one of the major components – quality of work life or quality of non-work life. Ergonomics, by definition, has the ability to increase someone's quality of work life, and consequently increase the overall quality of life of this individual. Ergonomics can also increase one's quality of non-work life, but the major effect that ergonomics can have in a worker's quality of life is by increasing quality of work life. This chapter aims to discuss ergonomics and its contribution to increasing the quality of life of workers.

2. QUALITY OF LIFE

Quality of life is an important aspect of one's life, and can be understood as the subjective analysis of how healthy, happy and satisfied a person is with his/her life in general. This value judgement will be dependent on the person's culture, education, aim in life and resources, for example, that are available to achieve the person's goals. Quality of life is measured by the use of indicators, such as overall health status (physical and mental), social status, family status (e.g. if married, satisfaction with marriage), leisure time and activities (social life), performance status, general comfort, emotional status and economic status.

Quality of life is a concept. It is the output of a process in which quality of work life and quality of non-work life are the inputs. The transformation here are the interactions between these two input components and the way the worker perceive – through cognitive appraisal – how good or not his/her quality of work and quality of non-work life are. One will have quality of life if quality is found in this person's work and non-work life. Lack of quality in each of these or in both represents a lack of quality of life or a limited quality of life. It is understood that human beings look forward to having quality of life. But to accomplish this, both work and non-work lives should also have quality in them.

Workers have their lives divided by participation in two different domains, which are the domain of work and the domain of non-work. The domains have the ability to affect each other, since the individual that has a life outside the organization is the same individual that has a life inside the organization. The individual brings to work what s/he is experiencing at home, and vice versa.

2.1. Quality of Work Life

Quality of work life can be described as the way an individual perceives and evaluates the characteristics intrinsic to his/her work. This evaluation is subject to the influence of past experiences, education, race and culture. Quality of work life is affected by the characteristics of the organization, workplace, job and tasks the individual performs, by the environment that surrounds the individual, and by other subsystems that exist in the work system. Although quality of work life is also influenced by factors from outside the organization, it is measured by evaluating the level to which people feels good, rewarded, and/or satisfied with their work.

Walton (1975: 92–7) proposed that quality of work life be analyzed by eight major conceptual categories: (1) adequate and fair compensation; (2) safety and healthy working conditions; (3) immediate opportunity to use and develop human capacities; (4) opportunity for continued growth and security; (5) social integration in the work organization; (6) constitutionalism in the work organization; (7) work and the total life space; and (8) the social relevance of work life.

Different researchers have measured quality of work life in many different ways. Some of the variables used to measure it are presented as they have been used in the literature. These measures are: availability of jobs, training, mobility, job security, fringe benefits, earnings, job safety, equitable distribution of wages, job opportunities, job challenge, job satisfaction, job future/career future, job involvement, autonomy, responsibility, advancement opportunities, job control, turnover rate, commitment to the organization, self-esteem, self-actualization, social support, challenge, stress, depression, burn-out, physical health, use of skills, etc. By far, the most used measure of quality of work life has been job satisfaction. It is also accepted that the working conditions affect quality of work life. Measures of these working conditions are, for example, noise, lighting, vibration, fumes, exposure to chemicals and heat/cold.

2.2. Quality of Non-work Life

Quality of non-work life can be described as the way an individual perceives and evaluates the characteristics intrinsic to his/her life outside the workplace. Here, again, previous experiences, education, race and culture affect the individual's appraisal of this domain.

Quality of non-work life has been measured as satisfaction with family life (e.g. marriage, social support, children), leisure activities and hobbies, social life, religious life, self-esteem, self-development, success in life, and financial security to cite some.

Recently, some researchers have looked at the conflict between work and non-work life as a factor that also affects the quality of either work or non-work life and, therefore, the quality of life. Conflict tends to occur when a person is not able to balance the work role and family role. This role conflict can originate from work or from home. Excessive demand from work is, for example, having to work extra hours because of work overload or understaffing, while an excessive demand from home is, for example, having to stay more hours at home due to no child care being available or sickness of a spouse. Conflict can also be caused by excessive demands from both work and non-work domains.

2.3. Final Remark on Quality of Life

The above has shown that the concepts of quality of work life and quality of non-work life are multi-dimensional. Beyond that, the components of each concept are believed to interact constantly. The complexity of the concepts of quality of life, work life and non-work life is evident and has been a challenge to researchers. The use of a systems approach –an approach that looks at all the components and its interactions at the same time, or that acknowledge these interactions and accounts for them during the analysis – is the most adequate way to go about to understand and measure these concepts.

3. ERGONOMICS

Ergonomics is the science or discipline that main goal is to fit, adjust, change, modify the artificial environment human beings work and live in to the limitations and capacities of the human beings. Some examples of how ergonomics accomplishes this goal are: (1) when a new software interface is designed to allow disabled people to use a computer or machine; (2) when new kitchen utensils are designed to allow that elderly people suffering from joint diseases be able to do their chores or still feed themselves more easily and safely; and (3) when a lifting device is designed and built to diminish the intensity and the repetition of movements in a work setting. Ergonomic principles are or can be present in all domains of human lives, from home, to transportation systems, at work or whenever a person is interacting with a machine or device. For example, when one uses an ergonomically designed chair or ergonomically designed controls in a car, one is experiencing the benefits of ergonomics in day-to-day activities.

Ergonomics developed much during this century, and it is acknowledge that its evolution is comprised of three distinct generations. Today, we see applications of all the so-called generations of ergonomics. The first generation was characterized by the design of human–machine interfaces, including control panels, knobs and dials, and the work environment, and it was heavily based on anthropometry. The second generation started with the study of human cognition. It became important to study the cognitive demands of the operator due to the advent of computer technology and its by-product, automation of jobs and tasks. Here ergonomics looks at human–interface design and software design, for example. The macro-ergonomic approach is considered to be the third generation of ergonomics. As proposed by Hendrick (1986) this approach is concerned with ergonomics research, development and application of organization–machine interface technology. It was proposed as a tool that would enable ergonomists to deal with the effects of new technologies in organizations, and consequently their effects on the job.

3.1. Macro-ergonomics

Hendrick (1986: 467) founded macro-ergonomics in the socio-technical systems approach. His definition of macro-ergonomics is: "Macro-ergonomics represents a top-down approach to system design based on a socio-technical systems perspective." Hendrick (1994) also states that the macro-ergonomic approach can satisfy three specific basic criteria for the correct organizational and work systems design, and related design of jobs and human–systems interfaces, which are: (1) be human-centered; (2) use an humanized task approach to function and task allocation; and (3)

adequately consider the relevant socio-technical systems variables in terms of their implications for the design. Macro-ergonomics, due to its conceptualization, has the ability to allow the ergonomist to use different ergonomic approaches and methodologies together, in an aggregated manner, better to design, re-design, change, fit or adjust a system to the human user. Macro-ergonomics helps better to analyze, understand and act upon a system by allowing for a broader understanding of the system being analyzed, which as a consequence increases the odds of having a positive outcome after the ergonomic intervention.

3.2. Using a Macro Framework for Ergonomic Interventions

Moro and Derjani-Bayeh (1996) showed an analytical framework that brings together the micro- and macro-ergonomic approaches, including tools such as participatory ergonomics and ergonomic work analysis under the broader macro-ergonomic intervention. It is believed that the macro-ergonomic approach is the most efficient way of looking into a system (e.g. organization) for the purpose of promoting organizational design, culture and structure changes, as well as new technology implementations. This approach also enriches micro-ergonomic interventions in which it allows the ergonomist to have a broader perspective of the organization and how it works. Macro-ergonomics also allow for the concomitant use of other ergonomic tools that are capable of contributing, in their own way, to the achievement of the goal of the ergonomic intervention. For example, the contribution from ergonomic work analysis (EWA) is the ability to analyze in depth the tasks and activities developed by the employees, while the contribution from participatory ergonomics is that of bringing together employees, supervisors and managers better to understand a work situation and to propose and implement changes. These approaches complement the macro-ergonomic approach and, together, can gather more information on the situation than the macro-ergonomic approach by itself, and also to promote more positive changes in the organization. Of course, these are not the only ergonomic approaches that are available. All other approaches developed in the field, each one in its own way, can contribute to the enrichment of the macro-ergonomic approach and can perform better an analysis of the situation and of all its components.

4. ERGONOMICS AND QUALITY OF LIFE

The contributions of ergonomics to improving one's quality of life will now be addressed. Here, the impact that ergonomics can have on one's quality of work life will be the main focus, since it is still in the work environment that ergonomics can provide major contributions to promote changes that will benefit the individual. Cherns and Davis (1975: 55) stated that "quality of working life permits a focus on a key problem of great significance – the dysfunction in the individual experience between work and the rest of life." One way of interpreting this statement is that the work experience individual's are having are not as positive as the experiences outside the work domain. Another is that the experiences individuals are having in the work domain are negatively affecting the rest of this person's life.

Ergonomics fits perfectly as a tool to help improve quality of work life by definition. Ergonomics has the goal of fitting the task, job and environment to human beings, thus allowing people

to have a better experience, to enjoy their work, to feel that there is quality in what they do and in how they perform their duties. By fitting the human–system interface to the characteristics of men/women, ergonomics is basically increasing the quality of the interaction between the humans and the system. In doing this, ergonomics is helping to increase the quality of the work experience and, consequently, is decreasing the dysfunction between work life and non-work life.

The different ways ergonomics can affect the quality of a worker's life vary from micro- to macro-interventions. Micro-ergonomic interventions are, for example, re-designing a work-station, providing for more ergonomically designed tools and accessories, and re-defining elements of the task to decrease muscular and cognitive demands. This kind of intervention is limited because deals with one or a few workstations or employees, usually in a sector or department. The benefits, although noticed and appreciated by the workers, will not necessarily solve the problem in its totality, since many other organizational characteristics – not accounted for by the intervention – can be playing a role in the situation the ergonomist is intervening on. The ergonomics' literature is full of examples of micro-interventions that increased satisfaction, decreased stress, diminished musculoskeletal problems, reduced turnover rates, decrease injury, etc.

Macro-ergonomic interventions, on the other hand, deal with more characteristics of the system being analyzed. With this kind of intervention, knowledge about the organization structure, culture and future goals is gathered. Together with the use of other ergonomic tools and approaches, the macro-ergonomic intervention can bring benefits like the ones mentioned above as well as benefits from changing the design of the jobs, the structure of the department, the culture of the organization, etc. In this case, the benefits brought about by the intervention are broader and have a greater impact in the situation as a whole. Benefits from this kind of intervention are, for example, a reduction in workload, an increase in control and participation, a reduction in psychosocial stressors, etc. There are many examples in the literature of how such changes benefit the employees.

The magnitude of the impact that the ergonomic intervention has is similar to the magnitude of the intervention itself, and because of that the impact on quality of work life tends to be greater when the ergonomist performs a macro-ergonomic intervention instead of a more limited micro-intervention. Nevertheless, micro- and macro-ergonomic interventions increase the quality of the employee(s) work life by changing job characteristics that are relevant to the concept of quality or work life.

5. IMPLICATIONS

In summary, ergonomics has the potential to affect quality of work life through different types of interventions, from micro to macro. The amount of change that the ergonomic intervention will bring is, of course, dependent on the magnitude of the intervention. A micro-ergonomic intervention will increase the quality of work life of a few individuals in a very limited manner. On the other hand, a macro-ergonomic intervention using participatory ergonomics and other analysis tools will increase the quality of work life of a lot of individuals, and this will happen in a broader manner. Of course, these increases can be more or less significant, depending on the nature of the "situation" and the "solutions" implemented. Nevertheless, it can be concluded that ergonomics can improve the quality of life of workers by improving their quality of work lives.

REFERENCES

CHERNS, A.B. and DAVIS, L.E. (eds), 1975, Goal for enhancing quality of working life. In *The Quality of Working Life: Problems, Prospects and State of the Art* (New York: Free Press), pp. 55–62.

HENDRICK, H.W., 1986, Macroergonomics: a conceptual model for integrating human factors with organizational design. In Brown, O. Jr and Hendrick, H.W. (eds) *Human Factors in Organizational Design and Management – II* (Amsterdam: Elsevier), pp. 467–77.

HENDRICK, H.W., 1994, Macroergonomics as a preventative strategy in occupational health: An organizational level approach. In Bradley, G.E. and Hendrick, H.W. (eds) *Human Factors in Organizational Design and Management – IV* (Amsterdam: Elsevier), pp. 713–18.

MORO, F.B.P. and DERJANI-BAYEH, A., 1996, Analytical framework for organizational interventions: an aggregate of the macroergonomic and microergonomic levels of organizational analysis. In Brown, O. Jr and Hendrick, H.W. (eds) *Human Factors in Organizational Design and Management – V* (Amsterdam: Elsevier), pp. 423–8.

WALTON, R.E., 1975, Criteria for quality of working life. In Davis, L.E. and Cherns, A.B. (eds) *The Quality of Working Life: Problems, Prospects and State of the Art* (New York: Free Press), pp. 91–104.

Forensic Human Factors/ Ergonomics

H. H. Cohen and C. A. LaRue

Error Analysis, Inc., Human Factors & Safety Consultants, 5811 Amaya Drive, Suite 205, La Mesa, CA, 91942-4156, USA

1. INTRODUCTION

Forensic human factors/ergonomics (HF/E) can be defined as the application of the science of HF/E to problems in the legal setting. The central role of the human in systems design is the same as traditional HF/E, as is the role of human performance capabilities and limitations resulting in human error. All that is unique to the practice of forensic HF/E is its application to legal issues, as opposed to other systems.

1.1. Reasonable Care

Another unique definition for forensic HF/E is the science of establishing "standards of reasonable care" in the design, maintenance and use of products, equipment and facilities. Establishing standards of reasonable care is what makes the field of forensic HF/E unique from not only all other applications of HF/E, but from other forensic sciences or disciplines.

Western civil law is concerned with standards of reasonable care. In order to live in modern society we must take precautions to insure that our actions and inactions do not harm other people or their property. *Negligence* theory generally holds all people to the same objective standard, despite their personal shortcomings, using a yardstick, "the man of ordinary prudence." The emergence of this theoretical level of ordinary prudence may be traced to an English case decided in 1837. In that case a man, contrary to his neighbor's advice, built a hayrick in a manner that allowed it to ignite and damage the neighbor's property. The court determined that the question facing the jury was whether the defendant acted as a "man of ordinary prudence" when he constructed the hayrick and not whether he had used his own best judgment (Keeton 1984).

1.2. Areas of Practice

Forensic HF/E has focussed largely on problems of civil, as opposed to criminal, litigation. Moreover, most applications have been directed to personal injury cases involving transportation, premises or products.

1.2.1. Transportation

It is logical that forensic HF/E emerged early to focus on a variety of transportation incidents since such systems design and safety have been studied by HF/E professionals since the field's beginnings. HF/E applications have evolved to include air transport, motor vehicles, industrial and commercial vehicles, rail and water transport, and recreational vehicles. Forensic HF/E cases in this category typically involve collisions between two or more motor vehicles (cars, trucks, etc.), between motor vehicles and individual means of transport (motorcycles, bicycles, etc.), as well as between motorized transport and pedestrians. Many forensic HF/E practitioners are qualified to offer expert opinions on such specific issues as visibility/conspicuity, perception–reaction, attention demand, skill level, human error, impairments to performance and decision-making.

1.2.2. Premises

By far, the largest focus of investigations regarding premises has been directed to the prevention or mitigation of fall injuries. This is not surprising considering the high relative frequency and severity of such incidents, second only to motor vehicle collisions. Specific premises accidents addressed by qualified HF/E professionals include: falls on level surfaces (e.g. slips, trips, missteps, slipperiness, hazard visibility/conspicuity), falls from elevated surfaces (e.g. stairs, ramps, ladders, work platforms, vehicle ingress and egress, handrails and guardrails), biomechanics (normal and disrupted gait, fall dynamics), and information-processing (eye scanning patterns, focus of attention, sensory cues, competing stimuli, user expectations). Typical systems issues in these applications are the design, construction, maintenance and use of the premises.

1.2.3. Products

Forensic HF/E professionals are often called upon to offer expert testimony with respect to issues of products liability. The most commonly addressed areas are alleged product design defects and failures to warn. Product design defects can include a myriad of core HF/E principles affecting a user's reasonable expectations including display–control compatibility, population stereotypes, lack of job performance or decision-making aids, and excessive demands. A large body of HF/E research literature has evolved in recent years concerning adequacy and effectiveness of warnings. Despite one's personal views on the merits of warnings, a growing database specific to forensic HF/E is a positive trend.

1.2.4. Other applications

Litigation in the realm of workplace health and safety (e.g. back injuries, repetitive strain) is, at least in the US, largely limited by workers' compensation laws unless a third party (e.g. a product manufacturer or premises owner) can be implicated. The common concern of other types of cases for the forensic HF/E professional is whether system failures attributed to human error are involved and whether the human error(s) were design or situation-induced and, therefore, preventable through technology.

2. THE PROCESS

Forensic HF/E requires the practitioner to analyze situations involved in litigation and offer opinions that may either support or refute legal theories being pursued. Courtroom proceedings are an adversarial process in which lawyers represent opposing parties; in civil cases, they typically represent either a plaintiff (the injured party) or a defendant (the alleged injuring party). The role of the expert witness is to offer opinions beyond the knowledge of ordinary persons, such as comprise a jury.

The steps typically followed by a forensic HF/E practitioner, with related client feedback, are shown in Figure 1. It should be noted that this process is based on forensic practice in the USA; other jurisdictions may differ.

2.1. Consultation with Client

The initial step in the forensic consulting process is a consultation with the client, who is generally an attorney or representative of
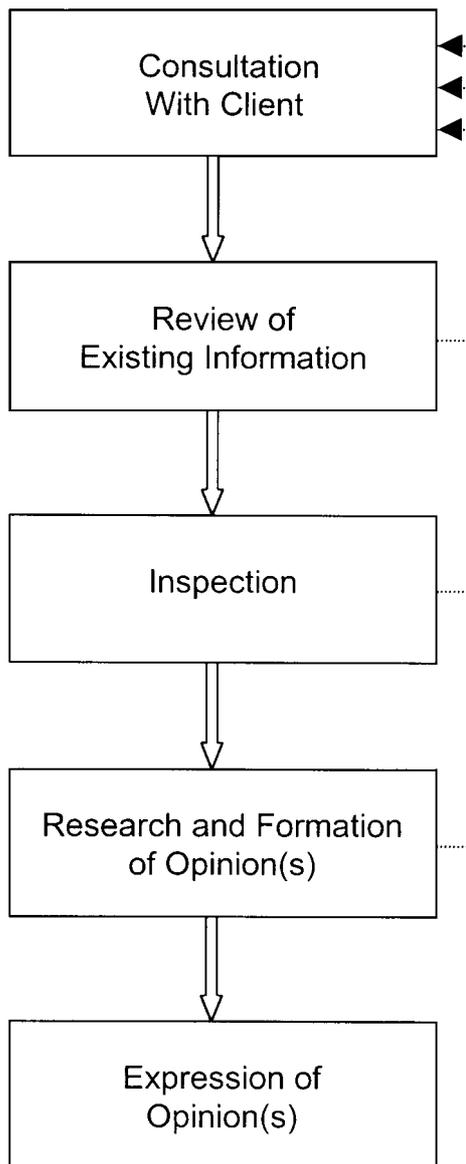
Figure 1. Process of forensic human factors/ergonomics

an insurance company. In this initial conversation, the client provides a description of the case and the issues involved. The forensic HF/E professional will want to confirm that the issues in the case are within his/her field(s) of expertise. When an expert stretches his credentials into areas in which he is not qualified, it adversely affects his reputation and credibility, and therefore, the effectiveness of his testimony. The client will want to confirm that the expert has experience and is qualified to offer opinions with respect to the specific issues of the case. Before work on a case begins, generally an agreement regarding fees is made. It is common for the client to pay the expert a retainer, an initial deposit for the expert to begin work on the case.

2.2. Review of Information

As in any scientific endeavor, the first true step is to review the existing information, which in the legal setting is referred to as *discovery*. It is important for the forensic HF/E professional to

review all pertinent discovery in order to understand the relevant background facts of the case.

In a typical case, there is a variety of information available regarding the plaintiff(s) in the form of an incident report, statements, medical records and both written and oral answers to legal questions in the form of interrogatories or depositions, respectively. Other important sources of information are witnesses to the incident and *persons most knowledgeable* about the circumstances who are associated with the defendant(s), such as employees, product designers, or premises managers. This information can also include historical data such as incident reports involving the same premises or product.

Reports and documentation prepared by other expert witnesses involved with the case (e.g. medical doctors, architects, design or traffic engineers, opposing HF/E expert) provide information as to the points both sides of the case will attempt to make.

2.3. Inspection

A very important step in the forensic process is the inspection of the accident location or product involved in the case. Although the purpose for the inspection is to gather as much first-hand data as possible, there are certain points to consider when conducting an inspection for the different types of cases. In all cases, photo-documenting the accident scene or equipment involved is imperative.

2.3.1. Premises case

It is important to measure relevant dimensions at the accident scene (e.g. stair riser, tread and handrail dimensions) depending on the nature of the case. Other testing at an accident location can include coefficient of friction to determine the slipperiness of a walking surface, slope and illumination measurements.

2.3.2. Products case

If the incident involved a product, then the inspection would include an evaluation of the item or exemplar including all controls, gauges, warning labels and guarding. It is also important to review the manuals, instructions and other safety information accompanying the product.

2.3.3. Transportation case

A case involving one or more vehicles would require an inspection of the vehicle(s) and the location of the accident, preferably under similar lighting and weather conditions. Measurements taken typically include scaled dimensions of the accident scene and illumination. A reconstruction of the accident using similar vehicle(s) can also be helpful in determining issues such as visibility/conspicuity, perception–reaction time, and vehicle performance.

2.4. Research and Development of Opinion

It may be beneficial to conduct research prior to developing a forensic HF/E opinion on a case. Research for premises cases could involve determining how similar facilities have handled the design, construction or maintenance issue at question. If a product is involved, research as to the design and safety features of competitors' products can be useful.

The forensic HF/E professional will integrate all the information gathered from the discovery, inspection and research
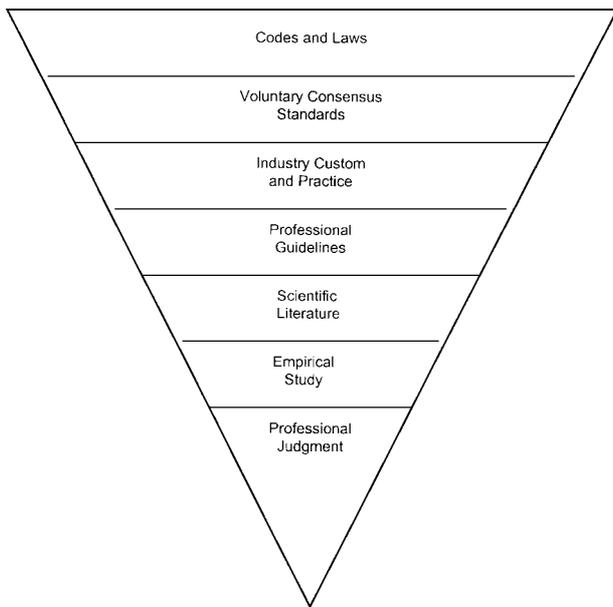
Figure 2. Hierarchy of the bases of forensic human factors/ergonomics opinions.

with his/her knowledge and experience to form an opinion. The expert's opinions are further strengthened when s/he can provide a sound authority as the basis for the opinion(s).

There are a variety of authorities on which these opinions may be based, some providing stronger support of the opinion than others, but all acceptable depending on the type of case (Cohen and LaRue 1997). Figure 2 illustrates the hierarchical nature of such opinion bases according to the relative ease with which they can be supported.

2.4.1. Codes and laws

Codes and laws are legal requirements governing a particular jurisdiction. The expert's opinion might be that the situation s/he has analyzed does or does not meet the required provisions of the applicable code. It is rare that the forensic HF/E expert is involved in a court case that has proceeded to the point where there is a clear violation of a local code or law. When this does occur, the issue is typically whether the violation has any causal bearing on the incident or was in effect at the appropriate time. Examples of these types of codes and laws are the Uniform Building Code (UBC) and the federal Americans with Disabilities Act (ADA).

2.4.2. Voluntary consensus standards

This category of opinion basis relies on a group of standards that is not legally binding, but has been arrived at through an industry consensus process by committees associated with standards setting bodies. These organizations include the American National Standards Institute (ANSI), the American Society for Testing and Materials (ASTM) and the International Standards Organization (ISO).

2.4.3. Industry custom and practice

An industry custom and practice is a way of doing business within a particular industry that has evolved over the years, usually based on experience. It is generally not formal and may not even be documented. In some cases, however, these practices may be

formalized as written company policies and procedures that may have evolved from some type of job analysis. These practices can be useful in forming an opinion especially when there is scientific evidence to support the practice and/or research that supports the proposition that it is the customary practice of the industry in question.

2.4.4. Guideline in a professional handbook

Most design-related professions have handbooks that offer design criteria. When none of the aforementioned bases are applicable, a citation from an authoritative professional handbook may be appropriate. Examples commonly used by HF/E professionals include the Illuminating Engineering Society (IES) *Lighting Handbook* (Kaufman and Haynes 1987), Woodson *et al.*'s *Human Factors Design Handbook* (1992) and *The Measure of Man and Woman* by Henry Dreyfuss Associates (1993). The major disadvantage of citing such an authority as the only opinion basis is that the source may not be widely known, especially by the responsible party such as a building owner or manager.

2.4.5. Scientific literature

Because of a background as research scientists, many HF/E professionals feel most comfortable when citing scientific literature as the basis of their opinions. However, the literature cited is typically of very limited circulation, usually far less so than the previously discussed bases. Consequently, citing HF/E scientific literature as an opinion basis may represent information readily available only to the HF/E professional.

2.4.6. Empirical study

Sometimes the issue at hand is not specifically addressed by any documented authority and can only be addressed through original empirical research. In addition to the usual areas of possible impeachment of research studies (i.e. issues dealing with the appropriateness of the experimental design, sample size, bias, confounding, and statistical power) this approach may not always be practical due to time and budget constraints. When feasible, however, research performed to answer specific issues not elsewhere addressed can be helpful in establishing a firm basis for forensic HF/E opinions.

2.4.7. Professional judgment

Sometimes it becomes necessary for a forensic HF/E professional to base an opinion solely on his/her professional judgment and experience when no other foundational source is available. It is helpful where professional judgment is being used as an opinion basis to extrapolate information from previously encountered situations. Generally the more similar the situation is to the current case, the more credible is the expert's opinion. An expert's opinion based solely on his/her judgment is only as good as the experience, credibility and credentials of the expert. Professional opinion is subject to attack, which may include impeachment of the expert's credentials, experience, and testimony in similar cases. Consequently, a legitimate and sometimes necessary basis for expert opinion, professional judgment alone is the hardest to support.

2.5. Expression of Opinion

The field of forensic HF/E requires the practitioner to analyze situations involved in litigation and offer opinions either

supporting or refuting the legal theories being pursued in the case. After the expert has formulated his opinion(s) based on relevant case information, inspections and research, there are a variety of methods in which an opinion may be rendered.

2.5.1. Reports
The first step is usually to provide an oral or, if asked, written report to the client indicating the initial opinion(s) including the strengths and weaknesses of the case. It is important that the forensic HF/E expert present both sides of the case so the client is aware of possible weaknesses. At this point, the client can decide if s/he is interested in pursuing the case further and the HF/E professional can determine if s/he wishes to be named as an expert in the case.

2.5.2. Deposition
A deposition is a formal question and answer period where the opposing attorney has the opportunity to question the designated expert witness under oath. The attorney is most interested in the expert's qualifications, his/her opinion(s), and the bases of the opinion(s). The entire proceedings are typed into a booklet by a court reporter and become part of the case discovery. The testimony carries the same weight as the testimony in a courtroom.

2.5.3. Testimony
If a case does not settle at any point in the legal process, it culminates with a trial before a judge and, typically in the United States, a jury. At this point, the expert witness will be sworn in and questioned by the attorney that retained him/her (direct examination) and the attorney for the opposing side (cross-examination). The cross-examination is usually an attempt to impeach unfavorable testimony. This is typically followed by redirect and recross examination. In some cases an arbitration will be held. An arbitration is a more informal proceeding, similar to a trial, but presided over by an impartial attorney without a jury or court reporter. Unless otherwise agreed, the decision at arbitration is typically non-binding, so the case would continue to trial if the attorneys were not satisfied with the outcome.

3. NEW AND FUTURE DEVELOPMENTS

3.1. Legal versus Scientific Terminology
An issue of increasing concern to forensic HF/E practitioners is the extent to which legal and scientific terminology may differ and the effect this may have on experts' opinions and their credibility. Examples of terms forensic HF/E practitioners are asked to consider in offering opinions include the following: defective, unreasonably dangerous, adequate/inadequate, foreseeable, duty, negligent, open and obvious, reasonable (and prudent) conduct, safe/unsafe, failure to warn, and "beyond a reasonable degree of scientific certainty."

The means for operationally defining constructs has been a cornerstone of HF/E since its very beginnings (Zackowitz and Cohen 1998). It may be argued that operational definitions are essential in order to strengthen the perception of a field's scientific credibility. A true scientific field must have, at the very least, a set of agreed upon terminology and methodologies that are understood by its practitioners and applied uniformly.

3.2. Accreditation and Certification
There is currently a strong movement toward accreditation and certification. Accreditation of academic programs has been established as a means to teach future professionals the core knowledge and skills that define the field of HF/E. As the profession grows toward a distinct knowledge and skill base, it is necessary to be able to demonstrate a minimal standard of professional competency. This is best accomplished through voluntary certification rather than regulatory licensing. A process such as this is taking place throughout the International Ergonomics Association (IEA) federated societies.

3.3. Professional Standards of Practice
Another important endeavor is developing a set of voluntary professional standards of practice. The Code of Practice recently adopted by the Forensics Professional Group (FPG) of the Human Factors and Ergonomics Society (HFES) is reproduced as Figure 3 and is a good first step.

3.4. Forensic HF/E Practice
To date, forensic HF/E practice is largely performed by individual practitioners, working alone. A model our firm has developed follows from more traditional HF/E practice, in that a team of forensic HF/E professionals, each with specialty interests, works within a multidisciplinary network of professionals outside the field, as required by the case. This network of professionals includes design engineers, architects and accident reconstructionists. The evolution of forensic HF/E consulting firms remains a challenge for the future. Another challenge for forensic HF/E practitioners is to continue to publish applied research studies in order to broaden the knowledge base of the practice, as has been well done thus far in the limited domain of product warnings.

Article V – Forensics Practice

Human factors scientists and practitioners do not allow the adversarial system of jurisprudence to affect the quality or integrity of their practice.

Principle 1. Members provide testimony objectively and without bias; their testimony is based on credible data and/or scientific principles; they are prepared to identify the merits and limitations of the data and principles as well as their own capability to interpret those data and apply those principles.

Principle 2. Members avoid impugning the integrity of other expert witnesses without a factual, reasonable, and substantive basis.

Principle 3. Members do not accept fees on a basis contingent on the outcome of the matter.

Principle 4. Members accept that the client is the attorney who engaged them and not the client of that attorney who is party of the suit.

Principle 5. Members avoid discussing the suit with others in a manner that would disclose the caption of the suit or parties involved, absent the permission of the engaging attorney, until the suit is resolved.

Principle 6. Members participating in a suit do not make public statements likely to influence or prejudice the judicial proceedings during their pendency.

Principle 7. Following suit resolution, members do not reveal information detrimental to the litigants' or client's interests, except where they believe silence would breach the greater duty of protecting public health and safety.

Figure 3. Proposed amendment to the Human Factors/ Ergonomics Society code of ethics.

REFERENCES

COHEN, H.H. and LARUE, C., 1997, Advice from an expert witness. *Ergonomics in Design*, 5.

HENRY DREYFUSS ASSOCIATES, 1993, *The Measure of Man and Woman: Human Factors in Design* (New York: Whitney Library of Design).

INTERNATIONAL CONFERENCE OF BUILDING OFFICIALS, 1997, *Uniform Building Code* (*UBC*)(Whittier: ICBO).

KAUFMAN, J.E. and HAYNES, H. (eds), 1987, *IES Lighting Handbook: Application Volume* (New York: Illuminating Engineering Society of North America).

KEETON, W.P. (ed.), 1984, *Prosser and Keeton on the Law of Torts* (5th ed.) (St Paul.: West).

US EQUAL EMPLOYMENT OPPORTUNITY COMMISSION AND US DEPARTMENT OF JUSTICE, 1992, *Americans with Disabilities Act (ADA)Handbook* (Washington, DC: US Government Printing office).

WOODSON, W.E., TILLMAN, B. and TILLMAN, P.L., 1992, *Human Factors Design Handbook* (2nd ed.) (New York: McGraw Hill).

ZACKOWITZ, I. and COHEN, H.H., 1998, Operationally defining a legal term for forensic human factors. In *Proceedings of Ergo Con '98*, San Jose, CA.

Job Demands and Occupational Health

R. Brauchler and K. Landau

Darmstadt University of Technology, Department of Ergonomics,
Petersenstrasse 30, D-64287 Darmstadt, Germany

1. INTRODUCTION

For many years epidemiological research has been trying to establish the association between stresses at the work place and illnesses suffered by workers. It has been customary to identify the workplace by a job title and the worker's illness by a medical diagnosis. This meant that the data were of only limited use, as no specific information, apart from the job title, was available on the tasks performed by the worker and the resulting stresses and strains.

This method of linking occupation and injury is of very limited usefulness for analyzing the health damage that can occur during a full working life. This type of simple association model ignores a variety of influences from both the occupational and private spheres, leaving a broad spread of residual factors that make it impossible to carry out a meaningful comparison of work stresses and state of health.

There is also the fact that industrial medicine has in the past concentrated mainly on creating information systems for industrial doctors, and has largely ignored the conditions at the work place and the stresses involved there. One reason for this is that the medical experts have frequently lacked the apparatus for determining and defining in ergonomic and technical terms the objective stresses present at the work place independently of the worker himself.

It would thus appear sensible to develop an indicator system for classifying occupational demands and medical diagnoses, not only for individual industrial operations or their departments, but also for whole branches of industry. This could serve as an early warning system in the form of a work-related micro-epidemiology for symptoms or diseases in which work is a contributory factor. There is also the question of how far EDP can be used to assist in the identification of industrial micro-epidemiologies.

2. JOB DEMANDS

2.1. Definitions

The term "job demands" is defined as the general work capabilities that a worker needs to perform a given job. Job demands thus indicate how hard and how difficult the job's work content is. They are complemented by non-job-specific, situative factors, e.g. those resulting from the environment.

At an International Labour Organisation (ILO) conference in Geneva in 1950 the four basic job demands were defined in a highly simplified form as:

- Ability, physical and mental.
- Stress, physical and mental.
- Responsibility.
- Environmental conditions.

This so-called Geneva Scheme has gained importance over the past 50 years as a basis for the monetary evaluation of work and
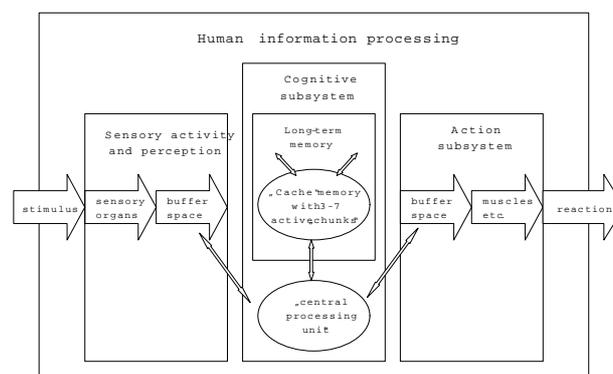
Figure 1. A concept of information processing and ensuing human action (Harmon and King, 1986).

for wage structuring. It is, however, too rough to be used for epidemiological purposes. It would be equally inappropriate to try to formulate demands in the form of personal variables, e.g. overall "ability" or "asset" factors like ability to achieve results or to think abstractly.

Job demand data obtained from the sphere of human functioning and reaction are considerably more useful for the purpose of comparing job analyses with occupational morbidity data (Figure 1).

Man possesses a subsystem functioning through sensory activity and perception. The way in which this system is used depends on:

- the type of information involved;
- the sensory organs involved;
- the sensory dimensions;
- the type of recognition (absolute or relative);
- the fine adjustment and accuracy of the information; and
- possible interference factors.

Determination of these variables for a given work system can yield data on job elements which can then be compared with the occurrence of occupational diseases.

The same applies to the human cognitive subsystem. This system has to select and code incoming information, which is then temporarily stored in "buffer memories." The sooner existing information is processed, the sooner is it possible to recall routine motor reactions that require no further cognitive processing. In contrast, in more complex work tasks, the cognitive system cannot rely solely on the short-term buffer memories. It also has to refer to the long-term memory that one could imagine to be structured in the form of scripts or chunks.

After information has been received and processed, the reaction which has been selected as the most sensible one is converted into the corresponding action for which motor processors are used to trigger muscular reactions. The action also has to be organized and directly controlled.

All three of these subsystems of this highly simplified human action model can thus be defined by more or less filigreed job elements that can be placed in relationship to morbidity data. The sum total of the intensity, duration, sequence and timing of these job elements is referred to as workload.

2.2. Analysis Procedures

For the statistical treatment of the relationship between job and possible disease it is necessary to have analytical data on demands
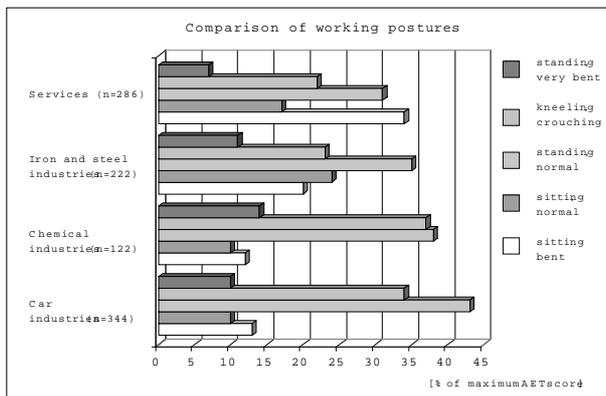
Figure 2. **Comparison of working postures in various branches of industry (Landau, Rohmert, 1992).**

and stresses on at least an ordinal scale. Various analytical procedures have been available for many years. The results obtained with them will vary according to the scientific discipline on which they are based (e.g. ergonomics, work psychology, work study, etc.) and the models used. In addition to job demands, these procedures normally collect data on:

- the work object;
- the work tools and materials;
- the work tasks; and
- the actions involved.

For scientific validity purposes we require this type of analytical procedure to:

- be based on a theoretical model (e.g. a work system model);
- record all job demands as far as this is possible;
- be as economical as possible in application, processing and evaluation;
- be standardizable in use; and
- enable a statement on its reliability.

Provided that they meet these criteria, the procedures are in most cases universally applicable and can be used to compare demands in different work systems. This is an essential condition for the epidemiological processing of data on job demands and stresses.

It will not be possible here to review the wide variety of analytical procedures now available. For further details please refer to the synopses published by Landau and Rohmert (1989) or Brauchler and Landau (1998).

2.3. Specimen Result

The following profile diagrams for various branches of industry (Figure 2) are an example of a demand analysis. The procedure used in this case was the AET, which is also described in this book.

This type of profile indicates the actual stress situations arising and can be used as an aid in the interpretation of medical findings and subjective symptoms.

3. OCCUPATIONAL DISEASE

3.1. Definitions

An illness or disease is generally defined as a malfunction of normal vital processes in organs or body systems caused by a pathogenic stimulus. The human and social costs caused by degenerative diseases and the disabilities resulting from them

are becoming an increasingly heavy financial burden in industrialized countries.

The direct and contingent costs caused by occupational disease can be minimized by the improvement of working conditions, but this can only be achieved by proper analysis of the job demands imposed on the worker and the stresses and strains resulting from them. Such analyses can be carried out not only on existing jobs but also on those that are still at the planning stage. This thinking is by no means new. Studies of the overall economic costs of occupational disease undertaken as far back as 1957 resulted in recommendations for increasing industrial productivity by showing greater understanding of the reasons for problems like low worker output, absenteeism and job changes and realization that the workers' behavior in such cases was frequently motivated primarily by health considerations.

The identification and evaluation of risk factors for specific diseases which may, at least partially, be of occupational origin needs to be performed by qualified ergonomists and industrial medical specialists. Whilst the ergonomist is primarily interested in workload or "stress factors", industrial medicine tends to think solely in terms of risk factors. "Risk" in this context is defined as the probability that a worker will contract a disease for reasons related to his or her job. The risk factors thus indicate the calculable health risk to which a worker is exposed as a result of certain activities or elements inherent in his/her job, e.g. physical traumata from lifting heavy loads or cardiovascular disease from the heavy burden of executive responsibility. The risk that these activities or elements could be causal factors behind an occupational disease which manifests itself after a certain period of time can be classified as either relative or absolute. The type and extent of the relevant disease are determined and characterized by the evaluation of subjective symptoms, medical examination and diagnosis. Although this purely medical approach can in some cases help to identify potentially high-risk jobs, a genuine early warning system enabling systematic action to prevent the occurrence of occupational diseases can only be successful if a clear correlation between ergonomic data on job stresses and clinical data on specific diseases is first established.

Although the term "job-related disease" or "occupational disease" has been legally recognized, the legal definition is both too narrow and inadequate and this has led to its being used too loosely. In 1976 the ILO turned its attention to job-related diseases and started an international project with the aim of formulating an exact definition of the term, investigating causal relationships and developing methods of early diagnosis and prophylaxis to reduce the incidence of these diseases. The ILO expert committee met in 1983 to discuss its findings and these were published in ILO Report TRS 714. This report noted the three following different interpretations of occupational diseases:

- They are a subcategory of job-related diseases.
- They have nothing in common with job-related diseases.
- They are the same thing as job-related diseases.

The most commonly held view is that "job-related disease" is a generic term covering not only occupational diseases and potential occupational diseases, but also borderline cases and non-specific occupational diseases. Other interpretations make similar types of differentiation, using "job-related disease" as a generic term for all scientifically recognized but not yet officially recognized occupational diseases, and also for those occupational diseases

which, although scientifically recognized, do not for various reasons qualify for a disability pension and those acute and chronic diseases where occupational stresses may be a contributory factor in either the emergence or the exacerbation of diseases, functional disturbances or merely general indisposition regularly associated with certain work activities.

Yet another interpretation excluded acute diseases altogether and restricted the use of the term to chronic diseases frequently causing temporary incapacitation where there is a clear association with specific job stresses.

The various scientific disciplines working on this problem have now drawn up a list of criteria for classifying a given disease as "job-related." This stipulates that four conditions must be satisfied:

- It must be the result of a specific influence.
- It must affect specific groups of workers.
- The work performed by the group must be the causal influence.
- The morbidity rate for this disease within the relevant group must be higher than the average for the population as a whole.

This definition thus postulates that job-related diseases are triggered or at least exacerbated by specific job demands or combinations thereof. Whether the demand is a major factor or only one of a number of contributory factors is irrelevant.

In the case of occupational diseases, it must normally be proved that a pathogenic influence emanating from a specific noxious substance is present. The question thus arose as to whether it is possible to apply this standard to job-related diseases where the cause/effect relationships between tasks, demands, stresses, strains and disease are frequently complex and may still be unknown. In order to obtain the required proof, it was first necessary to construct a simple basic model for the prediction of job-related diseases similar to that designed by Robert Koch for infectious diseases.

3.2. Models

Koch's model for predicting infectious diseases postulates that a person will become ill when the three basic variables — host, agent and environment — coincide in both time and place under specific conditions. The host would be the person and the agent could, for example, be lead, tobacco smoke or psychosocial factors. Environmental factors bring the host and the agent together and disease results.

In the case of chronic diseases, however, there will not normally be a single causal agent but rather a variety of factors, some of which could be interrelated with the environment.

Although existing knowledge on the etiology of individual job-related diseases is in many cases fairly sketchy, it can be assumed that e.g. time pressure, noise, recurrent dynamic work (undesirable movements or postures, etc.) or static work may initiate a pathological process that subsequently manifests itself as a occupational disease. The multifactorial nature of occupational diseases has prompted several authors to produce models with which to bring together the often-incongruous observations emanating from different sources on these diseases. One general model describing the possible course of the pathological process is outlined below:

The etiology of every occupational disease can be regarded as a classic pathological process consisting of the following elements:

- Causal factors (like job demands and work environment) which may ultimately lead to the emergence of a more or less specific pathological condition.
- In most cases a complex assortment of stimuli influencing the causal factors.
- Reactions of the worker's organism to the stimuli. At a certain point in time these reactions may develop into a manifest disease. Alternatively, the worker may recover or gradually develop resistance to repeated stimuli.

Both the Koch model and the analogous model described above are purely theoretical and there may therefore be problems in applying them in practice. For one thing, occupational diseases may not be the result of a process in which each successive phase can be predicted from the previous one (cf. dynamic modeling with Petri nets in Brauchler and Landau 1991). Second, it is not known in detail how the reparative processes function. Third, the causal relationships between the various factors are unknown.

Causal relationships between occupational stress and job-related diseases may either be dose-related or function on an all-or-nothing basis. Dose-related situations probably apply, for example, in most cases where physical or chemical risk factors are present; the all-or-nothing relationship, for example, in allergies. Only rarely do these causal relationships involve a single factor and one can therefore normally assume that the pathogenesis will be multifactorial. The individual factors can interact in a variety of ways and this again demonstrates the complexity of the problem:

- The causal factors can form a chain with either a fixed or a variable sequence.
- The factors can have an additive or a multiplicative (mutually potentiating) effect.
- A single factor may be insufficient to trigger a reaction on its own; specific combinations may have to be present.
- The individual factors may be pathogenic independently of each other, i.e. the same illness may be caused by several variables.
- The same factor may cause different illnesses.

Where several factors are present, the effect may correspond to the sum of these factors (the additive model), but could also be substantially higher as a result of synergy between the factors (the multiplicative model).

The Enderlein and Heuchert model is an example of the analogous model outlined above and is generally used for the planning and evaluation of analytical epidemiological studies. This model assumes the cause–effect relationship from first exposure to the stimuli until the emergence of the disease to be a fact. To determine the exact effect of exposure to the stimuli, each individual factor capable of causing the disease needs to be taken into consideration. These pathogenic factors are classified into three categories: effect-modifiers, non-confounders and confounders. The confounders distort the pathogenic process by correlating directly with the occupational exposure, e.g. reception of visual information from VDU screens at work followed by several hours of TV at home in the evening. In cases where there is no interaction between the confounder and the occupational exposure, the confounder effect can be eliminated mathematically. Effect-modifiers invariably interact with the exposure. A frequently recurring effect-modifier is, for example, a person's age. In cases where this is relevant, age-specific data

are needed to estimate the effect of the exposure. Finally, there are the non-confounders which do not directly correlate with the exposure but do increase the number of random variables and thus make it necessary to increase the number of spot checks.

4. EPIDEMIOLOGICAL EXPERT SYSTEMS

4.1. Relationship between Stress and Occupational Disease

For the purposes of predicting job-related health risk on the basis of data obtained from job analyses, the stress–strain concept postulated by Rohmert can be expanded to cover general lifestyle e.g. diet, habits, drugs, etc., and also the worker's social situation and general attitudes, i.e. redefinition of lifestyle and social situation. However, the concept needed to be dynamized to make allowance for the high degree of mobility between occupations and workplaces during the course of a person's working life and a system for determining stresses arising at both existing and previous workplaces had to be developed. In short, a model had to be designed in such a way that it could demonstrate the hypothetical association between job demands and occupational disease. It had to take account of:

- the current situation in the person's working and private life;
- the external factors influencing the worker both as a person and an employee (objective influences). These include the stresses arising from job demands and from the work environment, and also stress factors from the worker's private sphere, i.e. non-job-related;
- the worker's redefinition of the job situation. This will be strongly influenced by lifestyle (e.g. diet and habits) and social situation (attainments, aptitudes, skills and ability); and
- the need for dynamization of the model to take into account situations that have applied in the worker's previous professional career and private life.

The use of this type of multi-causal model to prove the hypothetical relationship between those determinants of the job demands causing stresses, strains and ultimately disease, and thus to predict job-related diseases is discussed in greater detail in publications by Brauchler (cf. Brauchler and Landau 1989, 1991,

1992, 1998, Brauchler *et al.* 1990, Landau and Brauchler 1994) (Figure 3).

The above diagram demonstrates clearly that complex calculations and simulation data will be required to relate clinical findings to the potentially relevant causal factors.

4.2. Instruments

The theoretical model now created for establishing causal relationships in occupational diseases and discussed here will, in conjunction with an appropriate indicator system which already exists in prototype form, provide information on industrial micro-epidemiology. The indicator system uses two highly standardized instruments that have been developed for the evaluation of the relevant data collected by industrial medical specialists and work analysts.

The first is a questionnaire for the objective assessment of stresses resulting from job demands. This uses an analytical approach (e.g. the AET method reviewed in this volume).

The second is a standardized questionnaire used for assessment of the worker's health status, habits and social situation at medical checkups. It contains 250 items covering clinical findings, diagnoses and laboratory results, plus sections for recording the worker's job history and stresses experienced in previous jobs and family and social history. The object of this examination is to minimize residual variances in the correlation between the health data and the job stresses.

The data obtained with these questionnaires make it possible to predict health risks with quantifiable accuracy. A database system named Epidemiological and Stress Estimation System ESES for compiling and analyzing this data has been designed and a special evaluation system using profile and cluster analysis, contingency analysis, multi-field panels and cross-cluster analysis (i.e. analyses of results from different, independently evaluated cluster analyses, for example, stress or disease clusters) has been developed (cf. Brauchler and Landau 1994).

The possibility of using stochastic procedures and the network theory has also been examined and the limitations thereby involved have been determined. These examinations have revealed that only rule-induction algorithms derived from knowledge-based systems are suitable for the creation of a prophylactic industrial health system that takes account of both multi-causality and retrospective aspects.

4.3. Prediction of Occupational Diseases

4.3.1. *Algorithm for determining causal relationship*
The use of an epidemiological early warning system for the prediction of job-related or occupational diseases will involve decisions based on theoretical evaluations and the use of established prophylactic methods to simulate the collection and evaluation of empirical data which normally takes place during ongoing routine medical examinations of workers in high-risk jobs. Expert systems are structured to handle problems by processing available data through appropriate algorithms to identify a strategy for prevention of the problem.

Algorithm-based systems process the potentially pathogenic factors involved in a given job and draw logical conclusions from their knowledge base The prediction system is object-oriented and the conclusions are drawn in the system's second step and based on the assessment of incidence of specific
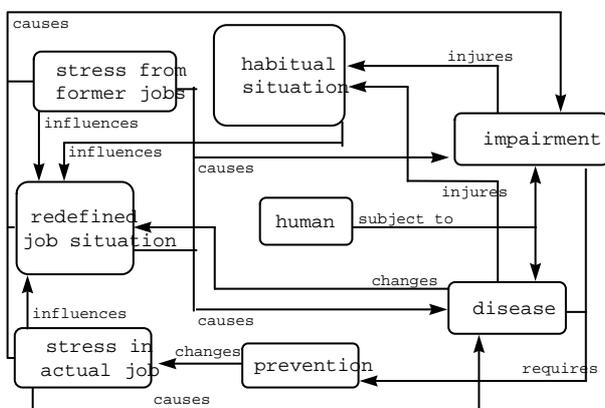
Figure 3. Possible relationship between stress and occupational disease.

diseases in analogous cases stored in the database. In cases where this predicts a potential health risk, a prophylactic strategy can be deduced from the information in the database. In cases where the there are sufficient similarities with previous cases on which data is available, it is also possible to predict the percentage probability with which a person starting work in a given job will contract a specific disease.

This method will only function correctly if two conditions are fulfilled. First, a detailed, standardized analysis of the case must be carried out. Second, an extensive data bank containing the results of previous investigations must be available for purposes of comparison. The prediction system discussed here fulfils both these conditions.

4.3.2. The ESES-prototype-system

Confirming to the model and the instrument shown in Sections 3.3 and 4.2 two data records (estimated stresses in the job and occupational health data) are linked in the ESES (Brauchler and Landau 1992). When predicting potential health risks in a given case, the user of the ESES-prototype-system will first select a disease, then retrieve the information on the causal factors stored in the database and finally obtain a statement of the potential risk in accordance with the following formula:

> *If* these specific causal factors are present, *then* stress X occurring with intensity Y will cause this disease with a percentage probability of Z.

The method establishes a meaningful link between stress data and clinical data and makes it possible to construct an epidemiological expert system which, in turn, facilitates the design and introduction of preventive measures. In larger companies this can be the responsibility of an expert committee of industrial medical specialists, work analysts, ergonomists and safety officers. In smaller companies lacking the necessary qualified personnel and expert knowledge, employee groups can be formed. The nucleus of an expert system is the knowledge base consisting of facts and rules. In accordance with the basic algorithm, a rule stating that performance of a given job for > 10 years increases the health risk is formulated as a pattern-driven system as follows:

> If << work duration > 10 >> then << health risk (person) = increased >>.

There are two possible ways of creating a knowledge base. The direct or inductive method is by the induction of rules. The authors and designers of the system described here gained experience of this method (Landau and Brauchler 1994) when using the ID3 algorithm to create a knowledge base in which rules are induced from available stress and clinical data.

The second or indirect method of creating a knowledge base can be based on:

- working hypotheses derived from the literature;
- the results of statistical evaluations of stress and clinical data; and
- medical precedents.

To ensure that all available data is used for assessment purposes, a comprehensive knowledge base must include data obtained by this second method.

The data contained in object-oriented systems is stored in the form of large numbers of object–slot–value triplets. The objects, i.e. in this case the test persons from the investigations,

are linked with the type and intensity of the potential pathogenic factors to which they are exposed. The links are established either by rule induction or from one of the three sources listed above under the indirect method.

The knowledge base that can be accessed by the user is built up from available case data, which is stored in another database. In the creation of this database special care is taken to ensure that it is kept dynamic by the addition on new case data and the correction of any errors in existing data.

Data on new cases, i.e. workers who are being checked for potential health risks, are input into the system and compared with the existing case database. The user first classifies the person's present and former stress situation on the basis of the stress and clinical data from the questionnaires described above and with the relevant parts of the epidemiological expert system. The system has appropriate menus for this classification procedure. The input data are then transformed automatically (lack of space makes it impossible to give more details of the transformation procedure here) and the system then calculates the incidence rates for specific diseases. The predictions are presented in the form of percentage probability rates for certain diseases and is supplemented by recommendations for preventive action which may range from redesign of the workplace environment to transfer of the worker another workplace.

It is also possible to use fuzzy set logic in knowledge-based systems to predict associations between stresses and diseases with quantifiable accuracy. Experience with this method is at present meager. Further investigations are necessary before any meaningful conclusions can be drawn.

5. CONCLUSIONS

The main objective of applied ergonomics and preventive medicine in industry is the early identification of job-related deterioration in a worker's condition before serious health damage occurs, and thereby the minimization of the human and social costs caused by degenerative diseases and the resulting disabilities. It is essential to collect all available data on potential health risks at the workplace and to evaluate these to produce a database that can be used to supply answers to ergonomic, technical and epidemiological questions.

Future studies carried out in industry should avoid the error of focusing primarily on achieving maximum cost-effectiveness in the evaluation and interpretation of occupational health and stress data, e.g. by limiting the volume of data collected, and should preferably aim to concentrate more on the multicausality of job-related diseases and on the collection of retrospective data on their causes. This will inevitably involve the use of cost-intensive expert system techniques.

REFERENCES

BRAUCHLER, R., BRAUCHLER, W. and LANDAU, K., 1990, ESES — A methodological approach to predict job-related diseases. In K. Noro and O. Brown Jr (eds) *Human Factors in Organisational Design and Management — III* (Amsterdam: Elsevier), 203–6.

BRAUCHLER, R. and LANDAU, K., 1989, Epidemiological analysis of workload data using ergonomic data bases. In *Recent Developments in Job Analysis: Proceedings of an International Symposium on Job Analysis*, University of Hohenheim, 14–15 March (London: Taylor & Francis), 143–54.

BRAUCHLER, R. and LANDAU, K., 1991, Setting up an epidemiological early-warning system — first steps of implementation. In Y. Quéinnec and F. Daniellou (eds) *Designing for Everyone* (London: Taylor & Francis), 640–2.

BRAUCHLER, R. and LANDAU, K., 1992, Implementation of an epidemiological early-warning system by using rule induction algorithms. In W. Karwowski and M. Mattila (eds) *Proceedings of the International Conference on "Computer Aided Ergonomics and Safety"*, Tampere, Finland, 249–54.

BRAUCHLER, R. and LANDAU, K., 1998, Task analysis: Part II — The scientific basis (knowledge base) for the guide. *International Journal of Industrial Ergonomics*, 22, 13–35.

HARMON, P. and KING, D., 1985, *Expert Systems: Artificial Intelligence in Business* (New York: Wiley).

LANDAU, K. and BRAUCHLER, R., 1994, Disease prediction using ergonomic knowledge bases. In *Proceedings of the 12th Triennial Congress of the International Ergonomics Association*, vol. 2 (Toronto: Ergonomics in Occupational Health and Safety), 159–60.

LANDAU, K. and ROHMERT, W., 1989, *Recent Developments in Job Analysis Association*, vol. 2 (Toronto: Ergonomics in Occupational Health and Safety), 159–60.

Legal Considerations for the Human Factors Specialist

G. A. Peters

Peters & Peters, 2001 Wilshire Blvd, Santa Monica, CA 90403-5664, USA

1. INTRODUCTION

The human factors specialist is a very likely candidate for some personal involvement with the legal system. It may be rather indirect, such as helping a company lawyer in answering written Interrogatories or Requests to Produce regarding human factors issues raised in a lawsuit. It may be as a company percipient (personal knowledge) witness who has to testify, under oath, at a Deposition or during Trial as to certain factual matters. It could be to assist (inform) lawyers in a lawsuit regarding human factors (liability mitigation), to recommend how to reduce existing human error or fault (liability correction), or to help develop a proposed human factors program within a company (liability prevention). However, the interaction with the law may be more direct and immediate either as a defendant in a lawsuit or when engaging in forensic activities such as that of an expert witness.

2. PERSPECTIVES AND APPROACHES

A proactive or early acquaintance with the legal process is an imperative for several reasons First, learning to understand and deal constructively with unfamiliar legal processes is best accomplished in a calm deliberate manner, whereas attempts at last-minute quick learning, under high stress and personal fear, can result in unintentional mistakes or a simple failure to comprehend the "big picture" and react appropriately to legal proceedings. Second, the old maxim "ignorance of the law is no excuse" has a special legal reality when such personal ignorance results in improper decisions, harmful actions and unnecessary mistakes (such illegal acts are known as errors and omissions). Third, there is a need to translate existing and future legal requirements into engineering specifications, appropriate research protocols and necessary "scientific" data (foundation) that could justify design decisions, technical conclusions and professional opinions. In essence, it is necessary to develop a fundamental understanding of the law so that it can be quickly enhanced when needed.

The *common* (judge made) *law* is based on some fairly universal basic principles whether applied at the community or local, state or provincial, national or international level. Its application is at the professional level involving discretionary or flexible decision-making. Thus, the "worst case" situation should be formulated for use by human factors specialists. The *regulatory, statute, or legislatively derived law* involves a "compliance" or purely technical (non-discretionary) approach in application. However, with increasing world trade and international harmonization of standards, the mere compliance with domestic regulations may not be sufficient. An attempt should be made to comply with the human factors requirements of all relevant, commonly accepted, and appropriate international standards, regulations, and statutes. Thus, a "best universal compliance" approach can help avoid legal complications (minimize legal exposure) for products, systems and services in the immediate future. For example, if there should be prospective legal fault (liability) for an unsafe design, the human factors specialist should determine the best approach to achieve and assure the highest practical level of safety without degradation by controllable human factors deficiencies (Peters 1996a, b, 1997).

3. BASIC CONCEPTS

The concept of *negligence* is central to most legal systems. Fault must be proven before there can be damage (legal redress). In general, negligence is defined as the failure to exercise ordinary or reasonable care, the kind of "due care" which persons of ordinary prudence would use to avoid injury to themselves and others. The exact words of the definition varies from one jurisdiction to another, but is usually read to the jurors by the judge, based on some form of judicially approved Jury Instructions or Charges (in essence, the operative law of the case). The mission of the jurors is to apply the law, as given to them by the judge, to the facts and circumstances of the case (the proof or evidence that has been presented to them).

The defenses to the negligence allegation (absence of due care) include contributory negligence and assumption of the risk. *Contributory negligence* is the kind of fault on the part of the plaintiff that contributes to the causation of the injury-producing event. It may bar or reduce liability. *Comparative negligence* involves the allocation of fault (often by percentage) to each party or all persons involved in causation. The more recent *pure comparative negligence* provides for the allocation of fault, then the assessment of damages in accordance with the degree of fault for all actors (parties and non-parties to the lawsuit). *Assumption of risk* may also bar or reduce liability (fault); however, the risk assumed by the plaintiff usually must be specific (not general) and voluntarily assumed (not coerced). A defense to a violation of a statute, regulation or standard is a good faith *compliance* (evidence of due care), whereas a relevant *violation* may be a rebuttable presumption of negligence (that is, something that must be refuted to avoid a finding of negligence or fault).

Strict liability is based on the allegation that a failure was caused by a "defect" that resulted in personal injury or property damage. A defect may be defined as "excessive preventable risk," a failure to meet "reasonable consumer expectations" or a similar criteria. There may be a new requirement for an *alternative design* that is economically and technically viable, does not introduce some new unacceptable risk and was known, knowable or available at the time of manufacture of the product. Strict liability was intended as a simplification of the proof necessary to sustain a negligence allegation, but in some jurisdictions it may have regressed to closely match traditional negligence causes of action. A *warranty* allegation may be made, where appropriate, or a claim of intentional or negligent *misrepresentation* (where there has been reliance and harm). In fact, there are many possible "causes of action" that are described in various legal publications or can be elaborated by attorneys at law based on their personal experience.

Engineering malpractice (professional liability, errors and omissions, or personal liability) is generally founded in negligence law (absence of due care). The human factors specialist may be personally named as a defendant if the "error" occurred during private practice. If the specialist was corporate employee there

may be legal protection (insulation from lawsuits) so that only the company is a named defendant. This is important as to insurance coverage. However, even though legally protected, bad practice or improper conduct may still have moral or ethical consequences and should be avoided by the exercise of due care. Avoidance is best practiced by the informed professional.

A key concept is *foreseeability*, which means whether the harm could have been predicted, was known or could have been known. The harm might be personal injury, property damage or some environmental problem. This is similar to predictions made in a failure mode and effect analysis, fault tree analysis, risk assessment or task analysis performed as part of the design process and that are relevant to the manufacture, transport, sale, use and disposal of the product, process, system or service. This serves to encourage efforts to identify and prevent defects, discrepancies and unsafe conditions by reasonable risk reduction efforts. *Reasonable* is a frequent and important modifier of legal concepts. Increasingly, formal risk assessment procedures (that include human factors) are detailed in trade practices and international standards.

4. RESIDUAL RISK

During the design process, a conscientious attempt should be made to eliminate all hazards and their associated risks of harm. Rather than rely upon just the subjective judgment of one engineer or department as to what constitutes a safe design, certain analytic techniques, design reviews and system testing may be prescribed. Hazard prevention is paramount, but situations may occur that require supplemental safety devices or the implementation of enforceable safety oriented procedures. In addition, there may be some level of risk that is deemed tolerable or acceptable. This *residual risk* generally requires the use of warnings, instructions and technical communications to alert, inform, minimize or safeguard against avoidable harm to the ultimate user (target of the warnings or possible victim of the hazard). The most common warnings are those found on the labels and packaging of products, material safety data sheets, insertions in owner's or operator's manuals, training instructions, and post sale letters, advisories, recalls and advertisements intended to reach and inform owners and users. It also includes visual and auditory displays, machine or process status indicators, emergency alarms and computer-generated troubleshooting instructions.

These attempts to control residual risk should be the result of a detailed engineering effort, appropriate human factors analyses and testing to determine their effectiveness. The legal issue is often whether proof exists that appropriate warnings and instructions would have significantly influenced relevant user behavior. If they fail to communicate the intended message they are functionally useless. Would proper instructions have increased the use of protective equipment, helped in the choice of accessory or optional safety devices or given a reasonable choice as to risk avoidance behavior? There is considerable scientific, engineering and legal information regarding the design and utilization of warnings and instructions (Peters and Peters 1999).

5. OPINION TESTING

When human factors specialists act as *expert witnesses* they are forced to function within a set of arbitrary rules, a subculture and a value complex of the court system. The conditional and qualified conclusions of an experimental research scientist may be inappropriate in such a system, where "scientific certainty" may be based on only a "more likely than not" criterion for expert opinions and conclusions. The term "reliable," as applied to evidence, may mean simply a credible basis or factual foundation for an opinion. The term "scientific" is generally very broadly defined under the law.

In the past, an expert witness was only someone who seemed able to help the trier-of-fact (judge or jury) reach a better-informed decision (verdict) by providing specialized evidence (testimony and exhibits). The judge determined whether the prospective expert was "qualified" based on credentials and relevant experience. This "gatekeeper" role has been expanded recently to include a focus on "methodology" as well as basic qualification. The trial judge may consider a variety of factors, as to the necessary foundation or reliability of an area of testimony, including peer acceptance, publications, testability and error rates. In other words, just how speculative or believable is the proffered testimony? It is important that jurors' reliance on the testimony can be justified. While this is usually the subject of cross-examination, the trial judge has been given greater responsibility as a gatekeeper to keep out testimony that lacks proper foundation.

The opinions of the human factors expert may be the result of a comparison with some commonly used criterion, the result of a balancing procedure or the result of reference to other information. In the courtroom, the testimony may be that a human error should be categorized as "foreseeable" and an "excessive preventable danger." However, there are some 30 different criteria utilized, specified or recommended throughout the world. Thus, the criterion might be a "reasonable certainty of no harm," the "best practicable means," etc. Common balancing criteria include "cost-benefit," "risk-benefit," and "risk-utility" analyses. In some countries, evaluations may be made in terms of "quality of life for the individual" and "universal design" for all people, all cultures, and all reasonably foreseeable situations. An example might be whether a warning or instruction is "adequate," when a readability test shows it to be at the college level and a review of the work force, to whom the warning is targeted, is it at or below the 6th grade reading level (Peters and Peters 1999).

6. SUBJECT–MATTER HARMONIZATION

Human factors expert testimony does not stand alone. Generally it is interpreted in the context of other testimony from different disciplines or activities. Similarly, in a company design context there are different academic disciplines, terminology and general knowledge being utilized. The effectiveness of the human factors expert, engineer or specialist depends on whether there has been sufficient harmonization as to the general factual context of the situation, the body of commonly used pre-existing knowledge, and the language (terminology) used for communication in the arena being used for the presentation of human factors knowledge.

There is a remarkable amount of highly specialized information available as background context or general subject-matter content, much of it is written by human factors specialists. There are books on slip and fall, driver vision and perception, alcohol and drug impairment, forensic aspects of highway safety, warnings and instructions, traffic control devices and accident

reconstruction for automobiles, trucks, bicycles, motorcycles, boats, snowmobiles, aircraft, trains, construction equipment and pedestrian accidents. There are company manuals for design, manufacture and testing. There are numerous trade publications and an unbelievable number of standards (including many that focus on human factors). These are a very efficient and low cost means for the rapid acquisition of background information and very helpful in the harmonization process (Peters and Peters 1984–93).

There are less widely used information sources, such as the neuropsychological assessment tests such as the Halstead–Reitan Battery, diagnostic criteria (DSM-IV) and trauma symptom checklists that may augment, supplement or help to explain the human factors testimony, report or communication. Similarly, socio-technical evaluations may be made of local cultures for jury selection and improved advocacy communication. These may be available from other expert witnesses or document files. However, the prime source of specific information for the human factors specialist should be from the experiment-based data, conclusions, principles and information contained in peer-accepted periodicals, transactions and books. An important caveat

is that the trial lawyer, plaintiff or defense may have access to such information to utilize during cross-examination for one purpose or another, so be prepared to some degree. In other words, an attempt should be made to translate the human factors or other information into a context that is acceptable and understandable to the target audience. Subject-matter harmonization may be the key to credibility.

REFERENCES

PETERS, G.A., 1996a, Human error prevention. In G.A. Peters and B.J. Peters (eds) *Sourcebook On Asbestos Diseases*, vol. 12 (Charlottesville: Michie), 207–34.

PETERS, G.A., 1996b, Liability prevention techniques. In *Proceedings of the JSME International Symposium on Product Liability and Failure Prevention*, Fukuoka, Japan, 167–84.

PETERS, G.A., 1997, Risk analysis. *Technology, Law and Insurance*, 2, 97–110.

PETERS, G.A. and PETERS, B.J. (eds), 1984–93, *Automotive Engineering and Litigation*, 6 vols (New York: Wiley).

PETERS, G.A. and PETERS, B.J., 1999, *Warnings, Instructions, and Technical Communications* (Tucson: Lawyers & Judges).

Low-cost Ergonomics Improvements

T. Kawakami

Institute for Science of Labour, 2-8-14, Sugao, Miyammae-ku, Kawasaki, Kanagawa 213, Japan

1. MOMENTUM OF THE LOW-COST IMPROVEMENT

Ergonomic improvements are sustaining and growing action in the place where people are at work. How to support the peoples' improvement initiative is the real agenda for ergonomics. Ergonomic improvements, which can be done at low-cost, have been playing vital roles to strengthen such peoples' improvement initiative in different countries and regions (Kogi *et al*. 1988). Low-cost improvement actions mobilize self-help efforts of workers and managers and cultivate sustaining and growing action. This is why low-cost improvements are essential in ergonomics. Visible achievements are always strong drive force for accelerating the further improvement action.

Cost constraints on upgrading ergonomic standard are commonly seen in many workplaces in different countries. Often important ergonomic improvement ideas are blocked by lack of financial initiative. Low-cost improvements have provided the way to break such cost barriers. Many ways to upgrade the working conditions in various ergonomic areas at low-cost have been paved and contributed to stepping up the existing ergonomic conditions in various workplaces. Even in large-sized enterprises having better financial background, low-cost improvements have been a key issue for increasing the workplace initiative for sustaining action.

2. WORK IMPROVEMENT IN SMALL ENTERPRISES (WISE) PROJECT IN THE PHILIPPINES AND THE ROLES OF LOW-COST IMPROVEMENT STRATEGY

Low-cost improvement strategy played fundamental roles in the Work Improvement in Small Enterprises (WISE) project in the Philippines from 1994 to 1996. The WISE project, technically assisted by the International Labour Organization (ILO) and financially supported by the UN Development Programme (UNDP), was aimed at joint improvements of working conditions and productivity in Philippine small enterprises (Batino 1997). Special emphasis was placed on building on local practice and focusing on existing achievements at low-cost in the Philippine workplaces. WISE identified nine priority technical areas for the improvement including: (1) materials storage and handling, (2) work-station design, (3) productive machine safety, (4) control of hazardous substances, (5) lighting, (6) premises, (7) welfare facilities, (8) work organization and (9) environmental protection. During the 3-year project periods of WISE, 2060 action ideas were proposed by the owners and workers in the participating small enterprises and as high as 1724 proposals (83.7%) were actually implemented at low-cost. Among the nine technical areas, 364 improvements were in materials storage and handling while 134 in workstation design. Ergonomics improve-

ments were the priority area for upgrading the working conditions and productivity.

Typical low-cost improvements carried out in WISE were:
- In "materials storage and handling," introduction of push carts, mobile racks, clear and marked passageways, even floors, multi-level shelves for storage, materials containers, hoists and conveyors
- In "work-station designs," adjusted work height, foot platforms and work item holders, placing frequently used tools and materials within easy reach of workers, jigs and fixtures for holding items, hanging tools and fixed tools, "home" for each tool, chairs of correct height with a sturdy backrest.
- In "productive machine safety," safety guards to moving parts, feeders, labels and signs, regular machine maintenance, clearly visible and easily reached emergency controls.
- In "control of hazardous substances," using safer substituting materials, decreasing the amount of hazardous substances in use, isolation of dust, chemicals, noise or heat out of the workplace, installing screens or partitions to reduce the harmful effects of dust, chemicals, noise or heat, providing workers with safety information of hazardous substances in use.
- In "lighting," maximal use of skylights, painting ceilings and walls in light color, local task lights, elimination of direct glare.
- In "premises," labels for chemical containers, covered containers for organic solvents, paints and glues, increased natural ventilation by having more openings, heat protection of the factory building, appropriate use of personal protective equipment, local exhaust ventilation, safe wiring connectors, fire extinguishers, emergency exits.
- In "welfare facilities," cool and safe drinking water, toilets designated separately for men and women, washing facilities, resting corners, hygienic place for eating meals, first-aid equipment, recreation and game facilities.
- In "work organization," job rotation to avoid monotonous and repetitive nature of work, buffer stock for allowing self-paced work, layout rearrangement for smoother flow of work between different work-stations, insertion of frequent short breaks.
- In "environmental protection," installing a gauge to water valve for saving, decreased amount of chemicals used, reuse or recycling of waste materials.

It was noteworthy that low-cost improvement actions carried out in WISE encouraged the workplace people to mobilize all available materials and skills for ergonomic changes. For example, drum cans and materials containers were made into pushcarts for moving heavy materials. Waste steel bars and wood pieces were used for establishing storage shelves or canteen tables. It was possible for many workers and factory owners to contribute to generating such improvement action by using their skills like carpenter's or welder's. Their practical experiences for developing tangible outputs in upgrading their own working conditions at low-cost have apparently enhanced their sense of participation and increased confidence of sustaining their action. Subsequent follow-up studies showed evidence that low-cost improvements carried out by the local people provided ergonomic benefits such as reduction of local muscular loads, decreased heart

rates during the work and shortened production time (Kawakami and Batino 1998).

3. EXTENSION OF THE LOW-COST IMPROVEMENT STRATEGY

In Thailand, the low-cost improvement programs learning from WISE have frequently been applied. The Occupational Health Training and Demonstration Center belonging to the Thai Ministry of Public Health trained managers and workers in four small enterprises in Samut Prakarn province in the south to Bangkok as a pilot intervention program (Tandhanskul *et al.* 1994). Practical training tools such as action checklists, slides showing local good examples and group work dynamics were used for the training. The follow-up visits to the participating factories after the training counted 63 improvements in multiple technical areas. It was noted that 22 improvements were carried out at no cost and another 23 cost < US$20. Budget limitation was not an absolute barrier for upgrading their working conditions. After confirming the visible benefits by low-cost ergonomic solutions, budget allocation for installing more expensive facilities for upgrading working conditions were facilitated in the participating small enterprises. The Thai Ministry of Public Health has been applying this ergonomic low-cost improvement program to other provinces and increasing the coverage.

The low-cost improvement approaches have been applied to the agricultural sector in Vietnam (Khai *et al.* 1996). Learning from local good examples and applying an action checklist, a variety of the improvement in agricultural working conditions have been attained. This program was named as the Work Improvement in Neighbourhood Improvement (WIND). The WIND program had two unique features. First, WIND is aimed at the application of ergonomic principles for the joint improvements in both living and working conditions. Second, WIND encouraged equal participation of male and female farmers to the improvement action. A follow-up study confirmed 88 improvements by the 20 pairs of husbands and wives who had participated in the pilot training course. The areas of the improvements implemented were: ergonomic kitchen and eating facilities; safe drinking water and sanitary facilities; improvements in work and housing environment; home economy planning and income generation; transporting agricultural product and child care. Among the 88 improvements, 20 could be done at no cost and another 37 < US$10. Ergonomic improvements such as changing methods of carrying heavy agricultural products, or newly designed kitchen facilities with appropriate height could be implemented in collaboration of men and women. It was confirmed that multi-faceted improvements relating to both living and working conditions in the agricultural sector in developing countries were possible using inexpensive, available local resources.

The low-cost improvement strategy have been breaking the cost barrier and facilitated continuing improvement action in industrially developed countries such as Japan as well as industrially developing countries. Large-scale enterprises having better financial sources have also experienced that the low-cost solutions could promote the active shop-floor participation and initiate growing ergonomic improvement action. The common steps taken in the low-cost improvement strategy consist of:

- learning from existing wise solutions and ready-to-use information;
- performing a walk-through survey with an action checklist for identifying feasible low-cost improvement options;
- organizing group discussion for proposing workable improvement ideas and increasing the sense of participation of all people in the workplace;
- finding the low-cost ways for stepping up the existing conditions first rather than aiming at complete solutions;
- implementing improvements by mobilizing available local resources and skills and confirming the improvement benefits; and
- sustaining and increasing participatory actions by regular walk-through and other practical activities.

4. CONCLUSION

The low-cost improvement strategy has provided local workplaces in different countries with large opportunities of participation and motivation. Application of sophisticated ergonomic risk-assessment methods needs experts and equipment. The sophisticated methods are useful for the improvement, but not always available for many local workplaces in different countries and regions. Conventional expert-centered ergonomic approaches have produced valuable recommendation for upgrading their working conditions, but they do not necessarily increase the initiative of the local workplace people. The low-cost improvement strategies apparently have an advantage for supporting locally growing improvement initiative. The sophisticated ergonomic improvements using expert-centered high technology would powerfully contribute to the ergonomic improvement when the low-cost strategy enhanced the dynamic improvement initiative of the local workplace people. It should be noted that the low-cost improvements relying on the local wisdom and initiative are never the second class improvements but the first priority decision for going beyond the existing standards to meet the people's changing needs.

REFERENCES

KOGI, K., PHOON, W.O. and THURMAN, J.E., 1988, *Low-cost Ways of Improving Working Conditions: 100 Examples from Asia* (Geneva: International Labour Office).

BATINO, J.E., 1997, Work improvement in small enterprises (WISE) project in the Philippines. In *Proceedings of the Fifth Southeast Asian Ergonomics Society Conference on Human Factors Vision*, SAES: Kuala Lumpur. pp. 536–41.

KAWAKAMI, K. and BATINO, J.M., 1998, Workplace improvements that can reduce the risk of musculoskeletal disorders – experiences of the WISE approach in the Philippines. *Asian–Pacific Newsletter on Occupational Health and Safety,* SAES: Tokyo.

TANDHANKUL, N., YUANGSAARD, S., PONGPANICH, C., PUNPENG, T., PUNNGOK, A. and KAWAKAMI, T., 1995, Experiences of successful action programmes for occupational health, safety, and ergonomics promotion in small-scale enterprises in Thailand. *Journal of Human Ergology*, 24, 105–15.

KHAI, T., KAWAKAMI, T., TOI, L.M. and KOGI, K., 1996, Improving safety and health of rural sugar-cane factories in the Mekong Delta Area in Vietnam. *Journal of Science of Labour*, 72, 14–22.

Management Perspectives for Workplace Ergonomics

C. F. Dillon

University of Connecticut School of Medicine, USA

1. INTRODUCTION

Ergonomics is the examination of human characteristics and abilities as they impact the design of equipment and jobs. Applied ergonomics aims to improve workplace efficiency, safety and well being. Management is the art of utilizing all available resources to accomplish a given set of production tasks in a timely and economical manner. This article provides an overview and perspectives on management roles in implementing workplace ergonomics programs. Business rationales and need for workplace ergonomics programs, and strategies pertinent to implementation are discussed from a management perspective.

2. GUIDING CONCEPTS AND ISSUES

2.1. The Spectrum of Workplace Injury

Historically workplace injury and safety concerns focused on acute injury, that is, injuries happening suddenly, and with a single obvious cause — for example, accidents from slips and falls, or overt physical trauma resulting from unsafe equipment. Adequate attention to such an obvious problem area came only over an extended time period. The development of scientific ergonomics, and the recognition of "cumulative trauma disorders" or chronically developing workplace musculoskeletal injuries attributable to multifactorial causes provides a more comprehensive perspective of the overall impact of the everyday configuration of working conditions, but, from a management point of view, seemingly adds a new layer of complexity to an already difficult situation. The reality, however is that both acute and chronically developing injuries have always characterized the working environment.

There are existing examples of excellence in corporate safety and loss prevention programs, and such industry leaders have generally experienced less difficulty in integrating the newer ergonomic concepts and technologies into their ongoing work process. Even at this late date, however, the implementation of site specific comprehensive company safety programs to address practical safety concerns relating to acute injury is far from uniform. For companies without a tradition of organized workplace safety efforts, the useful assimilation of ergonomic technology and perspectives to deal with chronically developing musculoskeletal injuries is at best problematic.

The existence of a broader spectrum of workplace injury is unwelcome news and it is understandably difficult for some business managers to accept. Skepticism has been fueled by a public and sometimes very partisan debate about the nature and extent of the problem both within business and public sector and in scientific circles. Business managers should know, however, that the essence of science is critical thinking and debate. On any specific issue, a diversity of opinion typically exists, and it is the overall balance of evidence from multiple studies that should be routinely employed as a benchmark. Current expert scientific consensus (National Research Council 1998) validates a wider view of the spectrum of both acute and chronically developing workplace injury. Further, both workplace and population-based surveys typically indicate that the prevalence of workplace injury is much larger than that ever reported to existing workers' compensation systems by at least an order of magnitude (Morse *et al.* 1998).

For the typical company, therefore, a large pool of individuals with work-related injury typically exists, and only a small portion of this is evident as reported workers' compensation claims. An obvious management fear is that to address such a situation properly, an expanded scope of spending for workers' compensation claims will be required, an unwelcome additional allocation of already scarce economic resources. This is not a new problem in the workplace, however, and the salient question to ask is how the costs associated with this wider spectrum of workplace injury is currently being paid by companies and their employees. The answer, in the typical case, is that there is substantial cost shifting to private medical insurance or other available benefit programs, and to the employee's own economic resources. Indirect costs associated with such injuries are difficult to determine, but are estimated at two to ten times the direct costs paid by workers' compensation.

A company's economic burden that may be engendered by the presence of adverse acute and chronic ergonomic hazards probably includes important contributions to such non-medically-related outcomes as increased absenteeism, rates of personal illness, employee turnover, and decrements in employee morale commitment and productivity. Although the magnitude of problems varies by company, such ergonomic-related factors appear to be widely pervasive in industry: at issue for each management team is whether there is value in rethinking the current system, and whether a more organized management approach based on ergonomic technology integration would be a more effective approach to deal with problems that may exist.

2.2 Historical and Traditional Ergonomic Views

In the early twentieth century, physical workload was a primary concern in the workplace, and the issue of physical fatigue was addressed primarily to increase productivity. Comprehensive motion time studies were wedded to this approach with the aim of creating optimal task organization and efficiency. Historically, general physical workload has been reduced in industry, a process enhanced by mechanization, although notable exceptions to this rule exist to this day. Despite such reductions, musculoskeletal injuries continued to be reported, and it was clear that the then existing injury risk concepts needed revision. Risk assessments made on the basis of whole body exposures were refined to stresses applied to local anatomic areas, and guidelines were further refined. Local estimates of magnitude and duration of local applied forces, postures, repetition rates, and task cycle recovery times were made. An appreciation of the hazards posed by prolonged static posture exposures with selective adaptation of muscle groups developed. These strengthened scientific approaches, while providing improved general understanding; however, with few exceptions, they have not yet provided highly precise guidelines for safe, permissible ergonomic exposure limits comparable, for example, to those currently available for industrial

chemical exposures. In an important sense, ergonomic recommendations, therefore, need to be tailored to the individual case.

2.3   Non–biomechanical Injury Risks

Considerable data has accumulated to indicate that worksite injury risk factors may be broader in scope than the traditionally described physical and biomechanical stresses. There is evidence suggesting that work organizational factors may be pertinent to musculoskeletal injury risk, either as effect modifiers amplifying the effects of purely physical risk factors, as factors that influence the frequency of reporting of workplace injury, or as primary risk factors in and of themselves. This means that a company's work organization and culture, and labor relations issues, will be pertinent to finding solutions for ergonomic problems. Many of such factors are under management control. For example, management controls the pace of production which powerfully influences repetition rates for physical factors. Viewed in this manner, injury resulting from workplace risks may be considered variable costs related to production and sales volumes. Factors such as piecework incentives, limited potential for job rotation, and extensive overtime emerge in many studies as risks for injuries. The possibility that incentives for higher production rates may produce more injuries and other undesirable outcomes, such as an increased rate of defective products and product returns, implies that net productivity is not always synonymous with greater production rates.

Current research indicates that the perceived support of supervisors and co-workers, and employee decision latitude in job performance are all important factors influencing injury rates. Discrepancy between workers and management with respect to their overall assessment of the safety of the workplace is also a very powerful injury risk. Variables consistently associated with lower injury rates in major studies are good working relationships between management and employees, an active role of management in safety programs, lower turnover and higher seniority of the workforce, and company commitment to ongoing safety training incorporating ongoing monitoring of unsafe behaviors and safety audits (Shannon 1998). These types of risk factors are not those associated with traditional engineering solutions to ergonomic problems, but appear to be emerging as powerful predictors of workplace injury. There are important implications for management, especially in the current rapidly changing business environment. For example, in a setting of staff integration in a merger or of planned staff reductions in downsizing, what practical steps can be taken to maintain supervisory and co-worker support systems? Ergonomic factors may have traditionally exerted more subtle influences on the structure of work organizations. For example, current worker seniority systems may have evolved in part as an adaptation to workload, in part functioning to decrease biomechanical exposures on longer-term employees in an organization.

2.4   Applications and the Sufficiency of Ergonomics

Ergonomics has been criticized for not offering pat, definitive solutions to workplace problems. Like all sciences, ergonomics has a body of theory and methodology. While bench lab and engineering research both powerfully contribute to ergonomic theory, ergonomics is, in an important sense, an observational science, meaning that such controlled experimentation is rarely possible in real-life industrial settings. Observational sciences depend on systematic observation of naturally occurring phenomena, and therefore have the handicap that multifactorial situations must be analyzed in toto, without direct control over pertinent causal variables. This is a similar situation to economics, medicine, or even business management. Observational, like experimental, sciences operate by building progressively more adequate explanatory theory; however, all theory, however well documented is considered provisional in the sense that future research is expected to eventually supersede and improve current ones.

Ergonomic theory, as currently developed, represents a powerful available technology, and is rapidly improving. It is currently capable of solving many important problems, though not all the problems one would wish. It will grow still more powerful as time and experience accumulates. Specific business applications of ergonomic theory and technology, however, must be individualized by site, and usually are not superior to the training and experience of those chosen to implement them. Business managers are often concerned about open-ended ergonomic commitments — i.e. they want to know, on a one-time basis, when sufficient attention to ergonomics has been given. If products, production processes, the workforce, and the general business environment were static and ergonomic science fully mature, there might in fact be an ultimate single solution to a company's ergonomic needs. Applied ergonomics, however, is a *process*, and not a fixed goal. Further, it is just one component of a complex, ongoing business process. The appropriate level of a company's attention to ergonomics must be continually rationalized according to an individual company's needs and ongoing local situation assessments, and continued improvements in ergonomic technology.

2.5   The Workers' Compensation Experience

Day-to-day experiences and difficulties with individual workers' compensation injury cases are a compelling reality to managers and businesses. Workers compensation injury management is an important, obligatory focus for management. In the workplace, ergonomics often becomes synonymous with a company's workers' compensation injury experience. As we have seen, this is not altogether true. Ergonomic factors do affect the safety of the workplace, and hence the likelihood of reported injury, but there is good evidence that substantial numbers of work-related injuries are never reported through the workers compensation system. Although they routinely provide important information on prevailing workplace risks, it is also possible that reported injuries may be somewhat unrepresentative of those generally occurring in a production setting, a factor which may introduce bias into plans using workers' compensation injury data to remedy pertinent workplace problems. Also, other important adverse outcomes exist, including employee job dissatisfaction, and effects on absenteeism, personal illness rates, productivity, employee turnover, and production quality problems. Insurance tracking and regulatory scrutiny of industry workers' compensation injury experience, however, in many respects virtually defines the ergonomic debate in both the public and business sectors. Thus, while important information is provided, it can nonetheless be a

narrow and somewhat limited perspective on the overall scope of workplace problems. The workers' compensation experience, however, represents a permanent fixture of the business landscape.

Both government regulators and business ultimately have a similar goal: to reduce workers' compensation injury to a minimum, given current knowledge of ergonomic technology. Government regulators typically perceive workplace injuries as a widespread and pervasive problem, and, based on available research data, feel that industry in the main has been too slow in addressing what is an obvious problem. A central theme is that regulators have used reported workers' compensation injuries to leverage what they see as needed general reform. On the business side, workplace injury is often perceived to be the exception, rather than the rule. Such injuries, and associated workers' compensation reports are often viewed as unpredictable aberrancies arising out of the exigencies of circumstance, with insurance purchased as a remedy for such unpredictable "acts of God". Injuries occurring are seen as somewhat expensive, inconvenient divergences from the everyday production process. There is some suspicion that the very existence of workers' compensation insurance benefits sets up a target for dishonest employees. For individual businesses, the psychological and personnel issues surrounding specific cases is viewed as substantial and a resource drain. An important facet of the business perspective on workers' compensation is that it can be highly individualistic, and thinking on this issue may be heavily influenced by a company's prior experience with individual "high profile" cases. Regulators are generally removed from the daily rigors of the workplace, which can contribute to ideological conflicts with business management. Details of the day-to-day reality of dealing with workers' compensation cases, however meaningful and compelling to business, are less important for regulators who generally hold to a population-based view, where individual experience on average is the principal consideration. From this perspective, the presence of generic, shared ergonomic risks surface as general industry problems to address, and experiences and risks common among injury reports are emphasized.

There are humanitarian and practical necessities to treat workers' compensation injuries when they occur. The frequency of reported injuries may wax and wane, with periodic lulls in reporting tending to reinforce the notion that injuries are the exception rather than the rule. Episodic increases in reported injury rates are dispiriting, especially for responsible managers routinely accustomed to success in other business activities. An intuitive response is to seek improved medical care for those injured. As it currently stands, therefore, the workers compensation injury experience has been heavily medicalized. For many companies, this scenario represents a form of "casualty bias." It should be obvious that medical management of workers' compensation injuries alone, however well perfected, cannot in and of itself be expected to "solve" the workers' compensation problems while pertinent causal workplace risks remain unaddressed. Rather, one must seek to learn from the injury experience, abstracting out lessons that will redress root causes in both the individual and the generic case, and then apply these lessons to production situations generally, so as to prevent future problems from developing, thereby managing the overall level of ergonomic risk. Medical providers treating injured workers must therefore not simply seek to alleviate suffering, but should be routinely expected to provide feedback to employers on the likely biomechanical causes for the observed injuries. The workers' compensation injury experience, therefore, loses substantial value if not coupled with ongoing workplace ergonomic assessments. Managers need to ask themselves with each injury if something new was learned from an ergonomic point of view, that is useful and productive. If this is not the case there is an obvious problem, and difficulties are virtually certain to continue.

3. PLANNING FOR ERGONOMICS

3.1. The Situation Assessment and Ergonomic Management Information Systems

A key question is the relationship, if any, of ergonomics initiatives to company profitability. To understand this question, it is necessary to assess accurately a company's existing current net liabilities, if any, including the scope of both reported and unreported injuries, and estimates of the true costs involved, including workers' compensation and the magnitude of cost shifting to other benefit plans. The magnitude of associated indirect costs should also be estimated. A key management problem, therefore, is to develop a reliable, informative, ongoing "management information" system to provide this ergonomic situation assessment. Traditionally, such situation assessments have been performed using reported workers' compensation data, but, as is already apparent, this results in a somewhat biased underestimate of the typical situation at hand.

The management challenge to develop more reliable ergonomic information systems can be approached in several ways. Businesses typically prefer passive acquisition of data (reviews of existing injury reports and benefits databases) fearing that data collection via active surveys of employees will produce more problems than any information gained may be worth. Planned, active data collection, through selective audits, surveys, or planned surveillance systems, however, is recognized usually to provide a higher quality of information. Naturalistic data collection, for example, by periodic management walk-throughs of plants can also be extremely useful. At a minimum, situations that should typically trigger workplace ergonomic assessments or monitoring include injuries, significant job work process, or task changes, plant equipment refitting, changes in equipment design or the purchase of new equipment, processes or facilities. Ergonomic risk assessments in these typical high risk situations are indicated in order to prevent ergonomic hazards from being inadvertently introduced into the workplace. Follow-up monitoring should be planned to allow feedback control over process and strategy revisions.

3.2. General Economic Perspectives

Conceptually, ergonomics is not a variable expense item in the external business environment, such as the cost of raw materials, that rightfully needs to be managed to the minimum. Nor, despite externally imposed government regulations, does ergonomics represent an external environmental force over which companies have little or no control, such as the prevailing economic conditions, or competitive developments within an industry sector. Rather, ergonomic practices are in fact incorporated into a firm's internal level of operating efficiency, which, in business models, typically represents a directly controllable cost. Ergonomic successes and failures are hence already built de facto

into existing production lines and operating budgets. As this has not typically been done in a systematic rationalized way, current company operating budgets may therefore hide inefficiencies which have heretofore eluded proper evaluation. Operative models for ergonomics in business must therefore move away from postures of denial (ergonomics as an external business environmental factor), and beyond the somewhat individualistic workers compensation model (a "write off" of the individual case as an internal overhead expense), towards a view of ergonomics as an integral part of the production and work organizational environment.

An important business consideration is the effect of ergonomic changes made on marginal productivity. In current business rationalization, each factor in production, including ergonomic ones, receives investment in proportion to its contribution to the production process. Given this operational paradigm, it can be seen that ergonomic changes can provide a return on investment, be economically neutral, or in fact represent an economic liability. In an era of increasing production rates and price competition, management fears are that ergonomic initiatives may constitute a threat to production via reduction in marginal profitability, putting a product "at risk," or even jeopardizing the continued existence of a product line. The skill of the management team and workforce, however, largely determines the economic success or failure of workplace ergonomic innovations, as is the case with developments in production capabilities generally. Such potential threats to marginal productivity appear to be poorly appreciated by many regulators, who tend to view corporations from the perspective of their total size and assets, rather than regarding them as decentralized and diversified entities. Often, from the regulatory point of view, marginal productivity theory appears to hold potentially beneficial ergonomic changes hostage to short-term profitability in the local production context.

3.3. Management Formulation of Ergonomic Programs

General guidance for overall conceptualization of management responsibilities in establishing systems for management of workplace musculoskeletal disorders has been provided in the recent peer reviewed ANSI draft consensus standard (ANSI 1996). Policy in the ergonomic area needs to be developed at the senior board level, and strategic planning coordinated with the existing corporate planning cycles. Major planned ergonomic interventions, for example, should be timed to coincide with new capital equipment acquisitions or annual shutdowns for maintenance and retooling. The management group responsible for organizing a company ergonomics strategy should represent a powerful guiding coalition: ideally, it should incorporate both elements of senior and departmental management and establish accountability, identify existing and needed resources, including experienced ergonomic consultants who can benchmark process operations against ergonomic best practices in industry generally and provide new options for the local ergonomic context. Management should further establish operational assignments, work practices, and responsibilities. Training and communications issues are significant: ergonomics issues typically require education to achieve "buy in" by management and

employees. The overall aim is to achieve an ongoing corporate infrastructure to monitor and respond to ergonomic issues.

Senior management roles include concept validation and facilitation of planned programs, including dealing with key blockers or obstacles to process change. While regulatory thinking places total and direct legal responsibility for workplace safety and ergonomics on businesses and their management, the practical business reality is that supervisors and workers on the shop floor have a direct, major day-to-day responsibility for implementation and success of ergonomics programs. These individuals should also participate in on-site planning of specific ergonomic changes by virtue of their practical experience and the necessity to have their "buy in" for any planned innovations to work. The power for success in workplace ergonomic programs is therefore, to a large extent, inherent in the work group itself, and will be affected by the existing climate of labor relations. Management has authority to institute changes in work practices and policies, and can strategize, lead, and furnish improved technologies, but, in the end, does not itself do ergonomics. There is, therefore, an inherent structural discrepancy between legal and actual practical working responsibility for ergonomics improvements. Guidelines are available for the development of worksite ergonomics programs on an operational level (NIOSH 1997).

Key requirements for success of ergonomic programs, therefore, include senior management planning and commitment, experienced ergonomic input, and supervisor and employee participation. Flexibility is necessary as iterative solutions are frequently required before final solutions are discovered. Major initiatives should be piloted before general introduction. Ergonomic plans must be in synchrony with both short- and long-term corporate goals and plans. The financial and production impacts of ergonomic changes should be continuously documented, and any planned interventions should be appropriate to the level of risk determined by prior situation assessments. Long-term maintenance of a coherent ergonomic initiative in a rapidly changing high production environment, with frequent organizational changes, should be considered.

3.4. Ergonomic Benefits and Cost Opportunities

For many companies, ergonomic cost opportunities exist. In situations where ergonomic and safety issues are poorly managed, insurance costs for worker's compensation represent one such opportunity cost. In fact, some workers' compensation insurers seek out high-risk companies with relatively poorly managed ergonomic and safety programs as an obvious opportunity to profit. Insurance ratings are typically held to prevailing, relatively high levels, while structured workplace safety/ergonomic and improved medical interventions are applied. Over a 2–3-year horizon, enterprising insurers can realize up to 50% profit against premiums ratings in a poorly managed situation.

In the realm of purchasing, many companies have found that, as a direct result of ergonomic assessments, specific standard equipment designs are preferable in terms of productivity, employee safety, and comfort. Standardization in purchasing contracts emphasizing ergonomically preferred products can represent a significant cost opportunity and also puts market forces into play as suppliers recognize ergonomic design as a customer-perceived added value in product differentiation. (It

should be recognized, however, that a certain percentage of individuals will predictably fall outside customary workstation design specifications, and will not be well served by this strategy.) When major capital equipment is purchased specifically to reduce musculoskeletal injuries, proper installation and employee training should be assured, and satisfactory warranties obtained, to help assure the equipment performs as expected.

Employee job satisfaction is an important intangible. From the point of view of employees themselves, general working conditions are a significant concern, and a primary factor in motivation, job satisfaction, and turnover. Ergonomic factors, as they may contribute to prevailing working conditions, are potentially of significance. In most major surveys, working conditions, surprisingly, typically rank somewhat above wages as factors of importance to employees. Development of a company's level of ergonomic sophistication relates not only to internal product stewardship, but can also be useful in product development and differentiation in the marketplace; ergonomic features represent value-added enhancements to products. Ergonomic knowledge can also function to reduce a company's consumer liability risks.

4. SUMMARY

Workplace ergonomics is a substantive issue worthy of management attention. Each management team must make its own positive assessments of need based on objective evidence and minimizing the exigencies of the individual workers' compensation case. An impartial perspective is required to develop and maintain the appropriate level of ergonomic sophistication, skills, and resources in a rapidly changing business environment. The scope of ergonomic problems varies by industry, by individual company, and over time. Some level of problems exist generally in industry, with some companies having major difficulties, and others little trouble. In an era of mature, highly competitive markets and shortened product life cycles, a central question for business is whether there exists a profitable basis for ergonomic technology integration. This question needs to be answered in the individual case. Investments are continually being made in this area, either on a planned, rational basis, or in an unplanned, de facto manner. Those making investments on a strategic basis will be more likely to profit from the experience, and are less likely to leave potentially important uncontrolled ergonomic factors operating in the production environment.

REFERENCES

AMERICAN NATIONAL STANDARDS INSTITUTE (ANSI), 1996, *Control of Work Related Cumulative Trauma Disorders* (Draft for ANSI Z-365), (New York: ANSI).

MORSE, T.F., DILLON, C.F., WARREN, N., *et al.*, 1998, The economic and social consequences of work-related musculoskeletal disorders: The Connecticut Upper Extremity Surveillance Project. *International Journal of Occupational and Environmental Health*, 4, 209–16.

NATIONAL RESEARCH COUNCIL, 1998, *Work-Related Musculoskeletal Disorders: A Review of the Evidence* (Washington DC: National Academy Press).

NATIONAL INSTITUTE of HEALTH and SAFETY (NIOSH), 1997, *Elements of Ergonomics Programs*. (Cincinnati, 117, US Department of Health and Human Services DHHS (NIOSH) Publication No. 97).

SHANNON, H.S., 1998, Organizational and psychosocial factors in work-related conditions. In HARRIS, J.S. and LOEPPKE, R.R., *Integrated Health Management* (Beverly Farms MA: OEM Press) pp.163–9.

WINKEL, J., WESTGAARD, R.H., 1996, A model for solving work related musculoskeletal problems in a profitable way. *Applied Ergonomics* 27, 71–7.

Production Standards and Performance Feedback: a Strategy for Improving Worker Productivity and Satisfaction

B. Das

Department of Industrial Engineering, Dalhousie University, Halifax, Nova Scotia, Canada

1. INTRODUCTION

In repetitive production tasks, the improvement of worker productivity and satisfaction is a major concern for management. Such tasks are often considered as monotonous, boring, and fatiguing, and this in turn may have a detrimental effect on worker satisfaction, job attitudes, and absenteeism. Furthermore, there may be adverse consequences on the physical and mental well-being of workers. The nature of such work often results in muscular tiredness due to stressing specific muscles and the concentration needed to perform a task. The application of suitable strategies to improve worker productivity and satisfaction, especially in repetitive industrial tasks, is therefore of the utmost importance.

A goal is a predetermined standard of success and end result. Goal or standard setting is believed to motivate operators in task performance through directing attention, mobilizing energy expenditure, and prolonging effort. The positive effects of goals on worker performance have been well established. Research studies have demonstrated that specific difficult goals lead to higher levels of performance than "do your best" or easy goals (Awdia *et al.* 1996; Locke and Latham 1984; Phillips and Gully 1997). However, some recent studies have indicated that the typical positive effect of specific difficult goal assignment may not hold for the performance of novel complex tasks (Gilliland and Landis 1992.

Feedback affects performance by motivating adjustment to work output, reinforcing response pattern and direction towards goal. It has been found effective in learning situations and in improving an individual's motivation in performance (Wofford and Goodwin 1990). For goal-setting to be effective, feedback is a necessary condition. Studies have shown that goal-setting with feedback is superior to goal-setting or feedback alone (Das 1982a; Erez 1977).

The goals can be either assigned or set participatively with the worker. Research studies have shown inconsistencies regarding the superiority of participative versus assigned goal-setting (Ludwig and Geller 1997). Some studies indicate that participatively set goals lead to better performances than assigned goals (Latham and Yulk 1975; Latham et al. 1978), while others have given the opposite results or effects (Dossett *et al.* 1979; Latham and Steele 1983). Worker participation in decision-making is the primary means of obtaining commitment to productivity and lowering resistance to change (Latham *et al.* 1994).

Worker satisfaction is an emotional reaction in terms of value responses. Both job content and context factors affect worker satisfaction on the job. Research studies on goal-setting have shown that difficult or hard goals generate less satisfaction and overall acceptance of tasks than easy goals, but that specific difficult goals produce more interest and less boredom than "do-your-best" goals (Kim and Hamner 1976). Locke and Latham (1984) showed that although hardest goal tasks produced the highest performance, they also resulted in the lowest degree of worker satisfaction. Performance feedback gives operators information about their actual performance which can be used by them for action. When feedback is present, it often — but not invariably — elicits a positive response that tasks are interesting and less tiring; it has also been found that worker satisfaction can improve significantly when a specific goal is provided with feedback (Das 1982b; Das and Shikdar 1990).

Monetary incentive is believed to have a positive impact on worker satisfaction, although in the laboratory setting there was no apparent effect on worker productivity and satisfaction (Das 1982a, 1982b). Task success also appears to be directly related to worker satisfaction: Hamner and Harnet (1974) found that subjects who exceeded their reference operators' performance were more satisfied that those who did not.

Goal setting and performance feedback are believed to have a positive impact in making the work more challenging and interesting, and less boring, creating more job attention especially in a repetitive production task. In addition, a task performed at an ergonomically designed workstation is likely to be less fatiguing (Das 1987).

In respect of assigned and participative goal-setting, it is generally believed that participation in goal-setting has a positive impact on worker satisfaction, although some studies have reported that satisfaction declined in both assigned and participative goal-setting conditions (Latham *et al.* 1978, Latham and Yulk 1976): Ivancevich (1976) reported that participative goal-setting was not superior to assigned goal-setting in terms of worker satisfaction.

Most of the studies in the past employed simulated or relatively simple tasks and goals were set arbitrarily or based on an operator's past performance; measured standards were not used. Consequently, the question of "how hard is hard" cannot be defined precisely. For the most part, studies dealt with worker performance, and worker satisfaction was seldom considered. Systematic controlled studies have seldom been performed with a repetitive manufacturing task employing measured standards under an ergonomically designed manufacturing work system to determine the effects of goal-setting and performance feedback, singly or jointly, on worker productivity and satisfaction.

In dealing with the production situation, the term "production standard" is often employed rather than "goal setting". The main objective of this article is to discuss the application of production standards, performance feedback, and monetary incentive for the purpose of improving worker productivity and satisfaction in a repetitive production task.

2. EFFECTS OF PRODUCTION STANDARDS AND FEEDBACK, SINGLY OR JOINTLY, ON WORKER PRODUCTIVITY AND SATISFACTION

For this study, a drill press operation was chosen to represent a
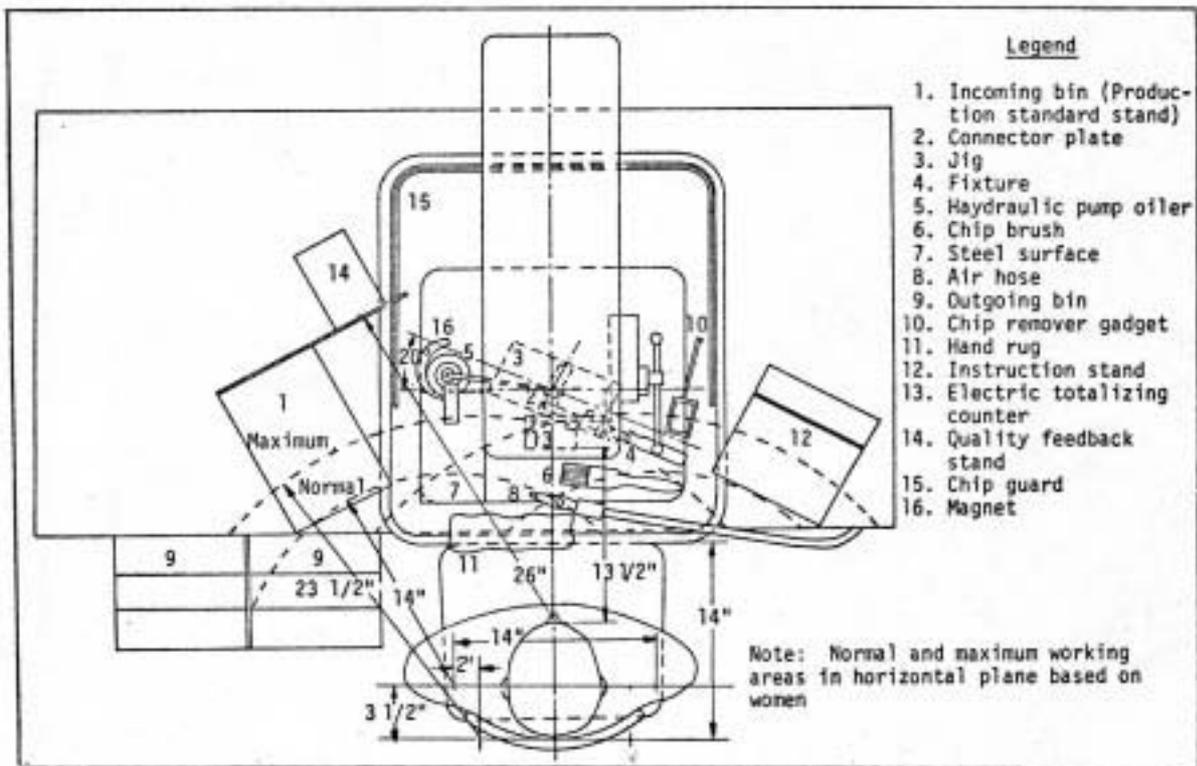
Figure 1. Workplace layout of drill press operation.

realistic repetitive manufacturing task. The task involved an operator drilling four holes into a previously prepared steel plate while he or she was in a seated position. A power feed drill press with jig and fixture was used to produce connector plates in the Machine Tool Laboratory, North Carolina State University (Figure 1).

The production standards were determined by means of methods–time measurement (MTM). The production/time standard for the operation was 240 holes/hour (100% normal). The percentage of cycle time that was machine controlled was 52%. The hard production standard was determined on the basis that external work elements would be performed at a pace of 130% of normal standard to achieve an overall production standard of 112% of normal or 268 holes/hour. An electric totalizing counter was used to provide quantity feedback and the production quality feedback was presented in terms of percentage of good holes drilled in a unit time. Appropriate ergonomic principles and data were employed to facilitate the design of a manufacturing work system (Das 1987).

The operators (56) were trained individually for one hour in the task performance and asked to perform the same task for another hour under a specific work condition. There were seven groups consisting of eight participants in each group. The seven groups were subjected to the following work conditions dealing with production standard (PS) and production feedback (PF):

(1) No PS/PF (Control)
(2) PS: 100% normal
(3) PS: 130% normal
(4) PF: Quantity
(5) PF: Quantity and quality
(6) PS: 130% normal + PF: Quantity and quality

(7) PS: 130% normal + PF: Quantity and quality + Monetary Incentive (MI): 1:1, piece rate.

Worker productivity was determined in terms of production quality (number of holes) and quality (number of good holes) output. Worker satisfaction scores were determined by employing modified Job Diagnostic Survey (JDS) scales (Hackman and Lawler 1971; Hackman and Oldham 1975). The modified JDS scales included the following job or work dimensions:

(1) skill variety
(2) task identity
(3) task significance
(4) autonomy
(5) production feedback
(6) production standard
(7) working conditions
(8) pay

Each participant was asked to answer the questionnaire, which consisted of 18 questions on a seven-point Likert-type scale, regarding his or her perception of the various job attributes that were actually present.

2.1. Production Standards

The provision of an assigned specific quantitative or 100% normal production standard had no significant effect on worker productivity, but had a significant positive effect on worker satisfaction. The hard production standard of 130% normal had no significant effect on worker productivity, but had a highly significant positive effect on worker satisfaction. The assigned hard production standard failed to generate a high level of effort from the operator. The non-acceptance of the assigned hard

standard by 37% of the subjects (Group 3) was probably responsible for such results.

2.2. Production Feedback

Quantity feedback had no significant effect on quantity output, but had a significant positive effect on quality output. Consequently, definitive conclusions could not be made regarding the overall improvement in worker productivity. This work condition had no significant effect on worker satisfaction. Quantity and quality feedback had no significant effect on worker productivity, but had a highly significant positive effect on worker satisfaction.

2.3. Combination of Production Standard and Feedback

The provision of a hard production standard in combination with quantity and quality feedback had a highly significant positive effect on worker productivity and satisfaction. The increases in quantity and quality output were 12.84% and 14.46% respectively, compared to the control group. The average increase in worker satisfaction score was 7.66, compared to the control group. This work condition was far superior to the hard production standard or the quantity and quality feedback alone in terms of worker productivity and satisfaction.

2.4. Combination of Production Standard and Feedback and Monetary Incentive

The conclusions stated above were found to be equally true when the operators were given the same experimental condition as before along with monetary incentive (1:1, piece rate for good holes only with a guaranteed base rate of $3.50 per hour, Group 7). However, a comparison of groups 6 and 7 showed that the monetary incentive had no significant effect on worker productivity, but had a significant negative effect on worker satisfaction. The possible reasons for such results were the perception of the participants that the monetary incentive was inadequate and the investigator was enticing them to high performance levels with insufficient monetary incentive.

3. DETERMINATION OF THE SPECIFIC PRODUCTION STANDARD WITH FEEDBACK TO MAXIMIZE WORKER PRODUCTIVITY AND SATISFACTION

This study was conducted in the machine shop, Technical University of Nova Scotia (Shikdar and Das 1992). Basically, this study employed a similar task and methodology to the previous study. The participants in this study were also college students who were paid $5.00 per hour. For the participative standard, the participant was asked by the investigator to decide upon a production standard above 100% normal that he or she would like to attempt. It was pointed out that operators in the past generally performed between 100% and 150% of normal.

Fifty-six operators participated in this study. The seven groups, with eight participants in each group, were given the following experimental conditions:

(1) No PF/PS (control)
(2) PS: 100% normal + PF
(3) PS: 130% normal + PF
(4) PS: 140% normal + PF

(5) PS: 150% normal + PF
(6) PS: participative + PF
(7) PS: 140% normal + PF + MI

It should be noted that production feedback (PF) included both quantity and quality feedback.

3.1. Assigned Normal Standard and Feedback

The incorporation of an assigned normal (100%) standard and feedback (Group 2) had a significant positive effect on quantity output, but no significant effect on quality output. No conclusive inference could be made, therefore, with regard to the overall improvement in worker productivity, although this work condition had a highly significant positive effect on worker satisfaction.

3.2. Assigned Hard Standard and Feedback

The provision of an assigned hard standard of 130% normal and feedback led to a significantly better performance than an assigned 100% standard and feedback, but no significant difference in worker satisfaction was found between the two work conditions. The incorporation of a still harder assigned standard of 140% normal with feedback led to a further improvement in worker productivity when compared to the provision of an assigned hard standard of 130% normal and feedback. The increases in quantity and quality output for Group 4 were 14.42% and 9.48% respectively, compared to the control group. No significant difference in worker satisfaction was found between the two work conditions. For Group 4, the average increase in worker satisfaction score was 4.00, compared to the control group.

The further increase in the assigned hard standard (150% normal) with feedback failed to increase worker productivity when a comparison was made with the provision of an assigned hard standard of 140% normal and feedback. In fact, a significant deterioration in performance took place. Again, no significant improvement in worker satisfaction was found.

3.3. Participative Standard and Feedback

A comparison between Groups 6 (participative standard + feedback) and 4 (assigned 140% normal + feedback) showed that the worker productivity of the latter group was significantly better than the former. Consequently, the assigned standard of 140% normal (with feedback) is regarded as the best standard for the present repetitive manufacturing task. The provision of a participative standard and feedback was not superior to an assigned normal or hard standard and feedback in terms of worker satisfaction.

3.4. Combination of Best or Assigned 140% Normal Standard and Feedback and Monetary Incentive

The provision of monetary incentive along with an assigned 140% normal standard and feedback (Group 7) had a significant improvement in quantity output when compared to the provision of an assigned 140% normal standard and feedback (Group 4), but the quality output deteriorated significantly. The probable reason for such an outcome could be that the subjects were working faster to reach the (quantity output) standard and did not give adequate attention to maintenance of quality level. Considering both quantity and quality output, the assigned hard

standard of 140% normal with feedback proved to be the best condition for improving worker productivity in this task.

The difference in worker satisfaction score of 1.61 between Groups 7 and 4 was not significant. Hence, monetary incentive had no impact on worker satisfaction. The possible reason for such result was the lack of opportunity to earn a substantial monetary gain in one hour of production time.

4. APPLICATION OF PRODUCTION STANDARDS, PERFORMANCE FEEDBACK AND MONETARY INCENTIVE IN INDUSTRY TO IMPROVE WORKER PRODUCTIVITY AND SATISFACTION

A large fish processing industry was selected to conduct the study (Shikdar and Das 1995). The selected task was a perch fish trimming operation, a highly repetitive production task (Figure 2). The task involved trimming and/or sorting perch fillets into four different product sizes. The work cycle time varied from 3 to 5 minutes. The cycle time involved processing one pan of incoming fillets of about 20 lb. The workers were unionized.

The existing method of operation was standardized for the purpose of the investigation. MTM was used to determine the normal (100%) production standard. The hard standard was set at 140% of normal. The participative standard (above 100% normal) was set by workers individually in consultation with the investigator. Performance feedback was provided in terms of production output (lb/hr, % of normal standard) every two hours on a feedback card in graphical form. Monetary incentive was based on a company policy (1 : 0.6).

The 48 participants (male and female), who had at least 6 months' job experience and a minimum of 7th grade education, took part in the study on a voluntary basis. The exact method of operation to be followed was explained to them. Operators were trained for one day to familiarize them with the standardized method. The participants were randomly assigned to six work groups with eight participants in each group.

The groups received the following work conditions with respect to PS, PF and MI:

Figure 2. Fish trimming workplace layout (all dimensions are in centimeters).

(1) No PF/PS (control)
(2) PS: Assigned 100% normal + PF (quantity and performance level)
(3) PS: Assigned 140% + PF
(4) PS: Participative + PF
(5) PS: Assigned 140% + PF + MI
(6) PS: Participative + PF + MI

The participants worked under the work conditions for 10 days over a ten-month period. Worker productivity was measured in terms of production output (lb/hr) and converted into percentage of normal standard. Worker satisfaction scores were determined by employing the modified JDS scales that included the following job or work dimensions:

(1) skill variety
(2) task identity
(3) task significance
(4) autonomy
(5) production standard
(6) performance feedback
(7) working conditions
(8) pay
(9) promotion
(10) supervision
(11) co-workers

Each participant was asked to answer the questionnaire, which consisted of 24 questions on a seven-point Likert-type scale regarding his or her perception of the various job attributes that were actually present.

4.1. Assigned Production Standards and Performance Feedback

The provision of an assigned normal (100%) standard with feedback (Group 2) was significantly superior to the control group. The improvement in worker productivity was 12.5% compared to the control group. The assigned hard (140%) standard with feedback (Group 3) improved worker productivity significantly beyond the normal standard with feedback. The worker productivity improvement was 8.9%. A comparison between Groups 2 and 1 showed that the provision of an assigned normal production standard and feedback had a highly significant positive effect on worker satisfaction. The provision of an assigned hard standard (140%) with feedback (Group 3) was significantly better than Group 1. No significant difference was found between Groups 3 and 2. Stated otherwise, an assigned hard standard with feedback did not improve worker satisfaction over an assigned normal standard and feedback.

4.2. Participative Production Standard and Performance Feedback

The participative standard with feedback (Group 4) was significantly better than the assigned hard (140%) standard with feedback. The improvement in worker productivity was 19.1%. A comparison between Group 4 and 1, showed that the provision of a participative standard with feedback improved worker satisfaction significantly. Highly significant differences were found between Groups 4 and 2 and Groups 4 and 3. Worker satisfaction for the participative condition was significantly better than the assigned normal (100%) or assigned hard (140%) standard conditions. Participation in setting a standard or decision making probably lead to a higher level of worker satisfaction.
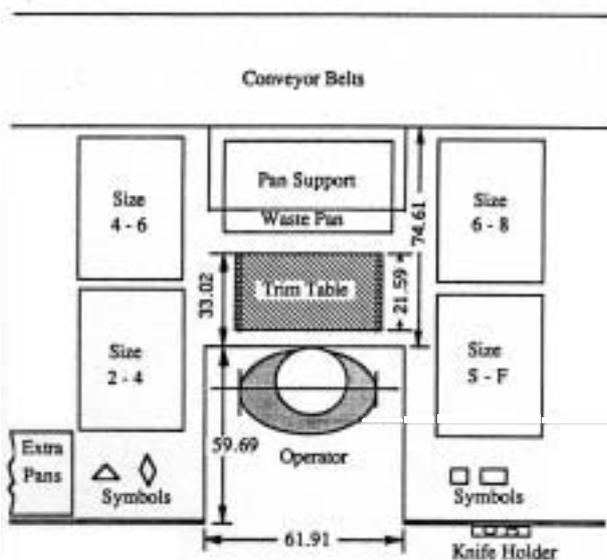
4.3. Monetary Incentive in Combination with Production Standards and Performance Feedback

The monetary incentive when combined with (1) an assigned hard (140%) standard and feedback (Group 5), and with (2) a participative standard and feedback (Group 6), produced further improvement in worker productivity compared to conditions (1) and (2) without monetary incentive. The respective worker productivity improvements were 14.7% and 8.3%. The provision of monetary incentive in conjunction with participative standard and feedback provided the maximum improvement in worker productivity of 57.9%, when compared to the control group.

As disclosed by the comparison of Groups 5 and 1, the provision of monetary incentive in conjunction with an assigned hard standard and feedback had a highly significant positive effect on worker satisfaction. However, no significant difference was found between Groups 5 and 3. Consequently, monetary incentive had no effect on worker satisfaction. A comparison between Group 6 (participative + feedback + monetary incentive) and Group 1 showed that worker satisfaction improved significantly as a consequence of a participative standard with feedback and monetary incentive. There was a significant negative difference in worker satisfaction between Groups 6 and 4. In other words, monetary incentive had a significant negative effect on worker satisfaction when provided with a participative standard and feedback.

5. SUMMARY AND CONCLUSIONS

Only the combination of an assigned hard production standard (130% normal) in the presence of production feedback on quantity and quality had a significant positive effect on both worker productivity and satisfaction. The increases in quantity and quality output were 13% and 15%, respectively, compared to the control group. The provision of an assigned hard standard alone failed to improve worker productivity; however, it had no adverse impact on worker satisfaction. The incorporation of quantity and quality feedback had no significant effect on worker productivity, but had a favorable impact on worker satisfaction. Monetary incentive added no incremental performance or satisfaction gain.

A progressive increase in the assigned standards (100%, 130%, 140%, and 150% normal) with feedback (quantity and quality) improved worker productivity significantly up to the provision of an assigned standard of 140% normal and feedback. The maximum increases in quantity and quality output were 14% and 10%, respectively, compared to the control group. No further improvement in worker productivity resulted as a consequence of an assigned harder standard of 150% normal and feedback. The provision of a participative standard with feedback was significantly inferior to an assigned 130% normal standard and feedback in terms of worker productivity. The provision of assigned and participative standards with feedback improved worker satisfaction significantly. No significant difference in worker satisfaction was found among the standards and feedback conditions. Monetary incentive had no beneficial effect on worker productivity and satisfaction.

An assigned normal (100% standard with feedback improved worker productivity significantly in industry. The assigned hard (140%) standard with feedback improved worker productivity

significantly beyond the normal standard with feedback. The participative standard with feedback was significantly superior to the assigned hard standard with feedback. The monetary incentive when provided with the hard and participative standards along with feedback improved performance significantly further. The maximum improvement in worker productivity was found for the monetary incentive, participative standard, and feedback condition. The improvement was 58% compared to the control group. The provision of assigned normal, assigned hard, and participative standards with feedback improved worker satisfaction significantly in a repetitive production task when compared with the control condition. The participative standard with feedback proved to be significantly superior to other work conditions in terms of worker satisfaction. Monetary incentive added no incremental satisfaction gain.

REFERENCES

AWDIA, G., BROWN, K.G., KRISTOF-BROWN, A. and LOCKE, E.A., 1996, Relationship of goals and microlevel work processes to performance on a multipath manual task. *Journal of Applied Psychology*, 81 (5), 483–97.

DAS, B., 1982a, Effects of production feedback and standards on worker productivity in a repetitive production task. *IIE Transactions,* 14(1), 27–37.

DAS, B., 1982b, Effects of production feedback and standards on worker satisfaction and job attitudes in a repetitive production task. *IIE Transactions,* 14(3), 193–203.

DAS, B., 1987, An ergonomic approach to designing a manufacturing work system. *International Journal of Industrial Ergonomics*, 1(3), 231–40.

DAS, B. and SHIKDAR, A.A., 1990, Applying production standards and feedback to improve worker productivity and satisfaction in a repetitive production task. *IIE Transactions*, 22(2), 124–32.

DOSSETT, D.L., 1983, Importance of participative goal setting and anticipated rewards on goal difficulty and job performance. *Journal of Applied Psychology*, 1978, 63(2), 163–71.

DOSSETT, D.L., LATHAM, G.P. and MITCHEL, T.R., 1979, Effects of assigned versus participatively set goals on employee behaviour when difficulty is held constant. *Journal of Applied Psychology*, 64(3), 291–8.

EREZ, M., 1977, Feedback: A necessary condition for goal setting-performance relationship. *Journal of Applied Psychology*, 62(5), 624–7.

GILLILAND, S.W. and LANDIS, R.S., 1992, Quality and quantity goals in a complex decision task: strategies and outcomes. *Journal of Applied Psychology*, 77(5), 672–81.

HACKMAN, J.R. and LAWLER, E.E. III, 1971, Employee reaction to job characteristics. *Journal of Applied Psychology Monograph*, 55(3), 259–89.

HACKMAN, J.R. and OLDHAM, G.R., 1975, Development of job diagnostic survey. *Journal of Applied Psychology*, 60(2), 159–70.

HAMNER, C.W. and HARNET, D.L., 1974, Goal setting, performance and satisfaction in an independent task. *Organizational Behaviour and Human Performance,* 12(2), 217–30.

IVANCEVICH, J.M., 1976, Effects of goal setting on performance and satisfaction. *Journal of Applied Psychology,* 61(5), 605–12.

KIM, J.S. and HAMNER, W.C., 1976, Effects of performance feedback and goal setting on productivity and satisfaction in an organizational setting. *Journal of Applied Psychology,* 61(1), 48–57.

KNAFER, R., ACKERMAN, P.L., MURTHA, T.C., DUGDALE, B. and Nelson, L., 1994, Goal setting, conditions of practice, and task performance: a resource allocation perspective. *Journal of Applied Psychology,* 79(6), 826–35

LATHAM, G.P. and STEELE, T.P., 1983, The motivational effects of participation versus assigned goal setting on performance. *Academy of Management Journal,* 26(3), 406–17.

LATHAM, G.P. and YULK, G.A., 1975, Assigned versus participative goal setting with educated and uneducated woodworkers. *Journal of Applied Psychology,* 60(3), 299–302.

LATHAM, G.P. and YULK, G.A., 1976, Effects of assigned and participative goal setting on performance and job satisfaction. *Journal of Applied Psychology,* 6(2), 166–71.

LATHAM, G.P., MITCHEL, T.R. and DOSSETT, D.L., 1978, Importance of participative goal setting and anticipated rewards on goal difficulty and job performance. *Journal of Applied Psychology,* 63(2), 163–71.

LATHAM, G.P., WINTERS, D.W. and LOCKE, E.A., 1994, Cognitive and motivation effects of participation: a mediator study. *Journal of Organizations Behavior* 15, 49–63.

LOCKE E.A. and LATHAM, G.P., 1984, *Goal Setting: A Motivational Technique that Works* (New Jersey: Prentice-Hall).

LUDWIG, T.D. and GELLER, E.S., 1997, Assigned versus participative goal setting and response generalization: managing injury control among professional pizza deliverers. *Journal of Applied Psychology,* 82 (2), 253–61.

PHILLIPS, J.M. and GULLY, S.M., 1997, Role of goal orientation, ability, need for achievement and focus of control in the self-efficacy and goal-setting process. *Journal of Applied Psychology,* 82 (5), 792–802.

SHIKDAR, A.A. and DAS, B., 1992, The effects of assigning progressively higher production standards with feedback on worker productivity and satisfaction. *International Journal of Human Factors in Manufacturing,* 2(4), 325–37.

SHIKDAR, A.A. and DAS, B., 1995, A field study of worker productivity improvements. *Applied Ergonomics,* 26(1), 21–7.

WOFFORD, J.C. and GOODWIN, V.L., 1990, Effects of feedback on cognitive processing and choice of decision style. *Journal of Applied Psychology,* 75(6), 603–12.

Psychosocial Work Factors and Work Organization

P. Carayon†‡, M. C. Haims‡ and C.-L. Yang‡

†Ecole des Mines de Nancy, Parc de Saurupt, F-54042 Nancy Cedex, France
‡University of Wisconsin-Madison, Department of Industrial Engineering,
1513 University Avenue, Madison, WI 53706, USA

1. INTRODUCTION

The emergence of macro-ergonomics has strongly contributed to the increasing interest in psychosocial work factors in the ergonomics field (Hendrick 1991, 1996). Work factors can be categorized into the individual, task, tools and technologies, physical environment and the organization (Smith and Carayon-Sainfort 1989). They can also be described as either physical or psychosocial (Cox and Ferguson 1994). The importance of psychosocial work factors in the field of ergonomics emerges from several considerations (Carayon and Lim 1998).

Physical and psychosocial ergonomics are interested in the same job factors. For instance, job rotation and job enlargement are job design strategies aimed at improving psychosocial work factors (increased task variety) and physical factors (reduced exposure to high physical load).

Physical and psychosocial work factors are related to each other. For instance, high work pressure may induce physical stressors, such as increased work pace and repetitiveness.

Psychosocial work factors play an important role in physical ergonomics interventions (e.g. participatory ergonomics).

Physical and psychosocial work factors are related to the same "outcome," for instance, work-related musculoskeletal disorders (see article on "Applications of Community Ergonomics").

2. DEFINITIONS

Within the past decade, the role of psychosocial work factors in ergonomics has gained much popularity. However, the term "psychosocial work factors" has been used "loosely" to define and represent many factors that are "a part of," "attached to" or "associated with" the individual. Some would consider what has been traditionally termed socio-economic factors such as income, education level and demographic or individual factors (e.g. age and marital status) as part of the psychosocial factors. To understand psychosocial factors in the workplace, one needs to take into account the ability of an individual to make a psychological connection to his or her job, and thus formulating the relationship between the person and the job. For instance, the International Labour Office (ILO 1986) defines psychosocial work factors as "interactions between and among work environment, job content, organizational conditions and workers' capacities, needs, culture, personal extra-job considerations that may, through perceptions and experience, influence health, work performance and job satisfaction." Thus, the underlying premise in defining psychosocial work factors is the inclusion of the behavioral and psychological components of job factors. below, we will use the definitions proposed by Hagberg et al. (1995) because they are most highly relevant for ergonomics.

Work organization is defined as the way work is structured, distributed, processed and supervised (Hagberg et al. 1995). It is an "objective" characteristic of the work environment, and depends on many factors, including management style, type of product or service, characteristics of the workforce, level and type of technology, and market conditions. Psychosocial work factors are "perceived" characteristics of the work environment that have an emotional connotation for workers and managers, and that can result in stress and strain (Hagberg et al. 1995).

Psychosocial work factors are multiple and various (Cooper and Marshall 1976), and are produced by different, interacting aspects of work. The Balance Theory of Job Design (Smith and Carayon-Sainfort 1989) proposed a conceptualization of the work system with five elements interacting to produce a "stress load." The five elements of the work system are: (1) the individual, (2) tasks, (3) technology and tools, (4) environment and (5) organizational factors. The interplay and interactions between these different factors can produce various stressors on the individual which then produce a "stress load" which has both physical and psychological components. The stress load, if sustained over time and depending on the individual resources, can produce adverse effects, such as health problems and lack of performance. Research and practice in the field of work organization has demonstrated that considering only a small number of work factors can be misleading and inefficient in solving job design problems. According to the Balance Theory, psychosocial work factors are multiple and of diverse nature. Table 1 shows a selected sample of the many different dimensions of jobs (e.g. job demands, control, social support) that have been studied extensively. Furthermore, each dimension is made up of different facets that define and operationalize that particular dimension. For example, as shown in Table 1, the dimension of job demands consists of various facets, such as quantitative workload, variance in workload, work pressure and cognitive

Table 1.   Selected Psychosocial Work Factors and their Facets

1. Job demands	quantitive workload
	variance in workload
	work pressure
	cognitive demands
	emotional demands
2. Job content	repetitiveness
	challenge
	utilization and development skills
3. Job control	task/instrumental control
	decision/organizational control
	control over physical enviroment
	resource control
	control over work pace: maching- pacing
4. Social interactions	social support from supervisor and colleagues
	supervisor complaint, praise, monitoring
	dealing with (difficult) clients/customers
5. Role factors	role ambiguity
	role conflict
6. Job future and career issues	job future ambiguity
	fear of job loss
7. Technology issues	computer-related problems
	electronic performance monitoring
8. Organizational and management issues	participation
	management style

demands; the dimension of job content includes repetitiveness, challenge on the job and utilization of skills.

3. MEASUREMENT OF PSYCHOSOCIAL WORK FACTORS

From the ergonomics point of view, the purpose of examining psychosocial work factors is to investigate their influence on and role in worker health and well-being. Thus, psychosocial work factors can be considered as predictors (i.e. independent variables) while worker health and well-being serve as the dependent variables or outcomes. The most often used method for measuring psychosocial work factors in applied settings is the questionnaire survey. Difficulties with questionnaire data on psychosocial work factors are often due to the lack of clarity of the definitions of the measured factors or poorly designed questionnaire items that measure "overlapping" conceptual dimensions of the psychosocial work factor of interest. Measures of any one facet typically include several items that can be grouped in a "scale." Reliability of the scale is often being assessed by the Cronbach a score method in which the intercorrelations among the scale are items are examined for internal consistency. In general, it is recommended that existing, well-established scales be used to ensure the "quality" of the data (i.e. reliability and validity) and to be able to compare the newly collected data with other groups for which data has been collected with the same instrument ("benchmarking").

The level of objectivity/subjectivity of the measures of psychosocial work factors will depend on the degree of influence of cognitive and emotional processing. For example, ratings of work factors by an observer cannot be considered as purely objective because of the potential influence of the observer's cognitive and emotional processing. However, ratings of work factors by an outside observer can be considered as more objective than an evaluative question answered by an employee about his/her work environment (e.g. "how stressful is your work environment?"). On the other hand, self-reported measures of psychosocial work factors can be more objective when devoid of evaluation and reaction (Kasl and Cooper 1987). Any data can be placed somewhere on this objectivity/subjectivity continuum from "low in dependency on cognitive and emotional processing" (e.g. objective) to "high in dependency on cognitive and emotional processing" (e.g. subjective) (see article on "Survey Design").

Presented are three different questionnaires that include numerous scales of psychosocial work factors. Validity and reliability analyses have been performed on all three questionnaires. Two questionnaires have been developed and used to measure psychosocial work factors in various groups of workers or large samples of workers: (1) the NIOSH Job Stress questionnaire (Hurrell and McLaney 1988), and (2) the Job Content Questionnaire (JCQ) (Karasek 1979). The NIOSH Job Stress questionnaire is often used in the Health Hazard Evaluations performed by NIOSH. Translations of Karasek's JCQ exist in many different languages, including Dutch and French. The third questionnaire, the University of Wisconsin Office Worker Survey (OWS), was developed to measure psychosocial work factors in office/computer work (Carayon 1991). It covers a wide range of psychosocial work factors of importance in office and computer work. In addition to many of the psychosocial work factors

measured by the NIOSH Job Stress Questionnaire or Karasek's JCQ, the OWS measures psychosocial work factors related to computer technology, such as computer-related problems (Carayon-Sainfort 1992). The OWS questionnaire has been translated in Finnish, Swedish and German. For all three questionnaires data exist for various groups of workers in numerous organizations of multiple countries. These data can serve as comparison to newly collected data and for benchmarking. Numerous other questionnaires for measuring psychosocial work factors exist, such as the Occupational Stress Questionnaire in Finland (Elo *et al.* 1994) and the Occupational Stress Indicator in England (Cooper *et al.* 1988). Other questionnaires are listed in Cook *et al.* (1981).

4. PSYCHOSOCIAL WORK FACTORS AND WORK ORGANIZATIONS

The role of "psychosocial work factors" in influencing individual and organizational health can be dated back to the early days of work mechanization and specialization, and the emergence of the concept of division of labor. Taylor (1911) expanded the principle of division of labor by designing efficient work systems accounting for proper job design, providing the right tools, motivating the individuals, and sharing of responsibilities between management and labor, and sharing of profits. This is known as the era of Scientific Management in which scientific methods measure work objectively with the objective of improving its efficiency. An analysis of psychosocial work factors in the Scientific Management system reveals that skill variety is minimal, workers have no control on the work processes, and the job is highly repetitive and monotonous.

In the next step of development in the Job Design field, we see the emergence of the Human Relations movement (Mayo 1945) which raised the issue of the potential influence of the work environment on an individual's motivation, productivity, and well-being. Thus, job design theorists incorporated worker behavior and work factors in their theories. The two theories of job enlargement and job enrichment formed the basis for many job design theories thereafter. Job enlargement theory emphasized giving a larger variety of tasks or activities to the worker. While this was an improvement from the era of Scientific Management, the additional tasks or activities could be of a similar skill level and content: workers were performing multiple tasks of the same "kind." This has been called "horizontal loading" of the job, as the opposite of job enrichment, which focused on the "vertical loading" of the job. Job enrichment aims at expanding the skills used by workers, while at the same time increasing their responsibility. Herzberg (1966) defined intrinsic and extrinsic factors (or motivation versus hygiene factors) that are important to worker motivation, thus leading to satisfaction or dissatisfaction, and psychological well-being. Intrinsic factors are related to the work (or job) conditions, such as having additional control over work schedules or resources, feedback, client relationships, skill use and development, better work content, direct communications, and personal accountability (Herzberg 1974). Extrinsic factors are related to aspects of financial rewards and benefits and also to the physical environment. Herzberg indicated that extrinsic factors could lead to dissatisfaction with work, but not to satisfaction, while intrinsic factors could increase

satisfaction with work. In a way similar to Herzberg's job enrichment theory, the Job Characteristics Theory (Hackman and Oldham 1976) focused on specific characteristics of the job (i.e. skill variety, task identity, task significance, autonomy and feedback) in combination with individual characteristics (growth need strength) which can determine personal and work outcomes.

The Sociotechnical Systems Theory recognized two inter-related systems in an organization: the social system and the technical system. The main principle of the Sociotechnical Systems Theory was that the social and technical systems interact with each other, and that the joint optimization of both systems can lead to increased satisfaction and performance. The social system focused on the workers' perception of the work environment (i.e. job design factors) and the technical system emphasized the technology and the work processes used in the work (e.g. automation, paced systems, and monitoring systems). It was Davis (1980) who provided a conceptual framework and a set of principles that formulated the Sociotechnical theory. His framework called for a flattened management structure that would promote participation, interaction between and across groups of workers, enriched jobs, and most importantly, meeting individual needs. The Sociotechnical Systems Theory proposed a new form of work organization: semi-autonomous work groups (Trist and Bamforth 1951). Recently, various forms of teamwork have been proposed and applied, from temporary teams (e.g. quality circles, project teams) to permanent teams (e.g. semi-autonomous work groups, self-managed teams).

Other trends in businesses include the widespread development of quality improvement strategies, such as Total Quality Management (TQM). TQM can involve important changes in the way work is organized (Smith *et al.* 1989). TQM can lead to positive and/or negative changes in psychosocial work factors (Sainfort *et al.* 1997). In addition, a basic principle of Total Quality Management is employee participation, which is considered as a positive psychosocial work factor.

The development of Information and Communication Technology (ICT) has also contributed to the emergence of new forms of work organization, such as telework and teamwork via computer-mediated communication. Telework or working at home is most common for clerical workers performing routine transactions and for autonomous professionals (e.g. writers, designers) (Sproull and Kiesler 1991). In terms of psychosocial factors, there are potential positive and negative aspects of ICT that can influence quality of working life and health. ICT can have both positive and negative effects on work demands. One example is in the context of telework or remote work. On the one hand, telework allows for increased control over work pace and variability of workload. It has been found, however, that electronic communication and telework have led to feelings of not being able to get away from work and to the augmentation (rather than substitution) of regular office hours (Sproull and Kiesler 1991, Phizacklea and Wolkowitz 1995). The work place and the home are one in the same, and constant access to technology eliminates time boundaries for work. Although there is a potential negative impact in terms of work demands, the freedom and flexibility offered by telework is a major advantage, especially for individuals with children and other non-work responsibilities (Phizacklea and Wolkowitz 1995). The ability to schedule work around other crucial responsibilities increases job

control and helps to reduce role conflict. Remote work does, however, lead to decreased social interaction with work peers and sometimes decreased career mobility because of the lack of informal, social networks developed within the organization (Sproull and Kiesler 1991).

Jobs in which people use ICT may require high mental effort. Some types of computer-mediated tasks may increase information-processing requirements and place great demands on attention, decision-making, and memory. Increased levels of cognitive demands due to ICT have been shown to influence employee stress and health (Lindstrom and Leino 1989, Czaja and Sharit 1993, Yang 1994). Studies have shown that several types of cognitive demands can be generated from the use of ICT: (1) great amount of information given in a certain unit of time, (2) abstract information being presented on the screen and (3) difficult and concurrent tasks being performed at the same time. In addition, the characteristics of the ICT, such as variability of system response time, can affect human's physiological and psychological responses. Cognitive demands can be increased when the system response time is poor and the nature of workflow is not transparent to the workers. In other words, unpredictable demands and interruptions of workflow caused by system breakdowns may be difficult to deal with because of the disruptive effect on cognitive control process. Overall, cognitive demands are associated with job characteristics such as intensity of computer work, the type of communication, and high speed/functions of computers. The implementation of ICT in work organization can lead to greater demands on cognitive resources in terms of memory, attention, and decision-making that may have a negative impact on worker health and work performance.

The positives and negatives of ICT are ultimately not inherent to the technology itself, but rather are effects of the organizational structures and policies (i.e. work organization) under which the technology resides (Smith *et al.* 1981, Sproull and Kiesler 1991, Carayon and Lim 1994). New computer-based technology has prompted some managers to reinforce hierarchical organizational structures by controlling information exchange, blocking certain channels of communication and enhancing surveillance (e.g. electronic performance monitoring). The same technology has prompted other managers to initiate a new management style characterized by a flexible, continuous learning work environment and culture to support information sharing and participation in decision-making (i.e. decision/organizational control). One of the biggest challenges of the future for organizational design and management will be to design for maximizing the positive and minimizing the negative potential aspects of ICT and its implementation in different types and forms of work.

5. CONCLUSION

It is important to recognize the multiplicity and variety of psychosocial work factors. According to the Balance Theory, various aspects of the different elements of the work system (i.e. the individual, task, technology, organization and environment) and/or of the interactions among the elements of the work system can be positive and negative from a psychosocial point of view. Psychosocial work factors have been shown to influence quality of working life and health, and therefore are important to consider in the ergonomic design of work. Finally, recent trends in businesses, e.g. team work, TQM and Information and

Communication Technology, have seen the emergence of new forms of work organization which can have both positive and negative psychosocial work factors. To prevent negative effects on people, we need to (re)design work to remove as much as possible these negative psychosocial work factors.

REFERENCES

CARAYON, P. 1991. *The Office Worker Survey*. Madison, WI: Department of Industrial Engineering, University of Wisconsin-Madison.

CARAYON, P. and LIM, S. Y. 1999. Psychosocial work factors. In W. Karwowski & W. S. Marras (eds), *The Occupational Ergonomics Handbook* (pp. 275–283). Boca Raton, Florida: CRC Press.

CARAYON, P. and LIM, S. Y. 1994. Stress in automated offices. In A. Kent (ed.), *The Encyclopedia of Library and Information Science* (Vol. 53, supplement 16, pp. 314–354). New York: Marcel Dekker.

CARAYON-SAINFORT, P. 1992. The use of computers in offices: Impact on task characteristics and worker stress. *International Journal of Human Computer Interaction,* 4(3), 245–261.

COOK, J. D., HEPWORTH, S. J., WALL, T. D. and WARR, P. B. 1981. *The Experience of Work*. London: Academic Press.

COOPER, C. L., and MARSHALL, J. 1976. Occupational sources of stress: A review of the literature relating to coronary heart disease and mental ill health, *Journal of Occupational Psychology*, 49(1), 11–25.

COOPER, C. L., SLOAN, S. J. and WILLIAMS, S. 1998. *Occupational Stress Indicator*. Windsor, England: NFER-Nelson.

COX, T., and FERGUSON, E. 1994. Measurement of the subjective work environment. *Work and Stress*, 8(2), 98–109.

CZAJA, S. J. and SHARIT, J. 1993. Stress reactions to computer-interactive tasks as a function of task structure and individual differences. *International Journal of Human–Computer–Interaction*, 5(1), 1–22.

DAVIS, L. E. 1980. Individuals and the organization. *California Management Review,* 22(2), 5–14.

ELO, A. L., LEPPANEN, A., LINDSTROM, K. and ROPPONEN, T. 1994. The Occupational Stress Questionnaire. In C. Zenz and O. B. Dickerson and E. P. Horvarth (eds), *Occupational Medicine* (pp. 1234–1237). St. Louis: Mosby.

HACKMAN, J. R. and OLDHAM, G. R. 1976. Motivation through the design of work: Test of a theory. *Organizational Behavior and Human Performance*, 16, 250–279.

HAGBERG, M., SILVERSTEIN, B., WELLS, R., SMITH, M. J., HENDRICK, H. W., CARAYON, P. and PERUSSE, M. 1995. *Work-Related Musculoskeletal Disorders (WMSDs): A Reference Book for Prevention*. London: Taylor & Francis.

HENDRICK, H. W. 1991. Human factors in organizational design and management. *Ergonomics*, 34, 743–756.

HENDRICK, H. W. 1996. Human Factors in ODAM: An historical perspective. In O. J. Brown and H. W. Hendrick (eds), *Human Factors in Organizational Design and Management-V* (pp. 429–434). Amsterdam, The Netherlands: Elsevier Science Publishers.

HERZBERG, F. 1966. *Work and the Nature of Man*. New York: Thomas Y. Crowell Company.

HERZBERG, F. 1974. The wise old turk. *Harvard Business Review* (September/October), 70–80.

HURRELL, J. J. J. and MCLANEY, M. A. 1988. Exposure to job stress. A new psychometric instrument. *Scandinavian Journal of Work Environment and Health,* 14(suppl. 1), 27–28.

ILO. 1986. *Psychosocial Factors at Work: Recognition and Control* (Occupational Safety and Health Series no. 56). Geneva, Switzerland: International Labour Office.

KARASEK, R. A. 1979. Job demands, job decision latitude, and mental strain: Implications for job redesign. *Administrative Science Quarterly*, 24, 285–308.

KASL, S. V. and COOPER, C. L. (eds) 1987. *Stress and Health: Issues in Research and Methodology*. Chichester: John Wiley & Sons.

LINDSTROM, K. and LEINO, T. 1989. Assessment of mental load and stress related to information technology change in banking and insurance. In F. Klix, N. A. Streitz, Y. Waern and H. Wandke (eds). *Man–Computer Interaction Research* (pp. 523–533): MACINTER-II.

MAYO, E. 1945. *The Social Problems of an Industrial Civilization* (3rd ed.). Andover, MA: The Andover Press.

PHIZACKLEA, A. and WOLKOWITZ, C. 1995. *Homeworking Women: Gender, Racism and Class at Work*. London: Sage Publications.

SAINFORT, F., CARAYON, P. and SMITH, M. J. 1997. Total Quality Management and quality of working life in a public sector organization. In P. Seppala, T. Luopajarvi, C. H. Nygard and M. Mattila (eds). *Proceedings of the 13th Triennial Congress of the International Ergonomics Association* (Vol. 1 – Organizational Design and Management, pp. 522–524). Helsinki, Finland: Finnish Institute of Occupational Health.

SMITH, M. J. and CARAYON-SAINFORT, P. 1989. A balance theory of job design for stress reduction. *International Journal of Industrial Ergonomics*, 4, 67–79.

SMITH, M. J., COHEN, B. G., STAMMERJOHN, I. W. J. and HAPP, A. 1981. An investigation of health complaints and job stress in video display operations. *Human Factors*, 23(4), 387–400.

SMITH, M. J., SAINFORT, F., CARAYON-SAINFORT, P. and FUNG, C. 1989. Efforts to solve quality problems: Secretary's Commission on Workforce Quality and Labor Market Efficiency, U.S. Department of Labor, Washington, D.C.

SPROULL, L., and KIESLER, S. 1991. *Connections*. Cambridge, MA: MIT Press.

TAYLOR, F. 1911. *The Principles of Scientific Management*. New York: Norton and Company.

TRIST, E. L. and BAMFORTH, K. 1951. Some social and psychological consequences of the long-wall method of coal getting. *Human Relations*, 4, 3–39.

YANG, C. L. 1994. *Test of a Model of Cognitive Demands and Worker Stress in Computerized Offices*. Unpublished Ph.D. dissertation, University of Wisconsin-Madison, Madison, WI.

Social Policy and Industrial Development

K. Vanwonterghem

CERGO International, Zavelvennestraat 64, B-3500 Hasselt, Belgium

1. INTRODUCTION

Ergonomics as a discipline focusing on the design and redesign of the "active human" systems is concerned with the living quality of human kind. It aims to bring equilibrium between human activities and environmental conditions that would increase human potential. A well-balanced "human-at-work" system (work in the widest sense of its content) should create conditions of acceptable risks with respect to occupational health and safety, and assure the human well-being, in combination with satisfactory efficiency of the human activity.

The above goals of ergonomics should be also valid in industrially developing countries (IDC), but many obstacles may destabilize development of human-oriented policies. Wars, threats of wars, revolts, riots, ethnic conflicts and general insecurity create climates in which the risks of loss of life and degenerating human integrity are put against human well-being and adequate quality of life. As the first priority should be striving for peace and avoiding conflicts, the knowledge and experience of ergonomics should be used to prevent such conflicts and to protect the working population.

A major obstacle to improving the quality of life is an impact of the non-controlled industrialization on the socioeconomic system of developing countries, especially as they strive to join the global economy. A need to develop and increase regional and local economies has to be balanced against the consequences of moving from the informal sectors and small size companies, where workers are often self-employed family members, including women and children with low educational background, to a well-organized industrial organization.

The expectance of such an evolution is an increased life quality, better housing facilities and living conditions, and increased financial means, which should serve to bring the social status to a higher level for all citizens. The underlying benefit should also be an improved health and safety. However, as shown by economically emerging regions in South America and Africa, the industrial development does not always bridge the gap between the wealthy and underprivileged populations. Such a gap becomes a global issue, and even in South East Asia and China, where economic growth is strong and sustained, many problems evolve. Whereas the living standards raised drastically for the privileged population, others (less fortunate) follow similar purchasing trends, buying technical gadgets such as sophisticated video systems or mobile phones, etc., while the essential values of health and safety are put aside to the lower priority level.

2. ERGONOMICS AND INDUSTRIALIZATION

In the industrialized world, there is a strong need to stimulate productivity to cope with the production costs, and the very expensive social security systems. One of the reasons for investing in developing countries is to approach those economies that need further development, but can also provide cheap labor and low, or nonexistent, social cost. The dominant strategy is often based on direct transfer of high tech systems, knowledge and experiences to developing countries. These new technologies (at least new for the IDC), such as advanced production systems, are imported and implemented as they are being used in the country of origin, where first priority is to increase system productivity within the shortest time possible. The responsibility to get the production system going is a true challenge for the local management that often puts priority on the technical development. Human, as part of the production process, is regarded as a technical tool, and classified as the necessary cost factor.

In this evolution, workers are trained to fulfill their tasks, while health and safety matters are handled through organizational techniques and formal working standards, procedures and legislation. Because of the fact that the directives are transferred along with the equipment and machinery, the IDC's often receive existing (international or national) standards. By classifying these formal prescriptions as valid in IDC, just because they are valid "over there" (in the industrialized countries) and, because of lack of knowledge that hinders productivity, other essential policy items are often neglected, denied or applied abusively.

The above industrializing principles may work for an undetermined time. This is because human beings have extremely powerful adaptation skills that lead to development of new cultures, i.e. the "industrial cultures", in a short period. The step from a farmer to a production worker in a semiautomatic electronic assembly plant may sometimes take only a few months. However, only few can take full advantage of this new situation. The remaining majority is often split into different categories, with some people just accepting the situation and others frustrated with unfulfilled ambitions. Drastic changes may also cause a social shock alienating, even separating, those who can follow the new possibilities from their "old" community.

3. ERGONOMICS AND SOCIAL POLICY

It is difficult to assess the impacts of expatriation, seasonal labor in neighboring regions or countries, learning to live with changing values, the introduction of new technologies in a daily life, respecting the variety of social structures (gender differences, elderly, social position), contacts with foreign cultures, the introduction of women's liberation, less rigid religion rules, the enlarged possibilities for leisure, or transfer of goods, drugs and other "created" needs, on the human well-being. Given the above, it is remarkable to see how local people bridge the distance between the rural-agricultural activities and the mechanized high-tech industrial activities. But developing an industrial culture is an ongoing process that takes time. In Europe, for example, it took ~200–300 years, during which health and safety evolved via experience, leading to the development of prevention policies.

Bridging the above gap is difficult because of the complexity of new working conditions. The superficial introduction of new technologies without taking into account the relevant human factors, is often ineffective and results in a high social cost. In the economic context, it can be observed that unemployment and the degree of workforce availability on the market are closely

related to the level of investments targeted for improving working conditions. Increased turnover rates and absenteeism illustrate these effects.

The divergent needs of industrialization and social development indicate that these problems cannot be solved by any single or universal method, that money does not cure everything, and that economic development must consider the social and human dimensions. Industrial development should also take into consideration the history, the reality of existing social systems of each country or region, and should pay respect to the fundamental values on which the culture is based on. These basic values, as is the case for the industrially developed world, are founded on factors such as life and professional expectances, socioeconomic and cultural status, and personal needs. They will steer the adaptation process of individuals and groups.

Risk-coping strategies, conscious or unconscious, are determined through the subjective interpretation of the acceptable risks, of feelings of comfort/discomfort, satisfaction, happiness, and the expected rewards. Ergonomics can help in the development of risk assessment methods, practical problem-solving measures, and the establishment of priorities at the micro- and macroeconomic levels.

The experiences of the developed industrial world should be utilized to help to create sustainable development in the newly industrializing countries. While "knowledge transfer" is the key, care should be exercised in practical implementation of such knowledge. Fortunately, it is believed that most of the basic ergonomics knowledge needed is already available in the IDC. This is because many professionals graduated from universities of the industrialized world. Furthermore, the global information networks keeps them informed about the current trends and new developments. The real problem, however, is translating the available knowledge and experience into a sustainable policy. In many developing countries there is a lack of national and regional prevention programs, and there are no necessary structures to bring them into the effective applications. This limits many of the experts as only individual actions are often possible. There are, for example, no prevention-oriented records of the useful health and safety solutions, and the experts are not, or only exceptionally, involved or consulted on the industrial and economic issues.

For these reasons, ergonomics in many IDC is seen as an interesting discipline, but one that remains limited to the academic cocoon. Furthermore, no funds are available to support the ergonomics programs, although the health and safety programs exist and are strongly supported. The "preventive" capacities of ergonomics at the early system design stage, before the advent of diseases and accidents, are not yet commonly recognized. The ultimate vocation of ergonomics is direct contacts with the shopfloor, as this is the place where prevention is most effective.

4. ERGONOMICS AND EDUCATION

The crucial challenge for ergonomics in IDC is to bring about an awareness about the risks, and ways to reduce their consequences in industrial practice, as well as providing the knowledge, both theoretical and practical, at all levels of society. Therefore, ergonomics should focus its activities in IDC on three basic types of activities, namely research, consulting, and training and education.

4.1. Research

With the concern to bring ergonomics as a theoretical and practical tool in IDC, ergonomics should join the economical world which evolves into an information-led system, in which communication becomes crucial for development. The existing techniques of telecommunication, such as the use of computers and the Internet, allow for sharing all the necessary knowledge between experts to increase their capacities. Is this the solution for ergonomics? Possibly but, as mentioned before, practical implementations have to take place in the real factories and workplaces. To realize this, the gap between the academic and practical problems must be bridged, and both the fundamental and applied research should be stimulated.

It is of great importance to agree about a common research methodology in ergonomics. This would create reliable, interpretable and comparable data that would ensure that all external factors are well described and/or measured. The assessment of such data should be put in the context of individual human reactions to the selected phenomena under study, and validated statistically for groups of individuals occupying the same activities. The human physiological, psychomotor and mental reactions, including objective and subjective experiences of the exposed workforce, should be measured to evaluate tasks, organization and environmental issues.

Knowledge about the specific capacities and differences in characteristics of the local workforce is very much needed. Therefore, where possible, local data banks about the anthropometric, biomechanical, physiological and psychomotor issues should be established. Some countries have developed their own basic data. Such data is also some are available from the International Labor Office and the World Health Organisation.

To fulfill the research requirements, skilled teams of ergonomists should be trained in the applied research methodology and data processing. Such teams should be able to operate nationwide as the specific problems are typically very large. Owing to the high complexity of the involved issues, the required expertise should be set at an academic level in one of the ergonomics-related disciplines, with an understanding of the multidisciplinary involvement of other disciplines. Such research teams should have strong interaction with, and the feedback from, industrial partners.

4.2. Consulting

The short-term projects (from a few days or weeks to the maximum of 6 months) should focus on actual problems occurring in industry. Project objectives should be formulated by industry, and assessed at various management levels as well as by workers at the shop-floor level. The teams should use a practical approach by identifying the problem, indicating its importance, formulating specific improvement measures (task, organization and environment), setting priorities by assessing the potential impact on the exposed workers, and following up implementation as a project evaluation.

Experts with good technical skills in developing and evaluating the technical and organizational aspects of the projects are needed. However, such practical skills are often absent in most of the industrially developing countries. Therefore, one of the first priorities should be to establish local training networks to fill up this gap. The trainers must have an appropriate

background in the ergonomics discipline, as well as close contacts with the industry and workers. They should serve as the bridge between the industry and academic world.

4.3. Training and Education

In terms of the required basic knowledge that is needed to realize such projects, ergonomics must be introduced at all levels of society, from the workshop floor up to the academic and political world. This should be done with the elementary principle that applying such knowledge in practice is essential for sustainable economic development.

At present, some attempts have been made to bring the basic knowledge in ergonomics to the workers at the shop-floor level by means of a series of "roving seminars" and "training for trainers" workshops, designed to reach bigger population in a short time. The main criticism of such attempts is that they typically use work design examples that originate in the industrialized world, instead of providing design solutions that meet local conditions and needs.

Similar phenomena can be seen at the academic level. In many cases, qualified ergonomists who graduated at universities in the developed countries, move back to their native universities. They often teach advanced (theoretical) material, having very limited industrial experience relevant to the local needs, and very few contacts with industry to teach practical implementation of their knowledge. The absence of managing structures and appropriate policies often leads to efforts which are not suitable for industrially developing countries because they are "too academic", removed from reality, and too expensive.

It should be noted here that the longer one is studying abroad the more s/he is alienated from the economic and social realities of the native country. Knowledge acquired in the developed world often does not match the limitations in the local financial mechanisms, advanced technology, slowly changing industrial policies, and the national standards and legislation.

5. CONCLUSIONS

The promise of ergonomics often relies on the belief in social, industrial and economic efficiencies. Therefore, if national authorities do not create a climate for sustainable development, all investments, new technologies, and efforts to improve working and living conditions may remain an interesting option for the academic exercise. An example of the above situation is the recent (February 1999) workshop organized in Thailand about the health and working conditions in South East Asia (Physical Workload and Heat Stress). About 20 experts participating in this workshop came to a conclusion that, even in a tropical area,

very little is known about the impact of heat on health and human behavior. To handle the issue for the working population, there is a tendency to introduce an ISO standard (WBGT, ISO: 7243) by legislatures in various IDC.

While the above may be a worthwhile strategy, two European Union-sponsored research projects (conducted in Thailand and Indonesia) concluded that there was an important difference in the physiological responses of the workers from the western world compared with South East Asia. These cardiocirculatory and thermoregulatory differences lead to the conclusion that particular attention should be paid to the assessment of physical workload. By strictly following international standards, without reference to the local thermoregulation capacities or physical characteristics of the local population, some unintended risks can inadvertently be introduced to the workers.

Sustainable development only can be realized by appropriate training and education in combination with fundamental and applied research, as well as practical consulting as a long-term growth factor. The local and national authorities have the responsibility to facilitate implementation, through political programs, of practical guidelines and directives. The quality of the ergonomic measures must be developed by pragmatic and scientific research. Management must recognize and understand the impact of the problem on the affected worker population, and make practical interventions or accept guidelines as needed.

The final goal is to combine the political world with the forces of economy and science in saving the human being's work and improving quality of life. As stated by James D. Wolfensohn, President of the World Bank: "There is no simple solution. Infrastructure capacity, agriculture and industrial productivity, economic policies and human investments are all essential, but none in itself is adequate to generate broad-based growth. Instead, we are learning that the most effective development strategies strike a balance among economic, social, financial and environmental factors. *The best begin and end with the dignity of every human being*" (*Herald Tribune*. 21 June 1999).

As for "economics of ergonomics" there is no single, simple solution. Developing, implementing and applying ergonomics knowledge on a large scale in industrially developing societies takes a long time before successful results become obvious. These may not even be recognized as a result of the economic or ergonomic program. However, ergonomics, in its multidisciplinary approach to "man-at-work" systems, has great potential to contribute to the socioeconomic growth of the industrially developing countries.

Socially Centered Design

K. M. Stanney†, J. Maxey† and G. Salvendy‡

†University of Central Florida, Department of Industrial Engineering &
Management Systems Orlando, FL 32816, USA

‡School of Industrial Engineering, Purdue University, West Lafayette, IN
47907, USA

1. INTRODUCTION

System design and the individuals who contribute to this activity have evolved to reflect the salient concerns of successive eras. The laboratory engineer as the principal designer has evolved into a team of designers including psychologists, social scientists and anthropologists. These latter additions to the design team are confronting the problem of designing for groups of users and organizations. This has led to socially centered design, which is concerned with "border issues," i.e. design variables that reflect socially constructed and maintained world views which both drive and constrain how people can and will react to and interact with a system or its elements. Socially centered design is a medial-ergonomic design that fills the void between traditional system centered design (i.e. macro-ergonomic design of the overall organization and work system structure, as well as process interfaces with the system's environment, people and technology) and contemporary user-centered design (i.e. micro-ergonomic design of specific jobs and related human–machine, human–environment and user system or software interfaces) (Stanney *et al*. 1997). Specifically, socially centered design takes a holistic approach to design, broadening the narrowly defined technical focus of past design approaches by considering users as their performance is influenced by their knowledge and understanding of the social context in which they perform their work. Following this design approach, naturalistic observation of human behavior as it is influenced by situational, contextual and social factors is used to assist in the formulation of system designs.

Socially centered design assumes that by taking into consideration organizational and social factors, such as informal work practices and shared artifacts, the system design process will be enhanced. This presupposes that social interaction is orderly and can therefore be understood. Goguen (1994) suggests that system designers can achieve this understanding by developing a sense of the intrinsic categories and methods (i.e. informal processes) members of an organization assign themselves during their interactions. In particular, participatory observation in which designers become immersed in and interact with target user populations in their natural work setting can be used to gain such understanding. Further, socially centered design recognizes that the roles and responsibilities of workers are situated (i.e. emergent rather than individualistic, local to a particular time and place, contingent on the current situation, embodied in a physical and social context, open to revision, and vague), being defined in relation to the social interaction in which they occur.

2. THE SOCIALLY CENTERED DESIGN PROCESS

The socially centered system design process begins with a contextual inquiry to determine the user, product, task and environ-mental parameters potentially influencing system design (Stanney *et al*. 1997). This investigation will generally involve some form of direct observation such as field studies or ethnographic observation. It is critically important that this observational period be conducted in an unbiased, non-leading manner (Nardi 1997). Designers are expected to become immersed in one or more communities of practice to observe, record and analyze what users actually do in applying artifacts to achieve goals. While such participatory observation leads to great insights, designers should avoid modifying the behavior of those observed by refraining from asking leading questions, or imposing biases where their perspectives differ from those being observed. In applying techniques like interviews, focus groups, detailed questionnaires or by measuring specific aspects of work performance, the immersed designers seek to adopt the user's frame of reference or view of reality to understand the how and why of artifact use. Once completed, the most common mistake of such observational studies is to overextend the interpretation of the results. Thus, it is essential that the rich abundance of observational data obtained be filtered, organized and interpreted in the larger context of the technology and science in which the artifacts exist.

3. SOCIALLY CENTERED DESIGN VARIABLES

The socially centered design variables for which data are accumulated during observational studies comprise those input or contextual variables that are expected to influence group processes (Stanney *et al*. 1997). These variables include the context itself, both as a global or comprehensive factor and the separate component variables that collectively comprise the context: artifacts, cooperative task activities, organizational structure, situational factors and interpersonal characteristics.

The context in which an activity is performed has both physical (e.g. locational and sensory cues) and social elements (e.g. presence of, relationship to and expectations of other individuals) that provide opportunities for or constrain individual behavior, particularly due to the schema (i.e. general purpose representations of concepts, situations or ideas) that develop over time.

Artifacts are the designed and created objects (e.g. paper and pen, computers) or systems (e.g. traffic lights, Internet) that humans apply and use to influence the world around them. Individuals within a given community generally use artifacts in particular ways according to past insights, common intuitions, shared understandings and activities, and current or recent experiences (Brown and Duguid 1994). Over time, however, different communities may develop "border" or non-central uses for peripheral or secondary attributes characteristic of particular artifacts. These border uses will likely be unique to a given community and may arise when functional fixedness is overcome by a perceptive user with a novel problem. While in the past these issues were recognized, their influence on design were generally thought to be inconsequential due to the tendency of artifacts to remain stable over time and populations of consumers (Stanney *et al*. 1997). With the de-massification of today's competitive market, however, designers have been forced to offer many customized products instead of a few standardized ones to cater to narrow demographic segments scattered in regional markets. The result is there may no longer be a guarantee of artifact continuity to include border uses. In turn, shared practices may not

be easily maintained, especially if customization is extended downwards to the individual user so that only a very few common core or border attributes exist to support member coordination of the community's internal behaviors (Gaver 1994). Such circumstances highlight the need for embracing socially centered design practices.

In organizational settings, cooperative task activities generally define or reflect the job or objective that artifacts are used in whole or part to accomplish. The influence of the task factor on group performance, when this is mediated in some manner by technological artifacts, manifests itself in terms of the degree to which effective group performance depends on both the transmission of information about the task and transmission of information about the values, interests, and personal commitments of group members (Hollingshead and McGrath 1995). These factors and the effectiveness with which system designs or artifacts support their performance can be identified via the observational studies of socially centered design.

The organization or group within which an activity is performed significantly affects the appropriateness of a system design. To implement plans and achieve common objectives, group members and the systems that support them must create and maintain meaningful relations (i.e. norms, power relationships, status relationships, cohesiveness and group density) which can be identified through observational studies. Thus, organizational requirements emerge from a system being placed in a social context and its associated group relations rather than those deriving from the functions to be performed or the tasks to be assisted (Dobson et al. 1994).

Situational characteristics reflect the various social networks and relationships that exist among group members, as well as the maturity of the group (e.g. just formed versus well established) (Stanney et al. 1997). In particular, cultural situational differences (e.g. collectivistic versus individualistic) are particularly influential in social and work settings, which in turn is reflected in the behavior of the members of these cultures.

Interpersonal characteristics (e.g. dominance, friendliness, authoritarianism, field-articulation, gender) reflect the attitudes, behaviors and motivations of individual group members. Groups whose members have compatible personalities tend to be more cohesive, more efficient and more productive than groups with incompatible members.

These socially centered design variables may not be readily noticed or recognized by designers as important without a contextual inquiry, yet they can influence the effectiveness of system and product design. Thus, through observational studies global context can be identified and the genre of artifacts that may be used to achieve particular goals (both central and border) within a certain range of situational conditions for a community of users are recognized. This information can then be used to design systems and products that directly meet users needs and are genuinely easy to use. There are several different methods used to elicit this information.

4. SOCIALLY CENTERED DESIGN TECHNIQUES

Socially centered design techniques reflect the methodologies applied to acquire data about users and their work context and related artifacts (Stanney et al. 1997). These techniques elicit identification of artifact attributes for both core and border uses and specify whether they facilitate, slow or impede activities; the scope and nature of relationships among individuals and the manner in which particular artifacts support these relationships; and how work, communication and information flow are mediated (Hollingshead and McGrath 1995). Several different techniques have been used in an attempt to elicit such contextual information.

4.1. Group Studies

Socially centered design can use a modified version of traditional sociological group studies, where predictive theories or even designers' intuitions are formulated and situationally tested in a variety of contextual settings (Stanney et al. 1997). This can be a timely and resource intensive approach.

4.2. Ethnographic Studies

The ethnographic approach identifies work practices and organizational interactions through anthropological studies of the work setting (Stanney et al. 1997). Ethnomethodology examines how competent members of an organization coordinate their behaviors, specifically in delimiting the categories, methods and artifacts used to render their activities intelligible to one another (Goguen, 1994). Ethnographic studies have two general focuses: (1) the moment-to-moment interactions of observational participants or (2) more global, longer-term activities such as how artifacts are used to distribute and regulate knowledge, communication and action, as well as identification of the personal objectives, beliefs, principles and practices of those observed (Nardi 1997).

The ethnographic approach often involves participatory observation in which designers intensely observe and interact with the cultural activities, belief systems, rituals, institutions and artifacts of those under observation. Using this approach, designers and intended users of the system or product develop a shared understanding of the nature of the work or activity to be performed, the users' needs in performing this activity, and the contextual components which support the completion of the activity. This mutual understanding is deeply tied to the context in which the work or activity is performed. Thus, design solutions based on the ethnographic approach may not generalize to other situations.

4.3. Cooperative Work Studies

Cooperative work studies involve observational investigations of groups working cooperatively in their natural social and cultural setting (Stanney et al. 1997). In general, the group under observation collaborates on a common goal that requires communication (either face-to-face or mediated exchanges) to support goal achievement, and who thus must in some manner coordinate their work activities. With this approach, structural variables reflecting the group (e.g. size, composition, interactions, internal processes) and external situational factors that reflect the natural environment are identified and used to inform the design process.

4.4. Anthropomorphic Studies

Anthropomorphic studies generally examine human-to-human communication patterns and then try to emulate these relationships in the design of interactive systems. The focus is to identify

the schema (e.g. communication modes, cognitive strategies) through which people perform activities or tasks.

5. APPLYING THE SOCIALLY CENTERED DESIGN APPROACH

Interactive system designs are more commonly being derived from contextual inquires, with more traditional design techniques being reserved for evaluation and refinement of the ensuing designs. Further, as advancing technologies emerge, designers must struggle with the issue of how to generate new innovative technological artifacts that meet users' requirements. The techniques of socially centered design may prove to be instrumental to this developmental process. In particular, socially centered design approaches are advantageous in the exploratory design stages before an initial design concept has been identified (Nardi 1997). At this stage, the information generated from observational studies can lead to creative design solutions that would likely have never been rendered without such insights. These approaches are also useful once a formal prototype has been generated when observational studies can focus on the intended users interacting with the system or product in its operational setting.

As for where socially centered design will have the greatest utility, the current surge of consumer products provides numerous opportunities. For example, while multimedia currently focuses on the technological integration of several input and output media, a socially centered approach would redirect focus from the media, which is transitory in nature, to the context in which this media is applied, uncovering its work practices, artifacts and environmental and cultural factors (Stanney *et al.* 1997). Socially centered design may prove to be particularly influential in the design of emerging interactive networks, particularly for interactive television (I-TV). Designers of I-TV, who are freed from the need to create media that appeals to the masses, will need to tap into the lives of their viewers to identify their unique skills and interests if they are to capture targeted audiences. The techniques of ethnographic studies will be well suited to such investigations. In addition, the emerging virtual environments (VE) technology is inherently more social than past non-interactive technologies. VE users are immersed in a particular context or virtual world, with its own set of interactive artifacts and situational characteristics. For such interactive products and systems to be effectively developed, the assumptions that designers make concerning the types of users who will adopt these systems, the types of tasks to be performed and the environmental character-istics in which these activities will be done become critically important. The effectiveness of such designs may thus heavily rely on the techniques of socially centered design.

6. CONCLUSIONS

System and product designers have come to the realization that it is not what a product or system can do that defines success but how well users can accomplish intended tasks with the aid of the product or system. Yet users are known to have idiosyncratic usage patterns that are not readily foreseen by designers. The challenge is to identify the artifacts, cooperative task activities, impact of organizational, cultural and situational factors, as well as the interpersonal behaviors that influence the effectiveness of a given system or product design. Failing to comprehend or consider these factors may make the acceptance and adoption of new system level solutions very unlikely or difficult at best. Socially centered design techniques can be used to systematically elicit an understanding of contextual factors and integrate this knowledge into the design process, thus providing a grounding framework within which a formal design can evolve.

REFERENCES

BROWN, J. and DUGUID, P., 1994, Borderline issues: Social and material aspects of design. *Human–Computer Interaction*, 9, 3–36.

DOBSON, J.E., BLYTH, A.J.C., CHUDGE, J. and STRENS, R., 1994, The ORDIT approach to organizational requirements. In Jirotka, M. and Goguen, J. (eds) *Requirements Engineering: Social and Technical Issues* (London: Academic Press), pp. 87–106.

GAVER, W., 1994, Commentary on borderline issues: 8. Grounding social behavior. *Human–Computer Interaction*, 9, 70–4.

GOGUEN, J.A., 1994, Requirements engineering as the reconciliation of social and technical issues. In Jirotka, M. and Goguen, J. (eds) *Requirements Engineering: Social and Technical Issues* (London: Academic Press), pp. 165–99.

HOLLINGSHEAD, A. and MCGRATH, J., 1995, Computer-assisted groups: a critical review of the empirical research. In Guzzo, R. A. and Salas, E. (eds) *Team Effectiveness and Decision Making in Organizations* (San Francisco: Jossey-Bass), pp. 46–78.

NARDI, B.A., 1997, The use of ethnographic methods in design and evaluation. In Helander, M. Landauer, T.K. and Prabhu, P. (eds) *Handbook of Human–Computer Interaction* (Amsterdam: Elsevier), pp. 361–6.

STANNEY, K.M., MAXEY, J. and SALVENDY, G., 1997, Socially centered design. In Salvendy, G. (ed), *Handbook of Human Factors and Ergonomics* (2nd ed) (New York: Wiley), pp. 637–56.

Sociotechnical Analysis of Work System Structure: Applying Empirical Models of Sociotechnical System Elements

H. W. Hendrick

Hendrick and Associates, Englewood, CO 80111, USA

1. INTRODUCTION

In order to systematically analyze the dimensions of a given work system's structure, we first must understand what those dimensions are. As we then shall see, utilizing empirically developed models, knowledge of the specific sociotechnical characteristics of a given work system can be used to analyze and optimize these dimensions for that work system.

2. THE STRUCTURAL DIMENSIONS OF WORK SYSTEMS

The organizational structure of a work system can be conceptualized as having three core dimensions. These are *complexity, formalization,* and *centralization* (Robbins 1983). Complexity, in turn, has two major dimensions: *differentiation* and *integration.*

2.1. Complexity: Differentiation

Work system structures employ three common types of differentiation: *vertical, horizontal,* and *spatial.* Increasing any one of these three increases a work system's complexity.

2.1.1   Vertical differentiation

This refers to the of the number of hierarchical levels separating the chief executive position from the jobs directly involved with the hands on work. Too much vertical differentiation for a given set of sociotechnical system conditions increases communication, coordination, and control problems, thus decreasing efficiency. Too little also decreases efficiency because of a lack of adequate coordination and control.

2.1.2 Horizontal differentiation

This refers to the degree of departmentalization and specialization within a work system. Horizontal differentiation increases complexity because it requires more sophisticated and expensive methods of control. In spite of this drawback, specialization is common to most work systems because of the inherent efficiencies in the division of labor.

The two most common ways to determine whether or not a work group should be divided into two or more departments are the degree of commonality of *goals* and *time orientation* To the extent that subgroups either differ in goals, or have differing time orientations, they should be structured as separate departments. For example, sales persons differ from R&D employees on both of these dimensions. Not only do they have very different goals, but also the time orientation of sales personnel usually is *short*

(one year or less), whereas it usually is *long* (three or more years) for R&D personnel. Thus, they clearly should be departmentalized separately, and usually are. (Robbins 1983)

2.1.3   Spatial dispersion

This refers to the degree an organization's activities are performed in multiple locations. There are three common measures of spatial dispersion: (1) the number of geographic locations comprising the total work system, (2) the average distance of the separated locations from the organization's headquarters, and (3) the proportion of employees in these separated units in relation to the number in the headquarters. In general, complexity increases as any of these three measures increase.

2.2. Complexity: Integration

Integration refers to the number of mechanisms designed into a work system for ensuring communication, coordination, and control among the differentiated elements. As the differentiation of a work system increases, the need for integrating mechanisms also increases. This happens because greater differentiation increases the number of units, levels, etc. that must communicate with one another, coordinate their separate activities, and be controlled for efficient operation.

Some of the more common integrating mechanisms that can be designed into a work system are formal rules and procedures, committees, task teams, liaison positions, and system integration offices. Computerized information and decision support systems also can be designed to serve as integrating mechanisms. Vertical differentiation, in itself, is a primary integrating mechanism (i.e. a manager at one level serves to coordinate and control the activities of several lower-level groups).

For a given work system's sociotechnical characteristics, too many integrating mechanisms stifle efficiency, and too few result in inadequate communication, coordination and control. Either too few or too many usually will increase costs and decrease productivity.

2.3. Formalization

Formalization refers to the degree to which employee decision discretion is limited by explicit job descriptions, extensive rules, and clearly defined procedures covering processes. Ergonomists can contribute to formalization by designing job, machine, and software interfaces so as to standardize procedures and allow little opportunity for operator decision discretion. By the same token, these interfaces can be ergonomically designed to permit greater flexibility and scope to employee decision-making.

2.4. Centralization

Centralization refers to the degree to which formal decision-making is concentrated in a relatively few individuals, group, or level, usually high in the organization. When the work system structure i`s highly centralized, lower-level supervisors and employees have only minimal input into the decisions affecting their jobs (Robbins 1983). In highly decentralized work systems, decisions are delegated downward to the lowest level having the necessary expertise.

Work systems carry out two basic forms of decision-making, *strategic* and *tactical*, and the degree of centralization often is quite different for each. *Tactical* decision-making has to do with the

day-to-day operation of the organization's business; strategic decision-making concerns the long range planning for the organization (Hendrick 1997).

3. SOCIOTECHNICAL ANALYSIS OF THE STRUCTURAL DIMENSIONS OF WORK SYSTEMS

The analysis of a work system's structure and related processes involves consideration of three major sociotechnical system elements that interact and affect optimal work system design. These are the (1) *technological subsystem,* (2) *personnel subsystem,* and (3) *relevant external environments* that permeate the organization and upon which its survival and success depends. Each of these elements has been studied in relation to its effects on the three organizational design dimensions of *complexity, formalization, and centralization,* described earlier, and empirical models have emerged. These models can be used as macroergonomic tools in analyzing and developing or modifying the design of a given work system.

3.1. Technological Subsystem Analysis

Technology, as a determinant of work system structure, has been operationally classified in several distinctly different ways that are useful in analyzing work system structure (e.g. see Hendrick (1997) for a summary of these). Perhaps the most generalizable and widely validated model of the technology-work system structure relationship is that developed by Perrow (1967).

3.1.1. Perrow: knowledge-based technology

Perrow's model uses a *knowledge-based* classification scheme. He begins by defining technology as the action one performs upon an object in order to transform the object. This action requires some form of technical knowledge. Using this approach, Perrow identified two underlying dimensions of knowledge-based technology. The first of these is *task variability*, or the number of exceptions encountered in one's work. The second concerns the type of search procedures one has available for responding to task exceptions, or *task analyzability*. These search procedures can range from "well defined" to "ill defined". At the "well defined" end of the continuum, solving problems can be accomplished using rational-logical, quantitative, and analytical reasoning. At the "ill defined" end there are no readily available formal search procedures. Instead, one must rely on the experience, judgement, and intuition of the decision-maker. Dichotomizing these two dimensions yields a two-by-two matrix having four cells. As shown in figure 1, each cell represents a different knowledge-based technology.

1. *Routine* technologies have well defined problems with few exceptions. They lend themselves to standardized coordination and control procedures and, thus, are associated with high formalization and centralization and moderately high vertical differentiation. Mass production units typify this category.
2. *Non-routine* technologies have many exceptions and difficult to analyze problems. As a result, these technologies require the flexibility that comes from being highly decentralized and having low formalization. Combat aerospace operations would be an example.
3. *Engineering* technologies have many exceptions, but can be

Figure 1. Perrow's knowledge-based technology classes

handled using well-defined rational-logical processes. Consequently, they lend themselves to moderate centralization, but require the flexibility that is achievable through low formalization. Engineering design firms or units typify these technologies.
4. *Craft* technologies involve fairly routine tasks, but problem solving relies heavily on the experience, judgement, and intuition if the individual craftsperson. Decisions thus must be made by those having the particular expertise. This requires decentralization, low formalization, and usually works best with low vertical differentiation.

3.1.2. Other technological considerations

During the 1980s and 1990s, major advances were, and continue to be made, in computer and communications technology. Two forms of this technology have major implications for work system design: these are advanced information technologies (AIT) and computer integrated manufacturing (CIM). AIT has tended to facilitate decentralizing operational or tactical decision-making, while enhancing the efficiency of centralized strategic decision-making (Bedeian and Zammuto, 1991). Computer-based AIT links employees electronically, thus enabling them to better participate in the tactical decision-making process. As a result, AIT enhances the efficiency of decentralization and greater employee professionalism (see 3.2.2 below). In addition, AIT enables lower level personnel to have a greater indirect influence on strategic decision-making. This occurs because they often are the ones who select and filter the information and structure the data bases that form the input for the highly centralized strategic decisions.

CIM, by its very nature, results in a very high level of integration of workflow processes and, thus, a high level of interdependence among differentiated units. This interdependence increases the need for designing in effective integrating mechanisms across functional units. In addition, CIM often increases the need for market-based unit grouping (e.g. project or product task team grouping) — especially during the product design phase (Bedeian and Zammuto 1991).

3.2. Personnel Subsystem Analysis

At least three major personnel subsystem characteristics are critical to an organization's work system design. These are (1) the *degree of professionalism*, (2) *demographic characteristics*, and (3) *psychosocial aspects* of the work force (Hendrick 1997).

3.2.1. Degree of professionalism

Formalization can take place either on or off the job (Robbins 1983). When done on the job, formalization is *external* to the

employee. The rules, procedures, and human-system interfaces are designed to limit employee *discretion*. This typically characterizes unskilled and semi-skilled positions, and is what is meant by the term "formalization". Professionalism, on the other hand, creates *internal* formalization of behavior through a socialization process that is an integral part of education and training. Thus, from an ergonomics design standpoint, there is a trade-off between formalization of the work system and professionalization of the jobs that comprise it. If the work system is designed to allow for low formalization and, thus, considerable employee discretion, jobs and related human-system interfaces need to be ergonomically designed to require persons with relatively greater professional training or education. In summary, in the absence of formal decision rules and procedures, employees need to have the necessary professional knowledge and skills to make the decisions. Most often, it is the need to have employees that can deal with unanticipated or unique situations that creates the need for low formalization and more highly professionalized jobs.

3.2.2. Demographic factors

The demographic characteristics of the workers that comprise the organization's personnel subsystem can interact with the work system's design. Those that are most striking within the USA are the (1) "graying" of the work force, (2) age group differences in the value system of workers, (3) broadening of the cultural diversity of the work force, and (4) increase in the number of woman in the work force.

Graying of the work force. Since the mid-1970s as the post-World War II baby boom bulge moves through their working careers, the average age of the work force has been increasing at the rate of about 6 months per year. As a result, we now have a more experienced, mature and better-trained work force than existed in the 1970s and before. In fact, according to the bureau of labor statistics, since the mid-1970s the number of persons with post-high school education in the labor force has doubled. If employees are to feel fully utilized and remain motivated, work systems must accommodate to this change by becoming less formalized and decentralizing more of the decision-making - particularly where high formalization and centralization traditionally have characterized the work system's structure.

Value system shifts. Based on extensive longitudinal studies of work force attitudes and values, Yankelovich (1979) noted that those workers born after World War II have very different views and feelings about work than their predecessors. He further noted that these conceptions and values would have a profound effect on work systems. Yankelovich refers to this group of workers as the "new breed". This "new breed" of workers has three common principal values that distinguish them from those of the mainstream of older workers: (1) The increasing importance of leisure, (2) the symbolic significance of the paid job, and (3) the insistence that jobs become less depersonalized and more meaningful. When asked what aspects of work are more important, the "new breed" person stresses "being recognized as an individual", and "the opportunity to be with pleasant people with whom I like to work" (Yankelovich 1979). In terms of work system design, these values translate into a need for less hierarchical, less formalized, and more decentralized organizational structures; and for an attendant greater

professionalism to be designed into individual jobs and human-system interfaces then found in traditional bureaucracies. In short, work system design characteristics that allow for greater individual recognition and respect for an employee's worth. They also can enhance meaningful social relationships on the job.

Cultural diversity. In a number of urban areas in the USA and elsewhere, such as the Los Angeles Basin, immigration has resulted in work forces that are far more culturally diverse than existed several decades ago. It now is apparent that unless organizations accommodate to this diversity, it will adversely affect employee motivation and commitment. Much of the accommodation required has to do with changing organizational cultures to be more inclusive. Decentralizing some aspects of decision-making to allow greater employee involvement and control appears to be one structural change that can facilitate this accommodation process. This can include the use of participatory ergonomics in designing or modifying work systems.

Women. Although women have been entering the work force in progressively increasing numbers, there is no clear indication as to how these demographic changes will, or should affect work system design. The one possible exception is that, as women have moved into traditionally male positions, they have tended to emphasize the importance of modifying work systems and jobs so as to allow for greater social interaction and more flexible work hours.

3.2.3. Psychosocial factors

Harvey, Hunt, and Schroder (1961) have identified the higher-order structural personality dimension of concreteness—abstractness of thinking, or *cognitive complexity* as underlying different conceptual systems for understanding reality. In general, the degree to which a given culture or subculture (1) encourages by its child-rearing and educational practices an *active* exposure or openness to learning from new experiences or *diversity*, and (2) through affluence, education, communications media, transportation systems, etc, provides *opportunities* for exposure to diversity, the more cognitively complex the persons of that culture will become. An active exposure to diversity enables persons to develop new conceptual categories and sub-categories in which to store experiential data, or *differentiation;* and to learn new rules and combinations of rules for *integrating* information and deriving deeper, more insightful conceptions of problems and solutions. Persons who experience an active exposure to considerable diversity thus develop a high degree of conceptual differentiation and integration, or *abstract* functioning. Conversely, relatively closed-minded approaches to new experiences and/or a lack of exposure to considerable diversity leads to a more limited development of differentiation and integration in one's conceptualizing, or *concrete* functioning.

Concrete functioning consistently has been found to be characterized by a relatively high need for structure and order and for stability and consistency, a low tolerance for ambiguity, closedness of beliefs, absolutist thinking, authoritarianism, paternalism and ethnocentrism. Concrete persons tend to interpret their world more literally and statically than do abstract persons. They tend to see their views, values, norms, and institutional structures as relatively static and unchanging. Conversely, abstract adult functioning is characterized by a low need for structure and order and stability and consistency, a high

tolerance for ambiguity, openness of beliefs, and relativistic thinking. They also demonstrate a greater capacity for empathy and related strong people orientation; consequently, they are less authoritarian, paternalistic, and ethnocentric. Abstract persons have a dynamic conception of their world and *expect* their views, values, norms, and institutional structures to change. (Harvey *et al.* 1961).

Not surprisingly, relatively concrete work groups and managers function best in work systems characterized by moderately high vertical differentiation, centralization and formalization; and in which the structure and processes are unambiguous and relatively slow to change. In contrast, cognitively complex or abstract persons prefer more dynamic work systems, characterized by relatively low levels of vertical differentiation, formalization, and centralization.

3.2.4. Personnel subsystem implications for future work system designs

Much of the available data on personnel subsystem determinants of work system design comes from attitude surveys and projections from psychosocial and demographic studies. In spite of their somewhat tenuous nature, there is considerable convergence of these data: They point to a need for work systems of the future to be as vertically undifferentiated, decentralized, and lacking in formalization as their technology and environments will permit. Given the rapidly developing trend towards highly dynamic virtual organizations, and attendant work systems, this indication is most fortunate.

4. EXTERNAL ENVIRONMENTAL CHARACTERISTICS

Both the survival and success an organization depends on its ability to adapt to is external environment or, more specifically, to its *relevant task environment*. Relevant task environment refers to that part of the firm's external environment that can positively or negatively influence the organization's effectiveness — the organization's critical constituencies.

4.1. Types of External Environments

Some of the most common types of external environments to which organizations must adapt are as follows.

- *Socioeconomic*. Particularly the degree of stability, nature of the competition, and availability of materials and qualified workers.
- *Educational*. Especially the availability of facilities and programs, and the educational level and aspirations of workers.
- *Political*. Particularly the degree of political stability and the governmental attitudes toward (a) business (friendliness

versus hostility), (b) control of prices, and (c) welfare of industrial workers.
- *Cultural*. Social status system, and values and attitudes held towards work, management, etc., and the nature of trade unions and union-management relationships.

4.2. Environmental Uncertainty

Of particular importance to work system design is the fact that all specific task environments vary along two highly critical dimensions: *Change* and *complexity* (Duncan 1972). Degree of change refers to the extent to which a given task environment is dynamic or remains stable over time. The degree of complexity refers to whether the components of an organization's specific task environment are few or many in number (i.e. does the company interact with few, or many government agencies, customers, suppliers, competitors etc.?). These two environmental dimensions of change and complexity, in combination, determine the *environmental uncertainty* of an organization. Figure 2 illustrates this relationship for four different levels of uncertainty.

With a high degree of uncertainty, a premium is placed on an organization's ability to be flexible and rapidly respond to change. With low uncertainty, maintaining stability and control for maximum efficiency becomes paramount for survival. The greater the environmental uncertainty, the more essential it is for the work system's structure to have relatively low vertical differentiation, decentralized tactical decision-making, low formalization, and a high level of professionalism among its work groups. In contrast, highly certain environments call for a comparatively high level of vertical differentiation, formalization, and centralization, such as historically are found in classical bureaucracies. Today, most large and/or high technology corporations are operating in highly dynamic and complex environments.

5. INTEGRATING THE FINDINGS

In weighting the results from the analysis of each major sociotechnical system element for a given work system, if we assign a weighting of '1' to the technological subsystem results, we would assign a rating of approximately '2' to the personnel subsystem results, and a weighting of approximately '3' to the external environment analysis. Based on the literature, this roughly approximates the relative importance of each in determining the design of work systems.

It also is important to note that, once the characteristics of the overall work system have been determined and installed; jobs and related human–machine and human–software interfaces need to be modified, as necessary, to harmonize. For example, if low formalization and decentralized tactical decision-making characterize the desired work system, then the individual jobs

Degree of complexity	Degree of change	
	Stable	Dynamic
Simple	Low uncertainty	Moderate uncertainty
Complex	Moderate uncertainty	High uncertainty

Figure 2. Environmental uncertainty of organizations.

will need to be designed or modified to require the necessary knowledge and skills to carry out necessary evaluations and make sound decisions. Similarly, information and, as required, decision support systems would need to be designed to provide the necessary information and support. Training programs also may need to be designed or modified to bring employees up to the required level of professionalism. By the same token, procedures may need to be modified to make them less formalized and allow for greater flexibility of action.

REFERENCES

BEDEIAN, A.G., and ZAMMUTO, R.F., 1991, *Organizations: Theory and design.* (Chicago: Dryden Press).

DUNCAN, R.B., 1972, Characteristics of organizational environments and perceived environmental uncertainty. *Administrative Science Quarterly,* 17, 313—27.

HARVEY, O.J., HUNT, D.E, and SCHRODER, H.M., 1961, *Conceptual systems and personality organization* (New York: Wiley).

HENDRICK, H.W., 1997, Organizational design and macroergonomics. In G. Salvendy (ed.), *Handbook of human factors and ergonomics* (New York, NY: Wiley) (pp. 594—636).

PERROW, C.,1967, A framework for the comparative analysis of organizations. *American Sociological Review,* 32, 194—208.

ROBBINS, S.R., 1983, *Organization theory: The structure and design of organizations* (Englewood Cliffs, NJ: Prentice-Hall).

YENKELOVICH, D., 1979, *Work values and the new breed* (New York: Van Norstrand Reinhold).

Sociotechnical Systems Theory: the Sociotechnical Systems Model of Work Systems

H. W. Hendrick

Hendrick and Associates, Englewood, CO 80111, USA

1. INTRODUCTION

The sociotechnical systems model of work systems was developed empirically in the late 1940s and 1950s by Trist and Bamforth (1951) and their colleagues at the Tavistock Institute of Human Relations in the UK. Follow-on research by Katz and Kahn and their colleagues at the University of Michigan's Survey Research Center, and by many others, served to confirm and refine the sociotechnical systems model. The basic tenets of sociotechnical systems theory, and the coining of the term *sociotechnical systems,* can be traced back to the classic studies by Trist and Bamforth relative to the effects of technological change in a deep seam Welsh coal mine (DeGreene 1973).

2. THE TAVISTOCK STUDIES

Prior to the introduction of new machine technology, the traditional system of mining coal in the Welsh mines was largely manual in nature. The work was carried out by small teams of miners who functioned relatively autonomously. To a large extent, control over work was exercised by the group itself. Each miner could perform a variety of tasks, thus enabling most jobs to be interchangeable among the workers. The miners derived considerable satisfaction from being able to complete the entire "task". Further, through their close group interaction, workers could readily satisfy social needs on the job. As a result of these work system characteristics, the psychosocial and cultural characteristics of the workforce, the task requirements, and the work system's design were *congruent.*

The technological change consisted of replacing this more costly manual, or *shortwall*, method of mining with mechanical coal cutters. Miners were no longer restricted to working a short face of coal. Instead, they now could extract coal from a long wall. Unfortunately, this new and more *technologically-efficient longwall* system resulted in a work system design that was not congruent with the psychosocial and cultural characteristics of the workforce. Instead of being able to work in small, close-knit groups, shifts of 10 to 20 men were required. The new jobs were designed to include only a narrowly defined set of tasks. The new work system structure severely limited opportunities for social interaction, and job rotation no longer was possible. The revised work system required a high degree of interdependence among the tasks of the three shifts. Thus, problems from one shift carried over to the next, thereby holding up labor stages in the extraction process. This complex and rigid work system was highly sensitive to both productivity and social disruptions. Instead of achieving the expected improved productivity, low production, absenteeism, and inter-group rivalry became common (DeGreene 1973).

As part of follow-on studies of other coal mines by the Tavistock Institute (Trist, Higgin, Murray, and Pollock 1963), this conventional longwall method was compared with a new, *composite* longwall method. The composite work system design utilized a combination of the new technology and features of the old psychosocial work structure of the manual system. As compared with the conventional longwall system, the composite work system's design reduced the interdependence of the shifts, increased the variety of skills utilized by each worker, permitted self-selection by workers of their team members, and created opportunities for satisfying social needs on the job. As a result of this work system redesign, production became significantly higher than for either the conventional longwall or the old manual system; and absenteeism and other measures of poor morale and dissatisfaction dropped dramatically (DeGreene 1973).

Prior to the Tavistock studies, there was a widely held belief in *technological determinism*. It incorporated the basic concept of Taylorism that there is one best way to organize work. It was a belief that the "one best way" for designing the work system is determined by the nature of the technology employed. Based on the Tavistock Institute studies, Emory and Trist (1960) concluded that *different organizational designs could utilize the same technology*. As demonstrated by the Tavistock studies, the key is to select a work system design that is compatible with the characteristics of: (a) the people who will constitute the personnel portion of the system; and (b) the relevant external environment; and then to (c) employ the available technology in a manner that achieves congruence.

Although any given technology can utilize different work system designs, technology, *once employed in the system*, often constrains the subset of possible work system designs. This has become even more evident with the introduction of computers and related automation into managerial, administrative, production, logistical, marketing, and other facets of our modern complex systems. With this progressively increasing automation, it has become increasingly important to determine first the optimal *macroergonomic* design of the work system before fully proceeding with the micro-ergonomic design of human–machine and human–software modules, subsystems, and interfaces. At least *conceptually,* a top-down macroergonomic approach is essential to insure that the "dog wags the tail", and not visa-versa.

3. JOINT CAUSATION AND SUBSYSTEM OPTIMIZATION

Sociotechnical systems theory views organizations as open systems engaged in transforming inputs into desired outcomes (DeGreene 1973). *Open* means that work systems have permeable boundaries exposed to the environments in which they exist. These environments thus permeate the organization along with the inputs to be transformed. There are several primary ways in which environmental changes enter the organization: through its marketing or sales function, through the people who work in it, and through its materials or other input functions (Davis 1982).

As transformation agencies, organizations continually interact with their external environment. They receive inputs from their environment, transform these into desired outputs, and export these outputs back to their environment. In performing this transformation process, work systems bring two critical factors to bear on the transformation process: technology in the form of

a *technological subsystem*, and people in the form of *a personnel subsystem.* The design of the technological subsystem tends to define the *tasks* to be performed, whereas the design of the personnel subsystem primarily prescribes the *ways* in which tasks are performed. The two subsystems interact with each other at every human–machine and human–software interface. Thus, the technological and personnel subsystems are *mutually interdependent.* They both operate under *joint causation*, meaning that both subsystems are affected by causal events in the external environment.

Joint causation underlies the related sociotechnical system concept of *joint optimization.* The technological subsystem, once designed, is relatively stable and fixed. Accordingly, it falls to the personnel subsystem to adapt further to environmental change. Because the two subsystems respond jointly to causal events, optimizing one subsystem and fitting the second to it results in sub-optimization of the *joint work system*. Consequently, maximizing overall work system effectiveness requires jointly optimizing *both* subsystems. Thus, in order to develop the best possible fit between the two, joint optimization requires the *joint design* of the technical and personnel subsystems, given the objectives and requirements of each, and of the overall work system (Davis 1982). Inherent in this joint design is developing an optimal structure for the overall work system (see Sociotechnical analysis of work system structure).

3.1. Joint Optimization vs. Human–centered Design

At first glance, the concept of joint optimization may appear to be at odds with human-centered interface design. It might seem that human-centered design would lead to maximizing the personnel subsystem at the expense of the technological subsystem and, thus, to sub-optimization of the work system. In fact, this has proven not to be the case. In human-centered design, the goal is to make optimal use of humans by the *appropriate* employment of technology so as to optimize jointly the capabilities of each, and their interfaces. When a human-centered design approach is not taken, invariably the capabilities of the technological subsystem are maximized at the expense of the personnel subsystem. In recent decades, this often has been exemplified by automating and giving the human the left-over functions to perform, which sub-optimizes the personnel subsystem and, thus, the total work system. To achieve the appropriate balance then, joint optimization is operationalized through (a) joint design, (b) a human-centered approach to function and task allocation and design, and (c) attending to the organization's sociotechnical characteristics (see Sociotechnical analysis of work system structure).

4. THE SOCIOTECHNICAL SYSTEM ELEMENTS

As is inferred above, the design of a work system's structure and related processes involves consideration of three major sociotechnical system components that interact and affect optimal organizational design. These are the (1) *technological subsystem,* (2) *personnel subsystem,* and (3) *relevant external environments* that permeate the organization.

Beginning with the Tavistock studies, cited earlier, a consistent finding — and one fundamental to sociotechnical system theory — is that the four basic sociotechnical system elements are

mutually interdependent. If any characteristic of one of the four elements is changed, it will affect the other three. For example, if one changes some aspect of the personnel subsystem, it will impact on the technological subsystem, the work system's interaction with the external environment, and the structure and/or processes of the work system. It is of critical importance for ergonomists if these impacts on the other sociotechnical system elements and their interfaces are not anticipated and planned for as the impacts are likely to be *dysfunctional* and affect the work system in unanticipated and sub-optimal ways.

5. WORK SYSTEM HARMONIZATION AND ORGANIZATIONAL SYNERGISM

A widely accepted view among system theorists and researchers is that all complex systems are synergistic: that the whole is more than the simple sum of its parts. Thus, like biological and other complex systems, sociotechnical systems should be synergistic. Theoretically, because of this synergism, the following should tend to occur in our complex work systems.

5.1. When Work Systems Have Incompatible Designs

When work system structures and processes are grossly incompatible with their sociotechnical system characteristics, and/or jobs and human-system interfaces are incompatible with the work system's design, the whole is less than the sum of its parts. Under these conditions, we can expect the following to be relatively deficient: (1) productivity, especially quality of production, (2) accident rates, lost time injuries, and adherence to safety standards and procedures, and (3) motivation, related aspects of job satisfaction, and perceived quality of work life (e.g. psychosocial comfort, stress, etc.).

Further, these detriments may be greater than a simple sum of their parts would indicate.

5.2. When Work Systems Have Compatible Designs

When, taking into account its particular sociotechnical system characteristics, a work system has effectively been designed macroergonomically, and that effort has been carried through to the micro-ergonomic design of jobs and human–machine and human–software interfaces, then the work system design is *harmonized.* As a result, synergistic functioning becomes possible; and system productivity, safety, employee satisfaction, commitment, and perceived quality of work life will be much greater than the simple sum of the parts would indicate.

5.3. Implications for the Potential of Organizations

Assuming the above two propositions are true, then a fully harmonized sociotechnical system has the potential to improve productivity, scrap rates, safety, health, employee motivation and commitment, and the quality of work life *exponentially* rather than linearly. Instead of the 10% to 25% improvements in these system effectiveness measures that many ergonomists have experienced from their successful micro-ergonomic interventions, we should see improvements of 50% to 90%, or more (Hendrick 1986). As has been documented with actual cases (Hendrick 1996) this prediction is proving to be accurate.

REFERENCES

DAVIS, L. E., 1982, Organizational design. In G. Salvendy (ed.), *Handbook of industrial engineering* (New York: Wiley), sections 2.1.1–2.1.29.

DEGREENE, K., 1973, *Sociotechnical Systems* (Englewood Cliffs, NJ: Prentice-Hall).

EMERY, F. E., and TRIST, E. L., 1960, Sociotechnical systems. In C. W. Chruman and M. Verhulst (eds), *Management Science* (Oxford: Pergamon).

HENDRICK, H. W., 1986, Macroergonomics: a conceptual model for integrating human factors with organizational design. In O. Brown, Jr., and H. W. Hendrick (eds), *Human Factors in Organizational Design and Management,* Vol.II (Amsterdam: North Holland), pp. 467–78.

HENDRICK, H. W., 1996, The ergonomics of economics is the economics of ergonomics. *Proceedings of the Human Factors Society 40th Annual Meeting* (Human Factors and Ergonomics Society: Santa Monica, CA), pp. 1-10. (Presidential Address).

TRIST, E. L., and BAMFORTH, K. W., 1951, Some social and psychological consequences of the longwall method of coal-getting. *Human Relations, 4,* 3–38.

TRIST, E. L., HIGGIN, G. W., MURRAY, H., and POLLOCK, A. B. 1963, *Organizational Choice* (London: Tavistock).

Telework

A. Buessing

Department of Psychology, Technical University Muenchen, Munich, Germany

1. INTRODUCTION

The first pioneers of teleworking worked in the 1970s. However, telework was not widely accepted at that time; rather it was rejected by some because it was associated with simple homework or with day laboring.

Why is telework on the upswing these days? The increase of information and knowledge work connected with a rapid growth in performance and network capability and an enormous decline in the cost of technology allows for new qualities of telework (e.g. Andriessen and Roe 1994). Globalization, increase of international competition and a consumer-oriented market demand for new flexible forms of organization (modular, distributed, virtual) into which telework as an "any-time-any-place" form of work organization easily can be integrated. New ways of living that result from a dissolution of core family structures, an increase of female employment and an individualization of biographies add topicality to the contract between work, family and leisure, and let the work-family challenge become a relevant field for organizations.

2. TELEWORK: FORMS, PENETRATION AND POTENTIAL

2.2. Forms of Telework

Many phrases have been coined to describe forms of work known as telework: telecommuting, remote work, distance work and, of course, telework are the popular ones. "Telework" is preferred here as it places emphasis on work being done at distance rather than placing emphasis on commuting. Besides the well-known home-based version, telework covers a wide range of different forms of work organization.

A consensus among the different approaches seems to be that telework is a media-supported distributed work process which is spatially decentralized. Based on selected criteria, some essential *forms* of telework can be differentiated (Wigand *et al.* 1997):

According to spatial factors one can distinguish between *home-based* telework, *center-based* telework (satellite, neighborhood offices, telecentres), *mobile* telework and *on-site* telework, which takes places at the location of the customer.

In terms of time factors one can classify in *full-time* telework and *alternating* telework (e.g. between home and central office).

Contractual rules can be applied to differ between, for example, telework based on an *employment* or *freelance contract*.

With regard to technical links one can make a difference between *asynchronous off-line work*, more or less synchronized information exchange and *synchronous on-line-work*.

From a broader perspective telework is part of a media supported telecooperation process (Wigand *et al.* 1997) among distributed employees, organization units and organizations including telemanagement ("media supported distributed coordination") and teleservices ("media supported distributed provision of services", e.g. telebanking, teleshopping, telemedicine, etc.).

2.2. Penetration and Potential of Telework

Today telework is a widespread form of work. However, peculiarities are found in its organization among different countries according to differences in culture, geography, economy and societal background. Therefore, and because of the different terms and definitions in use, estimations of telework penetration usually are not very reliable.

The USA is considered the mother country of telework. According to recent data it leads with ~11 million teleworkers, which represents ~9% of the labor force in 1997 (http://findsvp.com/prls/pr97/telecomm.html). In Europe telework is also a fast spreading form of work organization. The 1998 European Information Technology Observatory (EITO; http://www.fvit-eurobit.de/eito) provides data for the European telework penetration based on the TELDET study from 1994 (Korte and Wynne 1996). According to this study there is a total of ~1.5 million teleworkers in Europe, which equals a penetration rate of 1% in 1994 (in comparison in 1994 USA = 4.54% and Canada = 3.5%). However, among the 15 countries of the European Union (EU) one finds large differences. Five countries range above the average (Sweden, Finland, UK, Ireland, The Netherlands) while all 10 others countries fall more or less below this average of 1%. More recent data, for example for Germany, however indicate a strong progress in telework dissemination during the past couple of years: results from a 1997 study show a that ~0.875 million teleworkers, which would be ~2.4% in Germany at that time.

Estimations of telework potentials are even more vague than penetration figures. While the growth of telework in the USA between 1995 and 1997 reflected an estimated annual rate of 15% (which equals to an extrapolated rate of ~12% of the labor force in 2000) the European Commission sees ~10 million teleworkers, which is ~7% for the EU in 2000.

3. TASKS, QUALIFICATIONS, COSTS AND BENEFITS OF TELEWORK

3.1. Tasks under Telework

To a large extent telework today is *knowledge work*; it is carried out by qualified personnel, specialists and knowledge workers. Tasks like programming, management, research and development, steering and supervision of technical processes are in the focus of interest. Though, one finds all kinds of jobs under telework: counseling, consulting, management, marketing, monitoring, research and development, "project-based organization" type of telework where teleworkers are involved in different projects and different roles (e.g. consultant, trainer, salesman, information broker). Only a small proportion among teleworking tasks represent secretary and personal support functions because they require frequent and high rates of communication and tuning. Therefore, most appropriate are those tasks that need little regular and especially not much cooperation and spontaneous personal face-to-face communication.

3.2. Qualifications for Telework

In the beginning telework was dominated by low skilled work, fragmented and easily controllable working tasks, typical for word processing or data input. Since then a shift to more or less sophisticated *qualificational demands* under telework took place. While high work fragmentation and strong external control still can be found today's teleworking jobs cover a wide range of

demands, autonomy, necessary skills and qualifications, often reaching a high level of job enlargement and job enrichment.

Moreover, the *upkeeping of qualifications* is of importance: for people who are bound to the household because of family obligations or by being handicapped, telework is a chance to avoid dequalification or to keep up with the progress of new qualificational demands (Büssing 1998). Under such circumstances telework quite often functions as a bridge over to a "regular" non-telework job.

Self-leadership is considered a key qualifications of successful teleworkers. But the step away from being guided and led requires new competencies because insufficient qualification for self-organization and self-control can lead to time pressure, overtaxing, stress and reduced quality. Especially telework with its flexible working time arrangements and relatively high time sovereignty stands for more responsibilities and requirements in dealing with time because leaving the "golden cage" of standardized working times and the defined distinction between working and leisure time also implies the loss of organizational rules with all its relieving moments.

3.3. Costs and Benefits of Telework

From the companies point of view telework increases productivity and contributes to advantages in competition. For example Gray *et al.* (1993) report of increases in productivity of 40–70% in the USA. Cost reduction results from a saving of space, lower rent and running costs, and a decrease in labor costs, for example, by hiring personnel from areas with lower income level or by cutting back additional labor costs, e.g. for overtime work.

While it is fairly easy to measure cost reductions it is quite difficult to estimate benefits and economic potentials of telework. An important benefit of telework is the increase of efficiency through a more flexible and autonomous organization of work and through a concentrated and uninterrupted way of working. Another aspect of economic relevance are nearness and orientation towards customers due to decentralization and flexibility of teleworking. Moreover, telework allows a better use of human capital. According to many studies teleworkers show higher work motivation, satisfaction and overall well-being which is reflected in an increase in productivity, reduced absenteeism and turnover. Finally, telework seems to be a good way of binding qualified personnel especially in cases of family obligations (e.g. birth of a child, demands for care) and change of residence.

Besides cost reduction and benefits telework may cause new costs itself. These costs are to a large extent so called transaction costs (i.e. costs for new methods of coordination and management); however, it is also found that additional expenditure for hardware and software, payments for telecommunication and refunds for running costs especially when telework is home-based.

From the *employee* point of view telework allows a spatio-temporal independence of work and non-work, and, therefore, telework creates chances to integrate and balance work and private life. This provides opportunities for persons with family obligations, handicaps and limited mobility, and flexibility who otherwise would not be able to participate in the labor market. However, telework also gains attractiveness for regular job holders because telework permits a self-regulated and individualized work organization and working time, some time sovereignty, a reduction

of stress caused by commuting and an increase of family satisfaction. An important advantage of telework is seen in an improved quality of leisure. That is, no loss of time when changing between work and private activities and higher personal work efficiency under telework increase the actually available amount of private time, while time sovereignty ensures time for private activities at the right moment.

Besides these benefits teleworkers are regularly faced with some potential disadvantages. For example, they have to deal with the difficult demarcation between the different spheres of life, a stronger involvement in family affairs during working time, a loss of organizational transparency and a loss of resources like social support from colleagues, secretary assistance, informational support etc.

4. CHANGES IN ORGANIZATION, WORK AND PRIVATE LIFE UNDER TELEWORK

4.1. Changes in Management

Many organizational changes are associated with the introduction of telework: changes in human resources, structure, management style, culture, technology and tasks (Chapman *et al.* 1995); changes in the *style of management* are among those crucial for the success of telework. In most cases it is the middle management that has to deal with this new form of work organization, and their interests and actual possession of responsibility, hierarchy and control will undergo further change. The implementation of telework demands new ways of supervision and leadership according to the network system of telecooperation. To deal effectively with the organization of telework managers need — more than ever before — social and communicative competencies with respect to interpersonal relation and use of media to be able to build up trust, integration and identity; and, as a part of this process they have to accept the role of coaches and resource managers.

Moreover, and this is one of the important alterations, managers have to delegate many of their typical tasks to the teleworkers, e.g. coordination of dates, meetings, working time scheduling, information management. Therefore, the step from leading to *self-leadership of teleworkers* is ambivalent for management: it leads to less demands and a relief on the one side but on the other side it goes along with feelings of loss and uselessness; these feelings in turn may lead to reduced commitment, a lack of readiness and even reactance.

Under telework *management by objectives* and task-oriented management focusing on results instead of input and processes becomes the important way. And this forces managers to become more detailed and obliged, and to direct their communication more carefully to goal setting.

4.2. Changes in Work Motivation

A spatio-temporally flexible form of work organization like telework increases demands for *intrinsic motivation* on the side of the employee. Different work orientation of teleworkers compared to employees working at central offices are ascribed by van Sell and Jacobs (1994: 85): satisfied teleworkers are intrinsically motivated by their job and the challenges of their tasks while "office programmers derived most of their satisfaction from managing others and from having status in the office hierarchy."

However, it is known that work orientation and work satisfaction are also dependent on the social context (e.g. organizational climate), on the social support and on the feedback and appreciation of colleagues and superiors. Therefore, by accounting for motivation and satisfaction telework should be organized in a way that sufficient time and opportunities are available for task-related communication and exchange as well as for social contact in the central office.

4.3. Interrelation between Work and Private Life

There are many aspects of the intertwining of telework and private life to be mentioned (Büssing 1998). Therefore, we restrict to two important ones: the increased compatibility between work and private life by means of time sovereignty on the one and the reduced demarcation between the two spheres of life on the other side. In comparison with all other groups of employees teleworkers report the lowest time pressure and the highest degree of usability of free time according to the demands from family, children and friends. The multiple loads and the rhythms of family, spouses, children and all kinds of institutions (e.g. kindergarten, school, authorities) can more easily be matched. Moreover, they can make use of private activities not only during the evening, rather private life is much more equally distributed across a day. This is the good side. The bad side is the increased and hardly controllable spilling over of problems, conflicts and stress from telework to private life, and vice versa. With quite a few implications: for example children and other family members have to be urged not to interrupt and disturb, teleworkers miss the "switch off" and relaxing from work at home as well as from family responsibilities at work, role conflicts arise more often, and self-management takes a lot of time. However, altogether many studies find a better balancing between work and private life especially for those groups of teleworkers which need to closely coordinate their spheres of life because of young children and other indispensable family obligations.

5. FUTURE OF TELEWORK

The future of telework is difficult to forecast because present research has not yet developed models and theoretical approaches to explain the various impacts and outcomes of the different forms of telework and so far most of the research is on a relatively low methodological level (e.g. van Sell and Jacobs 1994). Besides advances in technology and work organization the future of telework depends largely on the readiness of companies as well as on the readiness, needs and attitudes of potential users (Handy and Mokhtarian 1996). Although we find many indications that demands and needs for telework tend to grow we know fairly little about the specific factors which support these tendencies.

First of all one can not be certain of the development of telework: will it follow the home or the center-based line of organization? Like in the past the dominating form of telework in the future will very likely be differ from country to country dependent upon geographical, economic, social and cultural circumstances. However, center-based telework can avoid or restrict potential disadvantages of home-based telework like for example social isolation, deficits in communication and cooperation, and incomplete working tasks without loosing some of the advantages of a highly flexible form of work organization close to the place of residence (Stanek and Mokhtarian 1998). Second, we lack specific knowledge about interaction and communication under telework, and their influence on trust and social relationships to colleagues, superiors and customers. Third, we need to know much more about changes of values and attitudes in work and family under telework to forecast its future. What we know by now is that telework not only leads to a dissolution of actual boundaries between work and non-work but also to a disintegration of symbolic confines between the areas of life with all their — quite often conflicting — implications for performance, appreciation, esteem on the one side and relaxation, social presence, and feelings of security on the other side (Büssing 1998).

In any case, the intra-organizational penetration of telework will be limited by transaction costs; especially the cost of coordination will make the practice of telework unreasonable beyond certain levels of dissemination. It is more likely that in the future telework will become an element of *telecooperation* of inter-organizational networks, networks of self-employed etc. as well as of *virtual organizations* of any kind (Wigand *et al.* 1997). This telecooperative embedding of telework will produce new challenges for both, management and teleworkers, and it will lead to new ways of communication, organizational cultures and identities.

REFERENCES

ANDRIESSEN, J.H.E. and ROE, R.A. (eds), 1994, *Telematics and Work* (Hillsdale: Erlbaum).

BÜSSING, A., 1998, Teleworking and quality of life. In Jackson and van der Wielen (eds) *From Telecommuting to the Virtual Organization* (London: Routledge), 78–97.

CHAPMAN, A.J., SHEEHY, N.P., HEYWOOD, S., DOOLEY, B. and COLLINS, S.C., 1995, The organizational implications of teleworking. In Cooper and Robertson (eds) *International Review of Industrial and Organizational Psychology*, vol. 10 (Chichester: Wiley), 229–48.

GRAY, M., HODSON, N. and GORDON, G., 1993, *Teleworking Explained* (Chichester: Wiley).

HANDY, S.L. and MOKHTARIAN, P.L., 1996, The future of telecommuting. *Futures*, 28, 227–40.

JACKSON, P. and VAN DER WIELEN, J. (eds), 1998, *From Telecommuting to the Virtual Organization* (London: Routledge).

KORTE, W.B. and WYNNE, R., 1996, *Telework: Penetration, Potential and Practice in Europe* (Amsterdam: IOS Press).

STANEK, D.M. and MOKHTARIAN, P.L., 1998, Developing models of preference for home-based and center-based telecommuting: findings and forecasts. *Technological Forecasting and Social Change*, 57, 53–74.

VAN SELL, M. and JACOBS, S.M., 1994, Telecommuting and quality of life: a review of the literature and model for research. *Telematics and Informatics*, 11, 81–95.

WIGAND, R.T., PICOT, A. and REICHWALD, R., 1997, *Information, Organization and Management: Expanding Markets and Corporate Boundaries* (Chichester: Wiley).

Trade Union Approaches to Workplace Improvements

J. Lund

School for Workers, University of Wisconsin-Extension, Madison, WI 53703, USA

1. INTRODUCTION

Labor unions are institutions whose primary function is to represent the interests of their members at both the micro- (enterprise) and macro- (societal) level. Unions receive the bulk of their operating income from membership dues and must receive policy and expenditure authorization from the memberships. Ergonomics is rapidly becoming a major Labor union issue, although not in any uniform fashion and approaches to ergonomics differ widely at both the macro and micro-levels.

At least four variables impact the ergonomics approach a labor union will develop: (1) employer acceptance of repetitive stress injuries (RSI) as work-related and compensable under workers' compensation regulations or statutes; (2) union member support for ergonomics relative to their other concerns (e.g. wages, job security, other working conditions); (3) employer willingness to invest resources in ergonomic improvements and to involve the union in that process; and (4) public policy (laws, regulations) concerning ergonomics. Obviously, all four variables are interrelated: some deal solely with the micro-level while others operate within both micro and macro-levels.

Here, we will briefly explore how these four forces together with union structure and function, interact to determine what approach a labor union will take vis-à-vis ergonomics. In doing so, we will briefly describe the policy environment and the actors themselves and how these factors and the environment interact to determine the menu of ergonomic approaches available to unions.

2. VARIABLES AFFECTING CHOICE OF UNION ERGONOMICS APPROACHES

2.1. Employer Recognition that an Ergonomics Problem Exists

The employer's first step toward correcting ergonomic hazards is to recognize that they are a workplace problem. While they may be costly to acknowledge because of legal liability under worker compensation law, they are also costly to ignore because of their negative impact on productivity and employee morale. According to the US AFL-CIO, RSI in the US account for > 700 000 cases annually, or about one-third of all serious reported injuries costing an estimated $7000–30 000 per case for a potential annual cost of $20 billion (AFL-CIO 1997). (Gathering truly comparable and comprehensive international data on the extent and costs of RSI is complicated by a number of factors, particularly the lack of a commonly accepted definition of RSI and a universally accepted list of RSI injuries and illnesses which are compensable. Notwithstanding the above, several trends can be discerned: (1) the number of RSI is increasing. The Australian Commonwealth

Banking Corporation reported an increase in 275% of the number of RSI from December 1983 to November 1994 (Melhorne 1998). RSI occurrences have increased by 100% in the UK from 1989 to 1993 (TUC 1996); (2) the proportion of RSI to total reported occupational injuries and illnesses is growing as evidence from the province of Ontario (Ashbury 1995), in Sweden, where musculoskeletal injuries account for every fourth accident; in France, they account for > 50% of occupational illnesses (Euro Review 1994); and (3) the cost of compensating RSI is substantial. In the UK, RSI injuries cost ~£1 billion (TUC 1996) and in New Zealand, where in 1989, NZ$16 million was paid in compensation to 6200 RSI sufferers (Beadle and White 1992).) Unions may be able to make the economic case for employer ergonomic interventions by aggressively pursuing RSI work injury cases at the same time they engage the employer in discussion about the total cost of such injuries.

2.2. Labor Union Recognition of the Problem

Labor unions cannot function at the micro-level without some form of recognition of the union's representative status from the employer; they may be able to function only at the macro-level but will then be able to influence micro-events indirectly. Winning and maintaining this representational status is the first objective of a Labor union (in some cases, ergonomic issues may be a key to successful union recruitment of new members). This may be likened somewhat to a hierarchy of needs. The next basic needs the union must fulfill are the job and economic security of the union's members. If members are at risk of losing their jobs or if wages are about to be drastically cut, no Labor union is about to launch a full-blown ergonomics campaign in the workplace. Next, the union may have other more pressing safety and health issues, particularly protection from traumatic injuries. A union representing timber cutters is more likely to direct its limited resources on protecting its members from being struck by falling trees and limbs or being cut by power saws *before* it will launch an attack on the underlying causes of Raynaud's syndrome. Union members may be reluctant to embrace ergonomics (a higher order need) if it negatively affects lower order needs; if it is necessary to mechanize jobs to reduce the risks of highly repetitive work, with resulting job loss, unions or their members may actually oppose ergonomics. On the other hand, some national- and international-level labor organizations have developed ergonomics policies and programs for their members. (By way of example, the Trades Union Congress in the UK has published *Five Steps at Your Workplace* which focuses branch unions on workplace RSI surveys, attempting to cost out RSI, analyzing jobs and workstations for RSI risk factors and other actions (TUC 1997). The International Labor Office guidelines also recommend performing workplace surveys using a checklist aimed at identifying risk factors and what corrective actions can be taken (Rosskam 1997). Unions may also use government-developed or -sponsored checklists and guidelines, such as a *Guidance Note for the Prevention of Occupational Overuse Syndrome in the Manufacturing Industry* developed by Worksafe Australia (National Occupational Health and Safety Commission 1996).)

2.3. Employer Mobilization to Action

Assuming the threshold employer acceptance of the work-relatedness of RSI, unions must gauge the willingness of the
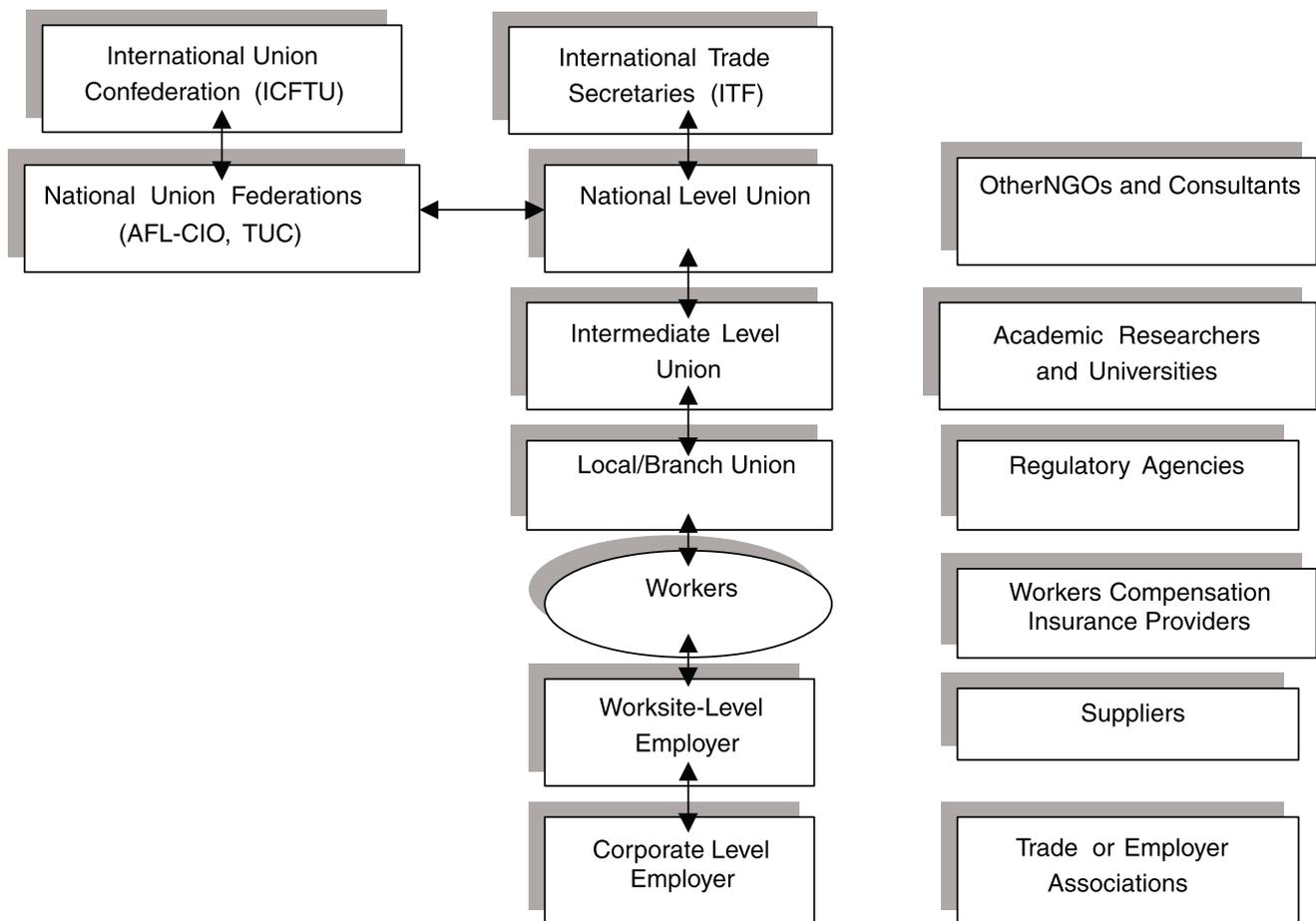
Figure 1. Schematic of actors.

employer to shift resources towards the identification and reduction of RSI risk factors; the profit motive or other strategic employer goals may drive employer reluctance to allocate what the union or workers feel are adequate resources to address the problem. Unions can play a very key role in both negotiating a larger allocation of resources as well as in enhancing the effectiveness of any employer risk reduction effort, but as management the employer must agree to commit resources in the first place. Employer willingness may be due in part to pressure from public policy in the form of mandatory standards, voluntary guidelines or perhaps pressure from third parties (Figure 1) such as a workers' compensation carrier, corporate officers, suppliers or customers. Employer willingness to commit may also be based in part upon the cost burden generated by RSI injury compensation and its impact on profitability. Unions may be able to focus the employer's attention in this area by persuasion or negotiation, or use of regulatory rights (e.g. filing complaints with regulatory agencies or claims for workers' compensation).

2.4. The Role of Public Policy

Public policy can play a significant role in determining the union's approach to ergonomics. Public policy may require the employer to identify and correct ergonomic problems via mandatory standards or voluntary guidelines; to compensate workers for various types of RSI; for the employer to consult with the union on what hazard identification and correction measures to deploy; or public policy may do none of the above. Unions may use

whatever policy exists and seek to change policy to meet its needs. The former can be a micro and/or a macro-effort, while the latter will generally be a macro-strategy.

Taking these four variables into account, the ideal environment for formulating an ergonomic approach for a labor organization would occur when the employer recognizes the existence of RSI problems and injured workers receive fair compensation; where the union and its members make ergonomics a high priority; where the employer commits significant resources to reducing RSI risk factors and is willing to collaborate closely with the union and where public policy supports and encourages such joint labor-management action and where both parties are able to build alliances with other third parties for assistance (insurance carriers, employer associations, suppliers, the government and universities). Generally, most cases will fall short of this ideal. Unions may analyze where shortcomings exist and develop appropriate strategies to achieve their ideal. A favorable macro-public policy can greatly assist unions in formulating an effective ergonomic approach; an unfavorable macro-public policy forces unions to focus on a more micro-level strategy, one workplace or industry at a time, with widely varying results.

3. ACTORS AND ENVIRONMENT

Labor unions are typically constituted both at the branch (workplace) and industry (national) level, although some organizations have intermediate levels at the regional or state

level. Internationally, many national level unions constitute International Trade Secretariats (ITS) such as the International Transport Workers Federation, which provide transnational research and support services to their affiliates. Unions also reach across branch and industry divisions to form federations, such as the American Federation of Labor-Congress of Industrial Organizations (AFL-CIO), Canadian Labor Congress (CLC) or LO (Landsorganisajonen in Norway, Sweden or Denmark). Internationally, there is a "federation of federations," the International Confederation of Free Trade Unions (ICFTU). National unions and their federations may also participate in a tripartite body (composed also of government and employer representatives) the International Labor Organization (ILO) headquartered in Geneva. These international bodies, the ILO in particular, make recommendations and conventions on a variety of subjects relating to safety and health, including ergonomics.

As Figure 1 suggests, employers are also organized at the branch (enterprise) level as well as the industry and national (corporate) level. Employers may also belong to employer associations that also may provide policy guidance and research in the areas of occupational health and safety.

There are a variety of "third parties" who can have a direct bearing on the union's ergonomic approaches, both at the macro- and micro-levels, which are shown to the right side of Figure 1.

Equipment suppliers and vendors can greatly assist unions and employers by developing ergonomically correct tools, workstations and other products which meet ergonomic criteria and reduce the risk of injury. Several branch unions in the US warehouse industry together with their employers have had success in convincing suppliers to reduce the weight of products stored in warehouses and to add better handles so as to reduce the risk of lifting injuries (obviously a more macro-level approach would have affected more workplaces). Other unions representing timber cutters have entered into direct dialog with power saw manufacturers to reduce the risk of vibration-induced white finger (Raynaud's syndrome). Telecommunications workers who install equipment are now entering into direct dialog with hand tool designers. Tool and equipment designers are motivated by the potential of increased sales and, depending upon national laws; the possibility of product liability suits if they fail to produce a safe product. Lawsuits against keyboard manufacturers in the U.S. may not have produced a jury verdict, but have resulted in at least one out-of-court settlement and a number of re-designed keyboards.

Workers' compensation insurance providers also have a role to play and may be important allies in formulating a union approach to ergonomics. In the USA, for example, many states require that these carriers offer "loss prevention" services to the enterprises to whom they provide workers' compensation insurance. Because of the significant increase in RSI, these carriers have a significant stake in reducing the number of injuries through ergonomic interventions. Several US unions have been able to work, at the branch level, with these carriers who provide training and technical assistance, and form a "tripartite" relationship with collaborating employers. Obviously, unless the employer is willing to enter into such an arrangement, unions cannot access these carriers since it is the employer, not the union, who is paying for the workers' compensation coverage.

Governmental agencies may also be a potential player in the union's ergonomics approach. At the macro-level, unions may lobby for mandatory standards requiring employers to develop and implement ergonomic interventions or to provide some form of compensation for RSI. The government may also sponsor or fund research identifying RSI risk factors or corrective measures. In the USA, the National Institute for Occupational Safety and Health (NIOSH), part of the Centers for Disease Control, has an Ergonomics Branch that conducts field evaluations of ergonomic problems at the request of unions or employers and it publishes a number of resources designed to reduce injury risk. It published a guide to reducing lifting injuries (NIOSH 1992) and many other helpful publications. In addition, the US Occupational Safety and Health Administration has, since the late 1970s, funded union and employer hazard recognition and abatement training efforts which have included ergonomics (OSHA's "New Directions" grant program). At the micro-level, union ergonomics approaches may utilize the government safety and health or labor inspectorates to enforce mandatory ergonomics standards, where they exist.

Universities and colleges may also provide important elements to union approaches to ergonomics. First of all, many universities in North America provide labor education services to unions, such as the University of Wisconsin's School for Workers, which at 75 years is the oldest university-based labor education program in North America. This school and many others have provided considerable ergonomic training and technical assistance to unions and employers in the USA In addition to such training and technical assistance, many unions have turned to university-based researchers to assist them in documenting the existence of RSI risks in their workplaces. Work by the University of Wisconsin-Madison's Department of Industrial Engineering has assisted the Communications Workers of America in documenting many RSI problems among video display terminal operators; the University of Michigan has assisted the United Autoworkers union similarly with the automobile industry. Building close relationships with labor unions is advantageous also to university researchers because of the access to workers and workplaces it affords.

Finally, there are other "non-governmental organizations" (NGO) which are potential players in a union ergonomics approach. This includes political parties who can assist unions in formulating policy at the macro-level, professional associations of occupational physicians, physical therapists and lawyers and in the USA, "Committees on Occupational Safety and Health" or "COSH groups." The latter are composed of many community activists, as well as health professionals and attorneys who can provide research and technical assistance to unions. One such example is the Massachusetts Center on Office Technology which helps workers, including those not represented by unions, in obtaining compensation for RSI and sponsors legislation and training programs.

Using the above set of actors and the four variables, it is possible to construct a model that can predict union approaches to ergonomics. The ideal case, as stated earlier, is where all four variables are positive. Where they are not, at least five scenarios can be discerned; these scenarios are described in more detail in Figure 2. As a general proposition, unions seeking to develop an ergonomic approach will turn both to public policy and to the third parties listed in Figure 1 to change the variable(s) from a negative to a positive value.

Figure 2. Union ergonomics scenarios and approaches.

Scenario	Description	Possible union approaches
Employer will not accept nexus between RSI injuries risk factors	Employer denies apparently valid compensation claims for RSI injuries filed by workers or has a policy announcing it will not compensate any or all RSI	• Sponsor/lobby legislation at state or national level to recognize some or all RSI as compensable or seek to change regulations concerning compensability with or without a coalition of like-minded third party groups • Using university, union and/or government researchers, document the extent of RSI and their relationship to work-related risk factors • Seek to work more closely with workers compensation carriers for more active involvement in risk factor reduction • Conduct membership awareness concerning RSI and their work-relatedness • Build alliances with NGO such as lawyers and medical providers who can provide assistance in documenting work relatedness and securing compensation
Union will not make ergonomics a priority	Union leaders and/or members see ergonomics as less pressing that other 'lower order' needs or union leaders	• Conduct training programs for officers, leaders and members to build awareness and review case studies of other unions • Document the extent and severity of the problem • Exchange information with other unions (national, international) about their response to similar situations
Employer will not commit adequate resources to address reduction of RSI risk factors	Employers are willing to recognize the problem, but not to address adequate resources to correct the problem	• Union can seek to negotiate the design of the RSI program • Union can start its own independent RSI program to document problems, attempt to secure intervention on a case-by-case basis and if not timely action, secure outside government intervention • Union can intervene with governmental agencies, universities, trade associations, suppliers and workers compensation carriers to provide low cost or free assistance • Union can turn to governmental enforcement agency to pressure compliance • Union can work with other third parties to start pressure campaign • Unions can develop their own model ergonomics programs or sponsor guidelines developed by others
Employer will not accept union as a collaborator in risk reduction program	Employer is committing to a risk reduction program but does not want significant union involvement	• Union can intervene with corporate organization of the employer • Union can petition for government intervention • Union can use other case studies as justification • Union seeks to negotiate a solution through collective bargaining

4. CONCLUSIONS

Before developing an ergonomic approach, labor unions must carefully examine the environment and actors and evaluate the four variables described above. At the micro-level, unions must determine employer acceptance of the work-relatedness nexus of RSI and employer willingness to reduce RSI risk factors and to accept the union as a collaborator in this process. Unions must also look within themselves to determine how much of a priority to assign to ergonomics. At the macro-level, the union must determine what support for RSI injury compensation and employer response to ergonomics problems are underwritten by public policy. To the extent that employer resistance is present on the first and third variable, unions need to shift attention to working with third parties as shown in Figure 1; the union may also need to combine forces with other unions and third parties to bring about a change in public policy. At the same time, employers need to better understand the environment in which unions operate and recognize the resources that can be brought to bear from such third parties. Unions by and large would prefer to work collaboratively with the employers at the micro-level, but where this is not possible, unions can and should have no difficulty in building alliances with third party groups and developing their own *independent* ergonomics approach.

ACKNOWLEDGEMENT

REFERENCES

http://www.aflcio.org/safety/factsheet.htm

ASHBURY, F. D., 1995, Occupational repetitive strain injuries and gender in Ontario, 1986 to 1991. *Journal of Occupational and Environmental Medicine*, 34 (4) 479.

BEADLE, C., and WHITE, S. 1992, Introduction to OOS (Occupational Overuse Syndrome). Sheet produced by Over Use Consumer Help awareness campaign. July, 1.

CARDY, S., 1998, Stop the strain drain. Pamphlet by the Trade Union Congress. March, 1.

European Foundation for the Improvement of Living and Working Conditions, 1994, Reducing RSI—a profitable measure. *Euro Review on Research in Health and Safety at Work*, 11.

GALLAGHER, J., 1996, Don't suffer in silence: designing RSI out of the workplace. *Conference Report of the TUC Ergonomics Society*, September, 4.

MELHORN, M. J. MD, 1998, Cumulative trauma disorders and repetitive strain injuries. *Clinical Orthepedics and Related Research*, 351, 112.

National Occupational Health and Safety Commission, Worksafe Australia, 1996, *Guidance Note For The Prevention of Occupational Overuse Syndrome in the Manufacturing Industry*, NOHSC: 3015(1996), July, 2.

ROSSKAM, E., 1997, Listening to our pain: preventing workplace injuries and illnesses through ergonomics. *International Labor Organization, World of Work.* No. 21, September/October, 3.

Types of Organizational Designs

H. W. Hendrick

Hendrick and Associates, Englewood, CO 80111, USA

1. INTRODUCTION

As described elsewhere in this encyclopedia, sociotechnical analyses can be carried out to help determine the optimal level of various structural dimensions to design into work systems. These include the degree of vertical and horizontal differentiation, the kinds and numbers of integrating mechanisms, where various kinds of decisions should be made, and the extent to which formal rules and procedures should direct activities and constrain decision-making. Ultimately, in designing work systems, these factors must be integrated into an overall structural form. A variety of *types* of structural forms have been successfully developed and used in organizations. Each has its advantages and disadvantages. The issue then, is to select the structural form with characteristics that best fit a given work system's needs.

There are four general types of organizational structure that are most commonly found. These are (1) classical *machine bureaucracy*, (2) *professional bureaucracy*, (3) *matrix organizations*, and (4) *free form designs* (Robbins 1983). It should be noted that large, complex organizations often do have relatively autonomous units with different overall forms. As a general rule, the smaller the organization, the greater is the likelihood that it will have a single overall type of work system structure.

2. CLASSICAL OR MACHINE BUREAUCRACY

This classical form of work system has its roots in two streams of thought: *Scientific management* and the *ideal bureaucracy*. Scientific management began with the work of Fredrick W. Taylor in the early 1900s. Although heavily influenced by Taylor's concepts, the classical bureaucratic design was actually conceptualized by Max Weber at the beginning of the twentieth century. Collectively, Taylor's and Weber's theoretical principles resulted in what today is known as the *machine bureaucracy* type of organizational structure. Its basic structural characteristics are as follows (Robbins 1983).

1. *Division of Labor.* Each job is narrowly defined, and consists of relatively routine and well-defined tasks.
2. *A well-defined hierarchy.* A relatively tall, clearly defined, formal hierarchical structure in which each lower office is under supervision and control of a higher one. Tasks tend to be grouped by function. Line and staff functions are clearly distinguished and are kept separate.
3. *High formalization.* Extensive use of formal rules, procedures, and detailed job descriptions to ensure uniformity and to regulate employee behavior.
4. *High centralization.* Decision-making is reserved for management. Employees have relatively little decision discretion.
5. *Career tracks for employees.* Members are expected to pursue a career within the organization; and career tracks form part of the work system design for all but the most unskilled positions.

2.1. Advantages

The primary advantages of the machine bureaucracy are efficiency, stability, and control over the work system's functioning. Narrowly defined jobs with a clear set of routinized tasks enable individuals to know their own function and the roles of others better, minimize the likelihood of error, require comparatively few prerequisite skills, and minimize training time and costs. Formalization ensures stability, control, and a smooth integrated pattern of functioning. Centralization further ensures control, reduces the education and training requirements for employees, and further enhances stability.

2.2. Disadvantages

There are two major disadvantages to the machine bureaucracy form of work system. First, this form tends to result in jobs that fail to utilize adequately the mental and psychological capacities of the workers. Thus they are lacking in intrinsic motivation. Second, these work systems are inherently inefficient in responding to environmental change and non-routine situations.

2.3. When to Use

Machine bureaucracy can be used when (1) system operations are repetitive or otherwise can be routinized, (2) the relevant external environments are comparatively simple and stable or predictable, or (3) the education and skill levels of the available personnel are relatively low.

To the extent that the above stated conditions do not exist, one of the other three forms of work system structure is likely to be more effective.

3. PROFESSIONAL BUREAUCRACY

Professionalism can be defined as the degree of training and education required by the design of specific jobs and related human–system interfaces. The professional bureaucracy is designed so as to rely on a relatively high degree of professionalism in the jobs that comprise the work system. Its differs from the machine bureaucracy design in that jobs are more broadly defined, somewhat less routinized, and allow for greater employee decision discretion (Robbins 1983). Accordingly, there is less need for formalization and *tactical* decision-making is decentralized. As with machine bureaucracies, positions are grouped functionally, are hierarchical, and *strategic* decision-making often remains centralized.

3.1. Advantages

As compared with the machine bureaucracy form of work system, there are at least three major advantages to the professional bureaucracy. First, professional bureaucracies can effectively cope with complex environments and non-routine tasks more efficiently. Second, jobs tend to be intrinsically motivating and better utilize the mental and psychological capabilities of employees. Third, they require less managerial tactical decision-

making and control, thus freeing management to give greater attention to long-range planning and strategic decision-making.

3.2. Disadvantages

They are not as efficient as machine bureaucracies for coping with relatively simple, stable environments. They require a more highly skilled work force and attendant greater training time. Thus, both salaries and training costs are relatively higher. Control is less tight and efficient. The distinction between line and staff functions is often less clear. The management skills required tend to be more sophisticated (i.e. a greater reliance on tolerance for ambiguity, and on persuasive and facilitation skills rather than on a simple and direct authoritarian style.

3.3. When to Use

A professional bureaucracy form of work system is to be preferred when the external environment is fairly complex and/or somewhat unstable, and there is an available applicant pool of professionalized workers. It is less optimal than machine bureaucracies for highly repetitive operations with a simple stable environment; or if the available management pool is highly authoritarian and concrete in its functioning.

4. THE CONCEPT OF ADHOCRACY

Although truer for machine bureaucracies, a major disadvantage of both forms of bureaucracy is that they tend to be inefficient in responding to highly complex, dynamic, or uncertain relevant external environments. In response to this shortcoming, two more recent forms of organization have evolved. These are the *matrix* and *free form* designs. Collectively, these two are known as *adhocracy* designs. An adhocracy can be described as a rapidly changing adaptive system organized around problems to be solved by groups having diverse professional skills. In terms of their structural dimensions, adhocracies are characterized by moderate to low complexity, low formalization, and decentralized decision-making (Robbins 1983). Because these work systems are staffed primarily by professionals, horizontal differentiation tends to be high. In contrast, vertical differentiation tends to be low. This low vertical differentiation reflects the low need for supervision because of the high level of professional staffing. Instead of formal rules and procedures, flexibility and rapidity of response are emphasized. It is for this reason that a high level of professionalism and an absence of administrative layers are needed.

4.1. Advantages

Adhocracies trace their roots to the task forces of World War II. The military created ad hoc teams to perform specific missions, and then disbanded them when the task was completed.

Team members' roles were flexible and often interchangeable. Sub-units would be added or deleted as required. These features enabled the teams to react quickly, efficiently, and creatively to a dynamic environment, and thus be highly effective. These are the same characteristics and resultant advantages that characterize modern adhocracies. When the primary organizational need is to be adaptive and innovative, and respond rapidly to changing situations and objectives, and when these responses require collaboration of persons possessing different specialties, the adhocracy forms of organization are more effective than the bureaucratic (Robbins 1983).

4.2. Disadvantages

All adhocracy forms have at least three major disadvantages: (1) *conflict*, (2) *social and psychological stress*, and (3) *inherently inefficient structures* (Robbins 1983).

4.2.1. Conflict

Since boss–subordinate relationships tend not to be clear, the lines of authority and responsibility are ambiguous. Thus, conflict is an integral part of adhocracy.

4.2.2. Psychological and sociological stress

Because of the dynamic nature of their environment and tasks, the structure of teams or units is temporary and work role interfaces are not stable. By contrast, the establishing and dismantling of human relationships is a slower psychosocial process that is stressed any time there is significant organizational change. Concrete functioning employees especially are likely to find it difficult to cope.

4.2.3. Inherent inefficiency

By comparison with bureaucracies, adhocracies lack both the precision and expediency that comes with routinization of function and structural stability. Accordingly, adhocracy designs are to be preferred only where these inefficiencies are more than offset by the gains in efficiency in terms of responsiveness or innovation.

5. THE MATRIX DESIGN

The matrix design is the most widely used form of adhocracy. It combines departmentalization by function with departmentalization by product line or project. As with bureaucracies, functional departments exist and tend to be lasting. Unlike bureaucracies, members of the functional departments are farmed out to project or product teams as new projects or product lines develop; and the combined technical expertise of the individual departments is required. When the need for a given department's professional input to the interdisciplinary team is no longer required, or when the level of effort reduces, employees return to their "home" department or transfer to another team. The product or project manager supervises the team's interdisciplinary effort, but each team member also has a functional department supervisor. Thus, the matrix design breaks the fundamental design concept of bureaucracy of *unity of command*.

5.1. Advantages

Matrix designs afford the best of two worlds: the stability and professional support of depth of functional departmentalization and the interdisciplinary response capability of ad hoc teams.

5.2. Disadvantages

In addition to those characteristics of all adhocracy designs that cause problems, the major disadvantage of matrix organizations is that employees must serve two bosses: their functional department head, who tends to be relatively long-term oriented and somewhat remote from the team member's immediate tasks, and the project team director who tends to be short-term oriented but immediate to the employee's present tasks. Serving two masters with overlapping supervisory responsibility and different

goals, responsibilities, and time orientations creates conflict and can disrupt organizational functioning.

A second major problem for the employee is that, when assigned too long to a project, he or she may have difficulty keeping technically current and may lose contact with his or her functional department. These consequences can both adversely affect the employee's career.

5.3. When to Use

The matrix organization is particularly well suited for responding to complex and dynamic external environments where both interdisciplinary responsiveness and providing for functional depth in individual disciplines are considered essential. Since many high technology organizations exist in these kinds of environments, the matrix form of work system is widely used.

6. FREE FORM DESIGNS

Of the four general types of organizational structure the *free-form* design is the newest and most rapidly growing. In its purest form, the free-form work system resembles an ameba — it continually changes its shape in order to survive (Szilagyi and Wallace 1990). The *raison d'être* of free form designs is responsiveness to change in very highly dynamic, complex and competitive environments.

In free-form systems, the functional departmentalization of bureaucracies is replaced by a *profit center* arrangement. Profit centers are results oriented and are managed by teams. Collectively, they constitute the work system. Free-form designs are characterized by very low formalization and vertical differentiation and highly decentralized decision-making. Very heavy reliance is placed on professionalism as expressed through participation and autonomy. As with matrix organizations, project teams are created, changed, and disbanded as required to meet organizational goals and problems. To function effectively in these work systems, managers and employees alike need to possess a great deal of personal flexibility, tolerance for ambiguity, and ability to handle change.

6.1. Advantages

The single major advantage of free-form designs is the ability to respond to highly competitive, complex, and dynamic relevant external environments with speed and innovation. This is the very purpose for which they exist.

6.2. Disadvantages

The free-form work systems have essentially the same units within disadvantages as matrix adhocracies, only to a greater extent. They thus require a highly professionalized work force to succeed.

6.3. When to Use

The free-form work systems can be used when (1) the organization's success or survival critically depends on speed of response and innovation, and (2) a highly professionalized work force and management pool are available. These features tend to characterize small to medium-sized high-technology organizations, and semi-autonomous, rapid prototyping "outlaw" large bureaucratic organizations, operating in highly dynamic, complex and competitive environments.

6.4. New Variations

The combination of new technological advances and the need to respond to complex, uncertain environments has led to new adhocracy forms of organization. Of these, two of particular note are the *modular* and *virtual* forms.

6.4.1. Modular organizations

The modular organization outsources non-vital functions while retaining full strategic control. Outside sources may be used to manufacture parts, handle logistics, and perform maintenance, housekeeping, or accounting activities. Structurally, the "organization" is a central hub surrounded by networks of outside suppliers and specialists. Modular parts readily can be added or taken away, as the situation requires.

6.4.2. Virtual designs

The virtual type of organizational structure is characterized by a continually evolving network of independent companies — suppliers, customers, and even competitors — to pursue common strategic objectives. These are linked together to share skills, costs, and access to one another's markets (Dees *et al.* 1995). In contrast to the modular firm, which maintains full strategic control, participants in the virtual firm give up part of their control and accept interdependent destinies. A virtual organization need not have a central office, organization chart, or hierarchy. Participating firms in a virtual organization may be involved in multiple alliances. They may form a virtual organization to attain specific strategic objectives, and then disband when those objectives are met. A major advantage of the virtual organization is ability to respond quickly to opportunities in a highly dynamic environment. Each subunit firm brings a particular set of competencies to the alliance, thus creating a more competitive, yet highly flexible entity.

REFERENCES

DEES, G. G., RASHEED, A.M., McLAUGHLIN, K. J., and PRIEM, R. L., 1995, The new corporate architecture, *The Academy of Management Executive, 9,* 7–20.

ROBBINS, S. R., 1983, *Organization Theory: The Structure and Design of Organizations* (Englewood Cliffs, NJ: Prentice-Hall).

SZILAGYI, A. D., Jr., and WALLACE, M. J., Jr., 1990, *Organizational Behavior and Performance* (5th ed.), (Glenview, IL: Scott Foresman).

Part 12

Methods and Techniques

Activity and Other Sampling Methods

J. R. Buck

Department of Industrial Engineering, University of Iowa, Iowa City, IA 52242, USA

1. INTRODUCTION

Sampling is a term typically used in industry to distinguish random or periodic observations from continuous observations. As an alternative to continuous observation, sampling has both advantages and disadvantages. On the advantage side, sampling is usually much less costly to perform. But it does have the disadvantage that results can be incorrect when sampling is improper. The key point in sampling is summed under the question, "can one obtain an accurate condition of the situation during a brief or at least short-term observation?" Unless an affirmative answer can be given, sampling should not be considered. It was that key observation that Tippett, the originator of activity analysis, made when he listened to the knitting mills in the midlands of England in the early 1930s. He noted that a momentary listening to the mills would disclose whether the machines were working properly. He called the sampling technique "ratio delay," and one occasionally hears that term yet today.

The most common form of industrial sampling is the case in which momentary observations are made at randomly determined times. This methodology is typically referred to as "work sampling" or "occurrence sampling." The basis of this methodology is simply binomial or Poisson sampling, which are well described in many books on statistics. In its simplest form, an event either occurs at an observation or not, and if that event occurs on 40% of those observations one would expect the event to occur with that frequency plus or minus sampling errors. Forty percent of 10 observations clearly has a much lower statistical precision than 40% of a 1000 observations. That precision is determined by the statistical confidence interval from the sampling.

There are many applications of activity analysis including: assessing the fraction of time that unproductive activities are occurring, machine interference studies in which one attempts to identify the relative frequency of various interference, setting time standards, and determining how various people spend their time. In most of these situations there is an advantage of identifying the high-frequency tasks as those can affect overall effectiveness the most and even small improvements there can make considerable improvement. Hence, activity analysis can be a useful tool in problem solving.

There are numerous variations of sampling used in activity analysis in addition to elementary random sampling. One is periodic sampling such as making observations every M minutes. If someone needs to be aboard a moving vehicle (e.g. plane, ship, car, bus), a high density of periodic observations can be very effective. In the 1950s, Christensen flew with the US Airforce and made observations of activities that the crewmember performed over the flight using periodic observations. He was investigating how crewmembers spent their time and if some of the activities were necessary. They ended up reducing some of the paperwork and reducing the crew size. About the same time MacFarland and Mosley observed bus drivers' hands and feet during the but trip to determine the frequency of left- and right-hand hand and feet movements. Their concern was whether these hand and feet activities were balanced between the left and right sides. As a result, they recommended an automatic shifting system for buses. The only real danger in a periodic set of observations is when the period selected is similar to the fundamental period of activities being observed.

2. TRADITIONAL ACTIVITY ANALYSIS

Activity analysis has traditionally consisted a person who is not part of the activity observing it, classifying the activity being observed, and recording it at each randomly selected points of time. These points of time are during a prescribed period. When a total of N observations are to be made during that period, the manager of the study needs to select N random numbers, order them in size and map those N numbers onto the period (Figure 1). Each random number should have enough digits so that each of the N can be unique. The remaining analysis follows the theory of binomial sampling. Poisson sampling is also used but infrequently.

2.1. Binomial Sampling

In basic binomial sampling, an event either occurs or not. If it occurs q percent of the time and N randomly timed observations are made, one would expect to see the event occurring on Nq observations. But with random sampling the actual number of observations will be more or less with a standard deviation of

$$V(X) = \sqrt{Nq(1-q)}$$

If X is the actual number of events observed, the probability of observing precisely that number is:

$$p(X) = \frac{N!}{X!\,(N-X)!}\; q^X (1-q)^{N-X} \tag{1}$$
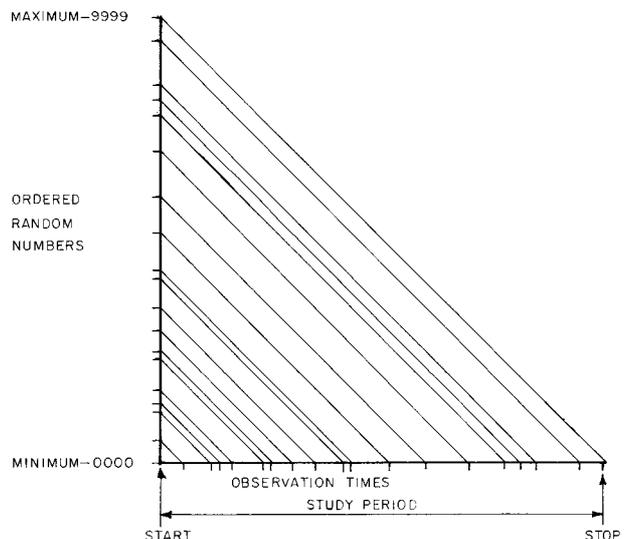
Figure 1.

for

$$0 \leq X \leq N$$

To find a confidence interval for q corresponding to the actual observations made, an integral of equation (1) over for the lower tail error and over q'' for as the upper tail error so that the sum of the two errors equals the confidence limit. However, that procedure is not as simple as recognizing that the binomial resembles the normal (Gaussian) distribution when , according to Dr Abraham Wald of Columbia University. To equate the binomial to the normal distribution, one should first recognize that the standardized normal deviate is:

$$Z_\alpha = \frac{X - \mu}{\sigma} \qquad (2)$$

Assuming that the binomial is being represented by the normal distribution, let us replace the normal mean m with the binomial mean Nq and replace the normal standard deviation σ with $\sqrt{Nq(1-q)}$ so that equation (2) becomes:

$$Z_\alpha = \frac{X - Nq}{\sqrt{Nq(1-q)}} \qquad (3)$$

Now dividing equation (3)'s numerator and denominator by $N > 0$:

$$Z_\alpha = \frac{[X/N - q]}{\sqrt{q(1-q)/N}} \qquad (4)$$

Note that the numerator of equation (4) describes the difference between the actual fraction of the time the event occurs q from the expected occurrence. Let the maximum acceptable difference between the actual q and the expected q be denoted as the acceptable error e. Solving equation (4) for N yields:

$$N = \left[\frac{Z_a}{e} \right]^2 q(1-q) \qquad (5)$$

When Z the normal standardized deviate for the desired confidence interval, say 1.645, 1.96 or 2.58 for 90, 95 and 99% confidence intervals respectively, the satisfying N shows the sample size that meets the confidence interval requirements for the specified error e. q can be estimated from the sample as

$$X/_N = q$$

Consider the numerical example where a 90% confidence interval and an error of 1% is acceptable in an activity analysis. Suppose that q is estimated as 30%. An appropriate sample size is:

$$N = \left[\frac{1.645}{0.01} \right]^2 0.3(1 - 0.3) = 5682 \text{ samples}$$

If during the sampling one finds that the estimated q differ from 0.3, for example, after 4000 observations, $X = 1120$, a better estimate of q is 1120/4000 = 0.28. Replacing 0.30 with 0.28 yields a smaller required sample size of 5455 samples. Repeating this calculation later in the sampling will allow the manager to assure a sample size that meets a desired confidence interval.

It should also be noted that nomographs are available for roughly approximating the sample size by entering the desired confidence interval and an acceptable error size relative to the random variable q.

2.2. Applications of Traditional Activity Analysis in Problem Solving

In plant maintenance operations a report of some malfunction or problem occurs and people go to the site. They diagnose the complaint, and send an individual or crew to perform the need maintenance or repair. Sometimes several crews are involved such as carpenters, plumbers and/or electricians. The manager of maintenance may be concerned with the effectiveness of the organization and ask for an activity analysis of the maintenance operations specifically looking for causes of being non-productive. Some of those causes occur when the crews are travelling to the work-site, waiting for spare parts, technical information or assignments, or experiencing delays due to interference within a crew or interference between different crafts. The specific improvements to be made depend upon the fraction of occurrences of these different causes of non-productivity. If travel is a large fraction, decentralization of maintenance should be considered. Delays due to spare parts, technical information or assignments implies the need to improve the support activities dealing with these three aspects. Interference within a crew may require crew composition or leadership and interference between crews involves crew scheduling.

Another application of traditional sampling occurs when a particular worker group is in short supply and management is considering aids for these workers in terms of assistants. An example may be nurses in a hospital. If management is to aid these nurses, they must find out how they spend their time in terms of activities that can be aided. If the activity analysis finds out that a very large amount of their time is spent in distributing bed-clothing to the rooms, this task may be handed-off to another group in the hospital that has extra time or some new personnel can be added to perform that activity.

An application of activity analysis in industry is determining the normal time required for a particular assembly operation. In this case, an activity analysis can be performed to see the fraction of time spend assembling this particular item. That time fraction coupled with other data allows one to determine a normal time (i.e. a time requirement for only performing the task, neglecting any time spent for personal needs, or resting from fatigue, or time lost due to failed support activities. The additional data needed are the total time worked as shown in the time records and the total number of assemblies performed. If only only a few persons performed the observed assembly activity, additional information of the speed of those persons observed is also needed. Numerous sources claim that the activity analysis procedure for estimating normal times are far less costly than traditional time study and that the accuracy is better because daily and weekly variations are factored into the data. Also many people report greater worker acceptance.

2.3. Some Speciality Forms of Sampling

There are several other forms of sampling that need to be identified for completeness. One is stratified sampling which is a procedure whereby sampling frequencies are apportioned to specified parts of the system. This technique is used for better representativeness or for greater precision on more important system features. Cochran (1957) describes this form of sampling in detail. It is used infrequently but some situations make it desirable.

Another form of sampling is where the external observer doing the sampling not only identifies the ongoing activity but also rates the speed of the observed person relative to the mythical "standard operator." Those estimated performance rates are often assumed to be normally distributed and the activity identifications as binomial. Hence, two analyses are usually made of such data. The technique is often referred to as "rated work sampling."

Another variation in activity sampling is to have the persons being observed to be their own observers. This sampling can be performed with a random reminder device that emits a signal at random points in time and at that time the person is supposed to enter the type of activity that is going on at time. Often, they are asked to denote where it took place and in some cases why. One source of random reminders is made by Divilbiss Electronics (Champaign, IL, USA). This form of sampling is particularly useful when outside observers might interfere with regular operations or when people under observation move from location to location extensively or dependent on variational factors so that outside observers would not be likely to know where the activities are going on.

3. PARAMETER MAPS

A useful graph in activity analysis is a parameter map of the statistical distribution being used. Parameter maps are simply rectilinear graphs where each axis represents a parameter of the probability distribution being used and plots on the graphs are the values of specific statistics. For example, rectangular distribution from parameter a to parameter b has a probability density function of:

$$p(x) = (x - a)/(b - a), \tag{6}$$

and the mean and variance statistics in terms of those parameters are:

$$E(x) = (a + b)/2 \tag{7}$$

$$V(x) = (b - a)^2/2. \tag{8}$$

The parameter map of this distribution is a rectangular graph with the x-axis representing parameter a and the y-axis representing parameter b. Isocontours of a specific $E(x)$ are straight lines of a unit slope from the upper left to the lower right starting at the y-axis at $y = 2E(x)$ and ending at the x-axis at $x = 2E(x)$. Variance isocontours start at the y-axis at $\sqrt{2V(x)}$ and this isocontour line increases linearly in an amount equal to a:

$$B = [2V(x)]^2 + a. \tag{9}$$

And Figure 2 describes this parameter map for this rectangular distribution.

Parameter maps are available for other distributions. One of particular importance, as will be shown below, is the beta
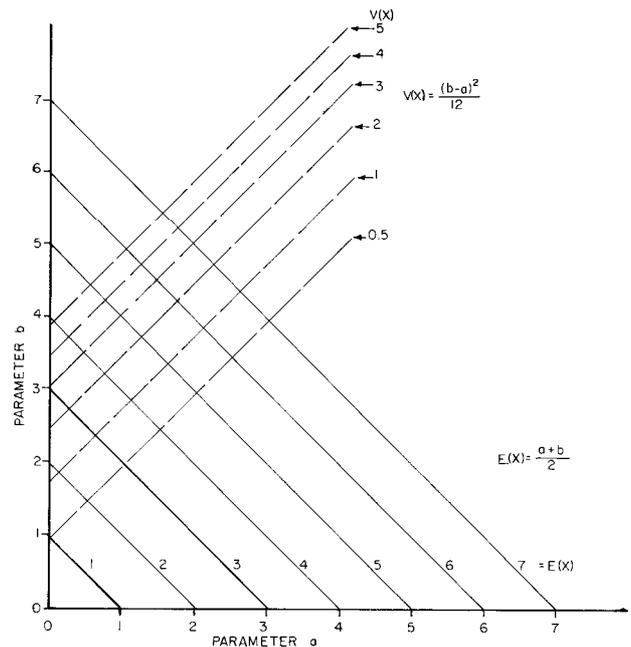
Figure 2.

distribution (of the first type) which has a probability function of:

$$p(q) = \frac{(a + b + 1)!}{(a - 1)!(b - 1)!} \, q^{a-1} (1-q)^{b-1} \tag{10}$$

where the parameters of this distribution are a and b. It follows that the mean and variance of the beta in terms of the parameters are respectively:

$$E(q) = \frac{a}{a + b} \text{ and } V(q) = \frac{E(q)[1 - E(q)]}{a + b + 1} = \frac{a}{a + b} \left[\frac{b}{a + b} \right] \frac{1}{a + b + 1} \tag{11, 12}$$

When estimates are available for the mean and variance, those estimates can be equated to equations (10) and (11) and solved simultaneously to find parameters a and b. It also follows that the mode of the beta (i.e. q that maximizes p(q)) is:

$$M(q) = \frac{a - 1}{a + b - 2} \tag{13}$$

Any two of these three statistics can be estimated to find parameters a and b. It follows that when the mean is fixed at $E(q)$, a contour of all combinations of parameters a and b that produce an equivalent mean fit the relationship.

In a similar vain, the contours for a constant mode $M(q)$ and variance $V(q)$ are defined by their respective relationships of:

$$b = \frac{a - 1}{M(q)} - a + 2 \text{ and } b = \frac{E(q)[1 - e(q)]}{V(q)} - a - 1 \tag{14, 15}$$
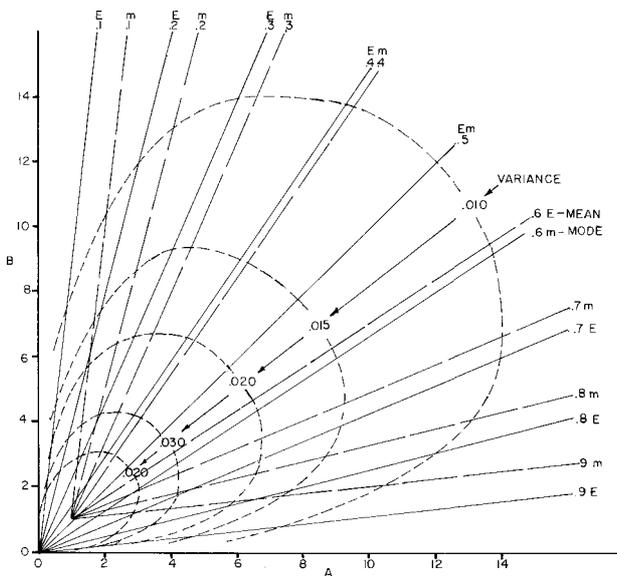
Figure 3.

Figure 3 provides a parameter map for the beta distribution showing all three statistics. Note that when $q = 0.5$, $E(q) = M(q)$ and so the mode and mean follow the same contour in that case. As q gets < 0.5, the mean is lower than the mode and as q gets > 0.5, the mode is lower than the mean. Also, both the mean and mode begin at $[a = 1, b = 1]$. Another important observation is that the constant variance contour curves around the map, emanating from the origin $[a = 0, b = 0]$ and returning there but crossing all $E(q)$ and $M(q)$ contours. Moreover, constant variance contours with smaller constant are farther from the origin.

4. SEQUENTIAL BAYESIAN ACTIVITY ANALYSIS

This type of activity analysis uses the binomial as does most traditional activity analysis but it starts off and ends up a bit differently. Bayes' theorem is the basis for this technique and this theorem states that an existing probability distribution can be modified under new evidence to a corrected posterior distribution. Accordingly, the starting basis of this Bayesian version of activity analysis may start with a prior distribution p(q) which may be a beta, as shown in equation (10). This prior distribution should be one that best fits the supervisor's or manager's current perception of the situation at the start. If neither of them have any prior knowledge, find some else who does. In a last resort, select a beta with $a = 1 = b$; an extremely conservative beta equivalent to the rectangular distribution where all variables q from 0 to 1 are equally likely. Thus, to start, one needs to estimate the expected value of the random variable q and the variance of q as envisioned at the start. One way to do that is to take estimates of $E(q)$ and $V(q)$, find the contours of these estimated statistics on the parameter map of Figure 3, trace those contours until they cross, and then find the parameters a and b that simultaneously match both. That beta distribution is the *a priori* or "prior" distribution. A second way is to insert these estimated statistics into equations (9) and (10) respectively and solve simultaneously. Results will be the same either way.

4.1. Can People Estimate Statistics Subjectively with Reasonable Accuracy?

The discussion about Bayesian activity analysis is briefly interrupted here to focus on an important question about starting

this analysis when the only available data are determined subjectively held by the people in the organization. Since Bayesian analysis can start with a model that is based on subjective estimates, it seems questionable that accurate final results can occur if the beginning is inaccurate.

A wide variety of studies was summarized by Sheridan and Ferrell (1970s), who have shown a few cases where people are presented with a relative frequency of an event and asked to estimate the fraction subjectively. Those studies demonstrated that people are accurate in their estimation but their precision is

for relative frequencies of q from ~25 to ~75%.

Above or below those limits precision deteriorates. So it is unreasonable to request greater precision than ~5% or to ask for relative frequencies that are very high or low.

4.2. Second Step in Sequential Bayesian Activity Analysis

As stated above, the first step is in establishing a prior beta. The second step involves sampling. Use binomial sampling by making N' observations, where N' is a comfortable day or half-day sample size. In reality, N' is the number of samples to be taken before updating the posterior distribution. During those N' observations ($N' > 0$), find X' events (observations when a particular activity is ongoing) through binomial sampling. It then follows through Raiffa and Schlaifer's (1961) conjugate distribution theory that:

$$A = a + X' \quad \text{and} \quad B = b + N' - X' \qquad (16, 17)$$

where A is the corresponding parameter of the posterior distribution and B is similarly the second parameter. The mean and variance of the posterior is:

$$E(q') = \frac{A}{A + B} \quad \text{and} \quad V(q') = \frac{E(q')\left[1 - E(q')\right]}{A + B + 1} \qquad (18, 19)$$

With respect to the beta parameter map, the prior distribution was at the point $[a,b]$ and the sampling moved one's knowledge
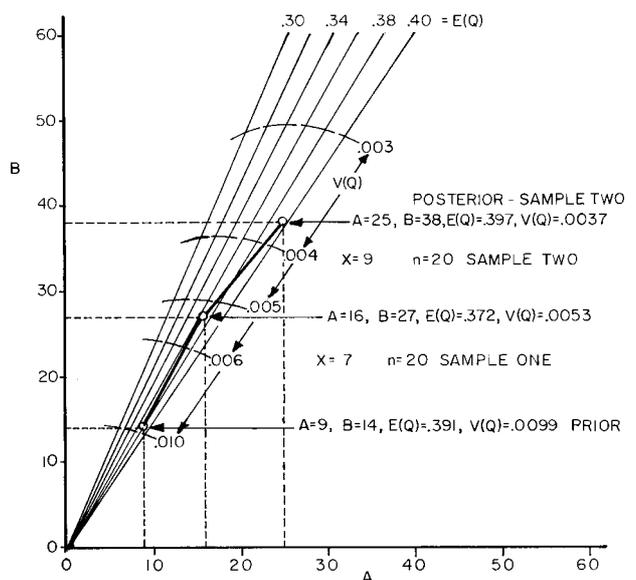
Figure 4.

about the situation to be represented as the point [*A*,*B*]. If further sampling is contemplated, that posterior point becomes a prior point to the next episode of sampling. In this way sampling evidence causes the prior distribution (as represented by the point [*a*,*b*]) to the posterior distribution (as represented by the point [*A*,*B*]). With more sampling, there is an incremental change following the same routine but with the prior point as the last sampling's posterior. Figure 4 illustrates the sequential chaining of points along a path on a parameter map from the initial prior to successive posterior points.

4.3. Numerical Example

Suppose that the prior beta has an estimated mean of 1/3 and a variance equal to 0.008 (or . . .). $\sqrt{V(q')} = 0.0891$). It follows from equation (18) that $2a = b$. From equation (19) one can calculate . $\frac{1}{3}\binom{2}{3}$ By substituting $2a$ for b in the last equation, one calculate $a = 8.93$ or ~9. Since $b = 2a = 18$. If 10 samples were made and the event was observed three times, $N' = 10$ and $X' = 3$ so that the posterior parameters are $A = 9 + 3 = 12$ and $B = 18 + 10 - 3 = 25$. The posterior mean is $E(q') = 12/(12 + 25) = 0.324$ and the variance is:

$$V(q') = \frac{0.324(1 - 0.324)}{12 + 25 + 1} = 0.00577$$

Note the reduction in variance from 0.008 to ~0.006.

4.4. How Does One Know When to Stop?

One can stop when one is satisfied, but if a statistically acceptable basis for stopping is desired, one should continue until the confidence interval of q is sufficiently small. Statistical sufficiency is determined by a stated acceptable error e and a specified probability that the error is not greater. The confidence interval for a beta distribution is, $q_1 \le E(q') \le q_2$ where the maximum acceptable error is $e_1 \le E(q') \le q_1$ and $e_2 \le q_2 - E(q')$. It is important to recognize that the beta distribution is not symmetrical about the mean $E(q')$ so that the lower interval does not necessarily equal the upper interval. If there is particular concern about being too small but less about being too large or vice versa, e_1 and e_2 may differ. A short table of beta confidence intervals are found in a number of sources (Schmitt 1989). Knowing parameters A and B, one can look up the confidence interval in Table 1.

Linear interpolation is usually an acceptable approximation if the parameters are between a pair in Table 1, but such interpolation is not simple unless one of the parameters is a tabled value.

Tabled confidence intervals for the beta distribution are really Bayesian high-density intervals that are defined at equal ordinates at the endpoints and an integral between endpoints equal to 0.90, 0.95, or 0.99 in the cases shown here. Please note that the mean $E(q')$ is not a midpoint between endpoints.

Table 1. 95% confidence interval table to find

a	b		
9	11	0.241	0.661
9	11	0.145	0.462

An alternative approximate method requires one to compute the skewness of the posterior distribution which is:

$$\alpha_3 = \frac{2(B - A)\sqrt{A + B + 1}}{(A + B + 2)\sqrt{AB}} \quad (20)$$

Once the skewness and variance are determined, the equation for the proper confidence interval and for the lower or upper partial confidence interval can be found and the standardized partial confidence and full confidence intervals computed. Each standardized full or partial confidence interval is a regression equation of tabled values so the resulting equations consist of a constant plus three terms: skewness, 1/*A** or 1/*B**, and the square of skewness, each with a coefficient. The standardized partial confidence interval at the 90% level is:

$$\frac{\sqrt{V(q')}}{E(q) - q_1} = 0.605 - 0.0437\,\alpha_3 - 0.0156\,\frac{1}{A^* \text{ or } B^*} + 0.0997\,\alpha_3^2$$

This particular regression equation had a squared coefficient of determination $R^2 = 0.98$ that shows that 98% of the variance in the tabled data was explained by the equation. Also shown, a mean squared error of prediction is 0.00000145, so the approximation is quite good. Whenever a blank is shown, that term does not need to be used. After the values are inserted and calculations are made, one equation computes:

$$\frac{\sqrt{V(q')}}{E(q') - q_1}$$

and the other:

$$\frac{\sqrt{V(q')}}{q_2 - q_1}$$

Note that the right hand side of the first equation is K and the second is K' so that:

$$E(q') - q_1 = \frac{\sqrt{V(q')}}{K} \quad \text{and} \quad q_2 - q_1 = \frac{\sqrt{V(q')}}{K'} \quad (21,22)$$

It easily follows that equations (22) − (21) = $q_2 - E(q')$

4.5. Numerical Example

Consider the case where $A = 120$ and $B = 240$ where the mean $E(q') = 1/3$ and the variance is $V(q') = 0.0006156$ and the SD = 0.2481:

$$Skewness\,\alpha_3 = \frac{2(240 - 120)\sqrt{361}}{362\sqrt{120(240)}} = \frac{4560}{61433.44} = 0.07423$$

$$\alpha_3^2 = 0.27245$$

The 90% confidence full and partial interval equations are:

$$\frac{0.0006156}{0.3333 - q_1} = 0.605 - 0.0437(0.07423) - \frac{0.0156}{120} + 0.0997(0.07423)^2 = 0.602176$$

$$\frac{0.0006156}{q_2 - q_1} = 0.301 + \frac{0.0155}{120} + 0.0321(0.074227) = 0.303512$$

Since

$$\frac{0.0006156}{0.602176} = 0.333 - q_1$$

it follows that $q_1 = 0.3323$

$$\frac{0.0006156}{0.303512} = q_1 \, 0.3323$$

and $q_2 = 0.33434$.

This makes the 90% confidence interval from 30.35 to 33.43% with 2.98% below the mean and 0.1% above.

While this is only an approximation but the regression $R^2 = 0.98$ at least and the mean MS of prediction error was at most $.0^4 145$.

4.6. Managing a Sequential Bayesian Activity Analysis

In sampling, the number of observations made each day is a management decision. More samples per day speeds up the study so that the results are available earlier but there are fewer daily variations so that the study's representiveness is decreased. Deadlines need to be met, however, and so managers must make a tradeoff between meeting deadlines and being more representative.

Available observers is an important resource for managing activity analysis studies and every observation an observer makes requires an amount of time for the person to leave the office and go to the observation site and return. The workday divided by the sum of the travel plus observation time represents the maximum possible number of observations made per day per observer less allowances for personal time and conflict. Those allowance are likely to be 10% or more. Sometimes managers have more assistance available and they can increase the rate of sampling. However, managers must be careful when increasing the sampling rate that more samples are not taken at only specific days of the week or hours of the day, as those increases can bias the study if the ongoing activities are not uniform over time.

A feature of sequential Bayesian activity analysis which can aid managers as they plan sampling activities in the future is to look ahead to see where the study is likely to meet the imposed conditions to complete the study. In most studies several different activities are under observation and so any critical activities are likely to occur less than 50% of the time. Accordingly, the beta distribution relative to each critical activity will likely have a long right tail and a mean well below 50%. The resulting confidence interval will probably have a larger $q_2 - E(q')$ interval than the interval $E(1') - q_1$ that are the acceptable errors at the study completion. Those maximum errors are approximately $2\sqrt{V(q')}$.

Consequently, find $V(q')$ which equates management's maximum error. That $V(q')$ contour and the intersecting $E(q')$ contour provides an estimated stopping point along the study path as plotted on the parameter map and it can be located as the expected posterior point $[A,B]$. If the current state of sampling is at the point $[a,b]$, the path of the remaining part of the study should follow near the isomean contour. Figure 5 presents a parameter map with these points marked out. If an equilateral triangle with legs parallel with the two axes is constructed from the current point $[a,b]$, the two legs will go from that point horizontally to $[a + n, b]$ and vertically from the current point to $[a, b + n]$. The satisfying value of n is the expected number of added samples needed. If n divided by the normal rate of sampling per day will exceed the deadline, it is clear that a greater sampling rate is required or a larger error may have to be accepted. Figure illustrates this look-ahead feature. This concept provides a look-ahead basis for better study management.

Another aspect for study management is to view daily plots on a parameter map and inspect for either strange behavior or specific patterns that do not conform to the statistical assumptions. One such case is the trend of constant change in the $E(q)$ over time. Binomial sampling is based on a constant q.

4.7. Other Bayesian Sampling

There is a development of Bayesian Poisson that is similar to the binomial sampling above which is given by Buck *et al.* (1995). Also, computerized systems for these Bayesian approaches are available from Phan Gia of the University de Moncton, New Brunswick, Canada.

5. FINAL REMARKS

Activity sampling (or work sampling) is an observational approach to finding out how people spend their time. Typically it is
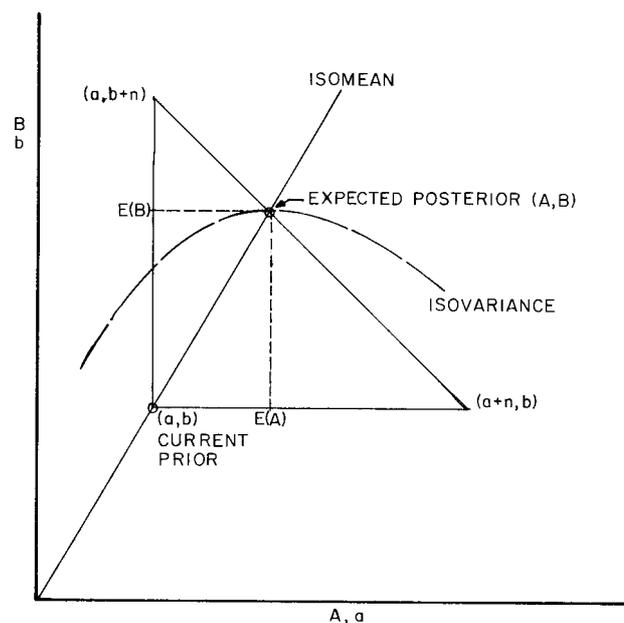
Figure 5.

performed with outside observers who look and classify the ongoing activity but in some cases the performing people can simply record their activity at random point of time. At the end of the study, the percentages of time the people spend on each specific activity can be estimated with reasonable accuracy and precision, depending upon the sample size. Traditional sampling techniques have been either binomial (Bernoulli) or Poisson sampling. The former is far more common and illustrated above. Also shown was Bayesian activity sampling and some of its advantages.

REFERENCES

BRACKEN, J., 1966, Percentage points of the beta distribution for use in Bayesian analysis of Bernoulli processes. *Technometrics*, 8, 687–94.

BUCK, J.R., ASKIN, R.G. and TANCHOCO, J.M.A., 1983, Advances in sequential Bayesian work sampling. *IIE Transactions*, 15, 19–30.

BUCK, J.R. and TANCHOCO, J.M.A., 1974, Sequential Bayesian work sampling. *AIIE Transactions*, 6, 318–26.

HEILAND, R. and RICHARDSON, W., 1965, *Work Sampling* (New York: McGraw-Hill).

RAIFFA, H. and SCHLAIFER, R., 1961, *Applied Statistical Decision Theory* (Boston: Division of Research, Graduate School of Business Administration, Harvard University).

SCHMITT, S.A., 1969, *Measuring Uncertainty: An Elementary Introduction to Bayesian Statistics* (Reading, MA: Addison-Wesley).

TIPPETT, L.H.C., 1935, Statistical methods in textile research uses of the binomial and Poisson distributions, a snap-reading method of making time studies of machines and operatives in factory surveys. *Journal of the Textile Institute Transactions*, 26, 51–5.

WHITT, W., 1979, A note on the influence of the sample on the posterior distribution. *Journal of the American Statistical Association*, 74, 424–6.

AET Ergonomic Job Description Questionnaire

K. Landau

Institute of Ergonomics, Darmstadt University of Technology,
Petersenstrasse 30, 64287 Darmstadt, Germany

1. INTRODUCTION

The origins of the AET (Abeitswissenschaftliches Erhebungsverfahren zur Tätigkeitsanalyse; ergonomic job analysis procedure) date back to a study (Rohmert et al. 1975) ordered by the German government to investigate discrimination against women at work with respect to pay. A job analysis procedure was required that allowed a detailed investigation of workload and strain within a given work system. At that time there was no job analysis procedure that could readily be used, although the position analysis questionnaire (PAQ) of McCormick *et al.* (1969) seemed to provide a basis for the psychological items.

Starting from that particular research project, the AET (Landau et al. 1975, Landau, 1978, Rohmert and Landau 1979, 1983) has been continuously developed over the last few years and applied in different industrial situations to analyze a wide range of shop-floor and management jobs (a total of about 7000 job analyses).

The AET has been developed for the universal analysis of work systems, the content of work ranging from "production of forces" to "production of information". The procedure can be used for selection, placement, training, job classification, rehabilitation, job design, occupational medicine, work safety, etc.

2. JOB ANALYSIS

This article describes job analysis as a subsection of a comprehensive system of work studies that covers the analysis of the individual components of person-at-work systems as well as the description and scaling of their interdependences. Job analysis starts from a model of work activity and assesses all relevant aspects of

- the work object
- the work resources
- the working environment
- the work tasks
- the work requirements

with regard to stress and strain considerations. Therefore, the basis of our understandig of job analysis is composed of the theoretical model of the person-at-work system and the concept of the simultaneous distinction and interdependence of stress and strain.

3. APPLICATIONS OF JOB ANALYSIS

The concept of job analysis as a universal "wideband" method allows the solution of quite a variety of practical and scientific problems. Applications are possible in the following fields:

1. Analysis of requirements and work design
   - Workplace documentation (e.g., for the investigation of accidents and analysis of industrial safety)
   - Work design
   - Design of work resources
   - Systematic reduction of stress
2. Industrial organization
   - Preparation of organizational changes
   - Design of work sequences
   - Organizational work design and work structuring
   - Organization of shift hours and rest allowances
3. Personnel management
   - Personnel recruitment
   - Personnel selection and placement
   - Integration of handicapped persons and other groups into the company
   - Basic and advanced education
   - Work evaluation and remuneration
4. Vocational counseling and research
   - Health status in different job classes
   - Job classification with respect to requirements
   - Vocational counseling and information.

According to work scientists, the main advantage of job analysis is its usefulness in adapting the work to the worker by constructive and methodical work design, and in adapting the worker to their work by adequate education, exercise and selection. In particular, the systematic application of job analysis procedures can either completely replace extremely expensive physiological investigations or at least restrict them to small bottleneck situations.

4. THE AET PROCEDURE

AET is oriented towards the elements and flows of the person-at-work system.

It has three parts (Table 1):
- Part A: analysis of the person-at-work system
- Part B: analysis of tasks
- Part C: analysis of demands

Besides the usual statistical requirements of an analysis procedure, a fundamental problem arises over theoretical justificationæasking for the selection of characteristics within the outlined classification of the procedure and for the scale level of the selected characteristics.

Particularly when selecting the items, it is important to consider the following basic facts:

- It is not possible that a fully completed catalog of items can be expected as part of the selected theoretical concept, especially in view of the economy of the procedure.
- Only those items which are important for numerous person-at-work systems should be included.
- The selection of items implies a certain judgment; hence "author-specific traits" influence the system of analysis, e.g., experience, standards of values, opinions.

So as the selection of the theoretical model forms the basis of the procedure, the selection, scaling and description of items within this model entails a series of subjective assumptions. However, these assumptions are acceptable as long as several analysts achieve reliable results by applying the procedure.

Parallel to the development of AET, an attempt was made to limit the subjective influences on the construction of items by using an iterative procedure development with respect to the experience of company specialists and ergonomists.

If one acknowledges the subjective, author-specific influence on the selection of a theoretical model substantiating AET on the

Table 1. Contents of the AET

PART A: WORK SYSTEM ANALYSIS

1. Work objects
   1.1 Material work objects
   1.2 Energy as work object
   1.3 Information as work object
   1.4 Human beings, animals, plants as work objects
2. Equipment
   2.1 Working equipment
       2.1.1 Equipment tools
       2.1.2 Means of transport
       2.1.3 Controls
3. Work environment
   3.1 Physical environment
       3.1.1 Environmental influences
       3.1.2 Dangerousness of work and risk of occupational disease
   3.2 Organizational and social environment
       3.2.1 Temporal organization of work
       3.2.2 Position in the organization of work sequence
       3.2.3 Hierarchical position in the organization
       3.2.4 Position in the communication system
   3.3 Principles and methods of remuneration
       3.3.1 Principles of remuneration
       3.3.2 Methods of remuneration

PART B: TASK ANALYSIS

1. Tasks relating to material work objects
2. Tasks relating to abstract work objects
3. Person-related tasks
4. Number and repetitiveness of tasks

PART C: JOB DEMAND ANALYSIS

1. Perception
   1.1 Mode of perception
       1.1.1 Visual
       1.1.2 Auditory
       1.1.3 Tactile
       1.1.4 Olfactory
       1.1.5 Proprioceptive
   1.2 Absolute/relative evaluation of perceived information
   1.3 Accuracy of perception
2. Decision
   2.1 Complexity of decision
   2.2 Pressure of time
   2.3 Required knowledge
3. Action
   3.1 Body postures
   3.2 Static work
   3.3 Heavy muscular work
   3.4 Light muscular work, active light work
   3.5 Strenuousness and frequency of movements

selection of items and on the determination of scale levels, codes and aids of classification, then one must also recognize the need to include technical criticism to achieve a progressive development of AET. The goals of the iterative AET development consisted of increasing its accessibility to the users while safeguarding or improving the statistical quality criteria.

The development of AET was oriented towards certain groups of users. Besides addressing ergonomically trained researchers and practitioners, it also addresses ergonomists. And AET is also aimed at industrial psychologists, representatives of labor and management, etc.—people who are interested in its results but who may not actually apply it.

AET part A describes the person-at-work system. The description and scaling refers to the objects of work and to the equipment and the working environment. The objects of work are analyzed under material, energy and information aspects. When a person is an object of work, it is essential to investigate the characteristics of that person's group. Finally, the qualities of material objects of work (like raw materials) are analyzed. The equipment, the work instruments (e.g., tools and implements) and other operating materials (e.g., hardware and software) are of interest. In this case the ergonomic system of classification is completed by technical aspects. The analysis of the environment is related to the physicochemical conditions in the workroom, the organizational and social working conditions and the principles and methods of pay.

There are 36 items for working and operating materials, 50 for the physical, organizational and social working conditions and 24 for the economic working conditions.

The object of investigation is the human activity in the person-at-work system; this is composed of the tasks which have to be carried out. The tasks stem from the purpose of the person-at-work system. In this sense part B of the analysis procedure represents a link between the tasks to be fulfilled and the resulting demands exerted on the working person. So the analysis of the person-at-work system is followed by the task analysis; this is carried out using 31 items subdivided according to the objects of work.

Task analysis is followed by demand analysis. Here the items are chosen to maximize the number of body functions considered. This is true for the function of power and energy generation in the different organs (stress during postural work and static work, heavy dynamic work and active light work) as well as for the functions of perception, decision and action in different mechanisms of information processing. The classification depends on deciding which area is required for receiving information, making decisions and taking actions. In demand analysis one distinguishes between the demands of reception of information (17 items), the demands of information processing (8 items) and the demands of information output or activity (17 items).

5. OBSERVATION INTERVIEW

The analysis of the job is done in the form of an observation interview, which means the necessary analytical data is collected first by observation of the job and working environment and second by interviewing the incumbent and the incumbent's superior (Figure 1).

Each AET item consists of a question outlining the state of affairs to be grasped and indicates the code for classifying this item. Sometimes examples are given as classification aids. The explanations clarify the questioning of the AET characteristic in view of extent, delineation and classification, but they cannot be taken as a complete and binding instruction for the rating. The classification of characteristics can only be done by using the corresponding code. The different codes are as follows:

- *Significance code (S)*: The importance or significance of this aspect for the task should be estimated in relation to other tasks or activities. Use a range from 0 to 5.
- *Duration code (D)*: This is based on a shift lasting eight hours. Eight hours is assumed even if the incumbent is a part-time worker, allowing comparisons between the work content of
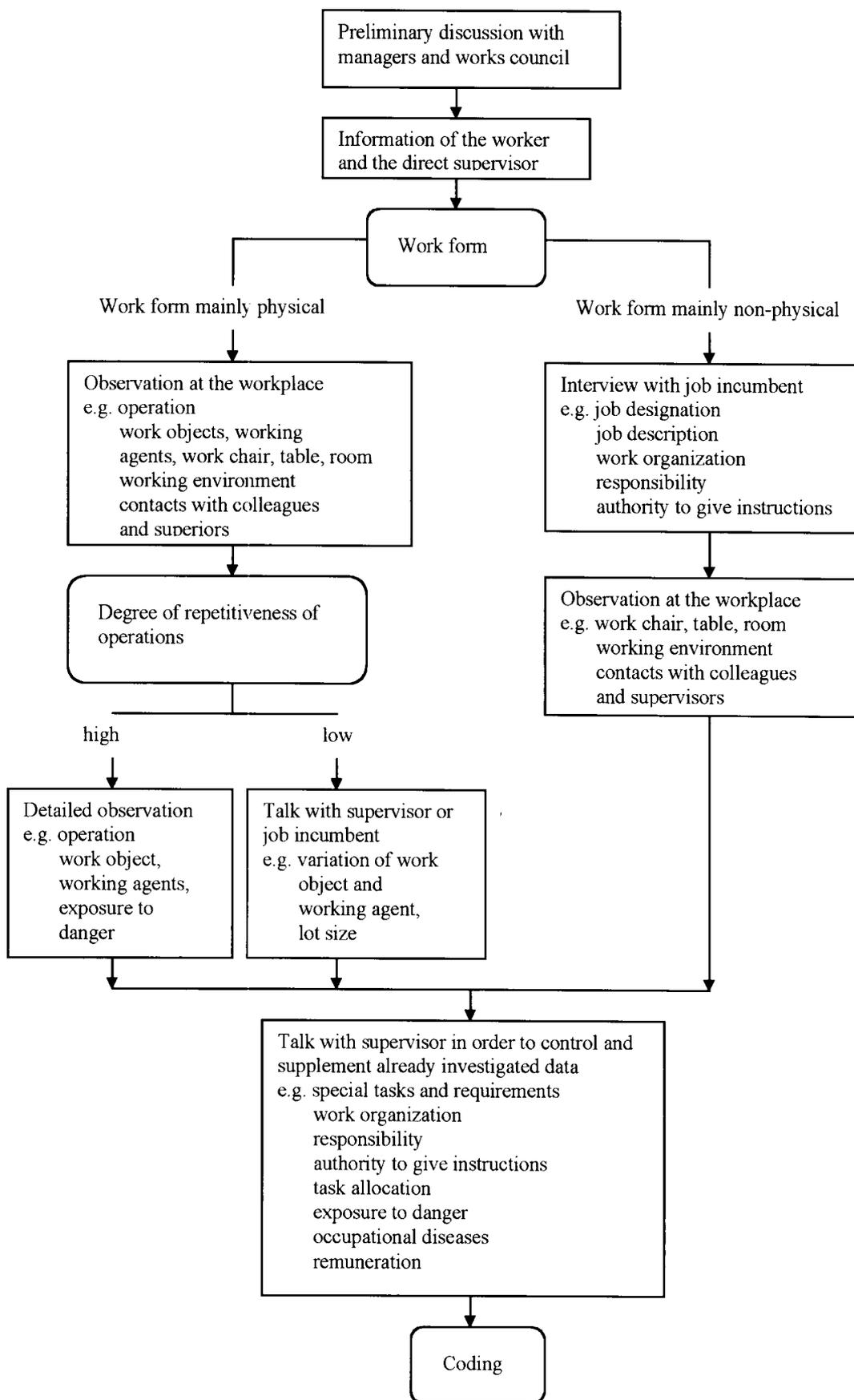
Figure 1. Observation interview

job profile: console operation		date of assessment:	branch: public service	incumbent: male
		0 %	50 %	100 %

working system:		
mode of assembly		
degree of technicalization		
degree of repetitiveness		
equipment:		
for transformation of material		
for energy and information	W	
for men as work objects		
means of transport: stationary		
means of transport: non-stationary		
other equipment		
positioning elements	H	
resources for state identification	W	
technical auxiliaries		
main tasks:		
behaviour-centred analysis		
fabrication, assembly	W	
operating, controlling	W	
checking		
supervising	W	
transporting, entering, arranging	W	
selling, negotiating, presenting	W	
planning, organizing	W	
coding, transmitting, arranging	W	
combining, analysing	W	
services	W	
demands:		
reception of information:		
visual	W	
auditory	W	
tactile, thermo-sensory		
olfactory and gustatory		
proprioceptive	W	
accuracy of the reception of inf.	W	
information-processing		
complexity of decision	W	
temporal scope of decision	W	
necessary knowledge	W	
activity		
postural work	Z	
static work	Z	
heavy dynamic work		
active light work	Z	
environmental influences/physical		
illumination	Z	
climate		
surface temperature		
vibrations		
noise	Z	
other environmental influences		
exposure to danger	S	
environmental influences/socio-organiz.		
temporal work organization	AS	
structural organization	HS	
contacts	H	

Figure 2. Example of an AET profile analysis

jobs with different shift hours. Use a range from under one-third of shift time to whole shift time.

- *Frequency code (F)*: This code characterizes the temporal distribution and position of stress sections. Use a range from 0 to 5.
- *Alternative code (A)*: The alternative code only asks about the presence of a characteristic. Enter 1 if the work characteristic in question does apply; enter 0 if not.
- *Exclusive code (E)*: The exclusive code is always related to only one specific question, e.g., multiple properties of a working instrument. Use a range from 0 to 5.

An AET analysis in the field is directly followed by the coding of the AET items. This coding has to be done on a standard form or in direct dialogue with the computer. The time required for an observation interview is usually 1 to 3 hours, depending on the analyst's practical experience, the type of job and the repetition rate of work processes.

6. AET EVALUATION

The evaluation of AET codings has to provide answers to the following questions:

1. What are the differences in the jobs regarding work content, objects of work, work instruments, workplace, working environment and work organization in different branches of industry, different enterprises, different departments, different wage groups?
2. What are the differences in the job characteristics of jobs requiring different education and job-related training?
3. To what extent do the job characteristics of native and foreign workers differ?
4. To what extent do the job characteristics of industrial employees, office workers, executive personnel and government workers differ?
5. Which jobs are particularly similar or differ markedly in view of different strain-relevant stress components?

For interrelated and basic requests, the best solution is through using univariate evaluation methods which are supplemented by a series of multivariate methods. For univariate methods it is possible to use well-known procedures of descriptive statistics and profile analysis. Appropriate multivariate techniques are cluster analysis, discrimination analysis, factor analysis and multidimensional scaling.

Having obtained the item group scores in the form of profiles, they can be used to give a graphic survey of the extent or the duration of stress experienced during the execution of jobs or groups of activities. This type of evaluation is known as profile analysis (Figure 2).

Running down the chart are the characteristics of the

workplace and the types of demand, and the horizontal bars show the maximum AET classification in percent.

Job analysis procedures are often criticized because of their unsuitability for a wide range of applications. But it is a costly business to develop universally applicable job analysis procedures. The optimal solution is to make the procedure as universal as possible within the financial limitations imposed. Universality is defined here as suitability for the interpretation and generalization of data obtained from different investigations, different sectors of industry or different companies.

REFERENCES

LANDAU, K., 1978. Das Arbeitswissenschaftliche Erhebungsverfahren zur Tätigkeitsanalyse — AET im Vergleich zu Verfahren der analytischen Arbeitsbewertung, *Fortschrittliche Betriebsführung*, 27 (1), 33–38.

LANDAU, K. and ROHMERT, W., 1989, Introduction to the problems of job analysis. In Landau and Rohmert (eds) *Recent Developments in Job Analysis*. Proceedings of an International Symposium on Job Analysis, University of Hohenheim, March 1989 (London: Taylor & Francis), pp. 1-24.

LANDAU, K. and ROHMERT, W., 1992, Evaluation of worker workload in flexible manufacturing industry. *International Journal of Human Factors in Manufacturing*, 2 (4), 369-88.

LANDAU, K., LUCZAK, H. and ROHMERT, W. 1975. Arbeitswissenschftliche Erhebungsbogen zur Tätigkeitsanalyse. In Rohmert, W. and Rutenfranz, J. (eds) *Arbeitswissenschaftliche Beurteilung der Belastung und Beanspruchung an unterschiedlichen Arbeitsplätzen*. Der Bundesminister für Arbeit und Sozialordnung. (Bonn).

McCORMICK, E.J., JEANNERET, P.R., and MECHAM, R.C., 1969, *The Development and Background of the Position Analysis Questionnaire (PQA)*. Occupational Research Center, Purdue University, Report 5.

ROHMERT, W., 1983, Determination of stress and strain at real work places: methods and results of field studies with air traffic control officers. In Moray, N. (Hrsg.) (1979): *Mental Workload, Its Theory and Measurement*. NATO Conference Series, Series III: Human Factors Col. 8 (New York: Plenum Press) pp. 423–443.

ROHMERT, W. and LANDAU, K. 1979. *Das Arbeitswissenschaftliche Erhebungsverfahren zur Tätigkeitsanalyse* (AET). *Handbuch und Merkmalheft* (Bern, Stuttgart, Wien: Hans Huber Verlag).

ROHMERT, W. and LANDAU, K., 1983, *A New Technique for Job Analysis* (London: Taylor & Francis).

ROHMERT, W., RUTENFRANZ, J., LUCZAK, H., LANDAU, K. and WUCHERPFENNIG, D. 1975. Arbeitswissenschaftliche Beurteilung der Belastung und Beanspruchung an unterschiedlichen industriellen Arbeitsplätzen. In Rohmert, W. and Rutenfranz, J. (eds) *Arbeitswissenschaftliche Beurteilung der Belastung und Beanspruchung an unterschiedlichen industriellen Arbeitsplätzen*. Der Bundesminister für Arbeit und Sozialordnung (Bonn), pp. 15–250.

Basic Ergonomics Checklists

K.Kogi

The Institute for Science of Labour, 2-8-14, Sugao, Miyamae, Kawasaki, Kanagawa, Japan

1. THE PLACE OF ERGONOMICS CHECKLISTS

Ergonomics checklists are widely used to evaluate ergonomics-related conditions of plans, designs, systems, equipment, facilities, and existing conditions of work and life. Various checklists have been developed for different purposes, usually as a tool for prioritizing action plans in ergonomic design or redesign, comparative studies, and training. In an ergonomics checklist, a selected number of items are presented in the form of a list with yes/no reply boxes or other types of noting down the evaluation result for each item. Such a checklist has an obvious advantage of taking into account important items relevant to the check purposes. A checklist for workstation design, for example, presents check items related to work space, layout, postures, controls and displays, varied operations, work outputs, physical environment, and work organization. In this way it helps its users go through relevant ergonomic aspects of workstation design.

From the results of applying a checklist, we can know items which are relevant or marginal for the work situation under consideration. We use the checklist results to improve the original designs or existing conditions. There are two basic types of ergonomics checklists. Differences in the design concept of these two types are listed in table 1.

The first type presents a list of items that should be analyzed and evaluated by the checklist users. These items are presented in order to evaluate whether each item under analysis is appropriate from ergonomics points of view. Often, certain evaluation criteria are described so that the checklist users can see if these criteria are met. For example, the checklist users are asked to see if "all controls related to a particular function or operation or having sequential relations are grouped together," then the reply may be yes or no, or alternatively to indicate which groupings of controls are satisfactory or unsatisfactory. The same item may be mentioned in a checklist as "grouping of controls," meaning that the checklist users should see if controls are adequately grouped. This first type checklists may be called "analysis checklists."

The second type presents a list of actions that could be taken to improve the existing designs or conditions. Usually, typical remedial actions are presented so that the checklist users can select those items which are useful to improve the conditions under consideration. In order to assist the users, actions that are relatively easy to take are given. For example, the checklist users are asked whether they propose an action so as to "place frequently used materials, tools and controls within easy reach" or "make displays and signals easy to distinguish from each other and easy to read'. This second type checklists may be called "action checklists." An example of such checklists is presented in *Ergonomic Checkpoints: Practical and Easy-to-implement Solutions for Improving Safety, Health and Working Conditions* (International Labour Office 1996). In this publication, developed jointly by the International Ergonomics Association (IEA) and the International Labour Office (ILO), each checkpoint is presented in the form of a low-cost improvement action. The attached checklist presents each action followed by a question "Do you propose action?" with reply boxes for "no", "yes" and "priority". The last box "priority" is for ticking if the reply is "yes" and if that action should be given priority in the user's opinion.

The design of an ergonomics checklist depends largely on the purpose of its use. Analysis checklists are suited for inventory purposes so that the users can check important aspects of a job, a workplace, or worker tasks from ergonomic points of view or check specific design aspects. For example, the Position Analysis Questionnaire (PAQ) (McCormick 1979) and the Ergonomic Job Analysis (AET) (Rohmert and Landau 1985) comprise items required for assessing job performance as a whole. Such checklists are particularly useful for knowing problem areas and comparing different jobs or workplaces.

Action checklists as represented by one incorporated in the IEA/ILO Ergonomic Checkpoints are useful for prioritizing practical improvement actions and for training in identifying practical improvements. By selecting those actions in the list which are applicable for improving the existing work situation, the users can compare selected practical improvements so as to find priority actions covering major areas or specific aspects of the work situation. Action checklists are not necessarily aimed at presenting all the relevant aspects, but guiding the users in finding applicable improvements and developing their own ideas. Action checklists are thus particularly useful in initiating improvement plans and in action training.

2. AREAS COVERED BY ERGONOMICS CHECKLISTS

Ergonomics checklists are to assist users in making informed decisions to enhance efficiency at work, safety and health, and physical and psychological well-being of people at work or using facilities or equipment. Many checklists therefore covers broad areas so that the users can examine work systems as a whole and identify necessary improvements. The selection of areas covered by each checklist depends on the specific approaches taken. Ergonomics checklists proposed to analyze work systems usually try to cover major areas of job characteristics, whereas there are

Table 1. Design concept of two main types of ergonomics checklists

	(a) Analysis checklist	(b) Action checklist
Purpose	Ergonomic job analysis orwork systems evaluation	Finding practical improvements in work systems or situations
Check items	Features of a job or a work system to be considered in better design	Actions to be taken to improve existing conditions
Check procedures	Examining whether given criteria are met or the item is satisfactory	Considering whether given actions can be proposed
Outcome	Analysis report identifying items requiring improvement	Priority actions to be taken for immediate imporvement
Emphasis	Profiling jobs or work systems by listing their important aspects	Practical nature of actions listed including simple solutions

Table 2. Examples of areas covered by analysis checklists (in parentheses are the number of items listed)

(a) Position Analysis Questionnaire (PAQ)	(b) Ergonomic Job Checklist (AET)	(c) Human Factors Checklist for air traffic control systems
Information input (where/how one gets information on the jobs to perform) (35)	The man-at-work system (work objects, tools and equipment, environment including physical, organizational, social and economic conditions) (143)	General (subsystems, necessary information) (10)
Mental process (information-processing and decision-making in doing the job) (14)		Visual displays (general, visual alerts) (66)
		Auditory alerts (general, speech messages) (29)
		Cognitive workload (general, automation) (58)
Work output (physical work done, tools and devices) (50)	The task analysis (different kinds of work object such as material and abstract objects, worker-related tasks) (31)	Data entry procedures (general, commands and command execution, menus, error messages and and user guidance (63)
Interpersonal relationships (36)		Data entry devices (general, keyboards, touchscreens, trackballs, grip devices, mice graphics, pushbuttons, pedals) (46)
Work situation and job context (Physical/ social contexts) (18)	The work demand analysis (elements of perception, decision and response/ activity) (42)	
Other job characteristics (work schedules, job demands) (35)	The supplement (body postures, movements in assembling)	Workstation design (user-centered design, control-room seating, communications equipment, environment (68)
		Human factors planning (30)

many other checklists covering only selected specific areas important for the intended improvements. As a result, general-purpose job analysis checklists such as the PAQ and AET have some common areas. Specific-purpose checklists such as traffic control systems checklists or products checklists may be designed to cover selected areas in detail. Examples of areas covered by analysis checklists are shown in table 2.

The PAQ has six major divisions consisting of information input, information processing, work output (physical work, tools and devices), interpersonal relationships, work situation and job content, and other job characteristics (work schedules and job demands). The AET consists of three parts dealing with the man-at-work system (work objects, tools, equipment and work environment), the task analysis, and the work demand analysis. Such a job-oriented approach is useful for undertaking task inventories and job analysis by considering important aspects of the given complex job tasks.

As shown in table 3, action checklists contrast in this respect with these analysis checklists. The checklist of the IEA/ILO Ergonomic Checkpoints mentions low-cost actions for various aspects of work activities, workplace facilities and equipment, working environment and work organization. The Work Improvement in Small Enterprises (WISE) checklist, developed for use by small enterprise owners/managers (Thurman *et al.* 1988) is used in checking small workplaces in order to identify simple improvements. In both these checklists, a special emphasis is placed on building on local practice and on low-cost solutions that are applicable immediately by using locally available skills.

Both checklists are accompanied with guidance about how to make simple improvements in low-cost ways.

New types of design support checklists devised for various ergonomic design purposes have diversified structures. Many such checklists are used for computer-aided design from ergonomics points of view. For example, a design support checklist for computer-aided designs (Aaltonen and Mattila 1998) consists of questions about layout design, mechanical design (location of equipment and controls, structural strength), control design, electrical design, maintenance design and installation design. Usually, these checklists belong to analysis checklists and are based on essential standards for ergonomic design. With the aid of these checklists, designers can document the standards used for the object in question.

3. THE PRACTICAL USE OF ACTION CHECKLISTS

The practical nature of checklist use is important. The users of the checklists expect to be able to have an overview of the ergonomics-related aspects of the work systems or objects considered for the purpose of identifying required improvements. Therefore, checklists should not be too long or too extensive. Analysis checklists tend to be long, and their users should devise appropriate ways of using them for their own purposes, for example, concentrating on certain aspects or trying to have a quick overview. Action checklists are easier to handle since they are relatively short. Action checklists can be used to initiate workable ideas by learning from the presented examples of actions. When action checklists are used in the course of a walk-

Table 3. Technical areas covered by typical action checklists

(a) IEA/ILO ergonomic checkpoints	(b) Work improvement in small enterprises checklist
Materials storage and handling (21)	Materials storage and handling (5)
Hand tools (15)	Workstation design (7)
Machine safety (20)	Machine safety (5)
Workstation design (15)	Control of hazardous agents (6)
Lighting (10)	Lighting (6)
Premises (heat, ventilation) (7)	Welfare facilities (5)
Control of hazardous substances and agents (6)	Premises (heat, noise, fire, passageways, wiring((7)
Welfare facilities (13)	Work organization (5)

through round, going through selected work areas or workstations, the time required for completing a checklist should be 30–60 minutes. Longer time can be used for analysis checklists, but we should keep in mind that lengthy check procedures may not be widely applied.

The format of checklists varies greatly. The reply entries are usually for yes/no, satisfactory/unsatisfactory or step-wise evaluations (for example, 1–5). There are both quantitative and qualitative criteria in using analysis checklists, and the users need to be guided about how to apply these different criteria. Items with quantitative criteria are clearer about evaluations to make, but not always relevant to the situation under consideration. As a rule, the checklist users should be allowed to make quick decisions based on the gained knowledge and their own experiences. The way each check item is presented is important in this respect. Each item should be clear about what aspects are to be checked. It is better to include clues for making quick judgment. For example, a question such as "Are displays and controls compatible with easy understanding and reactions?" is rather difficult to respond to quickly, while a question such as "Are controls color-coded for identification?" is easy to answer. Action checklists are thus easier to respond to as they mention practicable actions such as "use markings or colors on displays to help workers understand what to do."

The best way to use ergonomics checklists is to use them in group work. It is even better to plan the use of a checklist jointly and then apply it in the form of a joint inspection or a joint walk-through round. Each participant may fill in the checklist based on individual judgment. The results of checklist application can be discussed by small groups. Group presentations may follow if some groups join in the checklist application. We should try to summarize the results of such group discussions and agree on priority proposals for improving the work systems or objects checked.

Practical support for applying checklists is also important. The checklist users should be familiar with the work systems or objects to be checked. Knowledge and experiences in ergonomics principles are further necessary, and training in the use of checklists is needed. Such training can be combined with initial checklist exercises so that the users can learn how to use the checklist by doing themselves. Guidance materials and manuals on ergonomic principles and improvement actions are useful. Such support is effectively given when the checklist application is done by group work.

4. USEFUL CHECKPOINTS

Often, new checklists are designed for the specific purposes of the checklist users. This is encouraged as in this way the users can learn how to examine the various ergonomic aspects and how to utilize the check results. In designing new checklists, a compiled pool of "checkpoints" can be used as a database from which the checklist users can select those items which are applicable in their specific circumstances. The compiled points can also be used to get hints about new items to be added in the checklist by getting hints from these points. Table 4 gives examples of checkpoints useful for designing new checklists.

5. CHECKLISTS AS DESIGN AND TRAINING TOOLS

Ergonomics checklists are in the process of development. It is useful to exchange experiences in constructing and applying checklists. The design of new checklists adjusted to the checklist users' purposes is encouraged. This is important since the way work systems or facilities are used is increasingly varied on account of developing information technologies, flexible work systems, aging, and, increasingly, diversifying human needs.

One should note the shortcomings of general-purpose checklists in applying them to various situations. While knowledge about ergonomic standards is developing, there is a general lack of ergonomics norms and protocols of evaluation which are readily applicable in analyzing the varying aspects of work and objects. The evaluation of ergonomics aspects by

Table 4. Examples of useful checkpoints (from the IEA/ILO "Ergonomic Checkpoints")

Materials handling	Clear and mark transport routes.
	Use carts, hand-trucks and other wheeled devices or rollers when when moving materials.
	Use mechanical devices for lifting, lowering and moving heavy materials.
Hand tools	Provide a "home" for each tool.
	Use vices and clamps to hold materials or work items.
Machine safety	Make different controls easy to distinguish from each other.
	Use marking or colours on displays to help workers understand what to do.
Workstation design	Adjust the working height for each worker at elbow level or slightly below it.
	Allow workers to alternate standing and sitting at work as much as possible.
Lighting	Provide local lights for precision or inspection work.
	Relocate light sources or provide shields to eliminate direct glare.
Premises	Isolate or insulate sources of heat or cold.
	Isolate or cover noisy machines or parts of machines.
Welfare facilities	Provide drinking facilities, eating areas and rest rooms to ensure good performance and well-being.
	Provide a place for workers' meetings and training.
Work organization	Combine tasks to make the work more interesting and varied.
	Set up a small stock of unfinished products (buffer stock) between different workstations.

applying a checklist is based often on subjective assessment and subject to the lack of precision. We should keep in mind the limited scope of checklist results. It is important to understand the role of ergonomics checklists and use them as one of a range of practical evaluation tools for conducting social dialogue between employers, workers, users, and others concerned.

Two important uses of ergonomics checklists are in human-centered design and in training. In assisting designers in evaluating work situations and objects, ergonomics checklists can play a vital role. There are thus interesting proposals of design support checklists. Computer-aided application of such checklists is also developing. Design-support checklists usually contain ergonomics principles or standards useful for designers. The checklist users can use relevant check items and get useful advice about improving the specific aspects under consideration. The designers should use checklists as advisory tools in a flexible manner.

Another important use of ergonomics checklists is their application in education and training in ergonomics. Checklists help trainees to study complex interactions between the working situations and human characteristics in a more or less systematic manner. Analysis checklists are useful for training in standardized ergonomic analysis. Action checklists are useful for training in ergonomics interventions in local situations. Low-cost ideas contained in action checklists can motivate trainees to look for local solutions with practical benefits.

In using ergonomics checklists as design and training tools, the combined use of other information materials and databases is effective. Handy manuals, design standards, improvement guides, best practice stories, and illustrations or photographs of good examples are particularly effective. The use of checklists can be constructive when the checklist users pay due attention to data collection, learning through good examples and action-oriented interventions.

REFERENCES

AALTONEN, M. and MATTILA, M., 1998, Computer-aided design for ergonomics and safety. In W. Karwowski and G. Salvendy (eds), *Ergonomics in Manufacturing* (Norcross: Engineering and Management Press).

AHONEN, M., LAUNIS, M. and KUORINKA, T., 1989, *Ergonomic Workplace Analysis* (Helsinki: Finnish Institute of Occupational Health).

CARDOSI, K.M. and MURPHY, E.D., 1995, *Human Factors Checklist for the Design and Evaluation of Air Traffic Control Systems* (Washington DC: Federal Aviation Administration).

Di MARTINO, V. and CORLETT, N. (eds), 1998, *Work Organization and Ergonomics* (Geneva: International Labour Office).

EASTMAN KODAK COMPANY, 1983, *Ergonomic Design for People at Work: Volume 1* (Rochester: Eastman Kodak).

INTERNATIONAL LABOUR ORGANIZATION, 1996, *Ergonomic Checkpoints: Practical and Easy-to-implement Solutions for Improving Safety, Health and Working Conditions* (Geneva: International Labour Office).

KARWOWSKI, W. and MARRAS, W.S., 1998, *The Occupational Ergonomics Handbook* (Boca Raton: CRC Press).

KOGI, K., 1998, Collaborative field research and training in occupational health and ergonomics. *International Journal of Occupational and Environmental Health*, 4, 189–95.

KOGI, K., PHOON, W.O. and THURMAN, J.E., 1988, *Low-Cost Ways of Improving working Conditions: 100 Examples from Asia* (Geneva: International Labour Office).

McCORMICK, E.J., 1979, *Job Analysis: Methods and Applications* (New York: American Management Association).

NAG, P.K., 1998, Checklists. In *International Labour Office Encyclopaedia of Occupational Health and Safety* (Geneva: International Labour Office), pp. 29.14–29.24.

NORO, K. and IMADA, A. (eds), 1991, *Participatory Ergonomics* (London: Taylor & Francis).

ROHMERT, W. and LANDAU, K., 1985, *A New Technique of Job Analysis* (London: Taylor & Francis).

THURMAN, J.E., LOUZINE, A.E., and KOGI, K., 1988, *Higher Productivity and a Better Place to Work. Practical Ideas for Owners and Managers of Small and Medium-sized Industrial Enterprises: Action Manual* (Geneva: International Labour Office).

Biomechanical Modeling of Human Strength

D.B. Chaffin

Department of Industrial & Operations Engineering and Biomedical Engineering, University of Michigan, Ann Arbor, MI 48109-2117, USA

1. INTRODUCTION

Human strength is characterized by many experts as the maximum force or torque a well-motivated person can demonstrate for a few seconds without fatigue (i.e., with a minimum rest of at least one minute between repeated exertions). Many factors modify the resulting strength values. Some of the most important factors are as follows:

- Personal attributes
  - strength conditioning and genetics
  - gender
  - size (anthropometry)
  - age
- Task attributes
  - force direction (lift, push, etc.)
  - posture
  - speed of motion

Because these factors interact in a complex fashion, and the effect of each is only empirically known for some factors, biomechanical models have been proposed and used to assist in predicting the strength capability of a specified population of people (older men, small women, etc.) in a variety of ergonomic design scenarios.

2. WHY MODEL HUMAN STRENGTHS?

Besides the obvious curiosity people seem to have about strength (strength and aerobic training programs are big business), the major interest seems to come from a growing concern by employers for the following reasons:

- Manual exertions remain an essential part of many jobs today. It is believed these types of exertion account for a disproportionate number of serious injuries, mostly to the musculoskeletal system. Based on a Department of Labor Survey of work-related injuries and illnesses, overexertion accounted for 27% of injuries and illnesses in 1994. The direct medical and wage replacement costs associated with these injuries and illnesses are estimated at $15-20 billion dollars annually, with administrative and other indirect costs possibly quadrupling the total costs.
- When a job includes manual exertions, it is a job that often becomes difficult to staff, due to the large variability that exists in the strength performance capability of the working population. In fact a normal, healthy, mixed gender population of workers may have strength variations greater than 10:1 (Chaffin *et al.* 1999). If a manual task in a job is found to be particularly difficult for older workers, women or physically impaired individuals, legal remedies may mandate changes in the job to eliminate the discriminatory task, particularly if it is not an essential function in the job.

3. BASIC BIOMECHANICS OF STRENGTH PREDICTION

The musculoskeletal system can be viewed as a series of levers (bones) and mechanical actuators (muscles). The muscles produce a tendency for the bones at each joint to rotate. The magnitude of this tendency to rotate is called the rotational moment at a joint. During any physical exertion the muscles that surround each joint contract to produce moments in response to any external forces or torques required as one attempts to perform a manual task. Essentially at each joint a simple balance must be achieved between the moments produced by the muscle contractions and the moments caused by body segment weights and other external forces (e.g., hand loads). Mathematically this balance is expressed

$$M_j = S_j$$

where

M_j = the external moment at each joint

S_j = the maximum moment that can be produced by the muscles at each joint, (i.e., the muscle strength at a joint) normally specified for a particular population subgroup

The values for M_j normally depend on three general requirements:

- The load (force) acting on one's hands (weight lifted, push force, etc.).
- The posture one is in when performing the maximum effort.
- The size (anthropometry) of a person being considered for the task.

Given data describing each requirement, the moments at each joint can be easily computed using Newtonian mechanics, wherein external forces acting on the human body are multiplied by the perpendicular distance from the point of action to the joint. Population joint range-of-motion statistics are often consulted to assure that a given posture does not exceed the normal mobility at each joint. Figure 1 is a typical biomechanical model of a person used for computing major joint moments during exertions.

The joint moment strengths S_j are obtained by population measurements using standardized strength testing methods. Population values have been developed for most major muscle functions; remember that muscle strength moments vary over the range of motion for a joint. Thus joint angles must be known to predict S_j values. In fact, because muscles often span two joints, the angle at adjacent joints must be considered in predicting S_j values. A synthesis of values for S_j from the moment strength studies is presented in Chaffin *et al.* (1999).

Many calculations are needed to estimate the M_j and S_j values, so computer programs were developed for this purpose. The sagittal plane strength prediction program was first programmed in 1968 by Chaffin and was used to evaluate lunar exploration manual task simulations for NASA in the early 1970s. Nowadays it is known as the 2D static strength prediction program (2DSSPP). A three-dimensional version, known as the 3D static strength prediction program (3DSSPP), was written in 1975. It compares the population strength capabilities with 24 different muscle joint strengths, and allows asymmetric exertions involving one or both hands to be evaluated. The logic used in these programs is described in Figure 2.
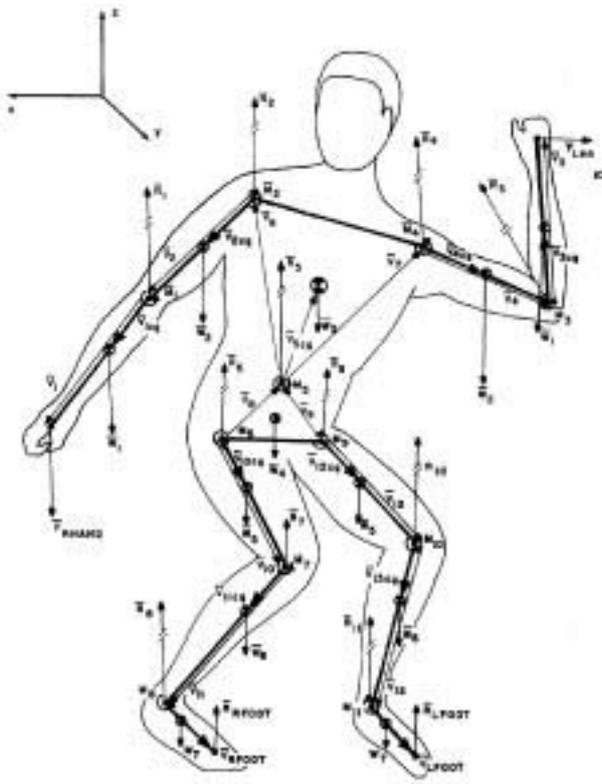
Figure 1. Three-dimensional distance, force and moment vectors used in the 12-link biomechanical model for strength prediction (Chaffin and Erig 1991)

In addition to using the muscle strength moments at each joint to predict the percentage of a population able to perform a specific exertion, the external moment loads at the lower back and feet are used to determine hand exertion forces that would either raise the risk of low back injury or of falling, respectively.

Thus strength models have become quite robust in their ability to evaluate high exertion tasks of many types.

4. STRENGTH MODEL VALIDATION AND LIMITATIONS

Three different validations of static strength prediction models have been performed over the years. In the first validation Garg and Chaffin (1975) had 71 male air force personnel perform 38 different maximum arm exertions (lifts, pushes, pulls, etc.) in a variety of arm/torso postures while seated. They found the predicted strengths were highly correlated with the group strengths when performing the 38 upper body tasks ($r = 0.93$ to 0.97). Chaffin *et al.* (1987) simulated 15 different whole-body exertions in the sagittal plane which were also performed by both men and women volunteers from a variety of industries. In some of these tests over 1000 people performed the exertions, though on average about 200 people performed each. Comparison of the 2DSSPP with the group strength data revealed a very high correlation ($r = 0.92$). This same study also included 3DSSPP simulations of 72 different one-arm exertions performed by five male army personnel. The correlations ranged from $r = 0.71$ to 0.83. But in this latter comparison, exact postural and bracing conditions were not available to use in the simulations. This and the small sample ($N = 5$) may have contributed to the lower correlations.

The last validation involved simulations of 56 one- and two-handed, whole-body exertions in 14 different symmetric, bent, and twisted-torso postures (Chaffin and Erig 1991). The simulation results were compared with the group strengths of 29 young males. Photographs from several views were available to assist in replicating the computer-rendered postures used by these subjects. The results indicated that if care is taken to assure the postures used in the model simulation are the same as the postures chosen by people performing the exertions, the prediction error
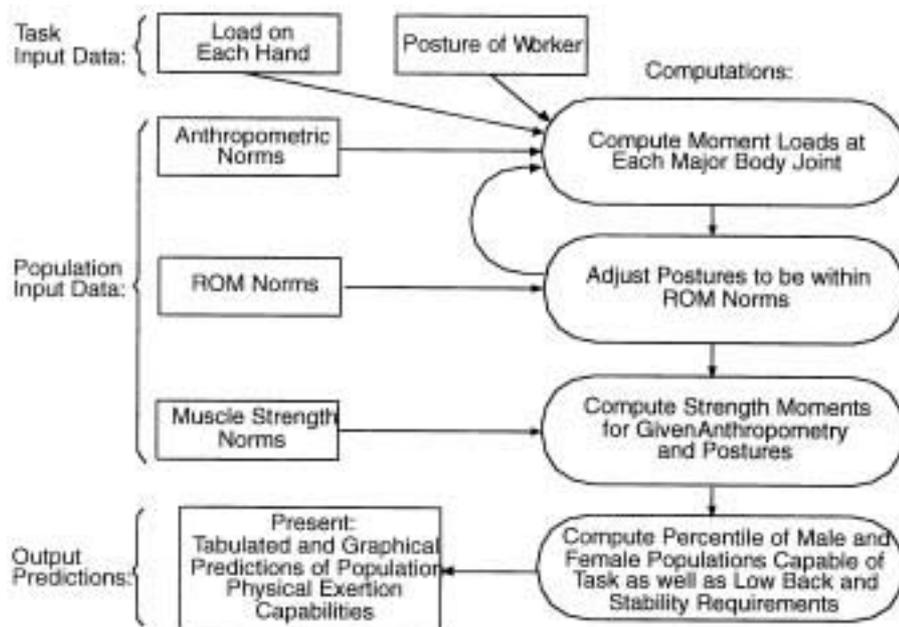
Figure 2. Biomechanical logic used to predict whole-body static exertion capabilities for given postures, hand force directions and anthropometric groups

standard deviation will be less than 6% of the mean predicted value.

At present the strength norms used as joint moment limits in the models are based on male and female populations who are relatively young (18–49 years). To improve the models further, I am now gathering strength values for older populations. In this regard, one comparison involving 98 men and women with a mean age of 73 years, showed a major decrease in strength performance in certain muscle functions. When these decreases were included in the 3DSSPP population database, it was found that some exertions which could easily be performed by younger people were predicted as impossible for most older people (Chaffin *et al.* 1994).

Perhaps the biggest limitation in strength prediction is that no dynamic models have been formulated or validated. Since many high exertion, manual tasks require a specific speed of motion (e.g., keeping up with a fast-paced production line) and since strength capability can vary significantly with movements, it is imperative to develop dynamic biomechanical models and dynamic strength data that improve future ergonomic evaluation methods.

5. USABILITY OF CURRENT STRENGTH PREDICTION TECHNOLOGY

To make them useful, the existing biomechanical models had to be easy to implement and easy to use on personal computers. Over the last 25 years several research engineers and computer programmers at the University of Michigan's Center for Ergonomics have worked to accomplish this goal, which resulted in the first 2D static strength prediction program licensed in 1984 by the university's Technology Management Office. This was followed in 1989 by the licensing of the first 3D static strength prediction program.

Figure 3 shows the main screen of the SSPP. The input values (i.e., body link angles, hand forces, and anthropometry) are shown in the upper left quadrant. A diagram depicting the body posture, the hand location, and the hand force direction, which are used as inputs, is depicted in the upper right quadrant. The predicted percentage of the male and female population having sufficient strength to perform the designated exertion (in this case lifting a 44 lb stock reel) is shown in tabular and graphical form in the lower left quadrant. The back compression force predictions for

men and women performing the 44 lb lift is shown in the lower right quadrant. In this situation, from inspection of the percent capable predictions (lower left), it is obvious that hip strengths are the most limiting muscle group strengths (only 66% of women and only 87% of men have sufficient hip strength to lift the 44 lb reel). Incidentally, these values are below that recommended by NIOSH, which believes jobs should accommodate 99% of men's strength and 75% of women's strength (or 90% of a mixed gender population). The lower right quadrant shows that the L5/S1 compression forces on the lumbar spinal disks of 924 and 845 lb are above the 770 lb recommended by NIOSH.

6. DISCUSSION

The existing biomechanical models are neither complete nor accurate for all types of exertion. If there are fast movements combined with maximum strength exertions, the static models will overpredict the population's capabilities. Where it is difficult to grip an object being manipulated, the models will overpredict the population's capabilities. For an older population, the models will overpredict its capabilities. These are issues that are being addressed by other biomechanics researchers. As improved biomechanical models and population databases are developed, they will be incorporated in subsequent versions of the program.

The important thing is that a great deal of information on population variability for strength and low back failure is now accessible in a form which can improve future job designs. I hope this chapter will help interested parties to understand the rapidly developing technologies needed to further improve manual tasks in industry.

REFERENCES

CHAFFIN, D.B. and ERIG, M., 1991, Three-dimensional biomechanical static strength prediction model sensitivity to postural and anthropometric inaccuracies. *IIE Transactions*, 23(3), 215-27.

CHAFFIN, D.B., FREIVALDS, A., and EVANS, S.M., 1987, On the validity of an isometric biomechanical model of worker strengths. *IIE Transactions*, 19(3), 280-88.

CHAFFIN, D.B., WOOLLEY, C.B., BUHR, T., and VERBRUGGE, L., 1994, Age effects in biomechanical modeling of static lifting strengths. *Proceedings of the Human Factors Society*, Nashville TN.

CHAFFIN, D.B., ANDERSSON, G.B.J., and MARTIN, B.J. (eds), 1999, *Occupational Biomechanics* (3rd ed.) (New York: John Wiley).

GARG, A.D. and CHAFFIN, D.B., 1975, A biomechanical computerized simulation of human strength. *AIIE Transactions*, 7(1), 1-15.
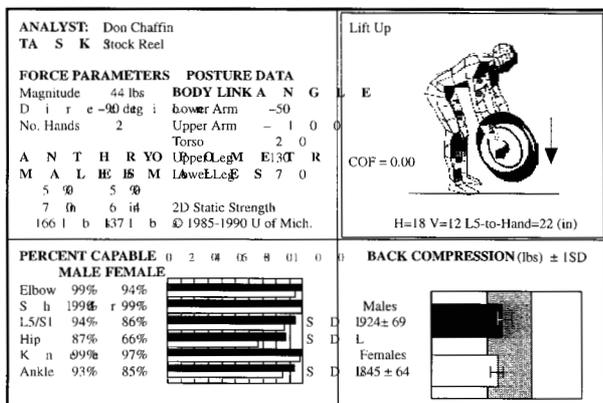
Figure 3. Static strength prediction program (Courtesy TMO Software, University of Michigan)

Biomechanics of Low Back: Guidelines for Manual Work

S. M. McGill

Occupational Biomechanics and Safety Laboratories, Faculty of Applied Health Sciences, Department of Kinesiology , University of Waterloo, Waterloo, Canada N2L 3G1

1. INTRODUCTION

Some of the well intended guidelines provided to workers to reduce the risk of low back injury often recommend to: "bend the knees and keep the back straight"; never jerk a load - lift slowly and smoothly"; "adjust your chair to keep the back upright and the hips, knees, and elbows at 90 degrees". These continue to be proliferated despite the fact that very few manual jobs can be performed by "bending the knees", and in fact some of these well intended guidelines can actually increase the risk of injury! Workers know this, and often return to their former habits. There is a need to examine the evidence to justify recommendations that reduce loading on the tissues at risk and develop recommendations that are applicable to a wide variety of workplace situations.

Having just indicated that "reducing the load on tissues at risk" is a laudable objective, a caveat is required here. While it often is common "wisdom" to direct effort to always reduce the demands of a job, some jobs may need to be made more demanding. Optimal occupational health requires some loading, but not too much. Too much of any one activity will lead to troubles, even sitting too much, standing too much, lifting too much, etc. However, some lifting, sitting, and standing is necessary to maintain health from the tissue and motor control level through to enhancing health at the systems level.

This chapter aims to integrate scientific evidence to build a foundation for the formation of a set of guidelines intended to reduce the risk of low back injury in a wide variety of occupational situations, ranging from heavy lifting to sedentary activities. Since some of these guidelines are based on recent research findings, they remain tentative until their efficacy under field trial is proven or disproven.

2. REDUCING THE RISK OF LOW BACK INJURY: ISSUES AND EXPERIMENTAL EVIDENCE

It is not the intent here to describe the detailed process of investigation of spine function and injury mechanisms in living workers. The interested reader is directed to McGill, 1997 for more information on this process. Rather, this section constitutes a discussion of some recent research findings as they relate to the issue of formulating guidelines for manual exertion for task assessment and implementation by ergonomists.

2.1. Should One Avoid End Range of Spine Motion During Exertion?

It is recognized that very few lifting tasks in industry can be accomplished by "bending the knees and not the back". Furthermore, most workers rarely adhere to this technique when repetitive lifts are required — a fact which is quite probably due to the increased physiologic cost of squatting compared with stooping. However, a case can be formulated for the preservation of neutral lumbar spine curvature (i.e., the natural relaxed curve observed when standing) while lifting, (specifically avoiding end range limits of spine motion about any of the 3 axes). This is a different concept than "trunk angle", as the posture of the lumbar spine can be maintained independent of thigh and trunk angles. Specifically, there has been much confusion in the literature between trunk angle or inclination, and the amount flexion in the lumbar spine. Bending over is accomplished by either hip flexion or spine flexion or both. It is the issue of specific lumbar spine flexion that is of importance here. Normal lordosis can be considered to be the curvature of the lumbar spine associated with the upright standing posture.

Using the tissue load distribution perspective, the following example demonstrates the shifts in tissue loading, predicted from our modelling approach, which has quite dramatic affects on shear loading of the intervertebral column and lends insight into the stoop-squat issue. First, the dominant direction of the pars lumborum fibres of longissimus thoracis and iliocostalis lumborum are noted to act obliquely to the compressive axis of the lumbar spine producing a posterior shear force on the superior vertebra. In contrast, the interspinous ligament complex acts with the opposite obliquity to impose an anterior shear force on the superior vertebra. Suppose a worker holds a load in the hands with the spine fully flexed sufficient to achieve myoelectric silence in the extensor muscles (reducing their tension), and with all joints held still so that the low back moment remains the same, then the recruited ligaments will add to the anterior shear to levels well over 1000 N, which is of great concern from an injury risk viewpoint. However, when a more neutral lordotic posture is adopted, the extensor musculature is responsible for creating the extensor moment, and, at the same time, will support the anterior shearing action of gravity on the upper body and hand-held load. Disabling the ligaments by avoiding full flexion greatly reduced shear loading. (For a more comprehensive discussion see McGill, 1997.)

The scientific evidence appears to point to the following positive effects if one lifts while avoiding full lumbar flexion: conscious control of lumbar musculature is retained, reducing shear load on the facet joints and providing neurological protection not present if one depends on passive tissue to support the load; the disc is loaded more uniformly permitting all of the annular fibers to share the stress rather than disabling some.

2.2. Should One Lift or Perform Extreme Torso Bending Shortly After Rising from Bed?

The diurnal variation in spine length together with the ability to flex forward has been well documented. Over the course of a day, hydrostatic pressures cause a net fluid out flow from the disc resulting in narrowing of the space between the vertebrae which in turn reduces tension in the ligaments. When laying down at night, osmotic pressures exceed the hydrostatic pressure causing the disc to expand. It has been estimated that disc bending stresses increased by 300% and ligament stresses by 80% and concluded there is an increased risk of injury to these tissues when bending forward early in the morning.

2.3. Should One Lift Immediately Following Prolonged Flexion Postures?

Prolonged flexion causes spine tissues to creep such that stresses accumulate with progressive tissue deformation over time which will lead to damage (see McGill, 1997). This is of particular importance for those individuals whose work or movement patterns are characterized by cyclic bouts of full end range of motion postures followed by exertion.

These data suggest that the spine has a memory, since the mechanics of the joints are modulated by previous loading history. Before lifting, following a stooped posture, or after prolonged sitting, a case could be made for standing or even consciously extending the spine for a short period. Allowing the disc nucleus to "equilibrate" and ligamentous tissues to slowly regain their rest length and stiffness reduces the risk of damage.

2.4. Should Intra-abdominal Pressure be Increased While Lifting?

It has been claimed for many years that intra-abdominal pressure (IAP) plays an important role in support of the lumbar spine, especially during strenuous lifting. This issue has been considered in lifting mechanics for years and, for some, has formed a cornerstone for prescription of abdominal belts to industrial workers and also has motivated various abdominal strengthening programs. Many have advocated the use of intra-abdominal pressure as a mechanism to reduce lumbar spine compression.

In fact increased IAP is produced with abdominal wall musculature activity which has actually been shown to increase spine compression. However, it appears that the spine prefers to sustain increased compression loads if intrinsic stability is increased. An unstabilized spine buckles under extremely low compressive load (e.g. approximately 20N). The geometry of the musculature suggests that individual components exert lateral and anterior–posterior forces on the spine which perhaps can be thought of as guy wires on a mast to prevent bending and compressive buckling. As well, activated abdominals create a rigid cylinder of the trunk resulting in a stiffer structure. Thus it appears that increased IAP, commonly observed during many activities including lifting, as well as in those people experiencing back pain, does not have a direct role to reduce spinal compression but rather is an agent used to stiffen the trunk and prevent tissue strain or failure from buckling.

2.5. Should Abdominal Belts be Prescribed to Manual Materials Handlers?

Readers are directed to the chapter on this topic in this textbook by McGill.

2.6. Should Workers Adopt a Lifting Strategy to Recruit the Lumbodorsal Fascia?

Studies have attributed various mechanical roles to the lumbodorsal fascia (LDF). In fact there have been some attempts to recommend lifting postures based on LDF hypotheses. Suggestions were originally made that lateral forces generated by internal oblique and transverse abdominis are transmitted to the LDF via their attachments to the lateral border. This lateral tension was hypothesized to increase longitudinal tension, from Poisson's effect, pulling in the direction of the posterior midline of the lumbar spine causing the posterior spinous processes to move

together resulting in lumbar extension. This proposed sequence of events formed an attractive proposition because the LDF has the largest moment arm of all extensor tissues. As a result, any extensor forces within the LDF would impose the smallest compressive penalty to vertebral components of the spine. However this hypothesis was examined by three studies which used quite different methods, all published about the same time, and which, collectively, questioned its viability. Given the confused state of knowledge about the role, if any, of the LDF, the promotion of lifting strategies based on intentional LDF involvement cannot be justified at this time.

2.7. Should the Trunk Musculature be Co-contracted to Stabilize the Spine?

The ability of the joints of the lumbar spine to bend in any direction is accomplished with large amounts of muscle co-activation. Such co-activation patterns are counter productive to generating the torque, necessary to support the applied load, in a way that minimizes the load penalty imposed on the spine from muscle contraction. Yet it has been recently documented that cocontracting torso musculature acts to stabilize the spine and prevent buckling (Cholewicki and McGill, 1996).

2.8. How Do People Hurt their Backs Picking Up a Pencil?

While injury from large exertions is understandable, explanation of how people injure their backs performing rather benign appearing tasks is more difficult — but the following is worth considering by the ergonomist. Continuing the considerations about stabilization from the previous section — a number of years ago, we were investigating the mechanics of power lifter's spines while they lifted extremely heavy loads using video fluoroscopy for a sagittal view of the lumbar spine. During the execution of a lift, one lifter reported discomfort and pain. Upon examination of his vertebral motion, one of the lumbar joints (specifically, the L4/L5 joint) reached the full flexion calibrated angle, while all other joints maintained their static position (2–3 degrees from full flexion). This is the first observation that we know of reported in the scientific literature documenting proportionately more rotation occurring at a single lumbar joint than at the other joints: the spine buckled! It would appear that this unique occurrence was due to an inappropriate sequencing of muscle forces (or a temporary loss of the normal motor control pattern). This motivated the work to investigate and continuously quantify stability of the lumbar spine throughout a reasonably wide variety of loading tasks (Cholewicki and McGill, 1996). Generally speaking, Cholewicki noted that the risk of such an event was greatest when there are high forces in the large muscles with simultaneous low forces in the small intersegmental muscles (a possibility with our power lifter) or when all muscle forces are low such as during a low level exertion. Thus, a mechanism is proposed, based on motor control error resulting in temporary inappropriate neural activation, that explains how injury might occur during extremely low load situations, for example, picking a pencil up from the floor following a long day at work performing a very demanding job.

2.9. Are Twisting Lifts Particularly Dangerous?

Twisting of the trunk has been identified as a factor in the incidence of occupational low back pain but the mechanisms of risk require some explanation. Some hypotheses have been based on an inertia argument in that twisting at speed will impose dangerous axial torques upon braking the axial rotation of the trunk at the end range of motion. Twisting has indeed been documented to damage the disc annulus and possibly the posterior ligaments if the joint is fully flexed prior to twisting.

Certainly the mechanisms of injury from torsional loads applied under twisting conditions remain inconclusive. However, it is clear that the increase in compressive load on the spine is dramatic if a comparatively small amount of axial twist torque is required in addition to the dominant extensor torque. These differences result from the difference in co-activation of the trunk musculature, combined with small moment arms in many cases, to generate the moments of force required. It appears that the lumbar spine pays dearly in order to support even small axial torques when extending during the lifting of a load.

2.10. Is "Lifting Smoothly" and Not Jerking the Load Always the Best Advice?

We have all heard that a load should be lifted smoothly and not "jerked". This recommendation was most likely rationalized on the basis that accelerating a load upwards increases its effective mass by virtue of an additional inertial force acting downwards together with the gravitational vector. However, this may not always be the case as it is possible to lift a load by transferring momentum from an already moving segment. The concept of momentum transfer during lifting which is a skill and feature of many skilled workers is known as the "kinetic lift". For example, if a load is awkwardly placed, perhaps placed on a work table at a distance of 75 cm from the worker, a slow-smooth lift would necessitate the generation of a large lumbar extensor torque for a lengthy duration of time — a situation that is most strenuous on the back. However, this load could be lifted with a very low lumbar extensor moment or quite possibly no moment at all. However, if the worker leaned forward and placed his hands on the load, with bent elbows, the elbow extensors and shoulder musculature could thrust upwards initiating upward motion of the trunk to create both linear and angular momentum in the upper body (note that the load has not yet moved). As the arms straighten, coupling takes place between the load and the large trunk mass (as the hands then start to apply upward force on the load) transferring some, or all, of the body momentum to the load causing it to be lifted with a jerk. This highly skilled "inertial" technique is observed quite frequently throughout industry and in some athletic events such as competitive weight lifting but it must be stressed that such lifts are conducted by highly practised and skilled individuals — the body momentum must be generated prior to the transfer to the load. In most cases, acceleration of loads to decrease low back stress in the manner described is not suitable for the "lay" individual when conducting the lifting chores of daily living. There can be no argument that reduction of the extensor moment required to support the hand load is paramount in reducing the risk of injury and that this is best accomplished by keeping the load as close to the body as possible.

2.11. Is Sedentary-seated Work Harmful?

Epidemiological evidence has documented the increased risk of disc herniation of those who perform sedentary jobs characterized by sitting. This has motivated occupational biomechanists to consider the duration of sitting as a risk factor when designing seated work in the interest of reducing the risk of injury. A recently proposed guideline has suggested a sitting limit of fifty minutes without a break, although this proposal will be tested and evaluated in the future. Further, the notion of making sitting a "dynamic task" with regular posture changes is gaining favour rather than the conventional textbook description of the ideal sitting posture being an upright torso with hips and knees to 90 degrees. Still, there is no substitute for regularly standing up from the chair — answering the phone for example provides such an opportunity.

2.12. What About Walking: Fast and Slow Occupational Gait?

Studies of walking mechanics have documented that slow walking ("mall walking") results in static muscle activity and spine loading while fast walking produces cyclical loading and unloading. This may explain the relief experienced by some from fast walking while slow walking exaggerates symptoms. Further, spine posture during standing and walking is quite different from the sitting posture supporting the recommendation to change from sitting to standing regularly.

3. TENTATIVE RISK REDUCTION GUIDELINES FOR OCCUPATIONAL INJURY

The following recommendations have been summarized from the biomechanical rationale developed in the previous section. Some are consistent with what has been advocated for years, while others contradict longstanding notions that were based on flawed, or unavailable, biomechanical understanding. They are more versatile and widely applicable than the commonly used instruction of "bend the knees — not the back" to reduce lifting stresses. These recommendations may have the potential to reduce tissue loads during the performance of a wide range of industrial exertion tasks and they are able to accommodate all tasks including those outside of the sagittal plane. Furthermore, the exact instructions issued to a specific worker should not be taken verbatim from the following list, but rather the biomechanical principle should be explained in a language and terminology which is familiar to the worker. In addition, often successful job incumbents have developed personal strategies for working that assist them in avoiding fatigue and injury. Their insights are the result of thousands of hours of performing the task and they can be very perceptive — attempts should be made to accommodate them.

3.1. Recommendations for Safer Work – A Tentative List

1. First and foremost, design work that facilitates variety.
   - Too much of any single activity leads to trouble. Relief of cumulative tissue strains are accomplished with posture changes, or better yet, other tasks that have different musculoskeletal demands.

2. Avoid a fully flexed or bent spine and rotate trunk using hips (preserving a neutral curve in the spine).
    - Disc herniation cannot occur.
    - Ligaments cannot be damaged as they are slack.
    - The anterior shearing effect from ligament involvement is minimized and the posterior supporting shear of the musculature is maximized.
    - Compressive testing of lumbar motion units has shown increases in tolerance with partial flexion but decreased ability to withstand compressive load at full flexion.
3. Choose a posture to minimize the reaction moment on low back so long as #2 is not compromised.
    - Neutral lordosis is still maintained but sometimes the load can be brought closer to the spine with bent knees (squat lift) or relatively straight knees (stoop lift). The key is to reduce the moment which has been shown to be a dominant risk factor.
4. Allow time for the disc nucleus to "equilibrate" and ligaments to regain stiffness after prolonged flexion (e.g. sitting or stooping) and do not immediately perform strenuous exertions.
    - After prolonged sitting or stooping spend time standing to allow the nuclear material within the disc to equilibrate and equalize the stress on the annulus, and allow the ligaments to regain their rest length and provide protective stiffness to the lumbar spine.
5. Avoid lifting or spine bending shortly after rising from bed.
    - Forward bending stresses on the disc and ligaments are higher in the early morning compared with later in the day (at least 1 hour after rising) causing discs to become injured at lower levels of load.
6. Prestress system even during "light" tasks.
    - Lightly co-contract the stabilizing musculature to remove the slack from the system and stiffen the spine, even during "light" tasks such as picking up a pencil (Cholewicki and McGill, 1996).
    - Mild co-contraction and the corresponding increase in stability increases the margin of safety of material failure of the column under axial load.
7. Avoid twisting and the simultaneous generation of high twisting torques.
    - Twisting reduces the intrinsic strength of the annulus by disabling some of its supporting fibers while increasing the stress in the remaining fibers under load.
    - Since there is no muscle designed to produce only axial torque the collective ability of the muscles to resist axial torque is limited and may not be able to protect the spine in certain postures.

- The additional compressive burden on the spine is substantial for even a low amount of axial torque production.
8. Exploit the acceleration profile of the load (rather than "always lift slowly and smoothly").
    - This is only for highly skilled individuals performing repetitive tasks.
    - Dangerous for heavy loads and should not be attempted.
    - It is possible that a transfer of momentum from the upper trunk to the load can start moving an awkwardly placed load without undue low back. Possibly, the viscoelastic property of biological material will safely absorb a momentary high load required to bring the load close to the trunk which reduces the reaction moment.
9. Avoid prolonged sitting.
    - Prolonged sitting is associated with disc herniation.
    - When required to sit for long periods, adjust posture often, stand up, at least every 50 minutes, and extend spine and/or walk for a few minutes.
    - Organize work to break up bouts of prolonged sitting into shorter periods that are better tolerated by the spine.
10. Consider the best rest break strategies.
    - Workers engaged in sedentary work would be best served by frequent, dynamic breaks to reduce tissue stress accumulation (for exercises documented to spare the spine, see McGill, 1998).
    - Workers engaged in dynamic work may be better served with longer and more "restful" breaks.

ACKNOWLEDGEMENTS

REFERENCES

CHOLEWICKI, J. and McGILL, S.M. (1996) Mechanical stability of the invivo lumbar spine: Implications for injury and chronic low back pain, *Clin.Biomech.* 11(1):1–15.

McGiLL, S.M. (1997) Biomechanics of low back injury: Implications on current practice and the clinic. *J. Biomech.* 30(5): 465–75.

McGILL, S.M. (1998) Low back exercises: Evidence for improving exercise regimens. *Physical Therapy* 78(7): 754–65.

Bivariate Anthropometric Design for Work Spaces and Products

J.G. Kreifeldt and K. Nah

Department of Mechanical Engineering, Tufts University, Medford MA, USA

1. INTRODUCTION

Fitting a work space or product to human geometry often requires that two or more of the human's anthropometric variables be accommodated simultaneously. For example, consider the simple bivariate problem in Figure 1 in which the overhead spacing (H) AND the knee clearance (K) are to be chosen so that together they may accommodate some desired percentage (e.g. 50%) of the design population. Approaching this or any other multivariate problem as a series of independent univariate ones can result in the accommodated percentage being severely incorrect because of the correlation between the variables.

This article will present a set of graphs and tables which simplify five types of bivariate design problems for specifying a pair of design values to accommodate some desired percentage of the population. Among the assumptions used in preparing the tables and graphs are that: (1) the relevant anthropometric dimensions of the user population are normally distributed, and (2) a given accommodation percentage can be achieved using equal or complementary percentile values of the anthropometric values in a manner to be explained.
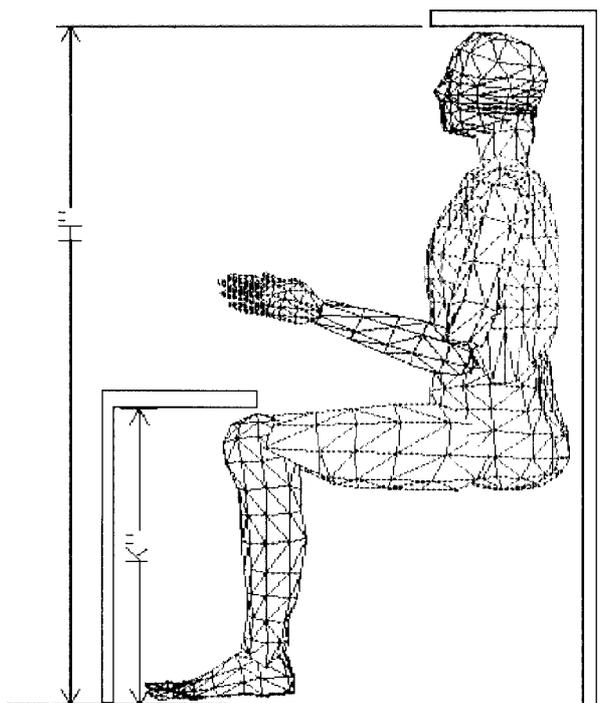
Figure 1. A simple work space design problem.

2. BACKGROUND

Ordinarily, a bivariate (or generally a multi-variable) design problem must be approached using the appropriate bivariate (multivariate) distribution. In the above example, the bivariate distribution for sitting knee height and sitting stature would be needed. Such a bivariate distribution for normally distributed variables would be expressed as the following equation in which if 'X' stands for the sitting knee height then Y stands for the sitting head height, ρ_{XY} is the correlation betwen the two variables and 'm' and 'σ' stand for the means and standard deviations of the variates.

$$p(X, Y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} \exp\left[-\frac{1}{2(1-\rho^2)}\left[\left(\frac{X-m_X}{\sigma_X}\right)^2+\right.\right.$$

$$\left(\frac{Y-m_Y}{\sigma_Y}\right)^2-2\rho\left(\frac{X-m_X}{\sigma_X}\right)\left(\frac{Y-m_Y}{\sigma_Y}\right)$$

In the problem given, the knee space and overhead space dimensions must be selected so that 50% of the population has knee heights AND sitting heights simultaneously less than these two values. In bivariate (and multivariate) design, there are an infinite number of pairs of values which will satisfy this requirement. (Although this may seem to complicate the problem, this actually is a strength of the multivariate approach because it can result in considerable design flexibility.) The approach presented here reduces the infinite number of choices by using the reasonable assumption that the selected values expressed as percentiles will be either equal or complementary depending on the problem type. This will be clarified in the following. The general approach for selecting the appropriate values of the variates in the example may be visualized in Figure 2 which represents a scatter plot of 2000 pairs of these two variables in which each point represents a particular person's pair of values.

The correlation between the values is 0.4. The figure has the measurement values on two axes and the z transformed values and equivalent GAussian percentiles (as a fraction) on the opposing axes. The general approach to the example problem is to select a value ("cut-off") for each variate so that 50% of the pair values are simultaneously less than those two values. These would then be the values to use for the respective design values. However, *any* point (pair) on the 50% contour shown superimposed on the scatter plot is a possible candidate pair. That is, a set of orthogonal lines (defining the cut-off values) through any point on the contour will contain 50% of the scatter points in the lower left quadrant of the crossed lines. In order to make a universal and rational choice of a single one from the infinity of possible point pairs on the contour, one can select that point which produces equal percentile (or z) values for the two variates. This point actually turns out to be the point of symmetry where a diagonal passes through the contour as well as lying on the best fit linear interpolation line to the bivariate data. Therefore, for the sample problem, the defining sitting knee height and sitting height are 21.4" and 36.6" respectively, meaning that if the K = knee space = 21.4" and H = head space = 36.6" in Figure 1 are at least these values, then these dimensions will accommodate at least 50% of the population. To generalize this problem for all normal bivariates with a correlation of 0.4, the
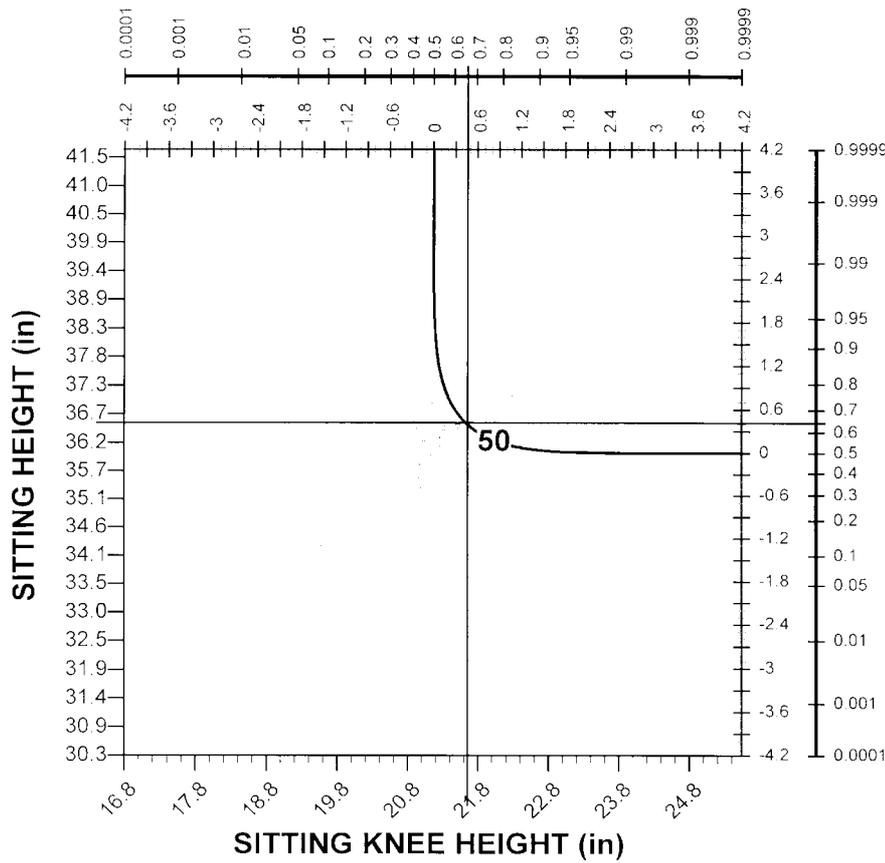
Figure 2. A bivariate scatter plot showing actual, Z and cumulative fractional values.

corresponding normalized values or the corresponding percentile values for this example would be: z = 0.43 or 66.6% for either variate. Obviously, any accommodation percentage could have been set such as 95% and the procedure would be the same.

3. METHODOLOGY

In order to generalize this to any pair of normally distributed variables with any bivariate correlation coefficient, the cumulative form of the normalized bivariate distribution can be back solved for the pair of equal z values (z_{eq}) which produce the desired percentage for the problem. The value of z_{eq} which solves the following equation produces the values in question in which P = the desired percentage and ρ = the bivariate correlation coefficient. The value z_{eq} can be found by a root solving algorithm. This procedure will be exemplified for the problem of Figure 1.

The normalized bivariate normal distribution is given in equation (2) where the triplet (z_x, z_y, r) refers to the two normalized normal variates and their bivariate correlation coefficient respectively.

$$p(z_x, z_y, \rho) = \frac{1}{2\pi \sqrt{1-\rho^2}} \exp[- \frac{(z_x^2 + z_y^2 - 2\rho \, z_x z_y)}{2(1-\rho^2)}]$$

Doubly integrating equation 2 from -∞ to values of (z_x, z_y) produces the fraction (or percentage if multiplied by 100) P of the distribution which has z values less than this pair of values. This is expressed

in equation (3) where the integrand is the normalized bivariate Gaussian distribution shown in equation (2).

$$\int_{y=-\infty}^{y=z_y} \int_{x=-\infty}^{x=z_x} p(x, y, r) \, dx \, dy = P$$

There are an infinite number of such (z_x, z_y) pairs defining the constant fraction (or percentage) values such as shown in Figure 1. To find the unique pair of z values which are equal and which produce the desired cumulative fraction, the z values are set equal to zeq as in equation (4) and 'back solved' by any root finding algorithm. This is done repeatedly for various values of 'P' and P in order to generate the graphs and tables for this paper.

$$\int_{y=-\infty}^{y=z_{eq}} \int_{x=-\infty}^{x=z_{eq}} p(x, y, r) \, dx \, dy = P$$

4. PROBLEM TYPES

Before providing these graphs and tables, we also consider the five typical generalized accommodation problems. These can be readily pictured as in Figure 3 showing five common accommodation problems as viewed on a bivariate plot of two variates normalized as their percentiles.
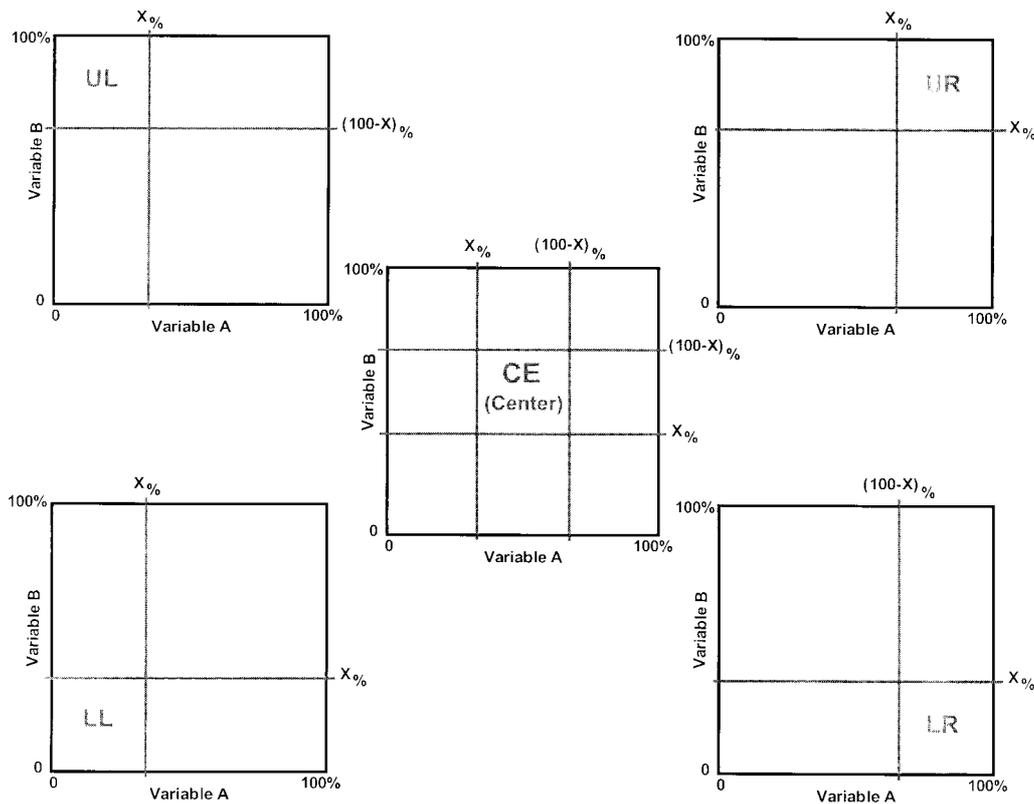
Figure 3. A bivariate chart showing five common accommodation problem types.

Given any desired percentage to be enclosed in one of the five areas, this article supplies appropriate values for the two variates expressed as 'z' values appropriate to equal or complementary percentile values. For example, for a specified percentage to be enclosed in the Lower Right (LL) section, the appropriate percentile values would be given as X and its complement as (100-x) both expressed as the corresponding Z and -Z values. The problem will define which variable to call A or B. Z values are given in the graphs and tables because of the need to convert ultimately to physical values, while percentile descriptions are used because of their clarity in explaining the theory.

The shape of Figure 3 can be considered a square because the axes are expressed in percentiles ranging from 0–100% and each of the five defined problem areas is also square because of the use of the equipoints to define the cut-off values. Thus, regardless of the percentage of the population to be included in one of the five given problem areas, the cut-off values will intersect on the ascending or descend diagonal.

This article uses the following nomenclature to distinguish the five problem types.

4.1. Type 1 – LL (Lower Left)

This refers to problems similar to Figure 1 in which both variates must be simultaneously less than some pair of values.

4.2. Type 2 – LR (Lower Right)

This refers to problems in which one variate must be greater than some value while the other is less than some other value. This problem is solved by a complementary pair of percentile values expressed as 'x' and 100-'x'). For example, consider specifying the sitting height and body weight of a manikin to be used to evaluate the adequacy of the height of a head restraint in

an automobile seat. Logic suggests that people with a *tall sitting stature* (head height above their seat) and a *light weight* are most at risk from a head restraint which is too low for them when sitting on a cushioned seat as this would cause them to sit higher than heavier people for that sitting height. Following a decision as to what percentage of people to allow at risk (e.g. 5%), it then follows that the sitting height and weight for the manikin must be specified. This problem can be solved by specifying the relevant percentile 'x' for the sitting height and its complementary percentile (100 - 'x') for weight which achieves the desired 'at risk' percentage given the bivariate correlation between sitting height and weight for the design population. The percentiles can then of course be converted into the appropriate sitting height and weight. If the height of the head restraint is designed to provide support at least for the manikin so designed, then only those who have a sitting height greater than the design sitting height and a weight less than the design weight are put at risk.

4.3. Type 3 – UR (Upper Right)

This refers to problems in which both variates must exceed some values. For example, it might be desired to find an arm length and a standing height which are both exceeded by only 5% of the population. Although Figure 3 shows equal complementary percentile values, it can be understood that these are the equal 'x' percentile values desired. Note that a UR problem can be turned into a LL one (and vice versa) by finding percentile cut-off values to achieve, for example, 5% in LL and then using the complementary percentile values.

4.4. Type 4 – CE (Center)

This is common in 'adjustment' problems. For example, assume that variate A = seat height of a chair as measured from the floor

and variate B = the arm height above a reference, and that a number of test subjects have been allowed to set these to their preferred comfortable values. (Further assure that a Gaussian bivariate distributioni adequately approximates the distribution of these pairs of values and their bivariate correlation is obtained.) If it is desired in a manufactured version to make seat height and arm height separately adjustable between lower and upper values so that at least 95% of users can find a comfortable setting, this would be a typical type 5 – CE problem. In this case, again the problem can be solved by specifying equal and complementary percentile values (lower and upper values) for each variate using the correlation found between the two variates.

4.5. Type 5 – UL (Upper Left)

This is a reflection of the LL problem which again is solved using complementary percentiles. Note that a UL problem can be turned into a LR one (and vice versa (simply by switching the roles of the variates. In the following tables and figures, only values for the LL, UR, LR, and CE types will be given because LR or UL are identical by interchaning the variates 'A' and 'B'. Instead of giving percentile values, the figures and tables will provide the z equivalents of the percentile values for ease of calculating the actual variate values.

5. CONVERTING FROM A Z VALUE TO A VARIATE VALUE

Once the required z values are found from the tables and figures provided, they can be readily converted to the actual variate values by any appropriate algorithmic or table format. Analytically, a z value can be transformed back to a variate value (v) through equation (5) in which mv and sv stand for the mean and standard deviation of the variate respectively:

$$y = (z)(s_v) + m_v$$

6. EQUATIONS FOR THE GRAPHS AND TABLES

The appropriate bivariate graphs and tables for the five types of problems are provided here in the 'z' form for reasons outlined previously. Negative correlation values are limited to those greater than -0.5 for brevity and because of the lack of reported negative correlations smaller than about -0.2 in human anthropometry. Note also that because of the symmetry for the Gaussian bivariate distribution, values corresponding to negative correlation values can be omitted for the Center (CE) type of problems because they are identical to those for positive correlations. Furthermore, the Upper Left table and graph have been omitted because they are identical to those of the Lower Right with a simple interchange of variables. The graphs and tables for the four problem areas were computed by the following procedures.

6.1. Lower Left (LL): Figure 4 and Table 1.

The equal percentile cut-offs (in normalized 'z' form) given a value of and desired accommodation fraction P are found by back solving equation (6) for z_{eq}.

$$\int_{y=-\infty}^{y=z_{eq}} \int_{X=-\infty}^{X=z_{eq}} p(x,y,r)\,dx\,dy = P$$

6.2. Lower Right (LR): Figure 5 and Table 2

The complementary cut-offs are found by back solving equation (7) for z_{eq} for a desired accommodation fraction P. Note that the cut-off values are complementary in that if one is -z (x%) then the other is z (100 - x%) as indicated by the opposite signs on the same absolute value of z_{eq}.

$$\int_{y=-\infty}^{y=-z_{eq}} \int_{X=z_{eq}}^{X=\infty} p(x,y,r)\,dx\,dy = P$$

6.3. Upper Right (UR): Figure 6 and Table 3

The equal cut-offs are found by back solving equation (8) for z_{eq} for a desired accommodation fraction P.

$$\int_{y=z_{eq}}^{y=\infty} \int_{X=z_{eq}}^{X=\infty} p(x,y,r)\,dx\,dy = P$$

6.4. Center (CE): Figure 7 and Table 4

The complementary percentile cut-offs are found by back solving equation (9) for z_{eq} for the desired accommodation fraction P. Note here again that this is equivalent to both variates having the identical *percentile* cut-offs although not necessarily the same actual values.

$$\int_{y=-z_{eq}}^{y=z_{eq}} \int_{X=-z_{eq}}^{X=z_{eq}} p(x,y,r)\,dx\,dy = P$$

7. EXAMPLES

Short examples of the uses of these tables will be given for the Lower Left, Lower Right and the Center type problems.

7.1. Lower Left (LL): Figure 4 and Table 1

Assume that it is desired to accommodate 95% of the population for the problem of Figure 1 using the means, standard deviations, and correlation coefficient given previously, then, from Figure 4 or Table 1, the appropriate z value for each variate is 1.93. Substituting the appropriate pair of mean and standard deviation values for each variate into equation (3) produces the required design values of: knee space = 23.14", and head space = 38.61".

7.2. Lower right (LR): Figure 5 and Table 2

Assume here that a product must have an opening through which air is forced by a fan blade rotating behind the opening. It is desired to make the opening (D) *small enough* and to recess the fan blade a distance (L) *far enough away* from the opening and so that, for example, 99% of children's fingers will be kept out of the opening and/or away from the blades. Thus, the values (D, L) are to be chosen so as to minimize the number of *long* and *thin* fingers which would be at risk from entering the hole and

touching the blade. This identifies the problem as a Type 2 – LR in which complementary cut-off values are appropriate. Another interpretation is that the design will put only 1% "at risk" implying that the "LR" corner of a bivariate finger-length–finger-thickness plot should contain no more than 1% of the data. It is assumed for this example that the correlation between finger thickness and finger length is 0.5. From Table 2 and/or Figure 5, the appropriate z values are:

D = minimum finger thickness = the lower z = 0.81
= 79th %tile value

L = maximum finger length- = the complementary z
= -0.81 = (100 - 79) = 21st %tile value

From which the values of D and L can be computed using equation (5) once the mean and standard deviation of the variates are specified. As mentioned previously, the *upper left* problem can be viewed as a *lower right* one with the variates switched and, therefore, the lower right example as given here suffices as an example for UL as well.

7.3. Upper right (UR): Figure 6 and Table 3

Assume in this case that it is desired to set values for height and weight such that only 5% of a population will exceed both of them. This is a Type 4 – UR problem for which Figure 6 and Table 3 are appropriate. The correlation between height and weight is approximately 0.2 which from the table or figure produces the z values of 0.893 for both variates corresponding to the 81.5th percentile value of each variate (again assuming the variates are normally distributed).

7.4. Center (CE): Figure 7 and Table 4

Assume that subjects have been allowed to set their own preferred

seat height (SH) above a vehicle floor and their seat distance (SD) back from a fixed heel reference point and that the data generated can adequately be described by a Gaussian (normal) bivariate distributions with the following univariate normal statistics of mean and standard deviation where "N" indicates a normal distribution.

Seat Height = N[12", 1.2"]
Seat Distance = N[s0", 2.0"]
ρ = 0.75

It is desired to set the "tightest" upper and lower bounds on both seat height and seat distance such that at least 90% of the population will find their preferred adjustment within these bounds. This is a Type 5 problem. Using the rationale of symmetrical and complementary cut-off values, these bounds can be found from Figure 7 and Table 4. Because this is a center problem, the z values for each variate are identical and for each variable, the upper and lower z values are identical except for sign. Therefore for each variate, the lower and upper z values are:

z = (-1.86, 1.86) corresponding to the percentiles (3.12%, 100 - 3.12 = 96.88%)

From which the actual bounding values are:

Seat Height$_{lower}$ = 1.2(-1.86) + 12 = 9.76"
Seat Height$_{upper}$ = 1.2(1.86) + 12 = 14.23"

Seat Distance$_{lower}$ = 2(-1.86) + 20 = 16.28
Seat Distance$_{upper}$ = 2(1.86) + 20 = 23.72

8. TABLES AND FIGURES

The tables and figures have been prepared in terms of the normalized variate "z" for selected correlation values and
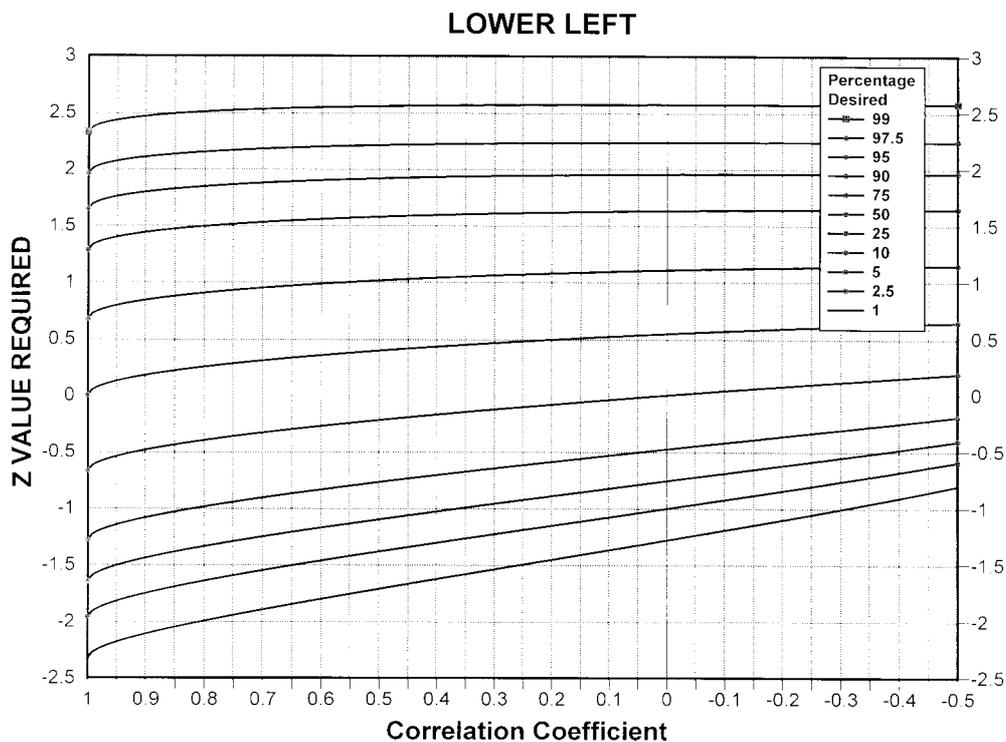
Figure 4. Values for the Lower Left (LL) problem type.

LOWER RIGHT
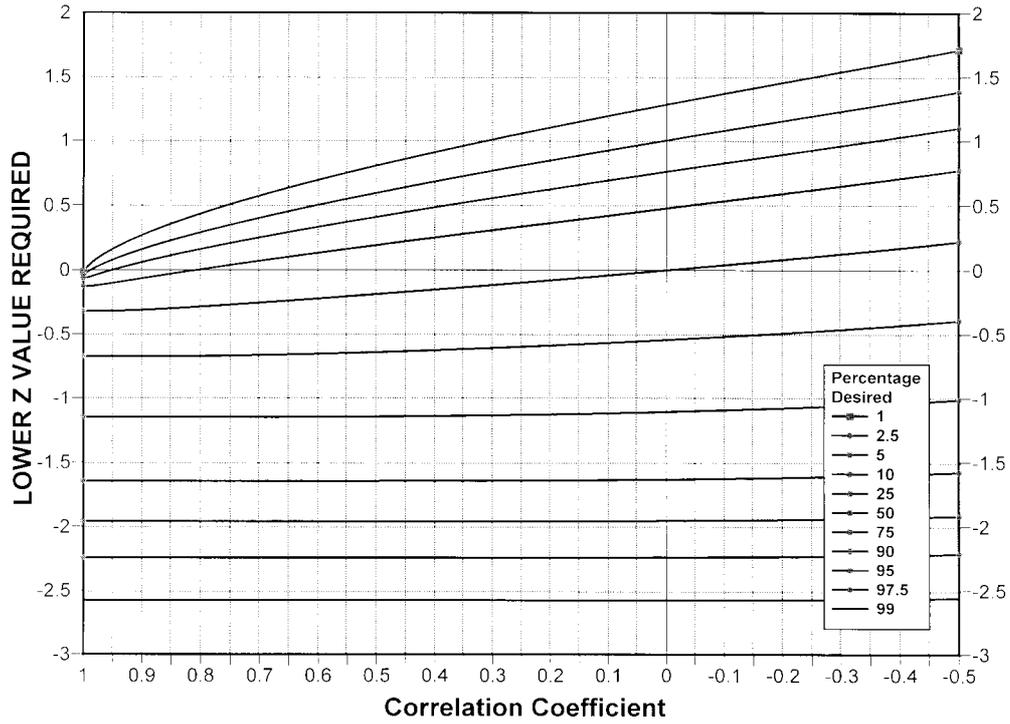
Figure 5. The smaller Z value for the Lower Right (LR) problem type.

UPPER RIGHT
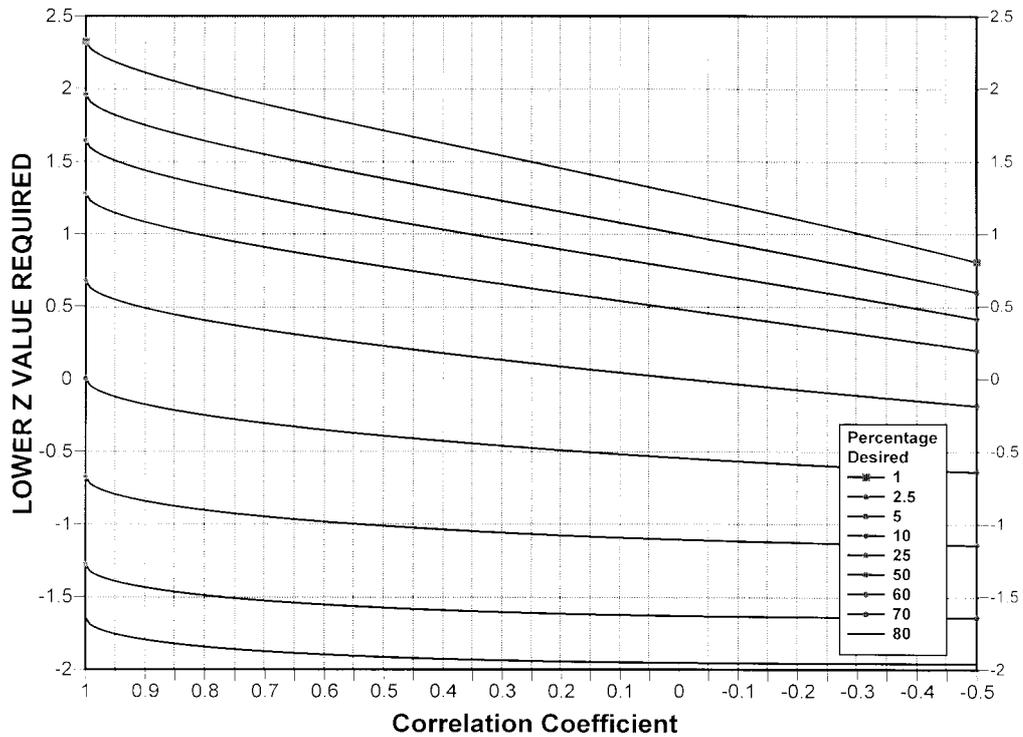
Figure 6. Z Values for the Upper Right (UR) problem type.

CENTER
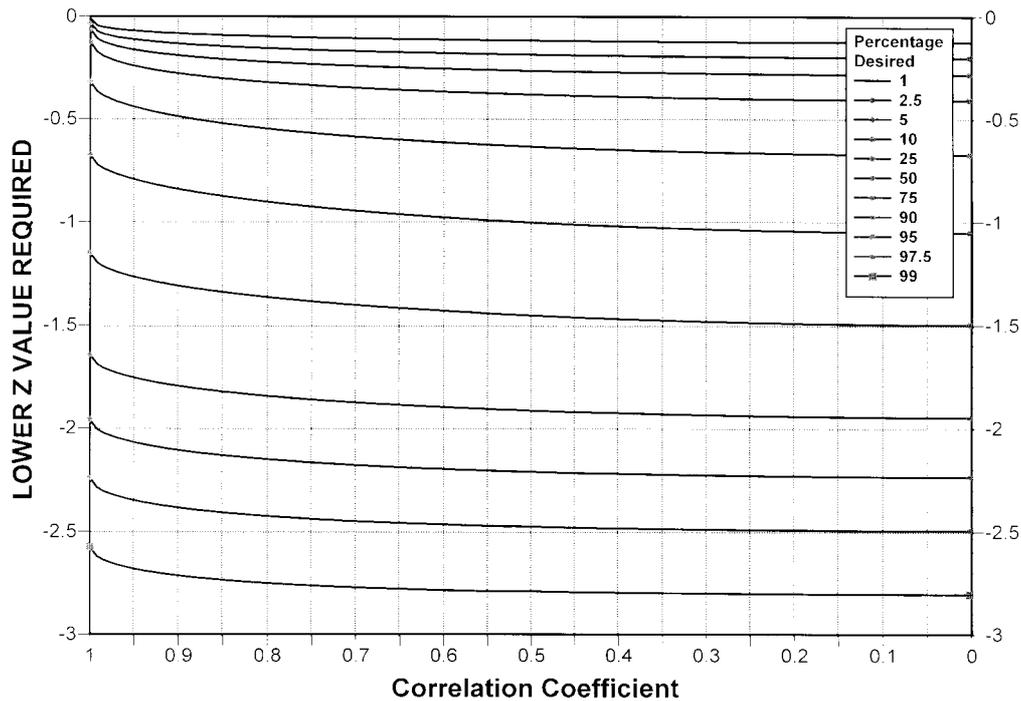
Figure 7. Smaller Z values for the Center (CE) problem type.

Table 1. Z value for the Lower Left (LL) problem type.

CORRELATION COEFFICIENT	LOWER LEFT QUADRANT										
	PERCENTAGE DESIRED										
	1	2.5	5	10	25	50	75	90	95	97.5	99
1	-2.325	-1.959	-1.644	-1.281	-0.674	0.000	0.675	1.282	1.645	1.960	2.328
0.95	-2.180	-1.817	-1.505	-1.145	-0.543	0.126	0.795	1.397	1.758	2.070	2.434
0.9	-2.108	-1.748	-1.439	-1.082	-0.485	0.178	0.842	1.439	1.797	2.108	2.468
0.85	-2.047	-1.691	-1.385	-1.032	-0.440	0.218	0.876	1.470	1.824	2.133	2.491
0.8	-1.992	-1.640	-1.336	-0.987	-0.401	0.251	0.904	1.493	1.845	2.152	2.508
0.75	-1.941	-1.593	-1.292	-0.946	-0.366	0.280	0.928	1.513	1.862	2.167	2.521
0.7	-1.892	-1.548	-1.250	-0.908	-0.333	0.307	0.949	1.529	1.877	2.179	2.531
0.65	-1.845	-1.505	-1.211	-0.872	-0.303	0.331	0.968	1.543	1.889	2.189	2.539
0.6	-1.800	-1.463	-1.172	-0.838	-0.275	0.353	0.984	1.556	1.899	2.198	2.546
0.55	-1.755	-1.423	-1.135	-0.804	-0.248	0.374	1.000	1.567	1.908	2.205	2.551
0.5	-1.711	-1.383	-1.099	-0.772	-0.222	0.394	1.014	1.577	1.916	2.211	2.556
0.45	-1.667	-1.344	-1.064	-0.741	-0.197	0.412	1.027	1.586	1.923	2.217	2.560
0.4	-1.624	-1.305	-1.029	-0.710	-0.173	0.429	1.039	1.594	1.929	2.221	2.563
0.35	-1.581	-1.267	-0.994	-0.680	-0.149	0.446	1.050	1.601	1.934	2.225	2.565
0.3	-1.539	-1.229	-0.960	-0.650	-0.126	0.462	1.060	1.607	1.938	2.228	2.567
0.25	-1.496	-1.191	-0.926	-0.621	-0.104	0.478	1.069	1.612	1.942	2.230	2.569
0.2	-1.453	-1.153	-0.893	-0.592	-0.083	0.492	1.078	1.617	1.946	2.233	2.570
0.15	-1.410	-1.115	-0.860	-0.563	-0.061	0.506	1.087	1.622	1.948	2.234	2.571
0.1	-1.367	-1.078	-0.826	-0.535	-0.041	0.520	1.094	1.626	1.951	2.236	2.572
0.05	-1.324	-1.040	-0.793	-0.506	0.020	0.533	1.101	1.629	1.953	2.237	2.572
0	-1.280	-1.002	-0.760	-0.478	0.000	0.545	1.108	1.632	1.954	2.238	2.572
-0.05	-1.236	-0.963	-0.726	-0.450	0.020	0.557	1.114	1.635	1.956	2.238	2.573
-0.1	-1.192	-0.925	-0.693	-0.422	0.039	0.568	1.119	1.637	1.957	2.239	2.573
-0.15	-1.147	-0.886	-0.659	-0.394	0.059	0.579	1.124	1.639	1.968	2.239	2.573
-0.2	-1.101	-0.847	-0.625	-0.366	0.078	0.590	1.129	1.640	1.958	2.239	2.573
-0.25	-1.054	-0.807	-0.590	-0.337	0.096	0.600	1.133	1.642	1.959	2.241	2.572
-0.3	-1.007	-0.766	-0.556	-0.309	0.115	0.609	1.136	1.643	1.959	2.241	2.567
-0.35	-0.959	-0.725	-0.520	-0.281	0.133	0.618	1.139	1.960	2.241	2.241	2.576
-0.4	-0.909	-0.693	-0.485	-0.252	0.151	0.627	1.142	1.644	1.960	2.241	2.576
-0.45	-0.858	-0.640	-0.448	-0.223	0.169	0.635	1.144	1.644	1.960	2.241	2.576
-0.5	-0.805	-0.596	-0.411	-0.194	0.186	0.642	1.146	1.645	1.960	2.241	2.576

z

Table 2. Smaller Z value for the Lower Right (LR) problem type.

| | LOWER RIGHT QUADRANT | | | | | | | | | | |
| | PERCENTAGE DESIRED | | | | | | | | | | |
CORRELATION COEFFICIENT	1.0	2.5	5.0	10.0	25.0	50.0	75.0	90.0	95.0	97.5	99.0
1.00	-0.01	-0.03	-0.06	-0.13	-0.32	-0.67	-1.15	-1.64	-1.96	-2.24	-2.58
0.95	0.16	0.08	0.00	-0.10	-0.32	-0.67	-1.15	-1.64	-1.96	-2.24	-2.58
0.90	0.27	0.16	0.06	-0.06	-0.31	-0.67	-1.15	-1.64	-1.96	-2.24	-2.58
0.85	0.36	0.23	0.11	0.03	-0.30	-0.67	-1.15	-1.64	-1.96	-2.24	-2.58
0.80	0.44	0.29	0.16	0.01	-0.28	-0.67	-1.15	-1.64	-1.96	-2.24	-2.58
0.75	0.51	0.35	0.21	0.04	-0.27	-0.67	-1.15	-1.64	-1.96	-2.24	-2.58
0.70	0.57	0.40	0.25	0.07	-0.25	-0.67	-1.15	-1.64	-1.96	-2.24	-2.58
0.65	0.64	0.45	0.29	0.10	-0.24	-0.66	-1.15	-1.64	-1.96	-2.24	-2.58
0.60	0.70	0.50	0.33	0.13	-0.22	-0.66	-1.15	-1.64	-1.96	-2.24	-2.58
0.55	0.75	0.55	0.37	0.16	-0.20	-0.65	-1.15	-1.64	-1.96	-2.24	-2.58
0.50	0.81	0.60	0.41	0.19	-0.19	-0.64	-1.15	-1.64	-1.96	-2.24	-2.58
0.45	0.86	0.64	0.45	0.22	-0.17	-0.63	-1.14	-1.64	-1.96	-2.24	-2.58
0.40	0.91	0.68	0.49	0.25	-0.15	-0.63	-1.14	-1.64	-1.96	-2.24	-2.58
0.35	0.96	0.73	0.52	0.28	-0.13	-0.62	-1.14	-1.64	-1.96	-2.24	-2.58
0.30	1.01	0.77	0.56	0.31	-0.11	-0.61	-1.14	-1.64	-1.96	-2.24	-2.58
0.25	1.06	0.81	0.59	0.34	-0.10	-0.60	-1.13	-1.64	-1.96	-2.24	-2.58
0.20	1.10	0.85	0.63	0.37	-0.08	-0.59	-1.13	-1.64	-1.96	-2.24	-2.58
0.15	1.15	0.89	0.66	0.39	-0.06	-0.58	-1.12	-1.64	-1.96	-2.24	-2.58
0.10	1.19	0.93	0.69	0.42	-0.04	-0.57	-1.12	-1.64	-1.96	-2.24	-2.58
0.05	1.24	0.96	0.73	0.45	-0.02	-0.56	-1.11	-1.63	-1.96	-2.24	-2.57
0	1.28	1.00	0.76	0.48	0.00	-0.54	-1.11	-1.63	-1.95	-2.24	-2.57
-0.05	1.32	1.04	0.79	0.51	0.02	-0.53	-1.10	-1.63	-1.95	-2.24	-2.57
-0.10	1.37	1.08	0.83	0.53	0.04	-0.52	-1.09	-1.63	-1.95	-2.24	-2.57
-0.15	1.41	1.12	0.86	0.56	0.06	-0.51	-1.09	-1.62	-1.95	-2.24	-2.57
-0.20	1.45	1.15	0.89	0.59	0.08	-0.49	-1.08	-1.62	-1.95	-2.23	-2.57
-0.25	1.50	1.19	0.93	0.62	0.10	0.48	-1.07	-1.61	-1.94	-2.23	-2.57
-0.30	1.54	1.23	0.96	0.65	0.13	-0.46	-1.06	-1.61	-1.94	-2.23	-2.57
-0.35	1.58	1.27	1.00	0.68	0.15	-0.45	-1.05	-1.60	-1.93	-2.23	-2.57
-0.40	1.63	1.31	1.03	0.71	0.17	-0.43	-1.04	-1.59	-1.93	-2.22	-2.56
-0.45	1.67	1.34	1.06	0.74	0.20	-0.41	-1.03	-1.59	-1.92	-2.22	-2.56
-0.50	1.71	1.38	1.10	0.77	0.22	-0.39	-1.01	-1.58	-1.92	-2.21	-2.56
											z

Table 3. Z value for the Upper Right (UR) problem type.

| | UPPER RIGHT QUADRANT | | | | | | | | | | |
| | PERCENTAGE DESIRED | | | | | | | | | | |
CORRELATION COEFFICIENT	1	2.5	5	10	25	50	75	90	95	97.5	99
1	2.325	1.959	1.644	1.281	0.674	0.000	-0.675	-1.282	-1.645	-1.960	-2.328
0.95	2.180	1.817	1.505	1.145	0.543	-0.126	-0.795	-1.397	-1.758	-2.070	-2.434
0.9	2.108	1.748	1.439	1.082	0.485	-0.178	-0.842	-1.439	-1.797	-2.108	-2.468
0.85	2.047	1.691	1.385	1.032	0.440	-0.218	-0.876	-1.470	-1.824	-2.133	-2.491
0.8	1.992	1.640	1.336	0.987	0.401	-0.251	-0.904	-1.493	-1.845	-2.152	-2.508
0.75	1.941	1.593	1.292	0.946	0.366	-0.280	-0.928	-1.513	-1.862	-2.167	-2.521
0.7	1.892	1.548	1.250	0.908	0.333	-0.307	-0.949	-1.529	-1.877	-2.179	-2.531
0.65	1.845	1.505	1.211	0.872	0.303	-0.331	-0.968	-1.543	-1.889	-2.189	-2.539
0.6	1.800	1.463	1.172	0.838	0.275	-0.353	-0.984	-1.556	-1.899	-2.198	-2.546
0.55	1.755	1.423	1.135	0.804	0.248	-0.374	-1.000	-1.567	-1.908	-2.205	-2.551
0.5	1.711	1.383	1.099	0.772	0.222	-0.394	-1.014	-1.577	-1.916	-2.211	-2.556
0.45	1.667	1.344	1.064	0.741	0.197	-0.412	-1.027	-1.586	-1.923	-2.217	-2.560
0.4	1.624	1.305	1.029	0.710	0.173	-0.429	-1.039	-1.594	-1.929	-2.221	-2.563
0.35	1.581	1.267	0.994	0.680	0.149	-0.446	-1.050	-1.601	-1.934	-2.225	-2.565
0.3	1.539	1.229	0.960	0.650	0.126	-0.462	-1.060	-1.607	-1.938	-2.228	-2.567
0.25	1.496	1.191	0.926	0.621	0.104	-0.478	-1.069	-1.612	-1.942	-2.230	-2.569
0.2	1.453	1.153	0.893	0.592	0.083	-0.492	-1.078	-1.617	-1.946	-2.233	-2.570
0.15	1.410	1.115	0.860	0.563	0.061	-0.506	-1.087	-1.622	-1.948	-2.234	-2.571
0.1	1.367	1.078	0.826	0.535	0.041	-0.520	-1.094	-1.626	-1.951	-2.236	-2.572
0.05	1.324	1.040	0.793	0.506	0.020	-0.533	-1.101	-1.629	-1.953	-2.237	-2.572
0	1.280	1.002	0.760	0.478	0.000	-0.545	-1.108	-1.632	-1.954	-2.238	-2.572
-0.05	1.236	0.963	0.726	0.450	-0.020	-0.557	-1.114	-1.635	-1.956	-2.238	-2.573
-0.1	1.192	0.925	0.693	0.422	-0.039	-0.568	-1.119	-1.637	-1.957	-2.239	-2.573
-0.15	1.147	0.886	0.659	0.394	-0.059	-0.579	-1.124	-1.639	-1.958	-2.239	-2.573
-0.2	1.101	0.847	0.625	0.366	-0.078	-0.590	-1.129	-1.640	-1.958	-2.239	-2.573
-0.25	1.054	0.807	0.590	0.337	-0.096	-0.600	-1.133	-1.642	-1.959	-2.241	-2.572
-0.3	1.007	0.766	0.556	0.309	-0.115	-0.609	-1.136	-1.643	-1.959	-2.241	-2.576
-0.35	0.959	0.725	0.520	0.281	-0.133	-0.618	-1.139	-1.643	-1.960	-2.241	-2.576
-0.4	0.909	0.683	0.485	0.252	-0.151	-0.627	-1.142	-1.644	-1.960	-2.241	-2.576
-0.45	0.858	0.640	0.448	0.223	-0.169	-0.635	-1.144	-1.644	-1.960	-2.241	-2.576
-0.5	0.805	0.596	0.411	0.194	-0.186	-0.642	-1.146	-1.645	-1.960	-2.241	-2.576
											z

Table 4. Smaller Z value for the Center (CE) problem type.

	CENTER										
	PERCENTAGE DESIRED										
CORRELATION COEFFICIENT	1	2.5	5	10	25	50	75	90	95	97.5	99
1	-0.01	-0.03	-0.06	-0.13	-0.32	-0.67	-1.15	-1.65	-1.96	-2.24	-2.58
0.95	-0.07	-0.11	-0.16	-0.24	-0.44	-0.80	-1.27	-1.76	-2.07	-2.35	-2.68
0.9	-0.08	-0.13	-0.19	-0.28	-0.49	-0.84	-1.31	-1.80	-2.11	-2.39	-2.71
0.85	-0.09	-0.15	-0.21	-0.30	-0.52	-0.88	-1.34	-1.82	-2.13	-2.41	-2.74
0.8	-0.10	-0.16	-0.22	-0.32	-0.55	-0.90	-1.37	-1.85	-2.15	-2.43	-2.75
0.75	-0.10	-0.16	-0.23	-0.34	-0.57	-0.93	-1.39	-1.86	-2.17	-2.44	-2.76
0.7	-0.11	-0.17	-0.24	-0.35	-0.59	-0.95	-1.40	-1.88	-2.18	-2.45	-2.77
0.65	-0.11	-0.17	-0.25	-0.36	-0.60	-0.96	-1.42	-1.89	-2.19	-2.46	-2.78
0.6	-0.11	-0.18	-0.25	-0.37	-0.61	-0.98	-1.43	-1.90	-2.20	-2.47	-2.79
0.55	-0.12	-0.18	-0.26	-0.37	-0.63	-0.99	-1.44	-1.91	-2.21	-2.47	-2.79
0.5	-0.12	-0.19	-0.26	-0.38	-0.63	-1.00	-1.45	-1.92	-2.21	-2.48	-2.79
0.45	-0.12	-0.19	-0.27	-0.39	-0.64	-1.01	-1.46	-1.92	-2.22	-2.48	-2.80
0.4	-0.12	-0.19	-0.27	-0.39	-0.65	-1.02	-1.47	-1.93	-2.22	-2.48	-2.80
0.35	-0.12	-0.19	-0.28	-0.40	-0.66	-1.03	-1.48	-1.93	-2.23	-2.49	-2.80
0.3	-0.12	-0.19	-0.28	-0.40	-0.66	-1.04	-1.48	-1.94	-2.23	-2.49	-2.80
0.25	-0.12	-0.20	-0.28	-0.40	-0.67	-1.04	-1.49	-1.94	-2.23	-2.49	-2.80
0.2	-0.12	-0.20	-0.28	-0.40	-0.67	-1.04	-1.49	-1.94	-2.23	-2.49	-2.80
0.15	-0.13	-0.20	-0.28	-0.41	-0.67	-1.05	-1.49	-1.95	-2.23	-2.49	-2.81
0.1	-0.13	-0.20	-0.28	-0.41	-0.67	-1.05	-1.50	-1.95	-2.24	-2.49	-2.80
0.05	-0.13	-0.20	-0.28	-0.41	-0.67	-1.05	-1.50	-1.95	-2.24	-2.50	-2.80
0	-0.13	-0.20	-0.28	-0.41	-0.67	-1.05	-1.50	-1.95	-2.24	-2.50	-2.80

z
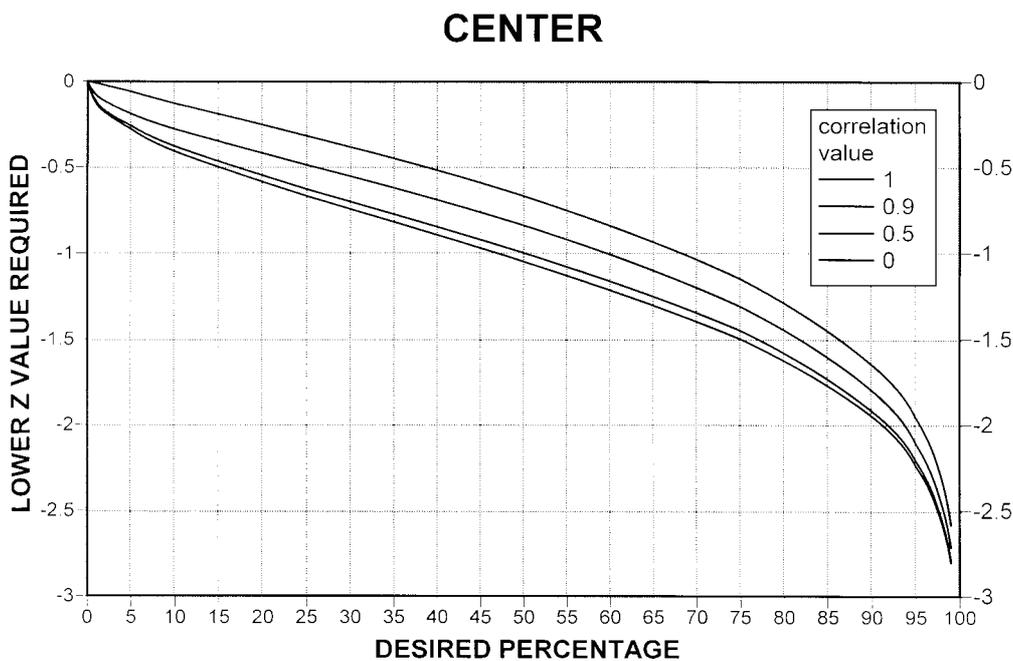
Figure 8. Data from table 4 replotted with the correlation value as the parameter in order to interpolate different percentages from the common ones used for the previous graphs and tables.

commonly desired percentage values. The tables give the z value in each cell for the indicated percentage desired with the relevant correlation coefficient between the variates.

The figures are especially helpful for determining the appropriate z value for any correlation value within the indicated range of [-0.5–1.0]. Or simple linear interpolation between neighboring correlation values in the tables should be accurate enough for most purposes. However, if it is desired to interpolate between the given percentage values, it is preferable to determine this by making a plot of z–vs–desired percentage for a selected

correlation value and then interpolating from the resulting graph. Figure 8 is such an example for four different correlation values with the data taken from Table 4. Alternatively, the appropriate equation as given previously for the problem type in question may be back solved with a suitable computer algorithm.

9. CONCLUSION

This paper has presented a series of graphs and tables suitable for bivariate design problems in which it is desired to specify

anthropometric values that together will achieve some specified accommodation percentage of the design population for five common design problem types. These values were computed by specifying the same (or complementary) percentile for each variate thereby reducing the infinite number of possible pair values to a single pair of equal percentile values. The percentile values to use are affected by the correlation between the variates. It is generally the case that the percentiles are not the same as the desired accommodation percentage. For example, if it is desired to accommodate 75% of some population in a Lower Left problem and if the correlation between the two variates is 0.4, then the percentile value to use for each variate is 85th rather than 75th percentile. Thus, approaching a bivariate design problem as two univariate ones can result in substantial errors of accommodation percentage. If, in fact, the 75th percentile values are used, the resulting accommodation percentage can be found to be equal to 61.04% rather than the desired 75%. A previous paper (Nah and Kreifeldt 1996) provides graphs and tables which illustrate the actual percentage achieved using equal percentile values for each variate for the problem types considered in this paper and for various correlation values.

REFERENCE

NAH, K. and KREIFELDT, J., 1996, The use of computer generated bivariate charts in anthropometric design. *Applied Ergonomics*, **27**(6), 397–409.

Cognitive Systems Engineering

E. Hollnagel and D. D. Woods

Ohio State University, 1971 Neil Ave, Columbus, OH 43210, USA

1. INTRODUCTION

In the second half of the twentieth century, one of the major challenges to human factors engineering and industrial psychology has been to understand the complexities of human–machine systems. These systems have become indispensable in the fabric of modern society as the technical processes that sustain production, services, and communication continue to grow in complexity and interdependency. Although technological developments, especially within information technology, have made it possible to build powerful, efficient, and highly reliable machines (such as aircraft, trains, power plants, factories, hospitals, etc.), the human operator remains an essential element. The conditions when the human operator is needed may have changed, but the need of humans in complex systems has not been significantly reduced. While humans may be required to do very little when everything goes as planned, the need to act is often extreme in critical situations. Furthermore, in many daily activities, such as buying a train ticket or getting cash money, human–human interactions have been replaced by human–machine interactions.

In order to understand the complexities of human–machine systems, it is necessary to have an appropriate basis, or conceptual foundation, for description and analysis. In this respect the study of human–machine systems is no different than any other scientific discipline. Each requires a set of concepts and a corresponding set of methods. The concepts are the basic hypotheses and assumptions about the domain, which in this case comprise humans working with technology. The concepts help identify what the important phenomena are and how they can be understood, and include the hypotheses and theories that are part and parcel of the scientific discipline. The concepts are the basis for the distinctions and analyses that can be made, and provide the "intellectual glue" that keeps everything together. The methods refer to the consistent and systematic ways in which the concepts can be applied, for instance, in the form of a classification system. The application can have a practical or utilitarian purpose such as in design, or a more scientific purpose, such as improving the understanding of the set of causes that have led to a specific consequence. The methods are intrinsically linked to the data, which constitute the empirical basis for the field and thereby provide the justification for the concepts.

2. THE BROKEN CIRCLE

The classical view of human–machine systems depicts a human and a machine that are linked by inputs and outputs (see figure 1). The control input to the machine determines whether the machine changes state or remains in the same state. As a result of this, some output is produced — for instance a set of measurements that indicate the state of the machine and the value of specific process parameters. The measurements, or the output from the machine, become the input to the human operator. According to the classical view, this input is "processed" by the
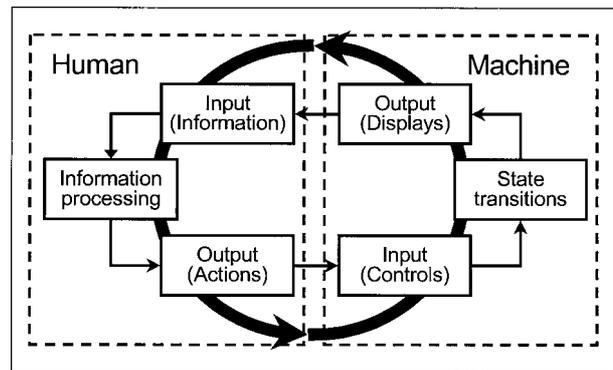
Figure 1: The classical view of human–machine systems.

operator and results in a response or output that becomes the control input to the machine. While the engineering sciences, such as control theory, have focused on describing how the machine works, the behavioral sciences have been more concerned with describing how the operator works, i.e. what goes on between receiving the input and producing the output.

When the challenge to describe the human operator was fully accepted by the behavioral sciences in the late 1960s, the focus changed from the human-machine system as a whole to the human operator as a separate system (see figure 2). In this way the circle or coupling between human and machine was broken. The link to the process was maintained in the sense that there was both an input to and an output from the operator, but for all intents and purposes these were considered less important than the processes that were assumed to take place within the operator's mind. Typically, only the operator part of the human–machine system was developed in any detail — or considered at all. The original cyclical model that described the coupling between humans and machines was transformed into a sequential or linear information processing model that mainly tried to account for the details of the human's responding to the input.

The conventional approach to the study of human–machine interaction is based on the notion of the human as an information processing system, either in the weak sense as an analogy or in the strong, metaphysical sense. This approach puts the focus on the interface or the interaction, i.e. on that which lies *between* the human and the machine, and studies the human operator as an information processing system interacting with a process or an artifact. This view is characteristic of, for instance, human factors engineering and ergonomics, human–computer interaction, cognitive science, and some versions of cognitive engineering.
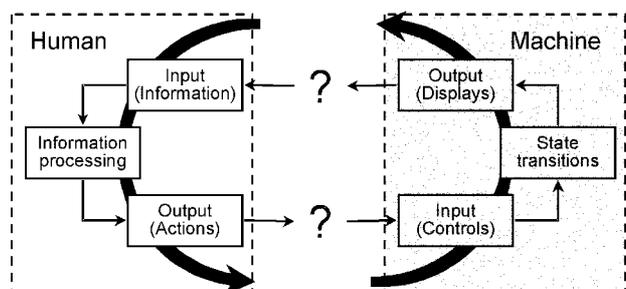
Figure 2: The broken circle with the human operator as a separate system.

Although it has been very successful as a basis for models, theories, and experiments, there is a growing consensus that it includes some fundamental limitations. Foremost among them is that it unavoidably separates or differentiates between human and machine, instead of seeing them together or as a whole. Yet understanding how a human–machine system works requires the ability to describe the system as a whole, hence to see it as more than a set of interacting parts. Since the concepts and methods of the classical view have proven insufficient for this, an alternative is required.

3. COGNITIVE SYSTEMS

Cognitive Systems Engineering (CSE) was formulated in the beginning of the 1980s (Hollnagel and Woods, 1983) to provide a consistent conceptual and methodological basis for research on human–machine systems, with design and evaluation as the two major activities. In CSE the focus is not on human cognition as an internal function or as a mental process, but rather on human activity or "cognition at work", i.e. on how cognition is necessary to accomplish effectively the tasks by which specific objectives related to either work or non-work activities can be achieved. Cognition is necessary to cope with the dilemmas, double binds, and trade-offs that arise from multiple and possibly inconsistent goals, organizational pressures, and clumsy technology. Rather than being isolated in the mind of a thoughtful individual, cognition at work typically involves several people distributed in space or time, which makes cooperation and coordination at least as important as human information processing. The interacting people are embedded in larger groups, professions, organizations, and institutions, which together define the conditions for work — the constraints and demands as well as the resources. Humans at work do not passively accept the technological artifacts nor the general conditions of their work, but actively and continuously adapt their tools and activities to respond to irregularities, disturbances, and to meet new demands. Cognition is part of an interconnected stream of activity that ebbs and flows, in which extended periods of lower activity are interspersed with busy, high tempo operations where correct and timely responses may be critical.

CSE proposes that composite operational systems can be looked at as single cognitive systems. Structurally they may comprise the individual people, the organization (both formal and informal), the high level technology artifacts (AI, automation, intelligent tutoring systems, computer-based visualization) and the low level technology artifacts (displays, alarms, procedures, paper notes, training programs) that are intended to support human practitioners. But functionally they can be seen as a single system. This is reflected by the main issues of CSE:

- *Coping with complexity*. The complexity is due to the multiple sources of information and control and to the possibly conflicting goals that characterize the working situation. People usually try to cope with complexity by reducing it, for instance by structuring the information at a higher level of abstraction with less resolution and making the required decisions at that level. The development of information technology has made it possible to provide computerized tools which to some extent accommodate the operators' needs, and thereby help their coping. This kind of support

clearly involves a replication of parts of human cognition, hence the use of an artificial cognitive system.

- *The use of tools*. Tools are artifacts that are used with a specific purpose to achieve a specific goal. Tools have traditionally been used to amplify human capabilities — in terms of physical performance (reach, force, speed, and precision), and in terms of perception and discrimination. More recently, tools have been introduced which are aimed at amplifying cognition. Although some cognitive tools have existed for ages, the use of computers has made it possible to design tools for more sophisticated functions, for instance decision making and planning.

- *Joint cognitive systems*. CSE recognizes that technological systems gradually have become "cognitive", in the sense that they are goal-driven and make use of cause-based (feedforward) regulation. Technological systems can thus be seen as artificial cognitive systems that interact with natural cognitive systems (i.e. humans). It is therefore appropriate to develop a view of joint cognitive systems, i.e. of cooperating systems which are described using a common set of terms — neither as machines nor as humans, but as cognitive systems.

Any discussion of CSE must obviously refer to a definition of what a cognitive system is. To avoid the thorny issues of what cognition is, and whether it can exist in non-human systems, a cognitive system can be defined as a system that is as able to control its behavior using information about itself and the situation, where the information can be prior information (knowledge, competence), situation specific information (feedback, indicators, measurements) and constructs (hypotheses, assumptions). The control can be complete or partial and will in the main depend on the ratio between expected and unexpected information. More formally, a cognitive system can be defined as a system that can modify its pattern of behavior on the basis of past experience in order to achieve specific anti-entropic ends.

4. CONSEQUENCES OF CSE

The notion of a joint cognitive system cannot easily be accommodated within the decomposed human-machine paradigm, and CSE can be seen as an independent alternative. Current methods mainly support a decomposition of a system into its parts (and in some cases also the reverse process of aggregation), but in a manner that implies partial independence between the parts. Some attempts have been made to develop methods that focus on the interaction and dependencies between sub-systems rather than on the components-elements. An example of that is multi-level flow modeling (MFM) which supports the goals-means analysis principle (Lind and Larsen, 1995). The overall framework for analysis must, however, be extended to recognize the dependency between data and interpretation, to account for the specific role of cognition (be it natural or artificial), and to highlight the consequences for design — supported by specific guidelines and design rules whenever possible.

In accordance with the intentions of CSE, the data must be found in situations that are representative of the real world (Hutchins, 1995). Any kind of systematic study carries with it some assumptions about what is being observed. These assumptions are relatively easy to understand when cognitive

systems are studied under controlled condition, which partly explains the preponderance of such studies. Yet it is a fundamental tenet of CSE that human action always is constrained by the context and studying cognition in the "wild" therefore does not release us from the obligation of understanding the assumptions that are made, even though they may be less easy to detect.

REFERENCES

HOLLNAGEL, E. and WOODS, D.D.,1983, Cognitive systems engineering: New wine in new bottles. *International Journal of Man-Machine Studies*, 18, 583–600.

HUTCHINS, E., 1995, *Cognition in the wild* (Cambridge, MA: MIT Press).

LIND, M. and LARSEN, M.N., 1995, Planning and the intentionality of dynamic environments. In J.-M. Hoc, P. C. Cacciabue and E. Hollnagel (eds*), Expertise and technology: Cognition and human-computer interaction* (Hillsdale, NJ: Lawrence Erlbaum).

Computer Simulation: Applications to Human Factors and Ergonomics

E.F. Fallon

Department of Industrial Engineering, National University of Ireland Galway, Galway, Ireland

1. INTRODUCTION

Simulation modeling is a powerful tool for human performance analysis in complex human–machine systems. In its broadest sense, the term simulation modeling can be applied to role playing, walk-throughs using mock-ups, aircraft training simulators, and virtual reality models of proposed systems, (Sanders 1991). However, in this chapter the emphasis is on modeling human performance or behavior in systems or as part of systems. The models referred to are abstract representations which can be coded on a computer. They are stochastic and therefore can be manipulated over time to simulate movement of the modeled system from state to state. Computer simulation of human performance in complex systems is presented as an integral part of systems design. Two categories of computer simulation models are highlighted, namely network models and microprocess models. A number of examples of software tools that implement both categories are presented.

2. COMPUTER SIMULATION

2.1. Definition

Computer simulation is the process of designing a mathematical–logical model of a real system and experimenting with this model on a computer. It encompasses the model-building process as well as the design and implementation of experiments involving the model and analysis of resultant data. Simulation modeling assumes that we can characterize a system in terms amenable to coding/computation. In this regard, a key concept is that of system–state description. If a system can be described by a set of variables with each combination of variable values representing a unique state or condition of the system, then manipulation of the variable values over time simulates movement of the system from state to state. This is precisely what computer simulation is — the representation of the dynamic behavior of the system by moving it from state to state in accordance with well-defined operating rules. Movement can be executed in a discrete manner, or a continuous manner, or a combination of these two. Experimentation with this representation allows inferences to be drawn about the real system being studied. The results obtained from running a simulation are observations of an experiment. This means that all inferences regarding the performance of the simulated system must be subject to appropriate statistical testing.

2.2. Steps in a Computer Simulation

Several stages in the development of a simulation model can be identified. At its most basic this development involves an iterative process which facilitates the evolution of an appropriate model from an initial simplified version. The main activities in the process are as follows (Chubb et al. 1987):

Problem formulation: The definition of the problem to be studied including a statement of the problem solving objective.

Model building: The abstraction of the system into mathematical–logical relationships in accordance with the problem formulation.

Data acquisition: The identification, specification and collection of data for model development and subsequent system analysis.

Model development and translation: The preparation of the model for computer processing.

Model verification: The process of establishing that the computer program executes as intended.

Model validation: The process of establishing that a desired accuracy or correspondence exists between the simulation model and the real system.

Design of experiments: The process of establishing the experimental conditions for using the model.

Experimentation: The undertaking of statistical experiments using the model.

Analysis of results: Consideration of results in the context of the objective formulated earlier.

Implementation: Implementation of the recommendations arising from the analysis of results.

The simulation modeling process may not adhere strictly to the sequence of stages outlined above as it may involve repeated evaluation and redesign of the model. Ultimately it should result in a simulation model that properly assesses the alternatives and enhances the decision-making process.

3. MODELS AND MODELING

3.1. Computer Simulation Modeling

Modeling or representing systems using abstractions is a difficult process that is made easier if the following criteria are satisfied (Pritsker and Wilson 1982):

Physical laws are available that are applicable to the system.

Pictorial or graphical representations of the system can be made. The variability of system inputs, elements and outputs is manageable.

Unfortunately these criteria are not immediately applicable to complex large systems which have been characterized as follows (Pritsker and Wilson 1982):

Few fundamental laws are available.

Many procedural elements are involved that are difficult to describe and represent.

Qualitative inputs are required that are difficult to quantify.

Random components are significant elements.

Human decision making is an integral part of such systems.

Computer simulation has been proposed as a modeling technique for tackling these problems. Complex human–machine systems can be characterized in the same manner and on this basis it would seem reasonable to propose computer simulation as a means of modeling complex human–machine systems, though the number of reported applications does not seem to support this.

3.2. Simulation of Complex Human–machine Systems

Computer simulation can provide a representation of the role of the human operator in systems. The question of interactions between tasks can be taken care of without having to resort to complex mathematical techniques. Computer simulation is

basically an expensive technique and there will necessarily be some trade-off between the complexity of the models used and the cost of the modeling effort. The simulation technique as applied to human–machine systems is usually employed in the following manner (Fallon *et al.* 1986). An underlying human performance model relevant to the current project is selected. A task analysis is then carried out and various task descriptors such as operating times, decision options, and workload requirements are determined. The model is implemented using a special purpose computer simulation package. Output requirements are determined, e.g. operator stress levels, evidence of failures, time required, and other variables that describe the human–machine system. The simulation model is then run many times in order to obtain statistically significant results.

3.3. Models

Models are representations of real or proposed systems. They are useful in describing, analyzing and designing systems and in enabling the modeler to understand the relationships between sub-systems and their interactions with other sub-systems. Models can be built to varying degrees of complexity depending on the needs of the modeler. It should be remembered that a model is a means to an end and not an end in itself, therefore, it is important to establish the purpose of the model before embarking on the study. The significance of an element of a system should be decided on the basis of the modeler's purpose. For example, if a modeler is interested in the number of units produced in a production system over a period of months, the fact that one small feature of the product can vary is unlikely to be significant. On the other hand, if the small varied feature necessitates a complex manual operation that is time-consuming and unreliable, then it may warrant further in-depth consideration. The purpose of the model will dictate the necessary action or otherwise. The ultimate success of a simulation study often depends on how well the significant elements in the model and their relationships can be defined. If the underlying relationships in models are only partly developed and understood, then the output of the simulation will reflect this.

3.4. Human Performance Models

The benefits of computer simulation can only be realized if appropriate and relevant models are applied. There are many different models of human performance and a number of different taxonomies for categorizing them, however they do not necessarily lend themselves to use in computer simulation. In the context of computer simulation, Meister (1995) identifies three relevant categories of models:

Task-network models: These models are top-down models and represent a detailed description of tasks and the relationship between them. They also require comprehensive data inputs at all subtask levels, e.g. performance time and probability of failure. They are compatible with representations of many task analysis techniques, e.g. Hierarchical Task Analysis, Decision Action Diagrams, and Functional Flow Diagrams.

Microprocess models: Meister describes this category as "bottom-up" models. Macros of human performance are synthesized from basic molecular human operations derived from research and theoretical studies, e.g. reaction times, time to move a limb, time to process x bits of information.

Control-theoretic models: These models are used to model continuous activities, e.g. tracking. Servo-control models represent interactions between the operator and the system. They are more quantitative than the other two categories of models above and address limited aspects of human behavior and performance.

Laughery (1997) only distinguishes between two categories of human performance models for computer simulation of complex human–machine systems:

Reductionist models: "Reductionist models use human–system task sequence as the primary organising structure", (Laughery 1997). This categorization is similar to the task network models referred to above.

First-principle models: "First-principle models of human behavior are structured around an organising framework that represents the underlying goals and principles of human performance", (Laughery 1997). This categorization is similar to the microprocess models of Meister.

The remainder of this article deals primarily with task network or reductionist models and to a lesser extent microprocess models used in computer simulation in human factors/ergonomics.

3.5. Network modeling

Network modeling is applied in the modeling of tasks that are mainly procedural in nature though there is no restriction on the nature of sub-tasks. A network model of a task is constructed by breaking up the overall task into a series of sub-task components, defining the sequence in which the sub-task components are performed and by specifying the selection rules for the next sub-task where a number of possibilities arise following the completion of a particular sub-task. The time to complete each sub-task must be available. Given this type of specification of the system a simulation can be performed which yields results related to system performance.

3.6. Siegel and Wolf Models

Siegel and Wolf (1968) first developed network models of human performance. These included, single and dual operator and small group models. The single operator model is time-oriented and considers operator activity as a series of discrete tasks. Parameters such as speed and accuracy (representing proficiency) and time stress threshold are used to reflect inter-operator variability. A probability of success is assigned to each task based on the goal aspiration of the operator, an accuracy factor, and time stress considerations. The goal of the Siegel and Wolf model was described as: "To allow a determination of where a man–machine system may overload or underload the operators while the system is in the early design stage", (Siegel and Wolf 1968). Essentially the state of loading of the operator is represented by time stress which is a function of the time available during the task to complete the remaining task elements.

4. NETWORK SIMULATION TOOLS

4.1. Systems Analysis Integrated Network of Tasks (SAINT)

SAINT — Systems Analysis of Integrated Network of Tasks — was a network modeling and simulation technique developed to assist in the design and analysis of complex human—machine systems. PC-based tools that contain enhanced and more comprehensive versions of its functionality have superseded it.

SAINT catered for the modeling and analysis of systems that contained procedural, risk, and random elements. It was designed to incorporate the best features of the Siegel and Wolf models and the network and symbol manipulation of the GERT simulation language. It employed a network of nodes and branches and a structured set of symbols within which it was possible to represent various task parameters. Developments of SAINT made it possible to incorporate human variability into its network models. A method was developed for building moderator functions (i.e. functions relating changes in the performance of a task to the independent variables known to affect the task) in a logical and fast manner. This was based on the premise that changes in human performance can be attributed to a number of possible factors other than simply random variation. Three classes of independent variables were identified: ability, environmental, and task variables. In addition, a set of 16 skill categories was selected in order to describe all operator task types. Quantitative relationships between performance using selected skills and the independent variables were established. Essentially this has the effect of modifying the parameters of task performance time distributions. SAINT has been applied to a broad range of problems from the military domain to evaluation of job decision aids and performance assessment of hot strip mills.

4.2. MicroSAINT

MicroSAINT is a network simulation software package for building simulation models of real-life processes. It was originally developed as a microcomputer version of SAINT and provided similar modeling capabilities without including the task moderator function facilities. MicroSAINT was initially presented as a human factors/ergonomics tool even though it differed very little from other simulation tools in terms of functionality offered. The primary difference was in the network modeling representation that was *activity on node* as opposed to the more prevalent *activity on branch*. This mapped well onto task analysis representations and possibly explains why it was more readily adopted by the human factors/ergonomics community than alternatives. MicroSAINT has found wide acceptance and is primarily used to model human service systems, training systems and human operator performance and interaction under changing conditions.

The main features of the system are: a network development tool with various routing decision types, a facility for developing mathematical expressions and functions (this can be used to implement the moderator functions described above in SAINT), a model animation facility, and automatic generation of statistics and other performance measures. MicroSAINT output can be accessed by standard statistical and analysis packages and can also be called from other software systems. MicroSAINT has been used on a wide range of applications within the manufacturing, health care and military domains.

4.3. WinCrew [1]

WinCrew is a discrete event simulation-modeling tool the development of which was sponsored by the US Army Research Laboratory's Human Research and Engineering Directorate (ARL-HRED). It combines the functionality of discrete event computer simulation from MicroSAINT with models of mental workload based on the Multiple Resource Theory (MRT) of attention. The

mission network diagram of nodes and branches is the basic building block of a WinCrew model. Task data in the form of operation times, conditions for task execution and consequences of failure are provided for each node. WinCrew also possesses a "Micro Model" feature that utilizes known human performance rules, principles and research findings to generate task mean times where test data or expert derived subject-matter data do not exist or are not available. A task allocation module that facilitates dynamic allocation while a task is being executed is also available.

The workload assessment component of WinCrew requires a number of inputs: mental resources, interfaces used, resource requirements of individual interfaces (resource/interface channels) and an assignment of resource/interface channels to tasks.

4.4. The Human Operator Simulator (HOS)

Meister (1995) classifies the Human Operator Simulator (HOS) as a microprocess model. He also describes it as a deterministic model. It is assumed that an operator's behavior is explainable and not random and that an operator's actions and the duration of these actions are determined fully by the state of the system and the operator's goals at any particular point in time. HOS works by building-up a model of human operations from a series of micro-models or fundamental human activities, e.g. reaction times, memory recall times and movement of limbs. Obviously individual micro-models can contain random components and therefore HOS is not truly deterministic. In theory, the micro-models of HOS could be combined with a stochastic network model. For example, the effect of the micro-models would be to modify the parameters of the distribution of task times. Such an approach, albeit with a restricted set of micro-models, has been implemented in SAINT and to a lesser extent in WinCrew. The Integrated Performance Modeling Environment (IPME) outlined below is a more comprehensive implementation.

The main weaknesses of HOS are that it has no mechanism for tackling error processes and only one operator can be modeled in detail. HOS was originally a research and development tool and its primary application domains were military and aerospace.

4.5. IPME (Integrated Performance Modeling Environment) [1]

IPME is a tool for analyzing human system performance in complex systems. It combines many of the capabilities of HOS with the MicroSAINT simulation engine. The key features of IPME include: facilities to model the effect of the environment, a module to model operator traits and states, a module to define user-defined performance-shaping factors, a workload measurement algorithm and simulation experimentation facilities. Unlike HOS, IPME has the capability to tackle operator task failures. It also has a novel feature that enables the definition of a physical workspace associated with a task network.

5. SUMMARY

In this article computer simulation as applied to analysis of human performance in complex systems has been defined and described. The relevance of task network and microprocess models to this activity has been established and a number of commercially available computer simulation tools that implement these types of models have been presented. It is clear that the capability to use this technology to analyze human performance in the context

of systems development exists. It is likely that further developments in the power and usability of computing systems will encourage their greater use.

REFERENCES

[1] http://www.radata.demon.co.uk/ipme.html

CHUBB, G.P., LAUGHERY, K.R., and PRITSKER, A.A.B., 1987, Simulating manned systems. In Salvendy (ed.), *Handbook of Human Factors,* (New York: Wiley*)*, pp. 1298–1327.

FALLON, E.F., BRENNAN, L., GALLWEY, T.J., and O'KELLY, M.E.J., 1986, Simulation modeling as a tool in human factors engineering. In Crosbie, Luker (eds), *Proceedings of the 1986 Summer Computer Simulation Conference*, 28–30 July, Reno, Nevada, USA, (San Diego: Society for Computer Simulation), pp. 689–95.

LAUGHERY, K.R., 1997, Computer modeling and simulation. In Salvendy (ed.), *Handbook of Human Factors and Ergonomics,* (New York: Wiley), pp. 1375–1408.

MEISTER, D., 1995, Simulation and modeling. In Wilson and Corlett (eds), *Evaluation of Human Work: A practical ergonomics methodology*, (London: Taylor & Francis), pp. 202–28.

PRITSKER A.A.B. and WILSON J.R., 1982, Computer Simulation. In Salvendy (ed.), *Handbook of Industrial Engineering*, (New York: Wiley).

SANDERS, A.F., 1991, Simulation as a tool in the measurement of human performance, *Ergonomics*, 34, 995–1025.

SIEGEL, A.I. and WOLF, J.J., 1968, *Man–Machine Simulation Models*, (New York: Wiley).

Content Analysis: Hypermedia Design

N. A. Stanton
Department of Psychology, University of Southampton, SO17 1BJ, UK

1. INTRODUCTION

It has been proposed that hypermedia and hypertext systems may have much to offer the information age (Nelson 1987; Neilson 1990). The idea that access to information will determine the success, or failure, of an organization is not new. Indeed, the initial ideas behind hypermedia can be traced back to a much-quoted article printed in the *Atlantic Monthly* in July 1945 by Vannevar Bush entitled "As we may think". Bush argued that the access to, and communication of, knowledge had made a significant contribution to the outcome of the Second World War. He argued further that traditional methods of transmitting the information (e.g. scholarly journals and books) were inadequate for the task. This observation was based upon Bush's task of orchestrating some six thousand American scientists in the application of science to warfare. He found problems associated with information management that could not be resolved by available systems. This led to his rather radical (in its time) proposal for the MEMEX system:

A . . . device in which an individual stores his books, records, and communications, and which is mechanised so that it may be consulted with exceeding speed and flexibility.

Bush was able to foresee this device as tool for augmenting human activity. However, there are two main problems of information management. The first problem is one of physical management— making the information easy to handle. Piles of books, reports, and scientific articles are not easy to use in quantity. Each is designed to be used as a stand-alone source. Whilst it is very easy to read one book and one article, problems occur when the numbers increase. The second problem relates to the way in which information is accessed and retrieved. This problem is different from the first because it refers to the user getting the appropriate information from a document in a timely manner, rather than the physical management of a large number of documents. It would seem that paper-based documents are relatively easy to use, in the sense that a shelf full of books of assorted sizes and colors makes it easy to identify the one required and, once selected, the chapter, section, page and paragraph are easy to locate. However, this assumes one knows effective mechanisms for retrieving the book (e.g. books shelved according to content) and accessing the material (e.g. contents listings and indexes). These retrieval mechanisms relieve the user from knowing the entire contents and precise positions of all of the books.

The first problem has largely been solved through technological development. Bush (1945: 103) prophesied this technological development. He wrote that:

The Encyclopaedia Britannica could be reduced to the volume of a matchbox.

There are four main technologies associated with digital media and electronic publishing: storage technologies (e.g. CD-ROM), communication technologies (e.g. the information superhighway), desktop technologies (e.g. personal computers), and interface technologies (e.g. Mosaic). The second problem relates to authoring and using the material. People have problems using and authoring electronic systems. There is no clear guidance, particularly when it comes to converting existing printed material into the electronic form. Some electronic systems do no more that their paper-based counterparts, i.e. electronic page turning. Whereas other systems make anything, and indeed everything, possible. In the first implementation, users are bound to feel constrained by the system, indeed the only advantage it has over a book is the reduction in physical space requirements. In the latter implementation, users may feel that they are unable to make sense of environment: the lack of a clear structure means that everything is apparently linked to everything else. Bush (1945) proposed the notion of trails:

associative indexing, the basic idea of which is a provision whereby any item may cause at will to select immediately and automatically together. ... The process of tying two items [or more] together is the important thing.

Hammond and Allinson (1988) suggest that hypertext will lead to more efficient production of electronic material and could empower students to take full control over what they choose to view. Nelson professes that each user in the hypertext network could also become an author by structuring and changing the material in their own individual way. This is a futuristic vision of hypertext. As Nelson (1974) wrote:

the structure of ideas is not sequential. They tie in together whichway . . . In an important sense there are no subjects at all; there is only all knowledge, since the cross-connections among the myriad topics of this world simply cannot be divided up neatly.

Stanton and Baber (1994) argue that many of the hypertext systems reported in the literature have been poorly designed, and suggest that there is a single root cause of this poor design. The basic concept of hypertext, in computer science terms, is of an "n–dimensional" space, which can be traversed by moving through links. "Space" as used in this context has a well-defined meaning, as the collection of objects and activities contained within a specific domain — for example, a similar fashion to its use in that of finite state architecture. However, much current research in hypertext appears to use the term "space" in its everyday sense, i.e. as a physical relationship between objects. This has led to the use of the concept of "electronic space", which can be represented in the form of a map, and through which users navigate. The translation of the concept of "mathematical space" into "electronic space" has led to the inappropriate adoption of a spatial metaphor; this adoption may have occurred unwittingly, but the notion of hypertext as space appears to have been accepted without question by the community. The much-cited phenomenon of getting lost in hypertext documents suggests that hypertext designers do not provide information in an appropriate format for users. The fact that most of the reported incidents of getting lost occur in very small documents suggests that, unless the problem is tackled soon, as systems grow larger, "lostness" will become more prevalent (Stanton, Taylor, and Tweedie 1992). The key problems of hypertext are twofold. The

first is that designers are often unaware of the problems users have in "navigating" hypertext. Consequently, they fail to provide adequate information on the screen. Second, because hypertext is currently designed as "electronic space" designers emphasize the development of links between nodes, and users have to traverse these links. Hence, the reports of getting lost (Stanton, Taylor, and Tweedie 1992).

Stanton and Baber (1994) suggest a simple but radical shift in hypertext design: rather than using links as programmable features, they suggest that nodes should be made more sophisticated. Nodes ought to be defined in terms of specific properties that would include defined links, such as those which relate to nodes containing similar information. This concept is similar to object-oriented programming; the objects in this example are node types. The properties of the nodes would then be the links, and would exhibit such characteristics as inheritance, membership, etc. Users would allocate certain properties to nodes, and these properties would form the basis of links. In practice, Stanton and Baber propose that the form of interaction with the text would be similar to content analysis techniques.

2. CONTENT ANALYSIS

Content analysis (Kirakowski and Corbett 1989; Robson 1993) is a technique that has been around since the beginning of the century for analyzing the content of documents. The term "documents" refers to all media: newspapers, diaries, speeches, letters, reports, books, journals, notices, films, photographs, videos, radio, and television programs. Content analysis consists of five main stages as follows.

2.1. Determine Objectives

The starting point of the process is to determine the objectives for the use of the material; for example, it may be intended for use as teaching or training material, as a journal, database, or an encyclopaedia, or to support a help desk, etc. Each of these different objectives may engage the user in different types of activity and therefore require the content to be structured in a different manner. Some material may be used for more than one type of activity and therefore require structuring in more than one way. Hypermedia has a distinct advantage over traditional media by virtue of enabling different structures to be placed upon it.

2.2. Define Unit of Analysis

Based upon the objectives and the use to which the material is to be put, the next task is to define the unit of analysis. In the case of text, the unit could comprise individual words, phrases, sentences, or paragraphs, etc., whereas in the case of pictures, the unit of analysis could be an object, frame, or sequence of frames, etc. It is also necessary to consider the context within which the unit occurs as this will affect the meaning of the unit. Software tools exist for identifying keywords, keywords in context and combined criteria lists. Such tools make the task of analyzing large documents much easier.

2.3. Construct Categories for Analysis

The construction of categories will be based upon the considerations above (objectives and unit of analysis). Generic categories might be formed from subject matter, objects, authors,

countries, etc. It is desirable that the categories are exhaustive and mutually exclusive. This ensures that everything can be categorized and reduces the ambiguity of categorization. The analysis will be no better or worse than the system of categorization.

2.4. Test Coding to Assess Reliability

To determine the objectivity of the coding and categorization system, it is necessary that the scheme is tested. Typically a sample of the material is taken and coded by at least two persons who have been trained in using the coding and categorization system by the person who devised it. The categorizations made by these persons should be tested for reliability. If reliability is low (i.e. there is little agreement on which units should be assigned to the categories) then the coding scheme may need revising or more training of the persons required. If reliability is high (i.e. there is a high degree of agreement on which units should be assigned to the categories) then the scheme is ready to be used.

2.5. Conduct Analysis

The content of the documents can be analyzed according to the devised scheme. The analysis is continually open to verification and checks of reliability. Indeed content analysis should be considered as a continual process. The analyst can change the objective, unit of analysis and construct new categories, but each change should undergo some form of testing. The advantages of using content analysis to assist in the design of hypermedia systems are that it can be used on existing documents, it enables re-analysis for reliability checks, and it is relatively cost efficient compared with designing a hypermedia system from scratch. However, one must bear in mind that bias or distortion could be introduced into the analysis and that the documents have originally been produced for another purpose.

3. PROPOSAL FOR ELECTRONIC IMPLEMENTATION

To begin with, text and graphics are entered into a hypertext system, but no links have been made. The author (or user) is now faced with the task of analyzing the information on the screen. This task is analogous to content analysis as typically performed on non-electronic, open-ended information. This framework could usefully support the construction of hypertext as it allows the author to place some degree of objective structure on the information, whereas traditionally links have been almost arbitrarily assigned. To assist in this task, the system contains a hierarchy of keywords and phrases that have been entered into the database. First of all the author picks an object from the screen. Second, the object that has been selected into either text, sound, graphic or movie is classified. Third, the author is required to indicate which of the associated keywords is appropriate to the context under analysis. Fourth, the author rates the "strength of association" of this object (through the use of a slider control, the strength assigned is relative, not absolute). Keywords also need to be defined (Waterworth, Chignell, and Zhai 1993); the author may also choose alternative classifications and/or add new ones.

Users would engage in a different kind of interaction. They would not see the classification forms used by authors; rather, they would be faced with the same information and, on selecting

an object, would be presented with a box allowing them to call upon further related information. This notion is an implementation of the idea proposed by Stanton and Baber (1992), which suggested that hypertext systems should not seek to constrain users. Indeed, the user-choice screen idea could be extended to enable users to call up information that has not been entered into the author's original classification. By altering the "strength of association", users could dynamically manipulate the list of potential information sources available to them. This proposition is closer to the original formulation of hypertext than many commercial systems currently allow. This proposal might seem to incorporate the spirit of Bush's MEMEX device outlined in the introduction. However, there are two important differences: the first is that the authors of the material have analyzed the content of the information to propose a keyword classification, and the second is the idea that some documents may have stronger association with some keywords than others.

The approach proposed serves two purposes: it forces designers to shift emphasis away from links in space and on to node design, and it means that new nodes can be added relatively easily to an existing application — i.e. there will be no need to redefine all the links when a new node is added. This approach can potentially be used with current software. It is possible, for instance, on HyperCard to copy the complete contents of one card and paste it into a new card. The properties of the buttons will remain the same unless the author changes them. If there was a provision for more global structuring, such as project scripts in SuperCard, then specific properties could be attached to objects. For instance, in a database, a new piece of information could be defined as an "interface card" and given the properties of other cards in that category.

The notion of "definable nodes", as Stanton and Baber (1994) term the concept, does not mean that the basic premises of hypertext need to be altered. Rather, we feel that they will be brought to the forefront of technological development. At present, designers appear to be caught up in trivial problems posed by lack of understanding; research is being carried out into the symptoms of problems in hypertext use, rather than into their causes. Emphasis is being placed on a metaphor, i.e. navigation, which at best is misleading and at worst is wrong. Finally, there appears to be little agreement as to how hypertext ought to be used. Some applications function well as education aids, others as database inquiry tools, but there does not appear to be a consensus as to how such packages should be designed. What is applicable to education need not be applicable to database inquiry. This means that, in addition to new approaches to software design, we suggest designers should consider issues of functionality and usability as well as computational elegance and ease of design. In any hypertext package a user will require only a small amount of information at any one time. Is it more important to provide that information in a usable form, or to show what other information could be accessed? If one opts for the latter, one will arrive at a scenario similar to "channel hopping" when viewing television — there will be a wealth of "information", but little knowledge.

ACKNOWLEDGEMENT

REFERENCES

BUSH, V., 1945, As we may think. *Atlantic Monthly* 76: 101–8.

HAMMOND, N., and ALLISON, L., 1988, Extending hypertext for learning: an investigation of access and guidance tools. In A. Sutcliffe and L. McCauley (eds) *People and Computers, vol. V* (Cambridge: Cambridge University Press).

KIRAKOWSKI, J., and CORBETT, M., 1989, *Effective Methodology for the Study of HCI* (Amsterdam: North-Holland).

NELSON, T., 1974, *Computer Lib* (Published by the author).

NIELSON, J., 1989, The art of navigating through hypertext. *Communications of the ACM* 33 (3): 296–310.

NIELSON, J., 1990, *Hypertext and Hypermedia* (Boston: Academic Press).

ROBSON, C., 1993, *Real World Research* (Oxford: Blackwell).

STANTON, N. A., and BABER, C., 1992, An investigation of styles and strategies in self directed learning. *Journal of Educational Multimedia and Hypermedia* 1 (2): 147–67.

STANTON, N. A., and BABER, C., 1994, The myth of navigating in hypertext: how a bandwagon lost its course! *Journal of Educational Multimedia and Hypermedia* (in press).

STANTON, N. A., TAYLOR, R.G., and TWEEDIE, L. A., 1992, Maps as navigation aids in hypertext environments: an empirical evaluation *Journal of Educational Multimedia and Hypermedia* 1(4): 431–44.

WATERWORTH, J. A., CHIGNELL, M. H., and ZHAI, S. M., 1993, From icons to interface models: designing hypermedia from the bottom up. *International Journal of Man-Machine Studies* 39: 453–72.

Design Methodology

W. W. Gasparski

Institute of Philosophy and Sociology, Polish Academy of Science, Warsaw, Poland

Methodology is nothing more than normative epistemology (Bunge 1983: 4) therefore *design methodology* is normative design epistemology. Such a characteristic needs the acceptance of a combination of the two concepts, namely: theory of cognition — i.e. epistemology, and a noncognitive but change oriented discipline — design. For some design theoreticians it would be easier to accept the combination in question if they were able to find references to the design oriented families of learned disciplines. In order to do this let us refer to Kotarbióski's praxiological inquiry and to Simon's "sciences of the artificial'. To refer to Kotarbióski is legitimized by the very fact that the *design methodology* we are talking about is of praxiological (i.e.. general methodological) origin. To refer to Simon is legitimized by the end products of any design activity which are conceptual models of manmade objects, i.e. *artifacts*.

"By creativity I understand inventiveness," says Kotarbinski, "whether in the domain of products or in the domain of methods. All efficient work leads to the rise of something what does not exist before, thus the work is creative to some extent" (Kotarbinski 1986: 42). Therefore the engineers (or engineering-like) creativity would be inventiveness in the domain of technical products and technological methods conditioned by the effective work of engineers of different specialties. Simon (1981: 129) says: "Historically and traditionally, it has been the task of the science disciplines to teach about natural things: how they are and how they work. It has been the task of engineering schools to teach about artificial things: how to make artifacts that have desired properties and how to design." Therefore design plays a twofold role: first, as a method characteristic of applied (practical) sciences oriented towards theories of particular classes of artifacts; second, a kind of activity serving the conceptual preparation of changes founded on the theories of applied (practical) sciences. Thus the subject of design methodology is about the totality of issues covered by the propositions on which the applied sciences are based, their reasoning, justification, and characteristics, and changes that have evolved from these theories.

Let us add that it is still an open question as to what degree the reasoning and justification for proposed changes are applicable only to the engineering sciences, and to what degree they are relevant to other practical sciences and design methods. These issues belong to general/generic design methodology (Warfield 1990). First, design methodology enables and simplifies the articulation of design knowledge, and, second, the conceptual apparatus elaborated by the methodology serves as a tool to do that. It is a meta-language in which a language of design is characterized and knowledge which a designer, as a reflective practitioner (Schōn 1983), produces as an additional product of his or her activity.

The language of design is one in which design solutions are formulated throughout the design process; it also includes a "language" of the particular design actions involved in that process. The first language is one of the general theory of objects (systems) adapted to the needs of particular design disciplines and to the multiple perspectives (Linstone 1995) which serve as a base for the assessment of design solutions produced by these disciplines. The second language is used to describe the inferential structure of the process of designing, actions contained within the process, and connections between them — for example, in the terms of *p-t-s* (problem-tasks-solution) networks (Gasparski 1989). This logic of the context of justification as related to design solutions is accompanied by a *sui generis* logic of discovery, or, rather, the generation of proposals from which the design solutions are selected. Zwicky's morphology, Altshuller's invention of algorithms, and so on, as well as iterative loops, may serve as examples of the second kind of design logic.

The creation of a language that is used by reflective practitioners as a means to describe design phenomenon is not, however, the main task of design methodology. The language is rather a by-product of the fundamental task of the methodology in which the primary task is to study the design process and combinations of particular design actions, in order to identify them and describe ways in which to perform them in a systematic manner — that is, to study methods of design.

Design methodology has two divisions: investigation of the design process is called the *pragmatic design methodology*, while the study of system designing is called *apragmatic design methodology*. Pragmatic methodology of design is of general character because the design process understood in terms of its inferential structure is invariant. It allows designers from one design discipline to take an advantage of the experience gained by designers involved in other design disciplines; design methodology serves as an effective intermediary to do this. The second division of design methodology is composed of several specialized subdisciplines studying design objectives of different design disciplines. General/generic apragmatic design methodology offers the subdisciplines (which are particular apragmatic methodologies) the general framework in which to investigate particular design objectives and general theses to concretize them in relation to particular cases. Among these are: (i) an outline of the general theory of change — for change is the task of design; (ii) an outline of the general theory of systems — since coping with complexity and simplifying it to manageable proportions is what design approach is about; (iii) *sui generis* ontology of practical situations, for the objects of design are collections of practical situations (from a pair of them on), the mapping of which are design problems (Æ *Design*). Design methodology, in relation to this issue, overlaps with the logic of interrogative sentences, since design problems are questions or sentences transferable to the form of interrogative sentences. The methodology of design is linked with the cognitive science, knowledge engineering, and the discipline called CAD (computer-aided design).

Design methodology of praxiological provenance is related not only to the methodology of science but also to the theories of action, practical, and applied philosophy, which study, among other things, the social context of human actions. In relation to this, we suggest that only the design which effectively prepares conceptually such impartial actions deserves the positive appraisal. "Thou shalt design — proclaims the *technological imperative -* or help implement only projects that will not endanger public welfare, and shalt alert the public against any projects that fail to satisfy this condition" (Bunge 1985: 310).

Changes conceptually prepared by design are evaluated against the criterion of their *relevancy*, and only such changes that pass the evaluation are consider *proper changes*. The relevancy concept comprises: intention, being real, rationality as well as utilitarian, ethical (moral), and aesthetical values of changes in question. This systemic assessment of design solutions needs to be guided by — among other things — the results of practical philosophy oriented toward social practice. It brings about participation of the users (of design solutions) in design processes as a method of manifestation of their values. It also brings about the so-called *technology assessment* as a method of evaluation of technical solutions from a social (resultant effects) point of view. Philosophical reflection on design pointing out design — understood in a proper way and performed in a proper manner — as a means of improving engineering, not only in a technological sense but also in a social sense, is becoming a *sui generis* designing of design. This "designological" approach is characteristic for Design Culture, the emerging stage of the Design Era (Gasparski 1993b: 156).

REFERENCES

BUNGE, M., 1985a, *Treatise on Basic Philosophy*, Vol. V: *Epistemology and Methodology* (Dordrecht: Reidel).

BUNGE, M., 1985b, *Treatise on Basic Philosophy*, Vol. VII: *Philosophy of Science and Technology* (Dordrecht: Reidel).

GASPARSKI, W.W., 1983, Design methodology: a personal statement. In: P. T. Durbin (ed.), *Philosophy of Technology: Practical, Historical and Other Dimensions* (Dordrecht: Kluwer), pp. 153–67.

GASPARSKI, W.W., 1993a, Design, science and philosophy: the praxiological perspective. In: M. J. de Vries, N. Cross, D. P. Grant (eds), *Design Methodology and Relationships with Science* (Dordrecht: Kluwer), pp. 165–89.

GASPARSKI, W.W., 1993b, The evolution of a design approach: a historical perspective and its relevance to the design of education systems. In C. M. Regeluth, B. H. Banathy, J. R. Olson (eds), *Comprehensive Systems Design: A New Educational Technology*, (Berlin: Springer), pp. 154–64.

KOTARBIÓSKI, T., 1986, *Thoughts about Humans and Human Affairs* (Wroclaw: Ossolineum).

LINSTONE, H.A., 1995, The multiple perspective. In A. Collen, W. Gasparski (eds), *Design and Systems: General Applications of Methodology* (New Brunswick, USA/London: Transaction), pp. 177–242.

SCHÖN, D.A.,1983, *The Reflective Practitioner: How Professionals Think in Action*, (New York: Basic Books).

SIMON, H.A., 1981, *The Sciences of the Artificial*, 2nd edn, (Cambridge MA: MIT Press).

WARFIELD, N., 1990, *A Science of Generic Design: Managing Complexity through Systems Design* (Salinas CA: Intersystems Publications).

DIALOG: Human Reliability Assessment

H. Bubb and I. Jastrzebska–Fraczek

Technical University of Munich, Boltzmannstraße 15, 85747 Garching
Germany

1. SYSTEM ERGONOMICS AND HUMAN ERROR

Most of the conventionally applied Human-Reliability-Assessment (HRA) methods are based on an event orientated categorizing system. They provide reliable prognosis for so-called skilled-based actions, which are often enough repeated. In the case of cognitive tasks they cannot support sufficient data. The HCR (human cognitive reliability) methods classify human error as belonging only to the ratio of time available to time required. As system ergonomics deals with the information flow in a man machine system (Figure 1) as well as with the information flow between different systems, this would serve as a base to assess human reliability independent to the level of information processing.

By the information flow of a singular man–machine–system the quality of human work can be calculated as the relation of the result of a work to the task. Performance in this context is: quality per time. Limits of quality are explicitly or implicitly defined in every system. If any human doing results in a jumping over such pre-defined limits, we call that *Human Error* (HE; Rigby 1970). The probability of non-occurrence of such HE's defines the *Human Reliability* (HR). This consideration is important not only for a simple singular man–machine–system but also for any complicated combinations of different singular man–machine–systems, which describes in its totality any arbitrary system as, for example, a power plant, an assembly plant or a traffic system. Apart from technical defects, especially human errors influence the reliability and dependability of such systems. After all experience primary technical, often unimportant, failures occur and initiate inadequate human reactions. As such effects cannot be treated only by measurement in the technical area more

profound treatments of human operation in the technological context and its possibilities of deviation are necessary. System ergonomics can be used as a base.

2. THE IDEA OF SYSTEM ERGONOMICS

System ergonomics starts with a general description of properties of every task and assigns experimental experience, and, from that, ergonomic recommendations to the partial aspects of the task can be made. The fundamental idea is that by the knowledge of the information transfer by the subsystems man and machine, the tasks to be performed by the operator may be designed. For this designing of the task that results from the system mission and from the specifically chosen layout of the system and the system components (e.g. the machine), the following fundamental rules are to be considered, formulated as questions:

1. Function: "What has the operator in view and how far is he assisted by the technical system?"
2. Feed back: "Is the operator allowed to recognize it, if he has effected something and what was the success of it?"
3. Compatibility: "How much effort has the operator to make in order to convert the code system between the different technical information channels?"

The *function* may be separated into the intrinsic task-contents and the influenceable task-design. Under this aspect the *task-contents* is essentially defined by the *temporal and spatial order* of the activities which are to be carried out to perform the total task. It may be described by the terms operation (= *temporal* organization), dimensionality (= *spatial order* of the task) and manner of control (= the *time and location window* within which the task must be finished (Bubb 1988). The *task-design* refers to the system structure that may be chosen to a large extent by the system planner. In this area the degree of difficulty may be influenced by the specifically designed layout. It can be distinguished between the manner of presenting task and result to the operator — the so-called *display (compensatory* respectively *pursuit task)*, and the manner of involving the operator in the total system — the so-called *manner of task (active* respectively *monitive task)*.

Feed back calls for a certain kind of redundancy, which asks for the *number of sensory organs*. A further aspect is the *time* that occurs between the input of information on the control element and the reaction of the system on the output side. Considering

Man-machine-system (MMS): Combination and entirety of interaction between man and working materials
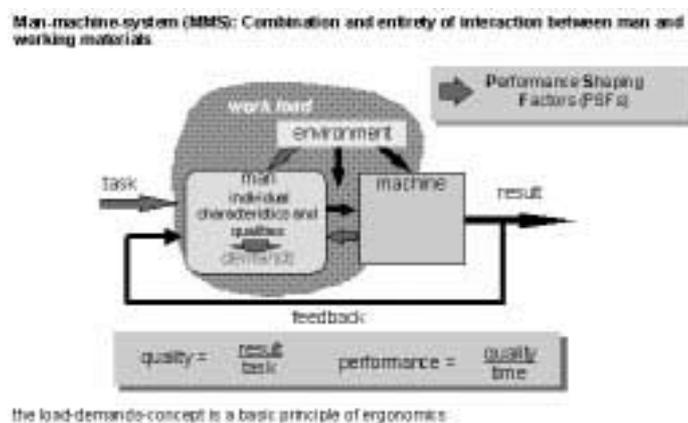
Figure 1.

the relevant recommendations a well-designed feed back allows the operator always to answer the question:

- "What have I done?"
- "In what a state is the system?"

Compatibility describes the facility that enables the operator to convert the code system between the different information channels that he has to handle. It characterizes the obvious between different areas of information, as reality indication by instruments, control elements, and internal models of the operator. Neglecting the rules of compatibility decreases the human reliability immensely (Spanner 1993).

3. HE SOFTWARE–TOOL DIALOG

As Sträter (1997) has shown, by the system ergonomic aspect an important number of human errors observed in the reality can be categorized and explained. Although he found some human error probability data from real live observations using a system ergonomic categorizing scheme, these results should be substituted by experimental investigations under controlled working conditions. For this reason the software DIALOG was developed. It allows to investigate to a far extent the different aspects of the system ergonomics described above. The task for the subject is to re-write a combination of several numbers between 1 and 20. This is an example for *sequential operation* of different degree of difficulty. Simultaneously further tasks can be displayed in form of traffic lights (with the color sequence "green," "yellow," "red"), which are to be switched off by a mouse click before the appearance of the color "red," up to nine traffic lights can be presented. That is an example for *simultaneous task* of different degree of difficulty. By an additional indicated circle different time windows can be investigated. This circle is originally "green." During the task a "red" segment appears increasing with time. That is a representation of a *static task* ("manner of control"). Figure 2 shows an example of the video screen, as it is presented to the subject during the experimental run.

Experiments were carried out using the described equipment. Thirty-two subjects had to accomplish different tasks presented by DIALOG during a session of ~1.5 h. All reactions of the subjects were recorded by DIALOG. In the following only a part of the possible analysis is reported.

4. RESULTS

In a first investigation the areas "operation" (sequential task of the difficulty "5 numbers" and "7 numbers" in combination "with" and "without simultaneous switching off of traffic lights") and "manner of control" ("with" and "without a temporal window of 4 s respectively 8 s," within which the tasks had to be accomplished) have been the objective in view.
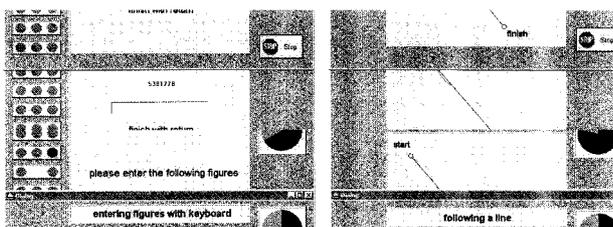
Figure 2.

It was found out that time pressure results in a more increased error probability than additional simultaneous tasks: In the case of recording five numbers the average error probability was ~0.03. This value increased with the factor 14 to 0.43 under the condition of time pressure. The simultaneous task of switching off traffic lights had no additional effect to the error probability, with a time window of 8 s. Without time pressure the increase of the difficulty of a sequential task of five to seven numbers had an effect of the factor 2.5. Under the condition of time pressure of 4 s the same increase had only an effect of the factor 1.3. A more difficult task seems to stimulate the attention. This suggestion is substituted by a more detailed investigation of errors, so one could separate the following kinds of error (Swain and Guttmann 1983):

- error of quality, z_1 (writing a wrong number, e.g. 123 instead 124);
- error of part omission, z_2 (omitting one number, e.g. 12 instead 124);
- error of omission, respectively negligence, z_3 (the task is totally neglected); and
- error of addition, z_4 (writing an additional number, not being asked for, e.g. 1234 instead 124).

In the case of time pressure all the types of error can occur in combination with a time error, i.e. the task is not finished within the time window, t (Figure 4).

Whereas the error of negligence does not appear without time pressure, this type crops up with time pressure and has a value of 0.076. Under the condition of time pressure the time error appears with a probability independent of the difficulty of the task of 0.07–0.09. Of course, all types of error are increasing with time pressure (see above). There is one exception: The error of quality disappears under this condition. The probability of all kinds of error becomes smaller when these appear in combination with time error. There is again one exception: the error of omission
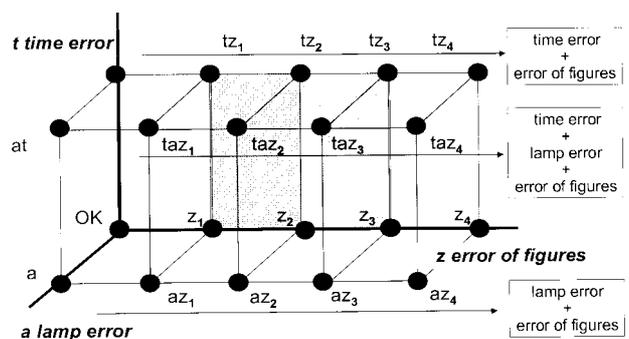
Figure 3.
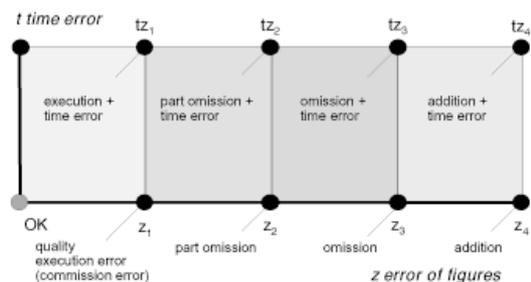
Figure 4.
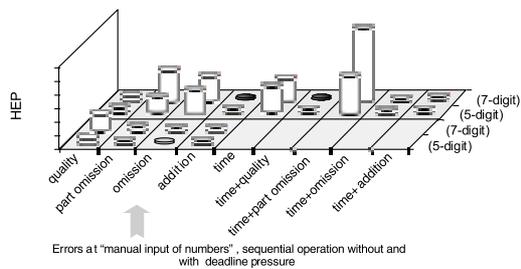
Figure 5.

Figure 6.

increases in connection with time error with the factor 16 for five numbers and 12 for seven numbers.

The observation shows that a more difficult (sequential) task under the condition of time pressure has an under-proportional ascent of error probability in relation to the behavior of more easier tasks (Figure 6).

REFERENCES

BUBB, H., 1988, System ergonomics as an approach to improve human reliability. *Nuclear Engineering and Design*, 110, S233–45.

BUBB, H. and JASTRZEBSKA-FRACZEK I., 1999, Dialog: a computer-based system for development of experimental data on human reliability. In W. Karwowski and W.S. Marras (eds) *The Occupational Ergonomics Handbook* (Boca Raton: CRC Press), 667–76.
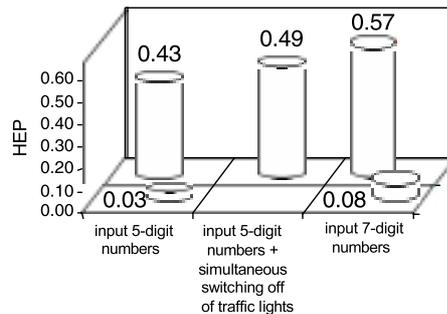
RIGBY, L.V., 1970, The nature of human error. In American Society for Quality Control, Annual Technical Conference, 457–66.

SPANNER, B., 1993, Einfluß der Kompatibilität von Stellteilen auf die menschliche Zuverlässigkeit. Fortschritts-Berichte VDI, Reihe 17 "Biotechnik" (Düsseldorf: VDI).

STRÅTER, O., 1997, Beurteilung der menschlichen Zuverlässigkeit auf der Basis von Betriebserfahrungen. Dissertation an der Technischen Universität München.

SWAIN, A.D. and GUTTMAN, H.E., 1983, *Handbook of Human Reliability Analysis with Emphasis on Nuclear Power Plant Applications*. NUREG/CR-1278 (Albuquereque: Scandia Laboratories).

Digital Human Models for Ergonomics

J. C. Woldstad

Texas Tech University, Lubbock, TX 79409-3061, USA

1. INTRODUCTION

Digital human models in the context of this section are computer-generated representations of human beings used in computer-aided design (CAD) or similar programs. These models are increasingly being used by ergonomists and other engineers to design both equipment and work environments to meet the needs of human operators. They have the advantage of allowing the designer to explore the potential advantages and disadvantages of different design configurations without requiring the construction of expensive physical mock-ups used in the past.

Using a digital human model, design engineers can position and manipulate operators of varying anthropometry within the simulated work environment. A variety of different analyses can be performed depending on the sophistication of the computer package. Currently available analysis modules include: reach and clearances, field of vision and visual obstructions, work posture and biomechanics, metabolic energy expenditure, time and motion, and others. In addition, the ability of these models to dynamically simulate human operators within proposed engineering projects has become a popular presentation and sales tools.

Joints
1 - Right Wrist
2 - Right Elbow
3 - Right Shoulder
4 - Right Sternoclavicular
5 - Left Sternoclavicular
6 - Left Shoulder
7 - Left Elbow
8 - Left Wrist
9 - Right Ankle
10 - Right Knee
11 - Right Hip
12 - Lower Back (L5/S1)
13 - Left Hip
14 - Left Knee
15 - Left Ankle
16 - Upper Back (T8/T9)
17 - Neck (C7/T1)
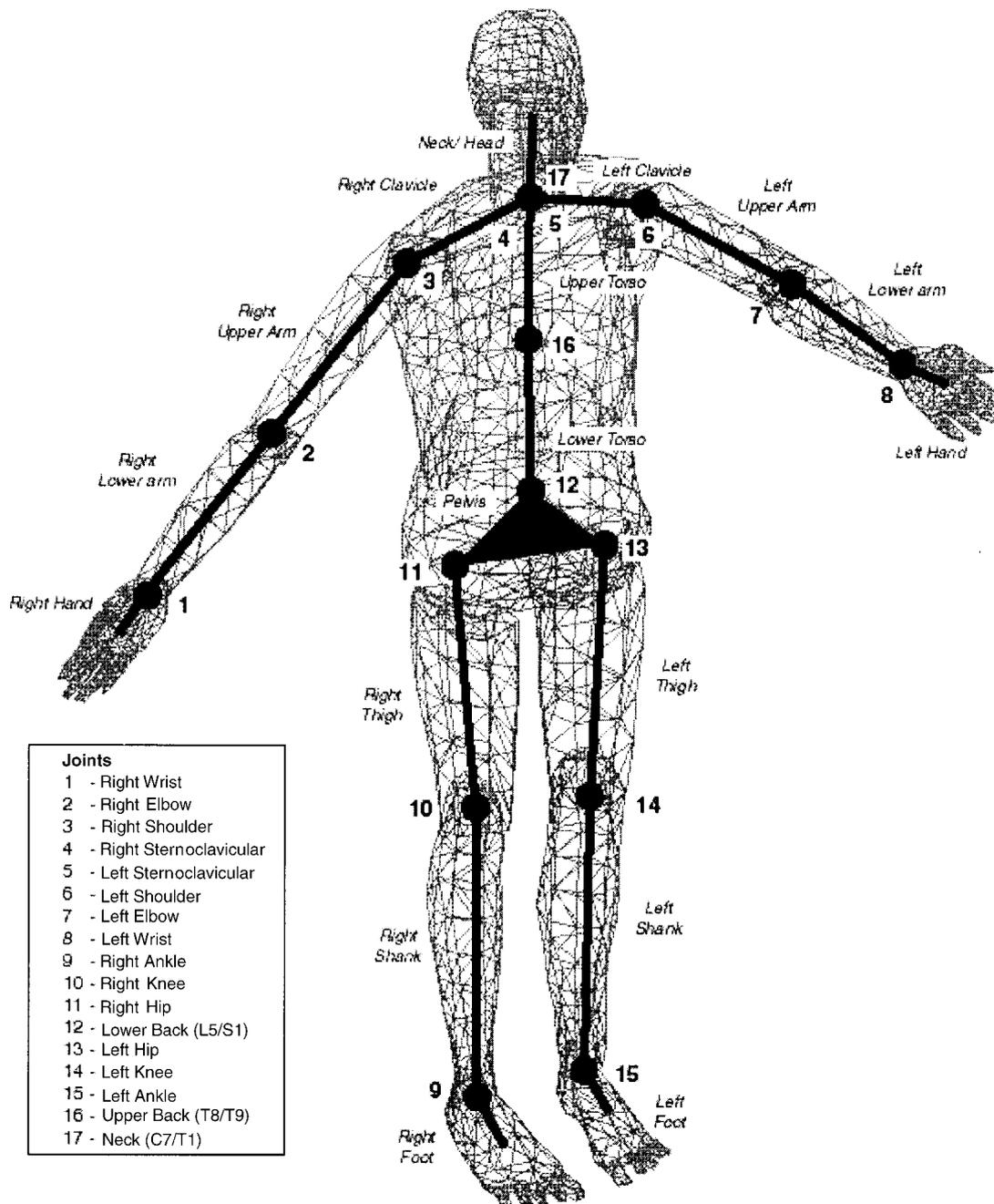
Figure 1. Typical simple three-dimensional link system for the human body.

2. MODEL STRUCTURE AND FORM

Most digital human models are composed of an underlying rigid-link framework similar to that used in many biomechanical models of the human body. This framework assumes that the body can be represented as a series of straight mechanical links connected at pin joints that roughly corresponds to the human skeleton. The sophistication of this framework will vary considerably from a relatively simple representation (Figure 1) to models that include more than 90 different links and 140 degrees of freedom. In addition, more sophisticated geometric models of the skeleton have been developed for joints that cannot be easily modeled using a rigid link structure. Most notably, a spherical model of the shoulder joint has been developed in an attempt to model the complexities of the shoulder complex (Badler *et al.* 1993).

The underlying skeletal structure of digital human models is almost always covered with solid geometric constructs used to represent the surface of the human body. Initially, these constructs or geometric primitives consisted of articulated polygons that would leave noticeable gaps between the segments when the model was positioned in extended postures. More recently, complex geometric models and smoothing algorithms have been developed that virtually eliminate the robot-like appearance of some of these models. In addition to surface geometry to represent the human body, most models come with a variety of different clothing options that can be applied to the model. Clothing options often interact with range-of-motion and collision detection components of the model to increase the realism.

3. ANTHROPOMETRY, POSTURE, AND MOTION

Anthropometric dimensions for most models can either be inputted directly or extracted from a number of existing databases. A fundamental limitation of digital human models is their inability to capture the large variation in human size and shape. Designers often desire models or manikins representative of various percentiles of the population (usually the 5th percentile female, 95th percentile male). Unfortunately, manikins of this type cannot be accurately constructed due to the inherent co-variance between human dimensions. The designers of digital human models have derived a number of different solutions to overcome this limitation. One approach has been to statistically model the relationships between different dimensions and then to select multiple dimensions that represent the boundary of a combined or joint confidence interval. Another has been to construct a family of representative anthropometries for the population under study. This group of manikins is either selected using Monte Carlo techniques or using data for whole-body surface scans of actual individuals.

To be effectively used, manikins not only need to be dimensioned, but they must be accurately positioned in the work environment. Similar to the anthropometric modules, most models allow the user to manipulate the posture of the manikin directly or use postures generated by the computer program. Postures generated by the program can be from a library of predefined "typical" human postures for the desired activity or generated using an inverse kinematics algorithm. Postures in most programs are limited by the known range-of-motion for the joint being positioned. Range-of-motion data is based on several

anthropometric surveys that have collected this information in addition to body dimensions. Accurately reproducing human body postures in a simulated environment has proven to be one of the more difficult tasks in constructing digital human models. As noted above, many human models have a large number of link segments and joints requiring the specification of many different parameters to fully specify the posture (some models have more than 140 degrees of freedom). Human body postures in general are also highly indeterminate or redundant, meaning that when the position of the hands and feet are fully specified, the rest of the body can still be positioned in many different ways. Research to understand how the human body selects from the many different potential postures available is currently ongoing at several universities and has not yet produced definitive results.

Human body motion algorithms are also included with many computer packages. These can produce typical reach, lift, and grasp motions for a stationary model, or walking and carrying motions for a model moving through the environment. Problems accurately predicting human motion are analogous to those associated with predicting postures, with even a higher level complexity involved. Similarly, our current understanding of human motion is not sufficient to allow truly accurate representation in a virtual environment. However, the algorithms employed often produce visually compelling animations of work activities and environments.

4. REACHES AND CLEARANCES

The principle use of digital human models by designers has been the comparison of the dimensions and limitations of the human body to the geometry of the proposed device or work area. Work envelopes can be easily calculated using these models to indicate the portion of the environment that can easily be reached by a worker in the simulated environment. Similarly, collision detection components to most models can be used to determine when and how a worker's body will come in contact with a component of the environment so that clearances can be determined. The determination of reaches and clearances depend upon the anthropometry of the individual manikin employed and the use of many different anthropometries that are representative of the population of interest (as discussed above) is usually necessary for accurate estimates. In addition, the posture and motion components to a model can play a key role in obtaining accurate estimates. Figure 2 below shows a typical use of a digital human model to determine both a reach and a clearance. In Figure 2, the operator must insert her arm into a limiting enclosure to manipulate a tool or device. The model is able to show how the enclosure limits the reach, and to some extent the motion of the operator.

5. VISION

Several models also allow the user to view the virtual environment from the perspective of the human manikin within the environment. This view allows the designer to explore the environment for potential visual obstructions. The view generated depends on the anthropometry, posture, and motion of the manikin, as well as, assumptions regarding the useful field of human vision. Figure 3 shows a typical scene from the point of view of a manikin within a virtual work environment. In this case the position of displays and controls for the seated operator are of concern in the design.
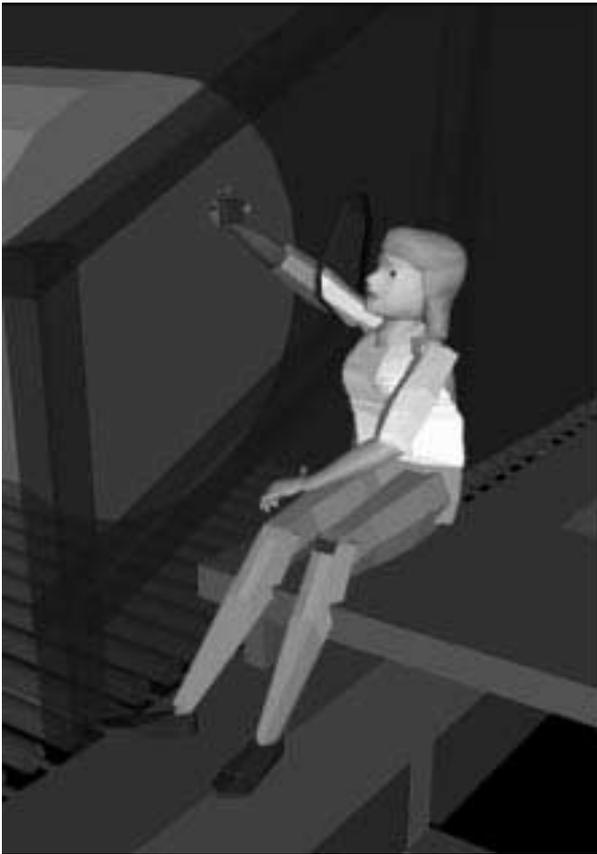
Figure 2. Reach and clearance evaluation (reprinted with permission from Engineering Animation, Inc., 1999).
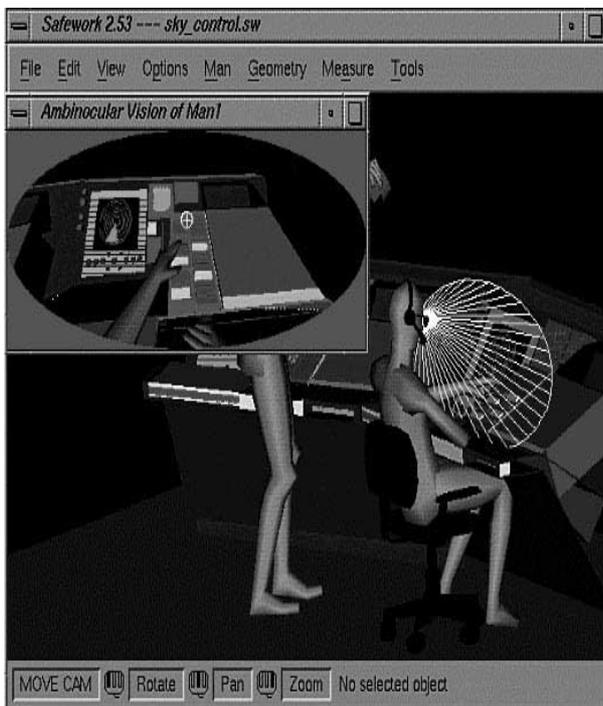
Figure 3. Field-of-view evaluation. (reprinted with permission from Safework, Inc., 1999)

6. BIOMECHANICS

Biomechanical models are often incorporated into modeling packages to allow the designer to estimate the mechanical stresses placed upon the internal structures of body during different activities. The objective is to estimate how work activities stress the bones, muscles and connective tissues of the body and to predict when these stresses will lead to damage of these structures. This approach is very popular in ergonomics because it closely corresponds with most expert views of the etiology of injury during manual materials handling (NIOSH 1981). Simple biomechanical models estimate the torque placed upon the joints of the body related to a work activity. More complex models will estimate parameters such as joint strength capabilities, internal muscle forces, and lower back intervertebral disc compressive force. A majority of models currently used are either based on or are similar to the 3D Static Strength Prediction Program produced by the University of Michigan (Chaffin *et al.* 1999). A typical output from this program is shown in Figure 4.

Biomechanical models are either two- or three-dimensional and either static or dynamic. For static models, the calculations require information on the orientation of the links in the model (subject posture), the length of each segment, the mass of each segment, and the location of the center-of-mass or each segment. Dynamic models require this same information plus the angular joint accelerations, linear acceleration of each segment at the center-of-mass, and the moment-of-inertia of each link through the center of mass.

7. OTHER MODEL FEATURES

A variety of other model analysis tools and features in addition to those described above are available for digital human modeling software, depending on the manufacturer and the intended use of the software. Some currently available features on different models include:

- Predetermined time study analysis based on the Methods Time Measurement (MTM) system.
- Work posture analysis based on the Ovako Working Posture Analysis System (OWAS) to analyze the relative discomfort of the back, arms, and legs in working postures.
- Metabolic energy expenditure rates based on the models of Garg *et al.* (1978).
- Analysis of manual materials handling tasks based on the National Institute of Occupational Safety and Health's (NIOSH 1994) lifting equation and the psychophysical limits recommended by Snook and Ciriello (1991).
- Upper limb analysis based on the Rapid Upper Limb Assessment (RULA) tool.

8. LIMITATIONS OF CURRENT MODEL

Current digital human models appear to be limited much more by our understanding of ergonomics and human behavior than by computer technology. Deficiencies in currently available data on three-dimensional human anthropometry and strength limit model accuracy. In addition, we do not have a clear understanding of the basic principles behind human posture and motion, as noted previously. In many cases, the methodology used by models is not clearly specified so that qualified users can judge the accuracy of the prediction due to the proprietary nature of many programs.
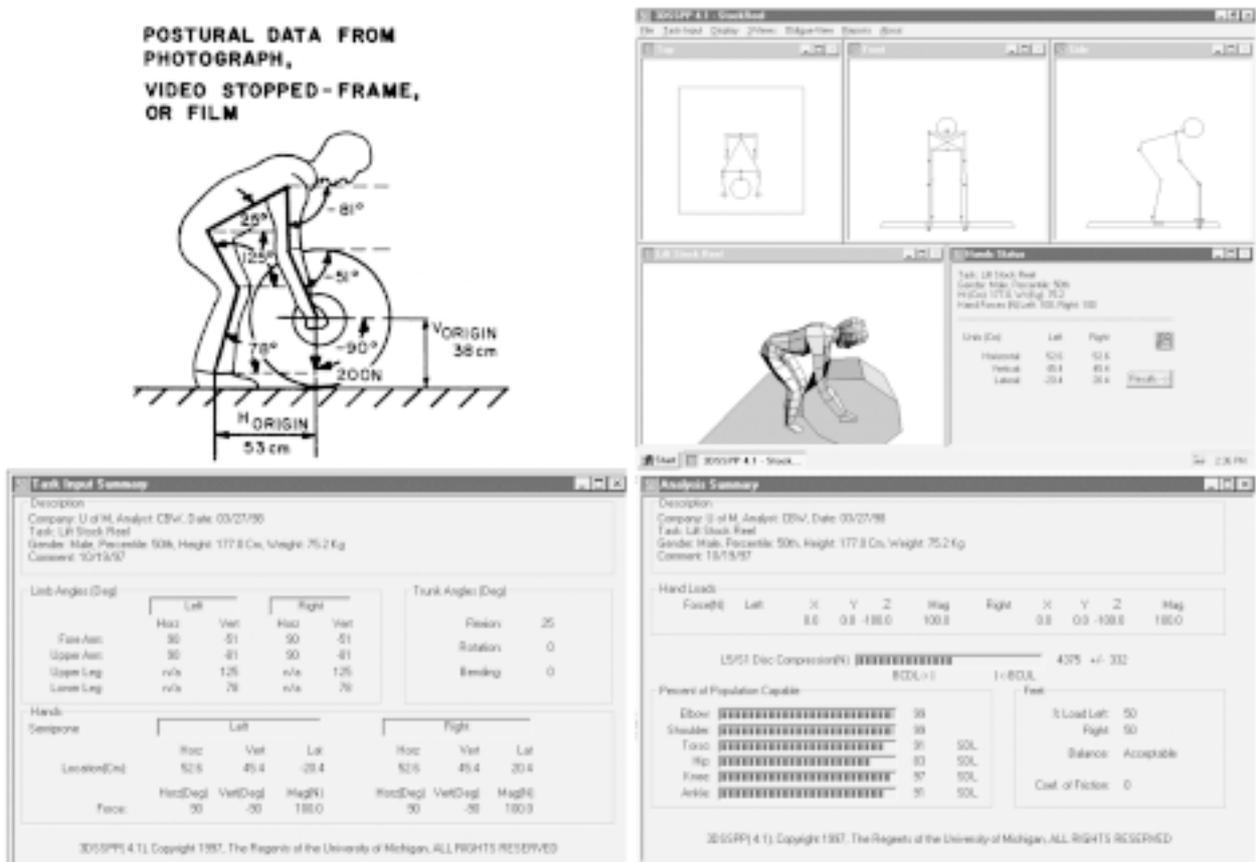
Figure 4. Work task and main output screens for the University of Michigan 3D Static Strength Prediction Program™ (reprinted with permission from the University of Michigan, 1999).

Standardization of language, computer file format and program structure is also clear needed in this technology. Undoubtedly, computer models will become more accurate as ergonomic tools advance, but in the interim, designers and ergonomics should attempt fully to understand the tools being employed in these models and use appropriate professional judgment.

9. RECOMMENDATIONS

- Digital human models are a useful tool for ergonomics analysis and will become increasingly more popular in the future.
- Models are currently available to assist designers in analyzing reaches and clearances, visual limitation, biomechanics, and a variety of other task functions.
- Limitations exist in our current anthropometric and strength databases, as well as, our current understanding of human posture and motion. Professional judgment should be exercised in interpreting model results that rely on this information.
- Additional information on human digital modeling can be obtained from the Society of Automotive Engineers (SAE),

Technical Subcommittee G-13 on Human Modeling Technology.

REFERENCES

BADLER, N.I, PHILLIPS, C.B. and WEBBER, B.L., 1993, *Simulating Humans: Computer Graphics Animation and Control* (New York: Oxford).

CHAFFIN, D.B., ANDERSSON, G.B.J. and MARTIN, B.J., 1999, *Occupational Biomechanics* (3rd ed.) (New York: Wiley).

GARG, A., CHAFFIN, D.B. and HERRIN, G., 1978, Prediction of metabolic rates for manual material handling jobs. *American Industrial Hygiene Association Journal*, 39, 661–74.

NIOSH, 1981, *Work Practices Guide for Manual Lifting*. NIOSH Technical Report No. 81-122 (Cincinnati: US Department of Health and Human Services, National Institute for Occupational Safety and Health).

NIOSH, 1994, *Applications Manual for the Revised NIOSH Lifting Equation*. NIOSH Technical Report No. 94-110 (Cincinnati: US Department of Health and Human Services, National Institute for Occupational Safety and Health).

SNOOK, S.H. and CIRIELLO, V.M., 1991, The design of manual handling tasks: revised tables of maximum acceptable weights and forces. *Ergonomics*, 34, 1197–1215.

Ecological Ergonomics: Theory in Context

N. Stanton
Department of Design, Brunel University, UK

1. INTRODUCTION

Ergonomics abounds with comparative studies that exhort the merits of one particular design over another, one particular environment over another, one particular training regime over another, and so on. While these are undoubtedly useful studies in their own right in that they answer questions to matters of current short-term concern, it is questionable whether they offer long-term advancements for ergonomics. Is the discipline to be chained to a cycle of testing?

Dowell and Long (1989) identified four deficiencies in the current status of the discipline, namely: poor integration, suspect efficacy, inefficient practices, and lack of systematic progress. These problems do not bode well for advancement. By way of softening the blow, Dowell and Long suggest that the trial and error approach, characterized by the conception of ergonomics as a craft, may be attributable to the relative youth of the discipline: its genesis being dated as 12 July 1949 (Oborne 1982).

The title "Ecological Ergonomics" might strike the reader as a little odd: surely all ergonomics studies are inherently ecological? This is not the case. Generalizing findings from one context to other contexts misses the essence of ecology: context has a direct bearing upon the phenomena being observed. In a different context, behavior may be very different. Intuitively we know this to be true, and this has led to calls for "user-centered design" and "field trials" in the development and evaluation of products.

2. THE ROLE OF THEORY

From the arguments raised by Dowell and Long (1989), it might be reasonable to suppose that the discipline would benefit from more theory building by its disciples. Certainly, this would tackle some of the deficiencies they have identified. The role of theory is to provide explanations that show causal linkages at work in the phenomenon we observe. In practice, theories are derived through the passage of time and considerable research effort. A life cycle of a theory might start with a conceptual framework. Conceptual schemes offer aid to theoretical advancement, for example, as a mechanism for classifying data and organizing ideas. Analyses of this information may lead to prototheories, which are typically loosely formulated, but offer claims regarding causal sequences. As data is gathered and organized, the theory can be modified reiteratively in the light of new evidence. The process of generating hypotheses based upon the theory, and testing these hypotheses and modifying the theory could go on *ad infinitum*. However, there are two points of exit. First, the evidence could become unequivocal, turning the theory into a law. Second, the theory could be overthrown by a paradigm shift within the discipline which makes the theory untenable. However, in most cases, theories tend to exist on balance scales: the weight of evidence presented either tips in favor of, or against, the argument.

3. MODELS OF HUMAN ACTION

In an attempt to predict human behavior with technological artifacts, theorists have tended to opt either for parsimony or for over-inclusiveness. An example of a parsimonious account is TOTE (Miller *et al.* 1960). Each TOTE unit contains four elements: Test–Operate–Test–Exit (see Figure 1). TOTE offers a functional, goal-oriented explanation of human behavior. For example, consider the goal for an operator of maintaining a target state for a process variable. To ensure that the variable is kept at its target level (e.g. level of a liquid in a vat) the operator is required to conduct a TEST (e.g. compare the value of the variable against its target). If the variable is above or below the target state, the operator might OPERATE a valve to return the variable to its target state. After operating the valve the operator would be required to conduct another TEST, to check that the desired state had been reached. If the variable had been returned to the target value, the operator could EXIT that task. The original conception of TOTE appears to suggest a fractal-like relationship between a multiplicity of units.
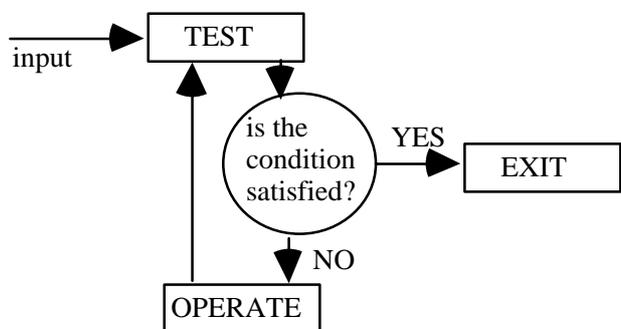
Figure 1.
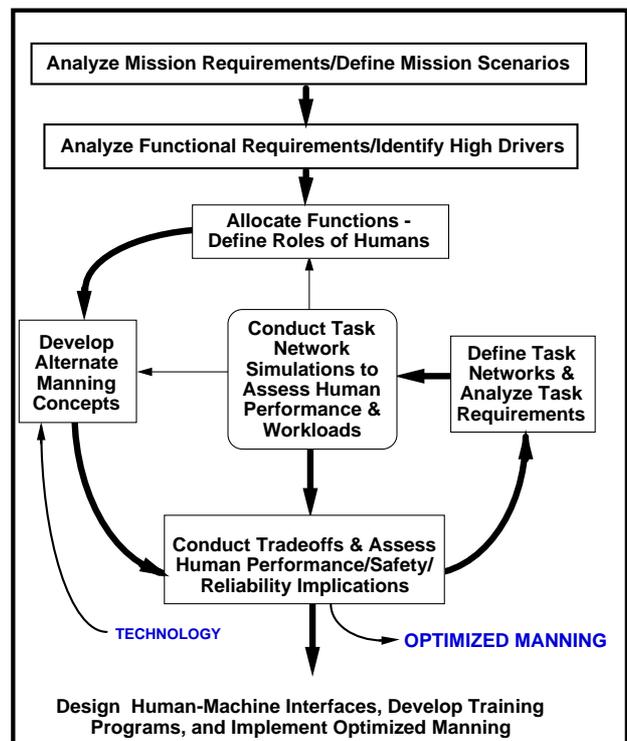
Figure 2.

An example of over-inclusiveness is the model of Performance Shaping Factors (PSFs: Miller and Swain 1987) where anything, and indeed everything, is thought to influence human behavior. PSFs include human–machine interface characteristics, task demands, task characteristics, instructions and procedures, stresses, environment, sociotechnical aspects, and individual factors. Literally hundreds of factors can be implicated as having an effect on human behavior.

In the comparison of these two approaches, the oversimplification of the first and the over-complexity of the second are striking. If we are to explain behavior, however, we must look to the simplest description that accounts for the phenomenon observed. Recent research on PSFs suggests that most of the variance in behavior can be accounted for by relatively few factors. In a study by Fujita (1993), 20% of the variation in operator performance was accounted for by training and experience. Similarly, in a study reported by Glendon *et al.* (1994), 30% of the variance was accounted for by work pressure. These data suggest that a small pool of factors could be identified to account for most of the influence on human behavior.

Carroll (1990) proposed the notion of a task-artifact cycle in order to explain the link between human activity, the design of artifacts, and the situation in which the artifacts were used. Carroll suggests that these elements are interrelated: the task sets the requirement for the artifacts and the artifact, in situations of use, defines the task. He argues that the artifacts cannot be fully understood outside the situations within which they are used. These propositions lead me to believe that context plays a central role in the task–artifact cycle, see Figure 2.

Norman (1988) illustrates a cycle of human action that makes implicit reference to the context in which the activities are performed. He describes a seven-stage model of action comprising: perception, interpretation, evaluation, goal formation, intention, sequencing actions, and execution of actions. Norman also draws the distinction between knowledge in the world and knowledge in the head. This is an important point; it suggests two types of context: external context and internal context.

4. THE IMPORTANCE OF CONTEXT

At present there appears to be a rather lively debate between cognitive scientists and ecological psychologists regarding the explanation of human action. Norman (1993) offers a caricature of these two approaches. The symbolic cognition stance taken by cognitive science focuses exclusively upon the symbolic representations of the mind and the processing structures of the brain, whereas the situated action stance taken by ecological psychology focuses almost exclusively upon the way in which structures of the world guide and constrain human behavior. At either extreme end of these viewpoints, something is missing from the explanation.

Human *action* occurs as an *event* within a *context*. By understanding the relationship between these three entities (i.e. the relationships between human action, device states, and context of the interaction), one might begin to propose a contextual theory of human action. In explaining behavior, we often point to the context in which certain decisions are made, but there are two types of context: internal and external. Internal context (cf. knowledge in the head) refers to the knowledge,

beliefs, experiences, and motivations of the individual concerned. External context (cf. knowledge in the world) refers to the situational, temporal, informational design and environmental characteristics present. To ignore either will result in an incomplete account.

5. AN ANALYSIS OF SITUATED ACTION

Gaver (1991) argues that information regarding the affordance of a device (i.e. the potential for action) is supplied not only by direct perception but also by exploration. Information about the operation of a device becomes apparent through interaction with that device. As the interaction progresses, new (previously unapparent) affordances are made available. Gaver calls these *sequential* affordances, when one action leads to a new affordance, and so on. For example, a door handle has the affordance of grabbing, the grabbing leads to the affordance of turning, and the turning leads to the affordance of pulling. Each action in turn leads to a new possibility that ultimately satisfies the goal of opening the door. Suchman (1987) refers to this type of activity as "situated" planning. Errors may occur when there is an incompatibility between the affordance offered by a device and its actual properties in two main ways. The first type of incompatibility occurs when a device does not allow an action that it appears to afford. The second type of incompatibility occurs when a device allows a certain action but it does not result in the desired goal.

In a recent paper, Baber and Stanton (1994) proposed a methodology for analyzing the situational component of human–artifact interaction. This approach suggested that by mapping a description of human activity onto a description of product states one could begin to identify situations where the product would allow erroneous activity. The methodology makes the relationship between product design, situational use, and errors explicit.

6. CONTEXTUAL ACTION THEORY

The value of an action performed, or a choice made, at one point in time often depends upon its relationship to the context within which it occurs. The coping model used in health psychology provides an analogous framework for a theory of contextual action. I am suggesting that it is necessary to consider how people cope with the operation of a device within a given context. This understanding will enable us to organize the existing research data to take the discipline beyond the last study to be published. Contextual Action Theory (CAT) would explain human action in terms of coping with technology within a context. To understand the theory, it is necessary to understand what enables people to cope, or fail to cope, with tasks.

In essence there are five main phases associated with contextual action. In the first phase, the actual demands and actual resources are presented to the actor. The actual demands are made up of the design of the device, the tasks to be performed on the device, environmental constraints (e.g. time), and so on. Factors relating to external context have impact here. Actual resources are represented by the training, knowledge, and experience of the actor with the device, and so on. Factors relating to internal context have impact here. The second phase consists of an appraisal of those demands and resources by the actor. It is noteworthy that actual demand and resources may differ from perceived demand and resources. A comparison of perceived

demand and perceived resources occurs at the third phase. It is proposed that an imbalance at this stage would lead to some form of degradation. Possible pathways that the degradation might take would occur in the fourth phase; these include emotional and behavioral responses or abandonment. Examples of emotional responses include decreases in user satisfaction and motivation. Examples of behavioral responses include increases in errors, increases in reaction time, increases in inefficient or inappropriate activities. The ultimate pathway would be for the user to abandon the interaction altogether. The effects of these responses on the interaction with the device are appraised in the fifth phase and fed back to the internal and external contexts.

It is possible to identify imbalance between actual (A) and perceived (P) demand (d) and actual (A) and perceived (P) resources (r). Hypothetically, imbalance would lead to some form of degradation in performance, as indicated by the use of pathways. Examples of the forms of imbalance are as follows:

Ad>Pd or Ad<Pd
Ar>Pr or Ar<Pr
Pd>Pr or Pd<Pr

In addition to the above framework, CAT incorporates the *coping hypothesis*: people cope with technology. A basic premise of the notion of comparing demands with resources is that, if demands cannot be met by existing resources, one way of compensating is to change the demands. Thus, a novice device user may interact in a different, less efficient manner than an expert. The novice reduces the demands by making the device less demanding. An obvious example is the use of icons and menus in preference to function keys and commands in computing tasks. If this pathway is not sought we may indeed observe some form of degradation in performance. One might propose that a well-designed interface helps novice users cope with a device, whereas experience helps expert users cope with a poorly designed interface.

REFERENCES

BABER, C. and STANTON, N. A., 1994, Task analysis for error identification: a methodology for designing error-tolerant consumer products, *Ergonomics* 37(11): 1923–41.

CARROLL, J. M., 1990, Infinite detail and emulation in an ontologically minimized HCI. In J.C. Chew and J. Whiteside (eds) *Proceedings of CHI '90*, (New York: ACM) 321–27.

DOWELL, J. and LONG, J., 1989, Towards a conception for an engineering discipline of human factors, *Ergonomics* 32(11): 1513–35.

FUJITA, J., 1993, What accounts for the operator performance? A simulator-based study of the influence of performance shaping factors on nuclear power plant operator performance. *Proceedings of Human Factors in Nuclear Safety* (London: IBC).

GAVER, W.W., 1991, Technological Affordances. *Proceedings of CHI '91* (New York: ACM).

GLENDON, A.I., STANTON, N.A. and HARRISON, D., 1994, Factor analysing a performance shaping concepts questionnaire. In S.A. Robertson (ed.) *Proceedings of the Ergonomics Society's Annual Conference* (London: Taylor & Francis) 340–45.

MILLER, D.W. and SWAIN, A.D., 1987, Human error and human reliability. In G. Salvendy (ed.) *Handbook of Human Factors* (New York: Wiley).

MILLER, G.A., GALANTER, E. and PRIBRAM, K.H., 1960, *Plans and the Structure of Behaviour* (London: Holt, Rinehart and Winston).

NORMAN, D.A., 1988, *The Psychology of Everyday Things* (New York: Basic Books).

NORMAN, D.A., 1993, Cognition in the head and in the world: an introduction to the special issue on situated action. *Cognitive Science* 17: 1–6.

OBORNE, D.J., 1982, *Ergonomics at Work* (Chichester: Wiley).

Ecological Interface Design: Applications

S. Chery and K. J. Vicente

Cognitive Engineering Laboratory, Department of Mechanical and Industrial Engineering, University of Toronto, 5 King's College Road, Toronto, Ontario Canada M5S 3G8

1. INTRODUCTION

Ecological interface design (EID) is a theoretical framework for human-computer interface design in complex work domains (see the article "Ecological Interface Design: Theory" by Torenvliet and Vicente) that is based on the ecological approach to human factors (see the article "An Ecological Approach to Human Factors" by Hajdukiewicz and Vicente). This article surveys the application domains that EID has been applied to so far. In addition, three particular applications of EID (conventional power plants, neonatal intensive care units, and hypertext information retrieval) are described in more detail. Given the insights obtained from the research conducted to date, the contributions of the EID framework are also summarized.

2. SURVEY OF APPLICATION DOMAINS

The EID framework was originally developed in the context of process control applications (Vicente and Rasmussen 1992). However, as the rest of this section will show, more recent research has considerably expanded the scope of the EID framework to a diverse set of application domains. Table 1 provides a representative list of references on EID, organized by application domain. Several conclusions can be derived from this brief survey. Most of the research is still in the area of process control. The most thoroughly investigated test bed is the process control microworld that has been the subject of a number of experiments conducted by Vicente and colleagues; see Vicente (1996) for a review.

There is some evidence of technology transfer from academe to industry. The most ambitious application has been implemented by Toshiba in Japan – an advanced control room for a future nuclear power plant (Itoh *et al.* 1993, 1995). This implementation is particularly noteworthy because it has been constructed on a very large scale; it is a fully functional, prototype control room connected to a full-scope plant simulator.

Also notable is the work of Pejtersen (1992), who designed an innovative full-scope document retrieval system that has been made available as a commercially available product and is currently being used in several public libraries in Scandinavia. It has been recognized by one international design award and one national design award, and it has been nominated for several others. These examples of technology transfer show that the EID framework is addressing problems relevant to the needs of industry.

The applications in table 1 span a relatively broad and diverse set of domains, including physical work domains that are primarily driven by causal constraints (e.g., process control) as well as social work domains that are primarily driven by intentional constraints (e.g., public library information retrieval). This diversity shows that the EID framework is fairly generalizable in terms of its applicability.

3. THREE EID APPLICATIONS

Section 2 was a broad survey of the work conducted on the EID framework. This section gives a deeper description of how EID can be applied in practice by describing examples from three different work domains (conventional power plants, neonatal intensive care units, and hypertext information retrieval). Each example begins with a brief description of the work then summarizes the abstraction hierarchy analysis, the interface design, and the empirical evaluation.

3.1. Ecological Display Design for a Conventional Power Plant

Computer displays are used for monitoring and control activities of large power plants with hundreds or even thousands of variables. Because of the need to navigate across many displays, operators may become disoriented and experience difficulty in remaining informationally coupled to the plant state and structure. Burns (1999) has investigated how an EID display may be implemented so as to support goal-directed problem solving while avoiding the problems typically associated with interface navigation in a large-scale system.

3.1.1 Abstraction hierarchy analysis

Burns (2000) applied the EID principles to a fossil-fuel power plant simulation obtained from industry. The simulation was of intermediate fidelity, containing 402 variables. An abstraction hierarchy representation was developed for the plant using five levels of constraint:

Table 1. Representative survey of EID applications

Application domain	Sectors	Sources
Process control	Power plants	Burns (2000); Dinadis and Vicente (1996); Hansen (1996); Itoh *et al.* (1993, 1995); Reising *et al.* (1998)
	Pasteurization	Reising and Sanderson (1996)
	Petrochemical	Jamieson and Vicente (1998)
	Thermohydraulic	Vicente (1996)
Medicine	Monitoring of cardiovascular system	Hajdukiewicz *et al.* (1998)
	Neonatal oxygenation processes	Sharp and Helmicki (1998)
	Hemodynamic monitoring	Effken *et al.* (1997)
Information retrieval	Clinical information system	Canfield and Petrucci (1993)
	General public library	Pejtersen (1992)
	Hypertext system	Xu *et al.* (2000)
Avaition	Engineering systems	Dinadis and Vicente (2000)
Software engineering	Intent specifications	Leveson (2000)

- *Functional purpose*: electrical energy production
- *Abstract function*: mass and energy flows
- *General processes*: volumetric flows and heat transfers
- *Physical function*: physical connection of the components of the plant
- *Physical form*: spatial arrangement and appearance of the plant components

3.1.2 Design

Display elements were designed for the top four levels of the abstraction hierarchy. The functional purpose view displayed the status of the plant relevant to its intended purpose. The abstract function view conveyed the mass and energy transformations throughout the plant. The generalized function view displayed the various flow rates and heat transfer rates. Finally, the physical function view showed the various components involved in the work domain operations, their settings and their physical connections. Three display suites were then constructed using the display elements just described. Each display suite contained all the display elements at the various levels of the abstraction hierarchy, but these elements were integrated in three different ways with varying implications for interface navigation.

In one display, only one window was visible at a time, allowing participants to view each level of abstraction serially. As a result, participants had to integrate information across displays by physically switching from one window to another. In a second display suite, four windows were visible at a time, allowing participants to view all levels of abstraction in parallel. However,

each level was displayed on a different window, so participants had to integrate information across displays mentally. Finally, in the third display suite, information at all four levels of abstraction was displayed at the same time and in the same location by superimposing display elements on top of each other. This configuration resulted in a seemingly very cluttered arrangement (figure 1).

3.1.3 Evaluation

An experiment was conducted to evaluate these three display suites with undergraduate engineering students as participants. Each participant performed search tasks, fault detection, and diagnostic tasks with one of the three display suite conditions. Surprisingly, the results revealed that performance was best in the seemingly cluttered, superimposed display. Despite slower fault detection times, participants in this highly integrated condition exhibited more accurate fault detection, as well as faster and more accurate diagnosis results. Thus, participants were able to overcome the visual complexity of the display and obtain a superior level of performance. These results have important practical implications for how to implement EID in large industry-scale work domains.

3.2. Ecological Display Design for a Neonatal Intensive Care Unit

Sharp and Helmicki (1998) applied the principles of EID to the development of an interface for the representation of infant tissue oxygenation processes in a neonatal intensive care unit. The
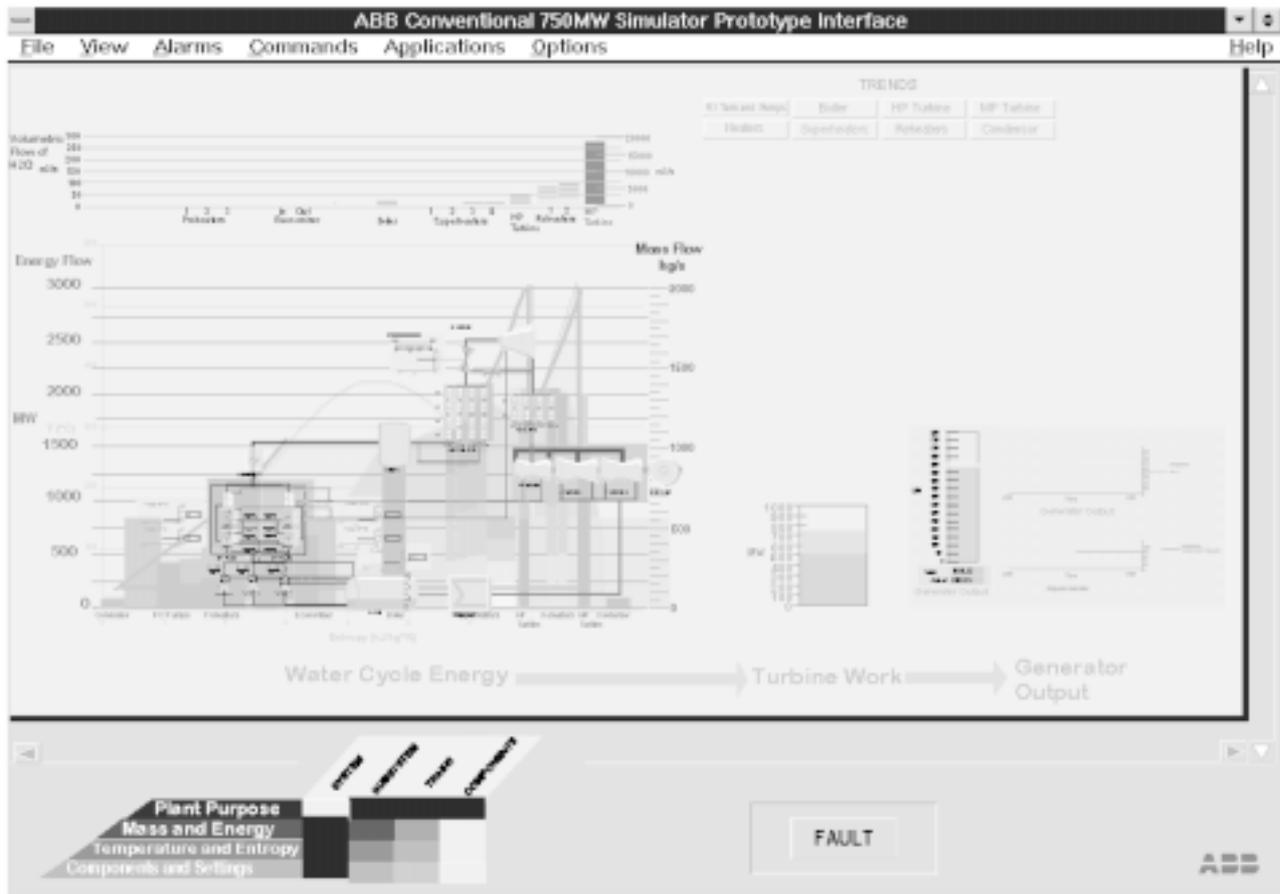
Figure 1. The highly integrated EID display designed and evaluated by Burns (1999).

purpose of their work was to assess the applicability of EID to medicine, a new domain with properties that are different from those of industrial process control.

3.2.1 Abstraction hierarchy analysis

Their first step was to develop an abstraction hierarchy representation of the physiological processes involved in tissue oxygenation. The levels of abstraction that had been developed for process control were not as relevant for medicine. As a result, Sharp and Helmicki (1998) identified a new set of levels of abstraction by identifying the nomenclature used by experts in this application domain. New domain-specific labels (purpose, balance, processes, transport/storage/control, and physical form) were used for the completion of an abstraction hierarchy by modeling tissue oxygenation at each level of abstraction.

3.2.2 Design

The second part of the project was to map the work domain constraints onto an interface. A significant obstacle was identified at this stage. Not all the variables that were identified by the abstraction hierarchy analysis could be sensed directly. As a result, these variables could not be directly mapped on the display, as the principles of EID require. In some cases it was possible to overcome this problem by deriving the required variables analytically or heuristically from sensed data; in others the required variables simply could not be displayed. The resulting EID display is shown in figure 2.

3.2.3 Evaluation

A comparative evaluation was performed using the EID display just described and the display currently used by physicians in this work domain. The accuracy of the diagnostic capabilities of

participants at three levels of experience (attendees, medical residents, and fellows) was assessed on both interfaces. The results showed that the attendees and residents performed better using the EID display. There was no significant difference between displays for the fellows, presumably because their greater level of experience allowed them to overcome the deficiencies of the existing display.

Sharp and Helmickis (1998) research is important because it shows that several significant obstacles may be encountered in applying the EID framework to new application domains. In this particular case, the obstacles were overcome by identifying a new set of domain-specific levels of abstraction and by analytically and heuristically deriving variables that could not be sensed directly. Despite the fact that some variables identified by the abstraction hierarchy analysis could not be displayed, the EID display produced a significant improvement in performance for the two less experienced groups of participants.

3.3. Ecological Interface Design for Hypertext Information Retrieval

Xu *et al*. (2000) applied EID to the domain of hypertext information retrieval. Like Sharp and Helmicki (1998), they too were interested in determining whether the EID framework could be applied to another domain besides process control. The domain chosen by Xu *et al*. was HyperErgo, a hypertext database consisting of ergonomic guidelines for computer workstation design. This domain differs substantially from process control in that it does not involve control of a dynamic system governed by physical laws. Instead it involves searching through a static database of information. Thus, the work of Xu *et al*. represents a strong challenge to the generalizability of EID.

The existing interface for HyperErgo served as the baseline
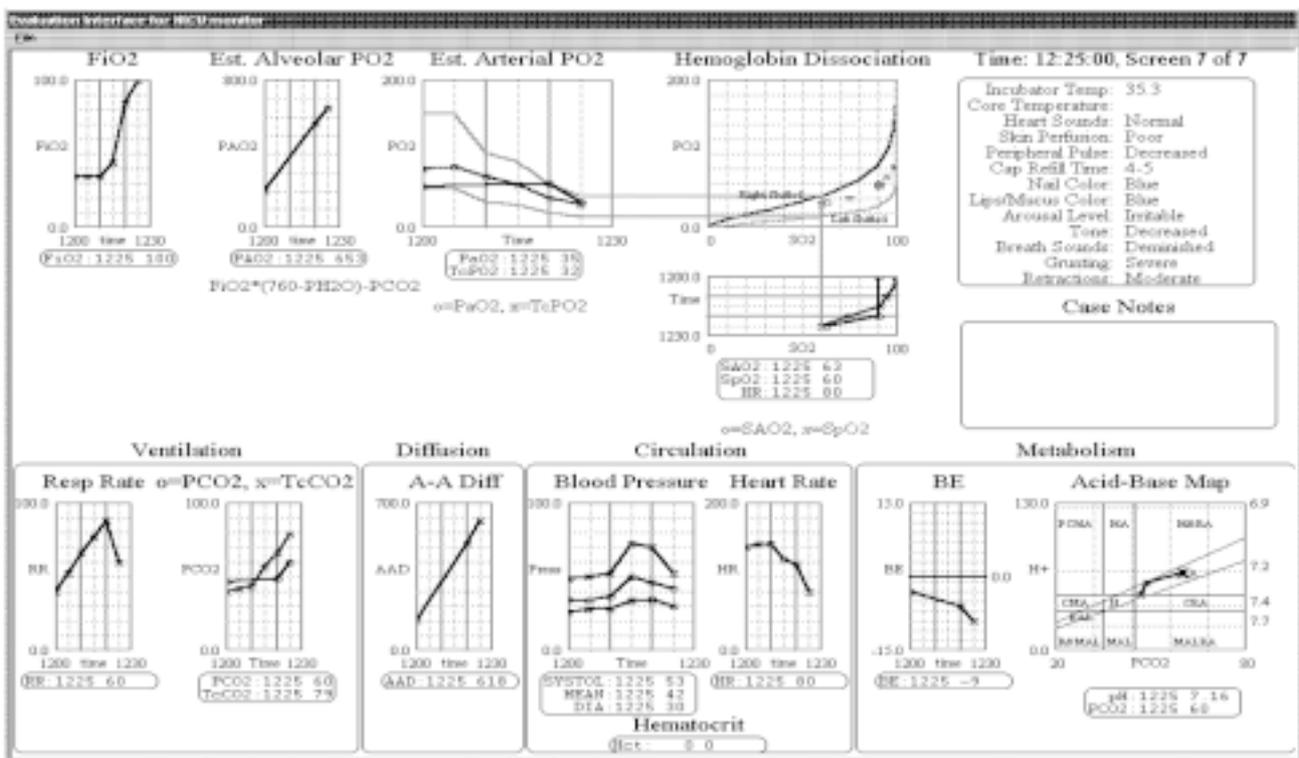
Figure 2. The EID display designed and evaluated by Sharp and Helmicki (1998).

control condition. That interface organized nodes in the hypertext database hierarchically according to a part-whole relationship that Xu *et al.* called a classification hierarchy. This type of relation is different from the means-ends relation defined by the EID-advocated abstraction hierarchy representation. Three levels of part-whole resolution were identified in this structure. Participants navigated in the structure by clicking on certain icons in the hypertext interface to search for information.

3.3.1 Abstraction hierarchy analysis

Xu *et al.* (2000) conducted an abstraction hierarchy analysis of the HyperErgo database. They identified five levels of abstraction with the following content:

- *Functional purpose*: to assure productivity, safety, health, and comfort
- *Abstract function*: first principles based on criteria of physiological efficiency, visibility, and reachability
- *Generalized function*: physical component interaction, physical component-body zone Interaction, and physical component-body area interaction
- *Physical function*: the components of the computer workstation
- *Physical form*: the appearance and spatial connectivity of those components

3.3.2 Design

An interface was then designed to convey information at four levels of abstraction (the functional purpose was deliberately omitted in the design). Note that the nodes in the database remained unchanged. The only difference was that they were connected according to the means-ends relations defined by the abstraction hierarchy analysis. In addition, the form in which information was presented was also kept constant. Figure 3 shows an example display from the redesigned interface for HyperErgo.

3.3.3 Evaluation

An evaluation was conducted to compare the classification hierarchy and abstraction hierarchy interfaces for HyperErgo. University students performed three types of search tasks (simple, complex, and problem solving) on both interfaces. A number of different performance measures were collected (search time, search accuracy, subjective ratings of preference, and disorientation measures). The results revealed no statistically
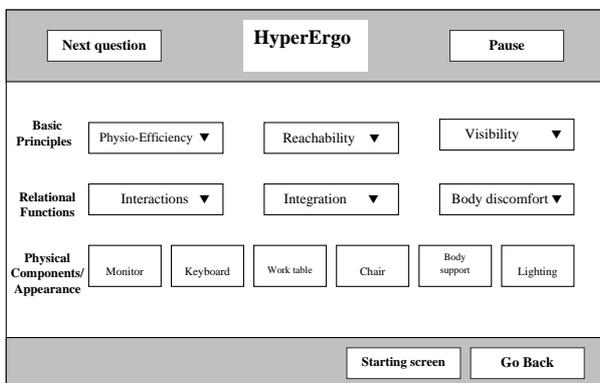
significant difference on the simple search task. However, as the complexity of the search task increased, there was a corresponding increase in the performance advantage of the EID interface (based on means-ends relations) over the classification hierarchy interface. For the complex and problem-solving tasks, a clear and statistically significant demarcation could be seen between conditions in favor of the EID interface. These empirical results show that the benefits of the EID framework generalize to domains whose characteristics are very different from those of process control or medicine.

4. CONTRIBUTIONS OF EID

Empirical evidence up to now indicates that an interface using EID can consistently outperform interfaces based on more traditional approaches. Perhaps just as importantly, this result has been obtained across a relatively diverse set of application domains, from physical work to social work. Thus, EID appears to represent a step forward in research on human-computer interface design for complex systems.

Despite these positive achievements, there are still comparatively few applications of EID to commercial products in industry. In part, this may be due to the fact that the EID framework is a relatively new approach to interface design. Regardless of the reason, much applied research still needs to be conducted to determine whether the benefits obtained in the research laboratory can also be obtained in operational settings. The empirical evidence gathered so far suggests this is a fruitful path to follow.

REFERENCES

BURNS, C.M., 2000, Putting it all together: improving display integration in ecological displays. *Human Factors*, forthcoming.

CANFIELD, K. and PETRUCCI, K., 1993, Interface design for clinical information systems: an ecological interface design approach. *Vienna Conference, Fin de Siecle: Human Computer Interaction* (Berlin: Springer-Verlag), pp. 391-402.

DINADIS, N. and VICENTE, K.J., 1996, Ecological interface design for a power plant feedwater subsystem. *IEEE Transactions on Nuclear Science*, 43, 266-77.

DINADIS, N. and VICENTE, K.J., 2000, Designing functional visualizations for aircraft system status displays. *International Journal of Aviation Psychology*, forthcoming.

EFFKEN, J.A., KIM, N.-G., and SHAW, R.E., 1997, Making the constraints visible: testing the ecological approach to interface design. *Ergonomics*, 40, 1-27.

HAJDUKIEWICZ, J.R., DOYLE, D.J., MILGRAM, P., VICENTE, K.J., and BURNS, C.M., 1998, A structured approach for patient monitoring in the operating room. *Proceedings of the Human Factors and Ergonomics Society 42nd Annual Meeting* (Santa Monica CA: Human Factors and Ergonomics Society), pp. 1038-42.

HANSEN, J.P., 1996, An experimental investigation of configural, digital, and temporal information on process displays. *Human Factors*, 37, 539-52.

ITOH, J., MONTA, K., SAKUMA, A., and MAKINO, M., 1993, An

Figure 3. One of the nodes in the hypertext EID display designed and evaluated by Xu *et al.* (2000).

ecological interface for intelligent man-machine systems for BWR plants. *Proceedings of the Topical Meeting on Nuclear Plant Instrumentation, Control and Man-Machine Interface Technologies* (LaGrange Park IL: ANS), pp. 97-103.

ITOH, J., SAKUMA, A., and MONTA, K., 1995, An ecological interface for supervisory control of BWR nuclear power plants. *Control Engineering Practice,* 3, 231-39.

JAMIESON, G.A. and VICENTE, K.J., 1998, Ecological interface design for petrochemical processing applications. Honeywell Technology Center Final Report CEL 98-04. University of Toronto, Cognitive Engineering Laboratory.

LEVESON, N.G., 2000, Intent specifications. *IEEE Transactions on Software Engineering*, forthcoming.

PEJTERSEN, A.M., 1992, The Book House: an icon based database system for fiction retrieval in public libraries. In Cronin (ed.) *The Marketing of Library and Information Services 2* (London: ASLIB), pp. 572-91.

REISING, D.V. and SANDERSON, P.M., 1996, Work domain analysis of a pasteurization plant: building an abstraction hierarchy representation. *Proceedings of the Human Factors and Ergonomics Society 40th Annual Meeting* (Santa Monica CA: Human Factors and Ergonomics Society), pp. 293-97

REISING, D.V.C., SANDERSON, P.M., JONES, B.G., MORAY, N., and RASMUSSEN, J., 1998, A direct perception display for rule-based behavior: supporting power plant startup with a "lattice" display. *Proceedings of the Human Factors and Ergonomics Society 42nd Annual Meeting* (Santa Monica CA: Human Factors and Ergonomics Society), pp. 224-28.

SHARP, T.D. and HELMICKI, A.J., 1998, The application of the ecological interface design approach to neonatal intensive care medicine. *Proceedings of the Human Factors and Ergonomics Society 42nd Annual Meeting* (Santa Monica CA: Human Factors and Ergonomics Society), pp. 350-54.

VICENTE, K.J., 1996, Improving dynamic decision making in complex systems through ecological interface design: a research overview. *System Dynamics Review*, 12, 251-79.

VICENTE, K.J. and RASMUSSEN, J., 1992, Ecological interface design: theoretical foundations. *IEEE Transactions on Systems, Man, and Cybernetics*, 22, 589-606.

XU, W., DAINOFF, M., and MARK, L., 2000, Facilitate complex search tasks in hypertext by externalizing functional properties of a work domain on the user interface. *International Journal of Human-Computer Interaction*, forthcoming.

Electromyography: Fundamentals

G. S. Rash[†] and P. M. Quesada[‡]

[†]BIO/MED Research & Consulting, Louisville, KY 40241, USA

[‡]University of Louisville Speed and Scientific School, Louisville, KY 40292, USA

1. INTRODUCTION

The study of muscle function through analysis of the electrical signals emanated during muscular contractions is termed electromyography (EMG). EMG is often misused and abused by many clinicians and researchers. Even experienced electromyographers fail to provide enough information and detail on the protocols, recording equipment and procedures used to allow other researchers consistently to replicate their studies. Hopefully, this chapter will clarify some of these problems and give the reader a basis for being able to conduct electromyography studies as part of their on-going research and practice.

2. WHAT ARE WE MEASURING?

Electromyography is measuring the electrical signal associated with the activation of the muscle. This may be voluntary or involuntary muscle contraction. The EMG activity of voluntary muscle contractions is related to tension. The functional unit of the muscle contraction is a motor unit, which is comprised of a single alpha motor neuron and all the fibers it enervates. This muscle fiber contracts when the action potentials (impulse) of the motor nerve which supplies it reaches a depolarization threshold. The depolarization generates an electromagnetic field and the potential is measured as a voltage. The depolarization, which spreads along the membrane of the muscle, is a muscle action potential. The motor unit action potential is the spatio- and temporal summation of the individual muscle action potentials for all the fibers of a single motor unit. Therefore, the EMG signal is the algebraic summation of the motor unit action potentials within the pick-up area of the electrode being used. The pick-up area of an electrode will almost always include more than one motor unit because muscle fibers of different motor units are intermingled throughout the entire muscle. Any portion of the muscle may contain fibers belonging to as many as 20–50 motor units.

2.1. Motor Unit

A single motor unit can have three to 2000 muscle fibers. Muscles controlling fine movements have smaller numbers of muscle fibers per motor units (usually less than 10 fibers per motor unit) than muscles controlling large gross movements (100–1000 fibers per motor unit). There is a hierarchy arrangement during a muscle contraction as motor units with fewer muscle fibers are typically recruited first, followed by the motor units with larger muscle fibers. The number of motor units per muscle is variable throughout the body.

2.2. Types of EMG

For the purpose of this chapter there are two main types of electromyography: clinical (sometimes called diagnostic EMG) and kinesiological. Diagnostic EMG, typically done by physiatrists and neurologists, are studies of the characteristics of the motor unit action potential for duration and amplitude. These are typically done to help diagnostic neuromuscular pathology. They also evaluate the spontaneous discharges of relaxed muscles and are able to isolate single motor unit activity. Kinesiological EMG is the type most found in the literature regarding movement analysis. This type of EMG studies the relationship of muscular function to movement of the body segments and evaluates timing of muscle activity with regard to the movements. Additionally, many of these studies attempt to examine the strength and force production of the muscles themselves.

3. RELATIONSHIP OF EMG TO BIOMECHANICAL FACTORS

There is a relationship of EMG to many biomechanical variables. With respect to isometric contractions, there is a positive relationship in the increase of tension within the muscle with regards to the amplitude of the EMG signal recorded. There is a lag time, however, as the EMG amplitude does not directly match the build-up of isometric tension. One must be careful when trying to estimate force production from the EMG signal, as there is questionable validity of the relationship of force to amplitude when many muscles are crossing the same joint, or when muscles cross multiple joints. When looking at muscle activity, with regards to concentric and eccentric contractions, one finds that eccentric contractions produce less muscle activity than concentric contraction when working against equal force. As the muscle fatigues, one sees a decreased tension despite constant or even larger amplitude of the muscle activity. There is a loss of the high-frequency component of the signal as one fatigues, which can be seen by a decrease in the median frequency of the muscle signal. During movement, there tends to be a relationship with EMG and velocity of the movement. There is an inverse relationship of strength production with concentric contractions and the speed of movement, while there is a positive relationship of strength production with eccentric contractions and the speed of movement. This means one can handle a larger load with eccentric contractions at higher speeds. For example, if a weight was very large and you lowered it to the ground in a fast but controlled manner, one handled a large weight at a high speed via eccentric contractions. One could not raise the weight (concentric contraction) at the speed it could be lowered. The forced production by the fibers are not necessarily any greater, but you were able to handle a larger amount of weight and the EMG activity of the muscles handling that weight would be smaller. Thus, one has an inverse relationship for concentric contractions and positive relationship for eccentric contractions with respect to speed of movement.

4. CHOOSING AN ELECTRODE AND AMPLIFIER

With regards to recording the EMG signal, the amplitude of the motor unit action potential depends on many factors which include: diameter of the muscle fiber, distance between active muscle fiber and the detection site (adipose tissue thickness), and filtering properties of the electrodes themselves. The objective is to obtain a signal free of noise (i.e. movement artifact, 60 Hz artifact, etc.). Therefore, the electrode type and amplifier characteristics play a crucial role in obtaining a noise-free signal.

4.1. Electrode

For kinesiological EMG there are two main types of electrodes: surface and fine-wire. The surface electrodes are also divided into two groups. The first is active electrodes, which have built-in amplifiers at the electrode site to improve the impedance (no gel is required for these and they decrease movement artifacts and increase the signal-to-noise ratio). The other is a passive electrode, which detects the EMG signal without a built-in amplifier, making it important to reduce all possible skin resistance as much as possible (requires conducting gels and extensive skin preparation). With passive electrodes, signal-to-noise ratio decreases and many movement artifacts are amplified along with the actual signal once amplification occurs. The advantages of surface electrodes are that there is minimal pain with application, they are more reproducible, are easy to apply and are very good for movement applications. The disadvantages of surface electrodes are that they have a large pick-up area and, therefore, have more potential for cross talk from adjacent muscles. Additionally, these electrodes can only be used for surface muscles.

Fine-wire electrodes require a needle for insertion into the belly of the muscle. The advantages of fine-wire electrodes are an increased band width, a more specific pick-up area, an ability to test deep muscles, isolation of specific muscle parts of large muscles, and an ability to test small muscles that would be impossible to detect with a surface electrode due to cross-talk. The disadvantages are that the needle insertion causes discomfort, the uncomfortableness can increase the tightness or spasticity in the muscles, cramping sometimes occurs, the electrodes are less repeatable as it is very difficult to place the needle/fine-wires in the same area of the muscle each time. One should stimulate the fine-wires to be able to determine their location, which increased the uncomfortableness of using this type of electrode. For certain muscles, fine-wires are the only possibility for obtaining their information.

Differences between the recording of surface and fine-wire electrodes, in part, are related to the differences in the bandwidths. Fine-wire electrodes have a higher frequency and can pick-up single motor unit activity as the fine-wire electrode band width ranges from ~2 to 1000 Hz. The surface electrode band width ranges from ~10 to 600 Hz.

4.2. Amplifier

Whether using surface or fine-wire electrodes, there are some electrode configurations that can also aide in decreasing unwanted noise. A monopolar arrangement is the easiest as it is a single electrode and a ground. However, this arrangement picks up more unwanted signals than any of the other potential configurations. Bipolar arrangements are widely used in movement analysis. In this arrangement, there are two active electrodes and a ground. The process is to look at what is common with the two active electrodes and to determine that this is noise and throw it away, keeping what is different in the two electrodes as the signal of interest. This is termed a differentially amplified system and is less prone to interference from adjacent and deeper muscles. A third arrangement is that of a double differentiated system. This is a system that has three active electrodes and one ground, therefore, possessing the ability to have two pairs of bipolar signals which are then again differentially amplified. This gives a smaller pick-up area, therefore, even less noise than the bipolar electrode

by itself. These electrode arrangements are unique to the amplified system purchased and much thought should be given when purchasing a system so that at minimum a bipolar system is acquired.

There are many other amplifier characteristics that should be noted. The first of which is the signal-to-noise ratio. This is the ratio of the wanted signal to the unwanted signal and is a measure of the quality of the amplified signal. The higher the ratio, the greater the noise reduction. Electrodes with on-site pre-amps (miniaturized and at the site of the electrode) are some of the best at providing a very large signal-to-noise ratio. The gain of the amplifier is also important. This is the amount of amplification applied to the signal and it should be sufficient enough to have output amplitude at 1.0 V. Another important characteristic of the amplifier is the bandwidth. This is simply the range of the collectable frequencies of the amplifier, and one wants this high enough to reject the low frequency movement artifacts and low enough to attenuate the signal as little as feasible. This means in general, one should be collecting in a range from 0 to 600 Hz for surface electrodes and 0 to 1000 Hz for fine-wire electrodes. Using the Nyquest Theorem, this means that one must sample at a minimum of 1200 Hz for surface electrodes and 2000 Hz for fine-wire electrodes to ensure capturing the entire signal. Once the signals have been recorded, then one could use a 10–15 Hz high-pass filter to eliminate the movement artifacts (some prefer to use an analog filter on the front end, but we prefer to filter movement artifacts after collection). One must make sure all applied filters have zero phase shifts. The measure of the ability of the differential amplifier to eliminate the common mode signal is termed the common mode rejection ratio. The higher the common mode rejection ratio, the better the cancellation of common signals (noise). A value of 10 000 (80 dB) is desirable. The input and impedance of the system should be $> 10^{12}$ ohms and have a low input bias current of the order of 50 pica-amps or less. A high input impedance allows for as much of the signal available for amplification to be amplified. Any signal input below the input bias current is not amplified. With these characteristics in mind, one should be able to purchase an amplifier that is sufficient for collecting electromyography signals.

There is also the potential for error introduced by the analog to digital board chosen. Most boards are only 10–12 bit boards and if the system does not allow full use of the collected range, one introduces error. This means if your collection is set-up for ± 10 V and you are collecting EMG, which is in the ± 1 V range after amplification, you are not optimizing your system and you will have quantization and sampling error. Therefore, one must be sure that the software and hardware arrangement purchased allows for optimization of the collected voltage range within the A-D range.

5. PLACING THE ELECTRODE

The electromyographer must have a very good understanding of the anatomy of the human body as electrode location and placement is very important. The first task is to clean the skin over the site where the electrode will be placed in order to reduce any skin resistance. This simple task can reduce the resistance of the skin by 200%. For many clinical applications of EMG, the belly of the muscle is used as a site for placing the electrodes. However, to assure repeatability of finding the specific site the

electrode was placed, the use of bony landmarks as a reference is a must. There are numerous books and publications describing the exact locations of placement of the electrodes for this purpose. Another widely accepted method of placing the electrodes or surface electrodes is to use the motor point. As with placing the electrodes over the belly of the muscle, there are numerous publications that give the general motor point locations as a starting location, then you can find the exact position by using the motor point finder or ohmmeter. Another specific issue that must be addressed is the interelectrode distance. Many electrodes have a constant interelectrode distance, but several have a variable interelectrode distance. This creates a potential problem in that one must be sure that this distance is kept constant throughout all subjects and trials to assure that the electrodes are over the same muscle fibers.

6. REDUCING NOISE

There are many sources of noise (any unwanted signal collected along side the wanted signal) and some of these sources are: electrostatic field (skin), electromagnetic fields (power lines), motion artifact due to loose electrodes at the skin interface or loose leads on the wires, involuntary reflex activity (clonus), and any other electrical devise that might be in the room when the studies are occurring. The majority of these artifacts can be removed from the system by a few simple means. Proper cleaning of the skin is one such measure. If site pre-amplified electrodes are not used, this becomes a more crucial task. Using bipolar differentially amplified or double differentially amplified systems also help dramatically in the removal of artifacts from the system. Attaching all loose electrode leads and making sure that there is some slack in these leads is important as well. If your system has the possibility to use the battery supply as opposed to line feed for the power source, this is a great advantage and should be utilized. Before beginning the collection of data, one should check that the electrodes is making proper contact and that there is no tension on the wires and that all of the wires are plugged into all connectors sufficiently. Once the electrodes are in position the subject should have manual muscle tests applied for the specific muscles being tested to make sure that the electrodes are picking up muscle activity appropriately. If certain electrodes seem to be working inappropriately, one can try switching the leads if possible with their system, or just switching electrode channels to see if this particular electrode works in another channel. If the signal is still bad after switching channels, one can switch electrodes to see if the electrode itself is malfunctioning. One must remember that there is a degradation of the signal as the amount of adipose tissue over the muscle being examined increases. Therefore, it may be difficult to pick-up any usable signal when dealing with obese individuals when using surface electrodes.

7. THE SIGNAL
7.1. Filtering the Signal

One disadvantage of using the newer computerized collection systems is that many do not afford one the ability to see a raw EMG signal in real time (like an oscilloscope). It is imperative that one somehow view the raw signal prior to any processing (except an analog anti-aliasing filter) as it is often difficult to differentiate between signal and noise in a raw EMG signal and usually impossible to differentiate if any processing has been done

to the EMG signal. Once the investigator looks at the raw signal, they need to determine if there is any filtering which needs to be done. The novice electromyographer may have trouble determining problems in the raw EMG signal. However, there are several items that can be looked at to help in this determination. Wavering base line is seen many times with low frequency movement artifacts. Large spikes can be indicative of abrupt movements of the electrode. Other things to look for are common signals across all channels and an underlying 60 Hz signal superimposed on the signal. If the signal does not look clean, the investigator may want to filter the data (some investigators say to always filter the data). There are three main types of filters applied to EMG data: high-pass, low-pass and notch filters. There are many types of filters that can be applied such as: Butterworth, Chebyshev, etc. These investigator's routinely use a fourth-order Butterworth high-pass digital filter with a 10–15 Hz cut-off, depending on the activity being analyzed (10 Hz for walking and 15 Hz for rapid movements) to remove movement artifacts. On the other end of the spectrum, we use an analog low-pass filter with a cut-off of 600 Hz for surface EMG and 1,000 Hz for fine-wire EMG as an anti-aliasing filter. If it was determined that 60 Hz signals were superimpose within the signals, we would use a digital notch filter to remove the signals within a 55–65 Hz range.

7.2. Processing the Signal

Now that we have a clean EMG signal, we can begin to look at the signal to gain information about the muscles. The primary information to be gained is on and off information. In most movement analysis situations, only the raw EMG is used. No processing other than that which is used for cleaning up the raw signal (high- and low-pass filters) is used. There are many common forms of processing that are done with EMG signals. The most common are: half-wave rectification (deletion of all negative aspects of the signal), full-wave rectification (absolute value of the entire signal), linear envelope (low-pass filtering of the full-wave rectified signal), root-mean-square (basically square the signal, take the mean of a timed determinant window ~100–200 ms, then take the square root), integrated EMG (area under the rectified curve can be determined for the entire activity or for preset time or amplitude values), and frequency analysis (typically determined via fast Fourier analysis and looking at the power density spectrum). Depending on your application, each of these processing techniques may have merit but each have disadvantages as well, since with any processing done to the data, information is lost.

8. NORMALIZING THE SIGNAL
8.1. Time Domain

For comparisons of EMG data from task to task or person to person, data needs to be presented in a common format. Thus, several means of normalization of the signal have been developed for both the time and amplitude domains. Probably the two most widely used time-base normalization techniques are to either normalize to a task/cycle or to phases within the task/cycle. As an example lets assume we want to look at the EMG of the back muscles with an individual who continually lifts items from the floor and places it in a bin. We can define a cycle as being from the initial movement of the object off the floor until the initial

movement of the object off the floor for the successive lift. One would then just simply divide the time-base by the total amount of time it took to perform the task and then all movements would be with respect to the percent of the cycle. This works well for many cyclic tasks, but has disadvantages if the task contains more than one phase. Dividing up the time-base to the percent of a phase works well for task with multiple phases. Using the same lifting task, lets now define the lifting phase as being from the point that the object begins to move from the floor until the subject obtains a fully erect standing position. The second phase would then be from the point at which the subject reached the standing position until the item is placed in the bin and a third phase would begin at the point when the object was placed in the bin until the subject is back in position to lift another object. Each one of these phases is handled as a separate event. Thus, the time it took to lift from the floor to the standing position would be used as the divisor to make a percent phase for the lifting phase, the time it took from the point when the body reached an erect standing position until the item was in the bin would be used as the divisor for the second phase, and so on for the third phase. This type of time-based standardization is very useful when the task has clear phases that can be determined. For the sake of this example, lets say that the maximum EMG activity occurred just prior to setting the item down in the bin. It is much more meaningful to be able to say that the maximum amount of the EMG was found at 95% of the second phase than to say the maximum EMG was found at 55% of the task. From this point you would have to go back and figure out what movement was going on at 55% of the task. Additionally, the intra and inter subject variability of setting the object in the bin at the same point in a multiphase cycle is typically large. For this reason, our lab prefers to use percents of phases when possible.

8.2. Amplitude Domain

Many times the amplitude of the signal is normalized as well. Probably the most widely used is to standardize to the maximum voluntary isometric contraction (MVIC) for the specific muscle being used. Based upon published references for manual muscle testing, the examiner then applies a force to the body part in sufficient magnitude that the subject is unable to maintain a static position while exerting against the examiner with a maximum muscle contraction. It is debatable if one can really ever obtain a true MVIC. Therefore, several other techniques have been devised. One of those is to use the maximum level of the signal across the entire task. In the lifting task previously described this would mean to take the maximum EMG level from each specific muscle during the entire task then normalize to this value. Many people prefer to use several peaks (4–5) and average these as the maximum so as to avoid the potential of using an erroneous high-spike as the maximum value. Another means of normalization is to use the mean level of the signal across the entire task. However, this is much less sensitive to any rapid peaks that were obtained during the task and would heavily skew the data if the majority of the signal contained times when the muscle was not active. A problem that exists when using the maximum or mean level across the entire task is that the EMG signal will vary based upon the velocity of the joints during the contractions. Therefore, unless one standardizes the velocity of the task, this method may not allow for comparisons across tasks. Another technique very

similar to the MVIC is to use a know level of force (e.g. divide by the amplitude of the EMG when lifting 20 lb at the specific velocity that the task will be performed). Another variation of this is to use the amplitude of the EMG signal when exerting a known force against an immovable object, therefore, eliminating velocity from the equation. All of these methods have positive and negative attributes and they are means of trying to compare amplitudes between muscles and individuals. Additionally, if the subjects being examined have any pathological conditions that involve the muscles you are testing (e.g. they have low back pain and you are measuring lifting), it will be virtually impossible to get a true MVIC and questionable whether the other normalization techniques are of any value as well. Regardless of the normalization technique used, whether it is time-based and/or amplitude based, one must remember that absolute information will be lost.

9. INTERPRETING THE SIGNAL

Now that we have cleaned up the data and completed any normalization that we may want to do, it is time to look at the signal and try to interpret its meaning. One must understand that there is a large variability of the EMG signal itself. Whether this is task-to-task variability within the same person or person-to-person variability within the same task, many combinations of muscle activity can produce the same movements because of the redundancy present in the neuromuscular system. EMG can be variable from task to task because of this normal redundancy, velocity or cadence changes, or slightly different movement patterns even though under observation they look the same. A normal range of EMG phasing will exist for a task but one must be very cautious of trying to define discreet points in the tasks where these patterns begin and end. This must be kept in mind when interpreting the EMG signals. Other factors enter into the equation with interpreting the EMG of individuals with pathological conditions that influence the task-to-task variability. The changes in velocity or cadence, the onset of fatigue, and the presence of pain can all effect the EMG patterns. Cross talk also makes interpretation of the signal difficult. Cross talk is interference of the EMG signals from adjacent muscles or deeper muscles that are within the pick-up area of the electrode. There are no fixed solutions available at this point and the size of the patient and size of the electrode lead does play a major role in the ability to decrease or increase cross talk. For example, if your system has electrodes with fixed active electrode distances which are large and you are working with a pediatric population you can be assured that your data will have large amounts of muscle information from adjacent and underlying muscles that is not wanted with your data. Many examiners utilize fine-wire electrodes in order to try to remedy this problem.

Now that we spent much time filtering and normalizing our data, it is time to discuss what the EMG signal can actually tell us. The muscle on and off timing patterns and relative increases and decreases in muscle activity are the two main parameters gained from the electromyography data. EMG data cannot tell us how strong the muscle is, if one muscle is stronger than another muscle, if the contraction is a concentric or eccentric contraction, or if the activity is under voluntary control by the individual. The strength of the muscle or determining one muscle to be stronger than another is one of the main areas that researchers

want to use EMG data for. The normalizing to the MVIC, the average, or max during the cycle are all attempts to allow us to compare from muscle to muscle within the same person and from muscle to muscle between individuals. This is routinely done but one must be cautious of the results due to the problems inherent with the collection techniques and variability among muscles, individuals, and tasks. Besides using EMG for determining the EMG patterns (times of activation and times of rest) many researchers use electromyography for evaluating the changes in the signals as the muscles fatigue. All of these are valuable uses of electromyography in occupational biomechanics.

With this overview of electromyography fundamentals in movement analysis, the reader should have a good idea of what the general process entails. By no means does this chapter give the investigator all the knowledge needed with becoming a proficient electromyographer. However, it gives a good overview and when augmented with other readings and material the reader should be able to utilize electromyography as a tool in their research and practice.

10. SUMMARY

10.1. Surface EMG

- Skin preparation
- Alcohol removal of dirt, oil, and dead skin.
- Shave excess hair if necessary. (Under ideal conditions this should always be done. However, it is not feasible in many cases.)
- If the skin is dry, some electrode gel rubbed into the skin can help.
- If the person is going to be sweating, spray an antiperspirant on the skin after cleaning with alcohol.
- Placement of electrodes
- There are specific references for different ways to measure for placement (Norris 1963, Johnson 1988, Goldgood 1984).
- General guidelines for large muscle groups:
- Best if over the largest mass of the muscle and align electrodes with muscle fibers.
- Use motor point and motor point finder to locate (general location charts are available).
- Cross talk
- Not a real problem with large muscle groups.
- Can sometimes be avoided my adjusting the electrode size, interelectrode distance (if an option on your brand of electrode), or by use of fine-wires.
- Application
- Skin placement.
- Avoid movement of electrodes by using straps or tape to firmly secure electrode in place.
- Avoid bending of leads, place leads pointing in the direction that you want the wire to continue in. (e.g. for electrodes placed on an extremity, have the lead pointing towards the proximal end of the extremity so that the wire will not have to be bent in order to go in the proximal direction).
- Avoid any stress on the wires by making sure that the wires are loose underneath the tape or wrap that is holding them in place. Be sure to check when the wires cross the joint that once the joint is fully extended the wires are not drawn taunt.
- Avoid placing electrodes over scars.
- Testing
- Do manual muscle tests to assure that you are getting a signal and that you are over the intended muscle.

- Do trial session to check signal and to get subject used to the setup and how instrumented.

10.2. Fine Wire EMG

- Indications
- For small muscles.
- For deep muscles not accessible by surface electrodes.
- To isolate specific muscles from a muscle group or adjacent muscles.
- Preparation
- Skin same as with surface electrodes.
- Electrode: fine-wires made as described by Basmajian and Stecko (1961) (wires typically connect to either the site pre-amplified electrode or fine-wire connectors and some type of ground is used).
- Can use topical agents: ethyl chloride, floromethane or EMLA patch or cream if desired (EMLA is a controlled substance in many countries).
- Assistant: needed to help in preparation, stabilize subject's extremity, or distract if needed. Having the subject blow out forcefully as the needle is inserted works well.
- It is usually easier to have the person in a laying position for insertion of the fine-wire electrodes.
- Inserter: needless to say, washing of hands and universal precautions are in order.
- Placement
- Use of a cross sectional anatomy book or EMG guide to locate insertion point and the direction of application is advised. We prefer Aldo Perotto, *Anatomical Guide for Electromyography: The limbs and Trunk*, 3rd ed. (Springfield: Charles C. Thomas).
- Can check placement of the fine-wires with electrical stimulation or my manual muscle testing if adjacent muscles would not be activated by the same movement needed for testing the selected muscle.
- Can use fine sandpaper on the wire ends to improve the quality of signal. (Only the end which is to be attached to the electrode, as the other end must be sterile.)
- After securing the wire to the electrode, tape the wire in place to the skin with small loop of wire to allow for movement.
- It is a good idea to recheck the placement of the wires via electrical stimulation or manual muscle testing if questions arise during the collection process.
- Removal of the wire
- To remove the wires, gently pull the wires out and check the ends of the wire to assure that no wire broke off inside the body.
- Special *Sharps* canisters are required for disposal of the needles. For disposal of the wire and gloves, you should check with your infection control or OSHA office to determine what needs to be done at your facility. At our facility, unless the gloves and wires have dripping blood, which could easily get on some one, they can be disposed of in a standard trash can. If the blood could drip off, therefore being a biohazard, then a biohazard container must be used.
- Processing of the fine wire electrodes — easiest means of obtaining fine-wire electrodes is directly from a vendor. We have used Nicolet Instrument Corporation (tel.: + 00 1 608 271 3333) for purchasing fine-wire electrodes. These

electrodes come in packs of 10 (using 27-gauge _ 33-mm needles) and can come in single or paired-hooked wires. One disadvantage to purchasing these needles already prepared is the limited number of size choices of needles. For example, in many individuals these needles are not long enough to reach the deep muscle being studied nor are the wires long enough once inserted to reach the electrode. In this case you can make your own electrodes as described by Basmajian and Stecko (1961) and then having them sterilized at your local hospital facility.

10.3. Data Collection for Surface and Fine Wire Electrodes

- It is preferred to do no processing on the original data during collection (except an analog anti-aliasing filter) as it is difficult or impossible to know if noise has corrupted your data.
- Collect surface EMG at a minimum of 1200 Hz.
- Collect fine-wire EMG at a minimum of 2000 Hz.
- A high-pass filter with a cut-off of 10–15 Hz works well for removing movement artifact.
- A front-end analog low-pass filter at 600 Hz for surface EMG and 1200 Hz for fine-wire EMG works well as an anti-aliasing filter.
- It is common to normalize the time-base to either a percent of the task involved or to percents of individual phases within the overall task.
- There are several means of normalizing the amplitude of the signal:
  - Percent MVIC.
  - Maximum level of signal across the task.
  - Mean level of signal across the task.
  - A known level of force.
  - Common types of processing are:
  - Half-wave rectification.
  - Full-wave rectification.
  - Linear envelope (also known as mean absolute value or mean rectified value).
  - Integrated EMG.
  - Frequency content analysis.

10.4. Additional Comments

When using EMG for research purposes:
- Care should be taken to use the appropriate units when reporting EMG data.

- When referring to the amplifier gain, the units should be in a ratio or dB.
- The input impedance should be expressed in ohms.
- The common mode rejection ratio should be as a ratio or in dB.
- When referring to filter bandwidths cut-off, the units should be in Hz.
- EMG, when a raw, average or rectified signal, should be referred to in millivolts.
- If the EMG has been integrated, then it should be expressed in terms of millivolt seconds with the specific period of analysis. If the integrated EMG was time reset or voltage reset, then the specific time or voltage should be indicated. Otherwise, individuals would not be able to reproduce the study that you have conducted.

REFERENCES

BASMAJIAN, J.V. and DELUCA, C.J., 1986, *Muscles Alive: Their Functions Revealed by Electromyography* (5th ed.) (Baltimore: Williams & Wilkins).

BASMAJIAN, J.V. and STECKO, G.A., 1961, A new bipolar indwelling electrode for electromyography. *Journal of Applied Physiology*, 17, 849.

GOLDGOOD, J., 1984, *Anatomical Correlates for Clinical Electromyography* (2nd ed.) (Baltimore: Williams & Wilkins).

JOHNSON, E.W., 1988, *Practical Electromyography* (2nd ed.) (Baltimore: Williams & Wilkins).

KIMURA, J., 1981, *Electrodiagnosis in Disease of Nerve and Muscle: Principles and Practice* (2nd ed.) (Philadelphia: F. Davis).

KOH, T.J. and GRABINER, M.D., 1993, Evaluation of methods to minimize crosstalk in surface EMG. *Journal of Biomechanics*, 26 (suppl.), 151–7.

NILSSON, J., PANIZZA, M. and HALLET, M., 1993, Principles of digital sampling of a physiologic signal. *EEG and Clinical Neurophysiology*, 89, 349–58.

NORRIS, F.H., 1963, *The EMG: A Guide and Atlas for Practical EMG* (New York: Gruve & Stratton).

WINTER, D.A., 1988, *The Biomechanics and Motor Control of Human Gait* (Waterloo: University of Waterloo Press).

WINTER, D.A., RAU, G., KADEFORS, R., BROMAN, H. and DELUCA, C.J., 1980, *Units, Terms, and Standards in the Reporting of EMG Research: A Report by the AdHoc Committee of the International Society of Electrophysiology and Kinesiology*.

Electromyography: Methods and Techniques

H. Strasser

Ergonomics Division/University of Siegen, Paul-Bonatz-Str. 9–11, 57068 Siegen, Germany

1. INTRODUCTION

Electromyography means the graphic representation of electric potential changes in the musculature (*myo* = muscle), i.e. biopotentials which can be picked up on the surface of the skin over a muscle, most favorably in a bipolar electrode configuration. These biopotentials can, in general, be interpreted as the spatial and timely summation of action potentials (e.g. Zipp 1982, Strasser *et al.* 1992). A burst of positive and negative peaks in the electromyogram (EMG) includes information in the frequency range between just a few Hz and ~1.5 kHz. However, spectrum analysis of raw EMG shows that the main energy is in an area under ~70 Hz. For details, see Kumar and Mital (1996). Fatigue of a muscle does lead to an increase of the lower frequencies and to a decrease of the higher frequencies in the spectrum. Changes in the frequency domain of the electromyogram are, however, not pronounced enough to reliably correlate them with definable degrees of fatigue of a muscle. Therefore, for occupational health and ergonomics purposes, usually amplitude and not frequency measures of the bioelectric activities are of interest. A fundamental understanding of the electromyographic activity patterns is provided by the work of Laurig (1970) or Tichauer (1978).

2. ELECTROMYOGRAPHIC ACTIVITY (EA) AS AN AMPLITUDE MEASURE

EA as the envelope of the rectified and low-pass filtered raw EMG (Figure 1) correlates with the exerted muscle force under static conditions.

Fatigue leads to an increase in excess of the corresponding characteristic line of the EA force course. EA can be used as a parameter for the "physiological cost" of manual work, i.e. for the quantification of the muscle activation which the human organism must exert to carry out a task under given working conditions. The "physiological cost" resulting from more or less favorably designed working tools which are operated by hand or also from more or less favorable manipulative work during motion-technical work design varies distinctly (Strasser and Ernst 1992, Strasser and Müller 1999).

3. STANDARDIZATION AND SPLITTING-UP OF EA TIME SERIES ASSOCIATED WITH REPETITIVE MANUAL MOVEMENTS INTO STATIC AND DYNAMIC COMPONENTS OF MUSCLE STRAIN

A careful processing of the raw EMG and calculation of standardized (normalized) electromyographic activity (sEA) is necessary for the evaluation of electromyographic time series (see tracings in the lower left part of Figure 2). To correlate EA time series with repetitive movements, smoothing procedures are first required. This can be done using computer software by taking the sliding average of actual digitalized values and immediately preceding and succeeding values. A sampling rate of, e.g., 16.66 Hz (1000 values/min in the analog–digital converter) suffices. Considerably more interpretable information than in an arithmetic mean can be found in the maxima and minima of EA time series, which then become much clearer and are the correlate of cyclical activation of a muscle. Minima represent the activity in the relaxation phase; maxima represent the contraction of a muscle. The mean difference between maxima and minima can therefore be attributed to the movements. It represents the dynamic part of the muscle strain. The mean minima are associated with the quasi-static base activation. To avoid artifacts, the extreme values should be determined by first averaging over a single cycle and then averaging over several cycles; this results in a physiological representation of repetitive movements. For example, if the extremes are determined from a range of 5%, then the thus calculated maxima and minima can be interpreted as the 95th and 5th percentiles (Hagberg and Sundelin 1986). An averaging over a longer period results in reliable extreme values. For the electromyographic evaluation of static work the above-described procedures of splitting-up EA into static and dynamic components, of course, are not to be applied.

EA values result as an amplitude measure from many uncontrollable influences (varying distances between the two active electrodes, the kind of electrode gel and the time needed
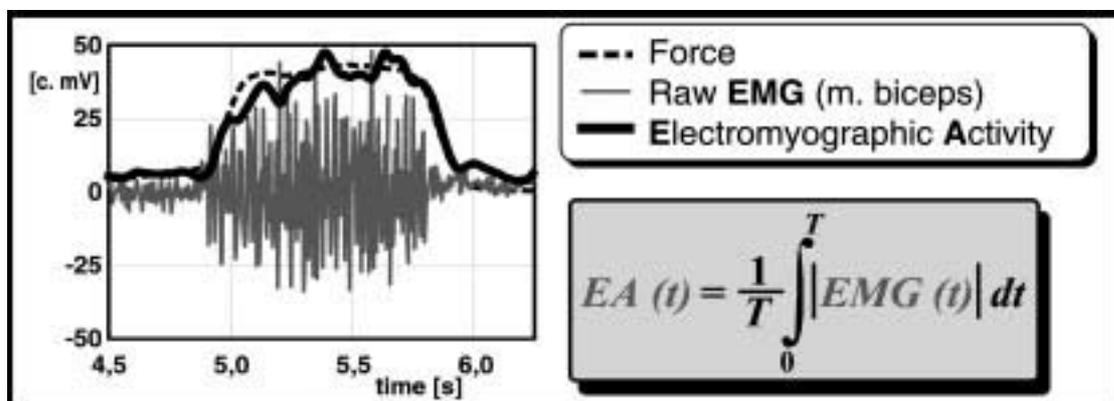
Figure 1. Burst of myoelectric signals (raw electromyogram) associated with the activation of the biceps, exerted force and electromyographic activity EA(t).

$$EA\,(t) = \frac{1}{T} \int_{0}^{T} \left| EMG\,(t) \right| dt$$
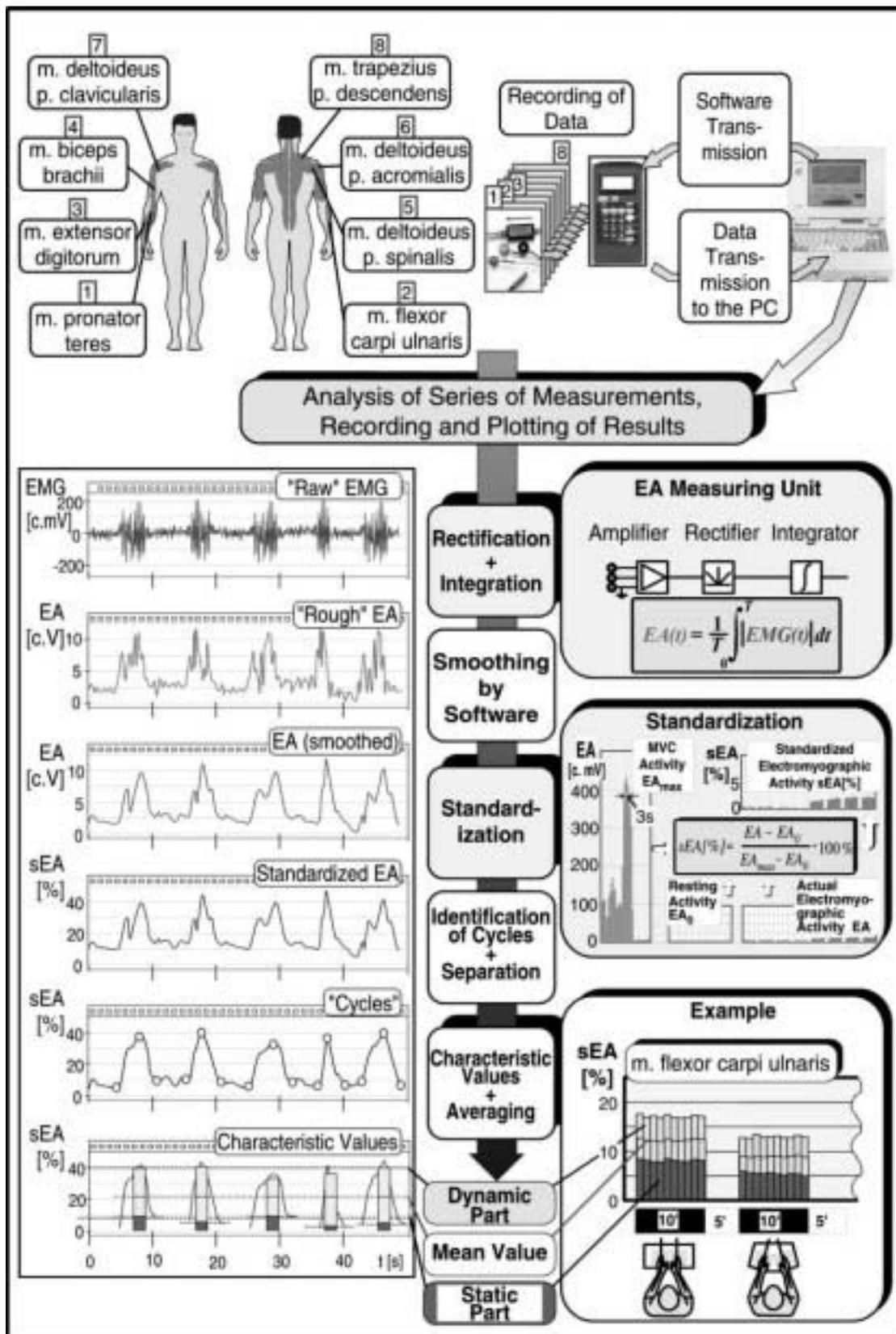
Figure 2. Outline of procedures for the evaluation of the electromyographic time series, starting with picking up myoelectric signals from muscles involved in work, recording time series of electromyographic activity EA(t) by a portable data recorder, and standardization of the A-D-converted EA.

for it to take effect, the impedance of the amplifier, the thickness of the subcutaneous skin layer over a muscle, etc.). The amplitude measures of the electromyographic data are therefore not directly interpretable. They vary greatly in repeat measurements under identical conditions even with the same test subject and even more when the measurements result from different tests after the electrodes have been reapplied even under the same operating conditions. Intermuscular comparisons are also not possible using electromyographic activity. Therefore a standardization of the EA values is necessary.

There are several possibilities for the standardization of EA data (Mathiassen *et al.* 1995, *passim*). For example, the measured values of a test series can be related to the globally highest or lowest measured value. These maxima or minima as reference values must, however, be the same for each test subject to ensure comparability of the data between the test subjects. Further, all individual EA values from predetermined test sections, e.g. 1 or 5 min mean values, can be related to the average value of a complete test series with several repeat measurements and several parameter variations. Depending on the amount of effort, these attempts at standardization allow at least interindividual comparisons of the measured data; however, they do not make intermuscular comparisons possible. In other words, the data are at most interpretable as stress data; they are not, however, interpretable as characteristic values of strain, i.e. they cannot be used to evaluate the degree to which a physical capacity is occupied by an imposed workload.

4. INTERPRETATION OF STANDARDIZED EA VALUES AS STRAIN PARAMETERS

Strain differs from stress in that stress (in this case the muscle force required) causes an individually differing degree of strain, depending on the muscle's capacity. If EA data is to be interpreted

as characteristic values of strain, then the actual EA values measured in certain stress situations must be related to the highest possible EA of a muscle resulting from maximum voluntary contraction (MVC). Even without external force demand, a muscle produces biopotentials that can be measured on the skin's surface as EA_0, which may differ from muscle to muscle. Thus, similar to the determination of the working heart rates as indicators for the total (whole body) strain during dynamic muscle work, it is advisable to "calculate" the resting value 'out' of the actual EA values and use the differences $EA_{max} — EA_0$ and $EA_{actual} — EA_0$ for the standardization. EA_0 — which are determined by the quality of the electrical amplifiers — should only be a few percent of the maximum activity EA_{max}. The resulting standardization approach that allows the interpretation of EA as strain data is shown in Figure 2. The respective maximum activity EA_{max} must be determined for each individual muscle for such standardizations.

5. NECESSITY OF MULTICHANNEL ELECTROMYOGRAPHY

Manipulating objects in a three-dimensional space requires, e.g., forces for the vertical lifting of an object against gravity, the moving of the object toward or away from the body (horizontally), and possibly inward or outward rotation forces of the lower arm during pronation and supination. Therefore, the specific maxima for the muscles involved in such activities must be determined individually for correct standardization. Thus, the muscle groups involved in working must be carefully selected already before the beginning of an electromyographic study. The same is true for the determination of the data for the standardization process. As an example Figure 3 shows EA of eight muscle groups of the hand–arm–shoulder system responsible for elementary arm movements during manual material handling. It becomes clear
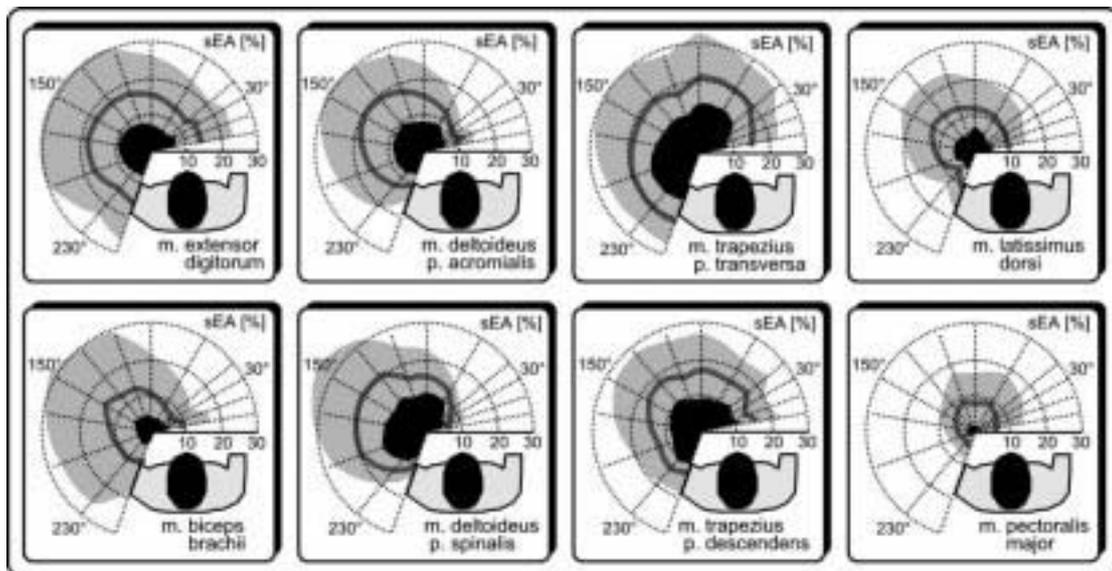
Figure 3. Static and dynamic components of the electromyographic activity (black and gray areas respectively of the circle diagrams; bold line indicates the mean EA) of eight muscle groups acting on the forearm, upper arm, shoulder and upper parts of the torso during manual materials handling in different directions between 10 and 250° (means from 11 Ss during manipulation of an external load of 1 kg over a distance of 38 cm at a working speed of 24 cycles min⁻¹) (from Strasser and Müller 1999).

that multichannel electromyographic registrations are necessary even for simple tasks.

The measuring of one or another muscle may lead to clear results and their interpretation may seem plausible. It is usually, however, inadequate for the complexity of the interactive "muscle play" of agonists and antagonists as well as synergists. It is usually not possible correctly to represent this via just a few electromyographic derivatives, i.e. singular electromyographic studies are often insufficient or invalid.

6. EXPERIENCE WITH ELECTROMYOGRAPHY AS A METHOD TO ASSESS THE ERGONOMIC QUALITY OF WORKPLACES AND WORKING TOOLS

In the past years a great deal of experience has been gained with computer-aided electromyography. Advanced methods as described above enable the measurement of the intensity of muscle exertions that are demanded when working with hand-held tools and controls. The same is true when repetitive manual movements during material-handling tasks have to be performed. Utilizing multichannel recording devices, physiological responses of those muscles involved in work can be quantified in figures and numbers, whereby more or less ergonomically designed tools, like masons' trowels, file handles, screwdrivers, wrist rests, and keyboards, lead to different physiological costs associated with work (e.g. Kluth *et al.* 1997, Strasser *et al.* 1996, 1998). So, after the pioneering feat of Tichauer (1978) electromyography meanwhile became a powerful ergonomic method all over the world (e.g. Eklund and Freivalds 1993, Kilbom *et al.* 1993, Marras 1990).

REFERENCES

EKLUND, J. and FREIVALDS, A. (eds), 1993, Special issue on 'Hand Tools for the 1990s', *Applied Ergonomics*, 24.

HAGBERG, M. and SUNDELIN, G., 1986, Discomfort and load on the upper trapezius muscle when operating a wordprocessor, *Ergonomics*, 29, 1637–45.

KILBOM, A., MÄKÄRÄINEN, M., SPERLING, L., KADEFORS, R. and LIEDBERG, L., 1993, Tool design, user characteristics and performance: a case study on plate-shears, *Applied Ergonomics*, 24, 221–30.

KLUTH, K., KELLERMANN, H.G. and STRASSER, H., 1987, Electromyographic and subjective methods for the assessment of the ergonomic quality of file handles. In P. SEPPÄLÄ, T. LUOPAJÄRVI, C.-H. NYGARD and M. MATTILA (eds), *From Experience to Innovation*, *Conference Proceedings IEA '97* (Helsinki: Finish Institute of Occupational Health), vol. 4, 515–17.

KUMAR, S. and MITAL, A. (eds), 1996, *Electromyography in Ergonomics* (London: Taylor & Francis).

LAURIG, W., 1970, *Elektromyographie als arbeitswissenschaftliche Untersuchungsmethode zur Beurteilung von statischer Muskelarbeit* (Berlin: Beuth).

MARRAS, W.S., 1990, Guidelines industrial electromyography (EMG), *International Journal of Industrial Ergonomics*, 6, 89–93.

MATHIASSEN, S.E., WINKEL, J. and HÄGG, G.M., 1995, Normalization of surface EMG amplitude from the upper trapezius muscle in ergonomic studies — a review, *Journal of Electromyography*, *Kinesiology*, 5, 197–226.

STRASSER, H. and ERNST, J., 1992, Physiological cost of horizontal materials handling while seated, *International Journal of Industrial Ergonomics*, 9, 303–13.

STRASSER, H., ERNST, J. and MÜLLER, K.-W., 1992, *Elektromyographische Untersuchungen des Hand–Arm Systems — Günstige Bewegungen für die ergonomische Arbeitsgestaltung. Schriftenreihe 'Arbeitsmedizin — Arbeitsschutz — Prophylaxe und Ergonomie'*, Band 11, C (Heidelberg: Haefner).

STRASSER, H. and MÜLLER, K.-W., 1999, Favourable movements of the hand–arm system in the horizontal plane assessed by electromyographic investigations and subjective rating, *International Journal of Industrial Ergonomics*, 23, 339–47.

STRASSER, H. and WANG, B., 1998, Screwdriver torque strength and physiological cost of muscles dependent on hand preference and direction of rotation, *Occupational Ergonomics*, 1, 13–22.

STRASSER, H., WANG, B. and HOFFMANN, A., 1996, Electromyographic and subjective evaluation of hand tools: the example of masons' trowels, *International Journal of Industrial Ergonomics*, 18, 91–106.

TICHAUER, E.R., 1978, *The Biomechanical Basis of Ergonomics: Anatomy Applied to the Design of Work Situations* (New York: Wiley).

ZIPP, P., 1982, Recommendations for the standardization of lead positions in surface electromyography, *European Journal of Applied Physiology*, 50, 41–54.

Ergograms

V. F. Venda†

†University of Manitoba, 75 Allendale Drive, Winnipeg R3T 5B6, Canada

1. INTRODUCTION

Theoretical and applied ergonomics studies need highly developed models of dynamic work processes and of interactions in complex human–machine–environment systems (Karwowski 1991). Verbal descriptions of human work provide not much assistance to ergonomists in prediction of work efficiency in different work environment, during training processes or when different work skills and technologies are being implemented (Karwowski and Mital 1986).

To study dynamics of work ergonomics and particularly its work dynamics oriented part called ergodynamics (Venda and Venda 1991) need special models, or ergograms. Ergograms are graphic models presenting work efficiency, Q, as functions of environment parameters, E, work factors, F, work strategies (skills, methods, technologies), S_i, and time, T.

A *single ergogram* has only two axes, F and Q. It shows work efficiency static characteristic curves, $Q_i(F)$. A double ergogram presents side by side work statics, $Q_i(F)$, and work dynamics, $Q_i(T)$. Usually ergonomist shows at the left side the characteristic curves of the work strategies (functional structures), $Q_i(F)$, and dynamics of work efficiency in time, $Q_i(T)$ at the right side of the double ergogram. The next type of ergograms is a triple ergogram, or a triogram. It has three axes, F, Q and T, and three quadrants presenting three interconnected diagrams: $F(T)$, $Q_i(F)$ and $Q_i(T)$.

The most advanced ergogram has four different axes, parameters (F, Q, E, T) and quadrants. We call this ergogram the quadrigram.

2. SINGLE AND DOUBLE ERGOGRAMS

We start description of all ergograms with the quadrigram. This the most advanced and complicated type of ergograms includes the single, double and triple ergograms as its parts which will be explained in a due course.

The quadrigram ("quadri" in Greek means four), as the name suggests, is ergogram which has four quadrants and four different axes; Q, T, F and E (Figure 1).

The T-axis measures the time of interaction; or simply the time allotted for a task or series of tasks. The Q-axis measures the work *output parameter*. This is the efficiency (safety, productivity, quality) of human work. This axis is usually normalized against some desired or recognized level of efficiency. The E-axis measures the *environmental parameter, which is an*
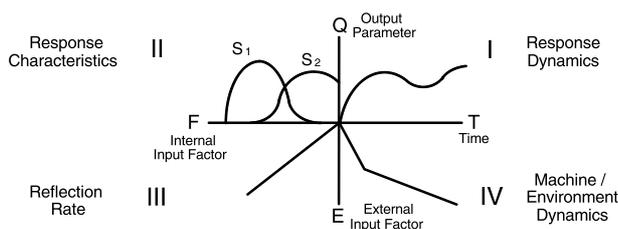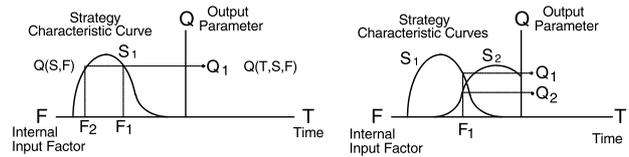
Figure 2. Characteristic curves for a single work strategy S_1 (a) and for two different work strategies S_1 and S_2 (b). Work efficiency is a function of S and F in statics (a) and of S, F and T in dynamics (b). Two different factor F, F_1 and F_2, may lead to the same efficiency (a); one F may lead to different efficiencies (b).

external input factor for the human. This factor is an objective measure of the task, such as size of the information input as a number of elements on the computer screen presenting the task.

The F-axis measures the *internal input factor, or work factor*. It is connected to the external input factor but presents internal, psychological reflection of the external parameter E. The E- and F-axes are related. They use the same units of measurement, but depending on the subject's psychological and physiological state (whether stressed or relaxed, for example) the value of the internal input factor may differ very significantly from that of the external input factor. We call F a work factor because its value is close correlated with value of work efficiency, Q. Correlation between E and Q may be very low.

Each quadrant is an individual plot showing the relationship between the two parameters on its axes. Each pair of adjacent quadrants shares an axis, thus they are functionally, and visually, related. The overall result is a diagram where the four parameters (on the axes) are displayed in four different graphs, and one can observe the relationships between them, both direct and indirect.

The transformation dynamics theory and ergodynamics (Venda and Venda, 1995) are the fundamental basis of the quadrigrams and other types of ergograms.

Figure 2 shows the top half of the quadrigram. It is the double ergogram, or a *transformation ergogram*. The left side of the upper diagram shows a strategy characteristic curve (S_1), which describes the relationship between the output parameter (work efficiency) and the internal input factor, F, for a given functional structure (or cognitive strategy). A functional structure (strategy) is the human's method or algorithm used for completing the task. In accordance with the first law of ergodynamics, $Q(F)$ is a bell-shaped curve if the full range of F was studied.

Figure 3 shows the complementary situation, where we have two different outputs for a single input. Here we have two different strategy curves (S_1 and S_2). These represent two strategies that one could use to complete a task. Given an input factor, the output
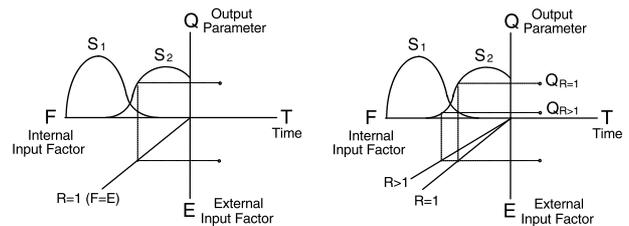
Figure 3. Different types of reflection: (a) a single normo-reflection; (b) two different reflection vectors.

Figure 1. Components of the quadrigram.

parameter may be very different, depending on which strategy is used.

Consider a simple task such as the alphabetical sorting of a set of cards with names on them. Sorting is a well-known and well-understood task. Although simple sorting is now in the domain of routine software design, it has more complicated analogies in human interaction systems. For example, an operator processing priorities of emergency signals at a power plant or similar facility is doing a type of sorting.

There are numerous documented strategies for sorting. One strategy might be to go through all the cards and find the lowest alphabetic name and then find for the next lowest and so forth until all the cards are in order. Another strategy might be to put each card into some subpile containing a smaller range such as A–E, F–H, etc., and then sort these subpiles. For a large set of cards the latter strategy would be much more efficient, but for a very small number of cards the former would probably be faster.

In accordance with the Second Law of ergodynamics, every work task can be done using a number of different strategies. Each strategy is presented with a certain bell-shaped characteristic curve $Q_i(F)$. Often there is no single best strategy for all situations. Certain strategies may suit a given range of information input volume better. Different strategies also play a role in learning situations, where a trainee may 'graduate' from a simple strategy to a more complex, more efficient one or may learn specialized strategies for specific situations.

3. THE QUADRIGRAMS

The quadrigrams demonstrate how complex the relationship can be between the external input and the human output of a task in human–machine–environment interaction. The quadrigrams help to further explain and predict this relationship.

Let us discuss several examples. Each example in this section describes a different feature of the quadrigram. We start with a simple example and add new concepts as we progress.

Consider the alphabetical sorting task described above. If we plot the information from this example on a quadrigram, then we measure the time taken to complete the task on the time axis. The external input factor of the task is a measure of the size of the task. It could be measured as the number of cards to sort or the number of characters in each name. This measure of the task is an external input, X. It gives a value on the E-axis (Figure 3). This allows us to plot a point in the fourth quadrant $E(T)$.

The internal input factor F is related to the external input factor E by the reflection rate R. Under normal reflection we have $R = 1$ and thus $F = E$. This means that normally the subject's response will be directly related to the external input of the task E. We call such case a normo-reflection (R_N).

Very often we start from the external signal E. If it was perceived as is (R_N), then we analyze the direct human–environment interaction: $F = E$. Thus instead of axis F, we may have a second axis E.

But sometimes (and very often) human performance may be initiated without any visible changes in the environment. The performance may be based on reminiscence of some prior information, new task, forecasting of future needs, etc. Here we have $E = $ constant ($DE = 0$) and $F > 0$. That means that measurement of the environment parameter E does not help to understand and model the performance. In this case we may

have two axes F and no axis E. This use of the triagram should be of special interest for psychological studies of human activity that do not depend directly on the dynamics of the environment.

Strictly speaking, when we talk about the environment's influence on human performance, we usually consider the influence of the environmental dynamics: changes (deviations) of environmental parameters from the level that human beings were adapted to. We all know from experience that a stable environment does not bring any new information. One no longer senses touch, sound, light, or other stimuli if s/he has adapted to them. That means that after enough time for adaptation, a stable E leads to $F = 0$. Therefore we may measure and draw the deviation of E from some constant, adapted level the ergograms.

The second quadrant $Q(F,S)$ in Figure 3 contains characteristic curves of the functional strategies (for example, of cognitive strategies) and gives us the output parameter Q as a function of F and the strategy S. For the sorting example the output factor would be a measure of efficiency or productivity. It could be measured in units.

Given the output parameter, we can plot a point in the first quadrant $Q(S,F,T)$. This quadrant gives us the final results of output versus time. It is labeled the response dynamics of the system.

So, to summarize the process: first we plot the task as a point in the $E(T)$ coordinates (fourth quadrant). Then we use a reflection vector R (in the third quadrant) to relate the external E and internal F input factors. When we find the actual value of internal input factor F and, if we know the human functional strategies and which concrete one is being used, S_i, then the F–S_i–Q analysis (second quadrant) yields the output parameter Q. We can then plot the task's response dynamics, solving for $Q(T)$ in the first quadrant.

This example describes the 'base case' for using the quadrigram. The other examples described below present applications of the quadrigram to explain more complex phenomena, but they all use the same approach in plotting the information.

4. REFLECTION VECTOR

The quadrigram is very convenient to model different stress levels. The second example deals with a situation where the psychological state of the subject changes. Instead of having a stable situation, consider a case where the subject is under great stress. Here the *reflection rate* ($R = F/E$), plotted in the third quadrant as a respective reflection vector, is no longer unity (Figure 3b). Instead of using a reflection rate $R = 1$ where the external input factor corresponds exactly to the internal input factor, the reflection rate is increased. When $R > 1$ the subject reacts as if the task was of greater complexity than is the case for the concrete strategy used. This is typical for the stress situations, for example in emergencies.

On the quadrigram, this is reflected as a high level of the internal input factor F. Here, the external input factor E, which is an objective measure of the volume of information, is much lower.

This example demonstrates the function of the reflection rate F/E in the third quadrant. The performance of the same subject doing the same task may vary substantially if the state of stress is different. Under high stress levels (performance anxiety, emergencies) the subject will be in a state of over perception.

S/he will react as if the volume of information is very high. The subject will try to internalize more information than need be considered. In extreme cases the subject may 'overload' and reach a very low state of efficiency. On the other hand, if an operator is sick, tired or inattentive ($R << 1$), s/he may process the problem using less information than is given. The subject here lacks some required inputs to the problem, and as a result cannot perform as effectively as possible.

Vector R in the third quadrant is a very universal element as well as the quadrigram itself. The vector may reflect the processing of information by an expert or a display system. The expert or display system may add some details to the initial environmental signal E, so information displayed F would be larger than the environmental information: $F > E$ and $R > 1$.

The information display may restrict the initial signal to large information chunks that the human operator could perceive simultaneously. Here the total number of information chunks would be smaller. Thus, $F < E$.

In practice vector R might be non-linear. Indeed, the operator could ignore small deviations of the control parameters ($F/E = 0$). Here the RE angle could equal 0. Middle size deviations could be taken objectively ($F/E = 1$). Very big, emergency deviations could be subjectively enlarged ($F/E >> 1$). This effect could be also reflected by rotating R in time according to the actual E and human operator psychological state dynamics that was analyzed earlier.

The quadrigrams are especially effective for dynamic modeling of human–machine–environment systems. A situation of great interest is when we want to model a dynamic system; i.e., a system with some degree of feedback. As a crude example, consider the task of memorizing the names of a group of people. The group is introduced and the subject must repeat the names. Initially the guesses will be mostly wrong. However, if the task is repeated, the subject will now be familiar with some names, and will likely do better on the second attempt. The success of the first attempt will have a definite bearing on the volume of information of the second attempt.

Suppose we use the E-axis to measure the number of unknown names (Figure 4).

Initially this will be the entire group. The people are introduced and the subject is asked to repeat the names. The subject will probably guess some fraction correctly. The second time the subject tries the task, it will be easier. Some information
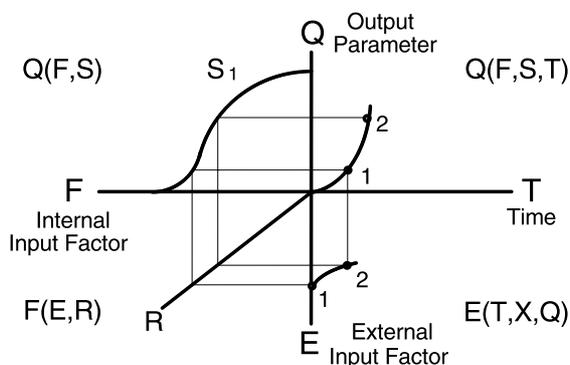
Figure 4. Modeling a dynamic human–machine–environment interaction using quadrigram.

would now be internalized. The volume of information input for subsequent attempts will thus be lower.

We could repeat the process a few more times until all the names were remembered and the system is in steady-state. In this example, the output parameter would be measured as the number of names remembered. The external input factor E would be a function of the time, the success of the previous try, and the size of the group $E(T,Q,X)$. This simple example can be considered to be a rudimentary dynamic system. In a dynamic system, the four quadrants contain the four curves $Q(T,S,F)$, $Q(F,S)$, $F(E,R)$ and $E(T,Q,X)$, and the output parameter Q of one cycle affects the external input factor E of the next cycle.

To conclude this short explanation of the idea and practical possibilities of the Ergograms, we recommend its use in the design, analysis and improvement of dynamic human–machine–environment systems. The most general and universal type of Ergogram is the quadrigram. With this versatile model, one can describe a wide variety of human interaction situations. With the graphical structure of the quadrigram one can visually relate input factors to the system reflection rate and see their influence on output through different strategies. All these parameters affect performance in human–computer interaction.

The use of the quadrigram is not limited to situations where hard quantitative data is available. It is also useful for visualizing the relationships between factors in theoretical problems. Even for case studies and single incidents, the quadrigram provides a framework from which to evaluate results and illuminate complicated causal relationships.

While the complete quadrigram may seem complex, the amount of information it contains more than justifies the effort required to understand it. Most potential users would already be familiar with portions of the model, such as output versus time. The quadrigram provides the opportunity to supplement familiar knowledge with a model that gives insight into both causes and effects of changes in the system components.

5. ERGOGRAMS IN MODELING COMMUNICATION DYNAMICS

In accordance with the Third Law of ergodynamics (the law of transformations and interactions), interactions and transformations between two work strategies S_1 and S_2 goes through the intersect point of the bell-shaped characteristic curves of S_1 and S_2 which reflects equal and common state for the two strategies. Efficiency of every functional strategy S_i is being characterized with the values of efficiency Q_i as function of the internal input factor F. F is often called also a parameter of human–environment mutual adaptation, or efficiency-complexity factor. If the efficiency of the two strategies is equal ($Q_1 = Q_2$) and the factor value is the same $F = F_{1,2}$ that means the strategies have an equal and common state. The experiments have shown that this common state is the best for the total productivity of the individuals using the two different functional strategies.

Thus, the most productive interaction between two different functional strategies goes through an environment state common for the interacting strategies. This is when the factor F and efficiency Q are equal for the strategies: $F = F_{1,2}$; $Q_1 = Q_2$.

Studies of interaction and transformation processes in communication between a teacher (professor) and students are very important for improvement of education process.

The central problem is a mutual adaptation between the
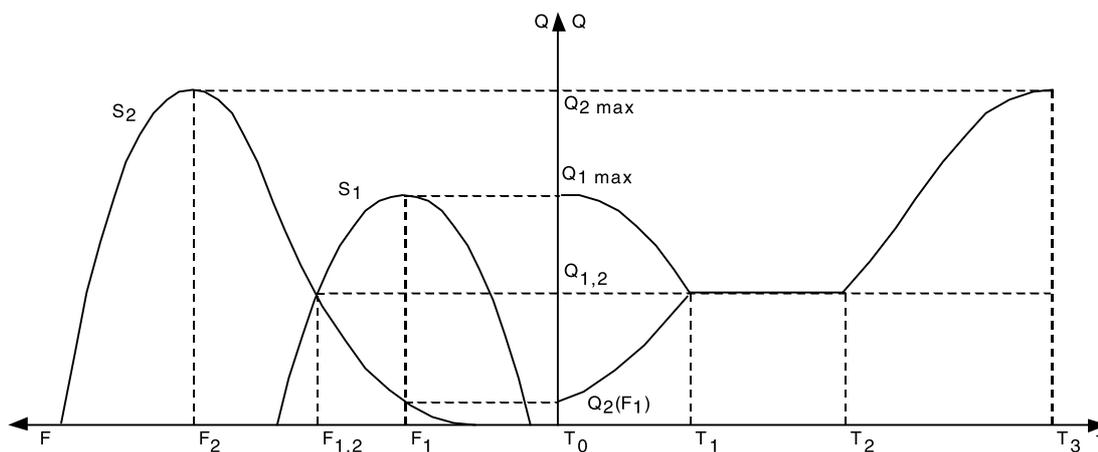
Figure 5. Process of professor–student mutual adaptation: the professor decreases the information presenting speed F from $F_{2\,opt}$ to $F_{1\,opt}$ and then leads students to the transformation S_1 into S_2 when $F = F_{1,2}$ and then to the maximal comprehension speed $Q_{2\,max}$. Q, Speed of knowledge transmitting (by professor) and comprehension (by students).

professor's speed of information presenting at the lectures and the students' perception and comprehension of the information.

In our experiments speed of presenting information was measured as the number of test questions (topics included into current tests and the final exam) explained by the professor during one lecture. The speed of the students' comprehension was measured as the number of questions they answered immediately after the lecture. Data were collected and given us by 27 professors from 15 Russian universities that took part in our program in 1986–90.

The functional strategy of students' managing of the course S_1 has a low optimal information perception speed $F_{1\,opt}$. At this perception speed, they reach the maximal comprehension $Q_{1\,max}$ (Figure 5).

The professor has some optimal information volume $F_{2\,opt}$. At this volume, s/he can manage the course with maximal efficiency $Q_{2\,opt}$.

A goal of the professor–student interaction is for the students to obtain strategy S_2, with a high information comprehension speed, equal to $F_{2\,opt}$. The student's starting functional strategy of understanding, based on the previous knowledge strategy, is S_1. If the professor presents information at his or her optimal speed $F = F_{2\,opt}$, the student's efficiency level of comprehension will be very low: $Q_1(F_{2\,opt}) \ll Q_{1\,max} = Q_1(F_{1\,opt})$. That means that the professor presents new knowledge too fast and students understand almost nothing. The efficiency of the education process in this type of registered and analyzed cases was close to zero: $Q_1(F_{2\,opt})$.

We measured F as the professor's information presenting speed in our experiments simply as the number of words the professor said, calculated with recorded audio tape. We measured F as the student's information perception speed as the number of words and symbols written during the lecture.

We measured Q as knowledge volume the professor presented as the number of topics (test questions) covered in the lecture. Q as knowledge volume comprehended by the students was the number of test questions the students correctly answered immediately after the lecture.

Please note that the scale of F at Figure 5 is reversed: the values of F increase from right to left. The ergograms give

ergonomists wide freedom in research, design and modeling human–machine–environment systems. One may change the axes, or their directions, to adapt the ergogram to the concrete needs.

Educational experiments have shown that if someone's comprehension efficiency decreases very much, then very strong negative emotional reactions and psychological resistance often appears. Lower efficiency means higher complexity. So, not only does low lecture time efficiency result for the students, but they also become hostile toward the professor. This resulted in obstacles for the professor–student interaction.

We decreased the complexity of the interaction and learning process by special organizing of the mutual adaptation between the professor and the students.

To improve students' comprehension at the start of the course, the professors decreased the volume of information presented at the lecture F from $F_{2\,opt}$ toward the F acceptable for the students. They started from a very simple short introduction, with minimal complexity and maximal efficiency of student comprehension: $F = F_{1\,opt}$, $Q_1(F_{1\,opt}) = Q_{1\,max}$. In fact we recommended that professors start from slow dictating of basic terms and definitions, and introduction into disciplines. By analyzing huge data on university educational processes, we found two main disadvantages: (1) the knowledge transmitted by professors and comprehended by the students did not equate; and (2) lecture information is not associated with concrete images of work processes (in ergonomics), technologies (in industrial engineering), or experiments (in psychology).

6. RECOMMENDATIONS

Here is how to use of ergograms to increase interaction efficiency. Even though we are talking about professor–student communication, this is only an example of many cases where ergograms may be used to improve communications in worker team, training center, in human–machine and human–computer interaction.

To reach mutual professor–student adaptation, we recommended that the professors who participated in the experiments should reach (in the beginning of the course by slowing down the teaching tempo) and maintain mutual

adaptation between the tempo of knowledge presented and its comprehension. After that, the professors gradually increased the volume of information presented from $F_{1\,opt}$ to $F_{1,2}$. Even with this very essential decrease of information presenting speed (if compared with usual $F_{2\,opt}$) in the interval of changing F from $F_{1\,opt}$ to $F_{1,2}$, the students' comprehension efficiency decreased from $Q_{1\,max}$ to $Q_1(F_{1,2})$. Not all students were motivated enough to make some extra effort to perceive the larger volume of information and overcome the added complexity of learning performance. We found that to be a crucial period of the professor–student mutual adaptation process. At the crossing point for both functional structures $F_{1,2}$; $Q_{1,2}$, the optimal conditions to transform S_1 into S_2 occurred. We recommended that the professors' speed of information presenting, F, should stay constant for the time the students need to change their previous functional structure from S_1 to S_2. This time was spent especially to help students to replace their previous cognitive strategy S_1, based on separate perception of different short topics, by the cognitive strategy S_2, based on understanding of general fundamentals of the discipline taught. Perception of it proceeded by big information chunks using associations between the topics.

After that time, when the new (for the students) structure S_2 (old for the professor) has been already formed by transformation of S_1 into S_2, increasing the information presentation tempo F led to fast increases in student comprehension efficiency, using the new structure S_2. Figure 6 shows this transformation, using a dynamic ergogram with an added F-axis. This axis helps to display both the dynamics of factor F and work efficiency Q.

The ergogram at Figure 6 has three axes F (two times, as $E = F$ and F), Q and T. This ergogram is called a triagram. It shows the dynamics of a factor in time $F(T)$, so one can evaluate not only the values of efficiency when different structures are used, but also times when those levels of efficiency would be reached according to the values of F at respective moments.

We should note an assumption is being used in building the "triagrams": human work structure is indeed inertial, so when F reaches some value F_1 the respective efficiency value $Q(F_1)$ will be reached, not at the same time, but later. This time delay could be easily assessed in the experiment as a reaction time. In using the quadrigram this assumption is not necessary. The time delay in human–machine mutual adaptation is easily displayed there. But in comparison with our ordinary ergogram (Figure 2) the triagram has unquestionable advantage and could be widely used.

A group of professors who participated in the experiments increased F monotonically as usually. The students did not have the time and conditions to transform S_1 into S_2. They continued to use the initial cognitive structure S_1. Then, when the professor's information presenting tempo became higher than $F_{1,2}$, the professor's performance efficiency, speed of presenting knowledge Q_2 increased, but ~40% of the students decreased their
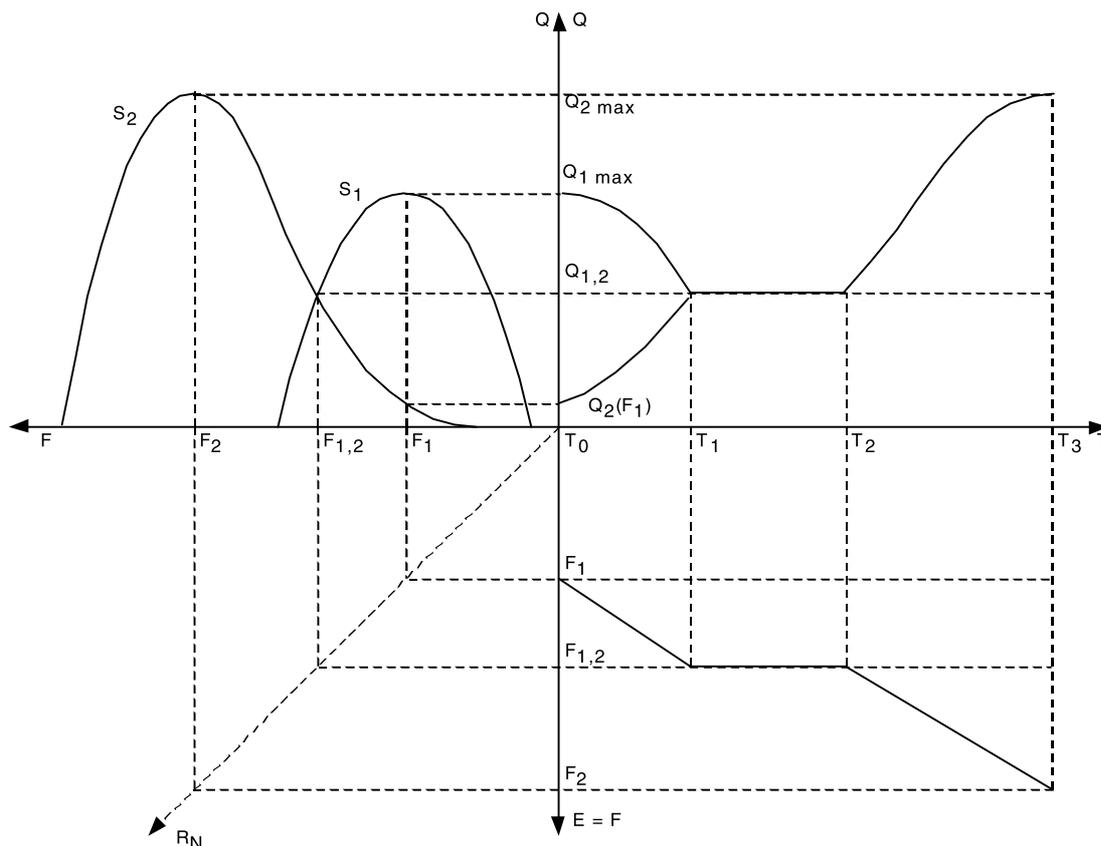
Figure 6. Triagram showing successful dynamics of professor–students interaction if there is a temporary stabilization of information volume presented F during time $T_1 - T_2$ to allow students to transform structure S_1 into S_2 and then increase performance efficiency Q. Figure 7. Ergogram showing unsuccessful dynamics of teacher–student interaction if the teacher presents information monotonously increasing a tempo of information presenting even starting from a tempo optimal for students' perception $F_{1\,opt}$ but without fixation of $F = F_{1,2} =$ constant to give students time for transformation S_1 into S_2.
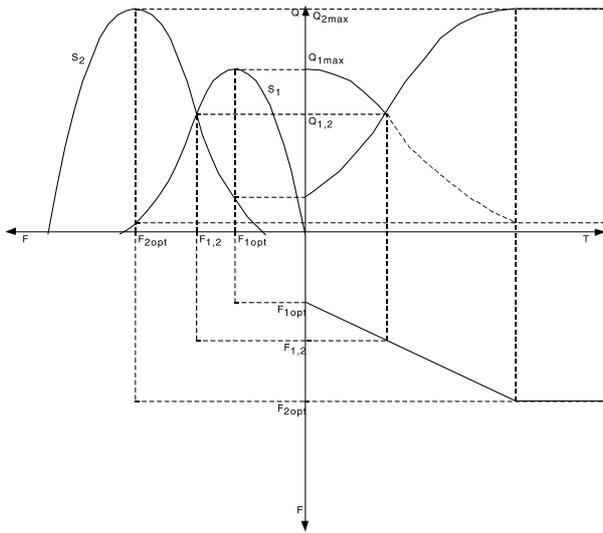
Figure 7.

comprehension efficiency Q_1 very quickly. This process is shown using quadrigram in Figure 7.

This analysis revealed that the current educational method typical for the universities and vocational training centers, based on monotonic increasing of the volume of knowledge presented to students, at a certain stage leads to almost zero lecture efficiency for a big part of the students. In our experiments the part was ~40%.

REFERENCES

KARWOWSKI, W. and MITAL, A., 1986, *Applications of Fuzzy Set Theory in Human Factors* (Amsterdam: Elsevier).

KARWOWSKI, W., 1991, Complexity, fuzziness and ergonomic incompatibility issues in the control of dynamic work environments. *Ergonomics*, 34, 671–86.

VENDA, V.F., 1990, *Hybrid Intelligence Systems: Evolution, Psychology, Ergonomics* (Moscow: Mashinostroenie). [in Russian]

VENDA, V.F. and VENDA, Y.V., 1991, Transformation dynamics in complex systems. *Journal of Washington Academy of Science*, 4.

VENDA, V.F., STRONG, D.R. and DROMEY, I.R., 1993, The quadrigram as a model for dynamic human–computer interaction studies and design. In G. Salvendy and M.J. Smith (eds) *Human–Computer Interaction — 19A* (Amsterdam: Elsevier), 654–9. VENDA, V.F. and VENDA, Y.V., 1995, *Dynamics in Ergonomics, Psychology, and Decisions: Introduction to Ergodynamics* (Norwood: Ablex).

Ergonomic Data for the Design and Evaluation of Technical Systems

H. Schmidtke and I. Jastrzebska-Fraczek

Department of Ergonomics, Technical University of Munich, Boltzmannstraße 15, 85747 Garching Germany

1. INTRODUCTION

Even a considerable amount of ergonomic data and recommendations and their didactical presentation in various educational publications, e.g. Schmidtke (1993), Luczak (1993), Bullinger (1994), or data collections (van Cott and Kinkade 1972, Woodson 1981) is available today. And experts frequently note that some work places and tools partly display decisive ergonomic deficiencies. One clear cause is that the dissemination of ergonomic knowledge is not as widespread as would be beneficial for the employee or system user. The designer of work places and tools often has not as much experience in this field as necessary and is guided by his intuition only in ergonomic issues. However, in many cases, even the ergonomic expert finds it rather difficult to compress the available know-how into tables in a way that renders it adequate for each individual application. This leads to an performance or health protection, where a consequent application of ergonomic experience in this field would result in considerable improvements. This is the indication where marked benefits can be gained from an appropriate application of a computer.

2. THE ERGONOMIC DATABASE (EDS) AS AN EXAMPLE FOR AN ELECTRONIC DATABASE FOR THE ERGONOMIC DESIGN OF MAN-MACHINE-SYSTEMS

2.1. Modules and Targets of the Ergonomic Database

As early as during the planning and developing phase, and subsequently during the designing and blueprint phase, the ergonomic database (EDS) confronts the designer of equipment and work places with the most important ergonomic demands posed on the components he works on and their interaction in the work tasks. In contrast to CAD systems, EDS is a pure database system. Figure 1 depicts the main menu mask of EDS.

EDS consists of two different modules:

- basic — contains the ergonomic data compiled for technical components, environmental factors and work tasks; and
- individual — contains data on body measures, human forces, sizes of movement and methods for strain/stress analysis (man module), ergonomic requirements for specific types of work place (consulting module) and a module for ergonomic examination and evaluation. Finally a checklist for task and product analysis is included.

In addition to the data base system, each menu mask shows a direct access to literature search and a definition of ergonomic
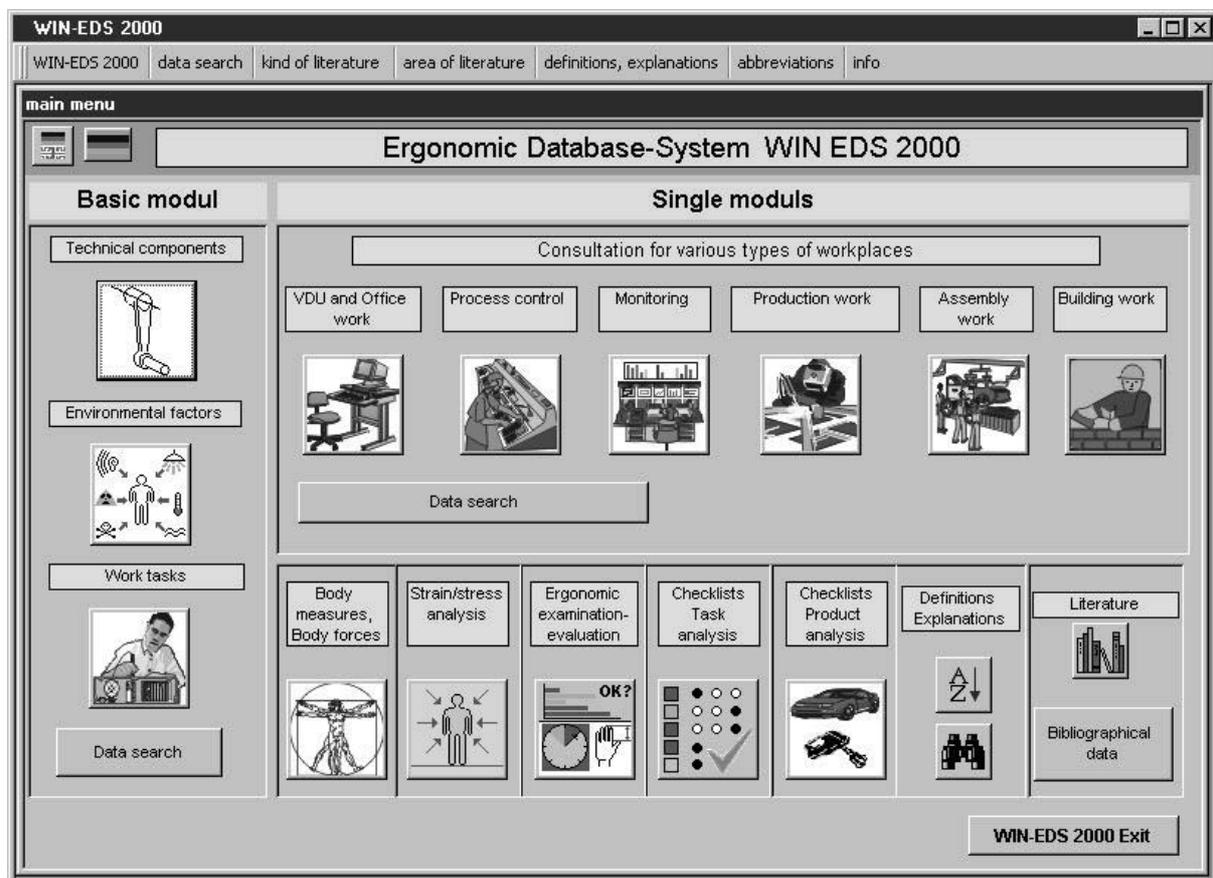
Figure 1: Main menu EDS

terminology. The access to the database, literature search, and explanation of ergonomic terminology is guided by menu support. EDS is self-explanatory to an extent that the user can work with standard commands.

The data compiled in EDS comprises requirements on man–machine systems classified according to the following categories:

- Technical components — 15 submenus, 107 data sheets, 1080 single positions.
- Environmental factors — five submenus, 18 data sheets, 142 single positions.
- Work tasks — six submenus, 23 data sheets, 198 single positions.
- Additional data on:
- body measures — 11 data sheets, 108 single measurements;
- human forces — eight data sheets, 46 examples for force activities;
- size of movements — six data sheets, 31 individual measurements;
- strain/stress analysis — three methods for analysis;
- consulting on task — six submenus, 180 data sheets, 1871 single positions) categories;
- checklist for task analysis — six types of work, 520 questions; and
- checklist for product analysis — four types of products, 490 questions.
- The literary search is classified according to fields (e.g. technical components) and types (e.g. books, standards).
- The collection of definitions for ergonomic terminology presently contains > 450 alphabetically listed definitions.

The majority of databases available so far are oriented upon checklists, similar to those already used in various areas of technology. The database concept presented differs from the checklist concept. It compares defined positions to specified values. Even if not all ergonomic requirements can be translated into numbers, verbally formulated demands can be useful and important for the designer of work or tools However, practical application has proved it to be useful to integrate a module "Checklist for task analysis" and "Checklist for product analysis" into the EDS besides the pure data collection. By means of this modules, it is possible to compile a specific checklist for a testing task without much effort.

2.2. Technical Components

The term "Technical components" covers a very wide scope. It can be used to describe a single key in a keyboard as well as a large switchboard. This means for the understanding of this database system, all components at the interface to the human user are included. Consequently, technical components not representing an interface to the user, as a rule, are not an object of ergonomic observation. The following classification for technical components has been applied:

- Access to workplace.
- Work tables.
- Consoles.
- Seats.
- Body support.
- Equipment for optical and acoustic information.
- Displays.
- Control devices, general requirements.

- Control devices for rotary motion.
- Control devices for linear motion.
- Pedals.
- Transport, lifting, conveying and storage equipment.
- Lifesaving and protection equipment.
- Facilities for test and maintenance.
- Workplaces and means for production.

The subsequent lower level — as an example from transport, lifting, conveying and storage equipment — contains the relevant characteristics:

- Design of lifting equipment.
- Safety devices on lifting equipment.
- Handcarts, forklift trucks.
- Elevators for persons and loads, ladder shaft.
- Design of trailers.
- Transport containers and packing materials.
- Mass of goods manually transported.
- Angular speed of goods to be inspected.
- Handles on packing materials, equipment and drawers.
- Shelves.
- Deposit shelves for office materials, records and regulations.

2.3. Environmental Factors

The group of ergonomic requirements for environmental factors lists demands and permissible limits basing on laws and regulations, as well as provisions and definitions in standards and regulations, as long as they do not contradict generally accepted ergonomic know-how. The following environmental factors contained in EDS at the present:

- Illumination.
- Noise effects.
- Climate, ventilation.
- Vibration.
- Toxic substances and radiation.

2.4. Work Tasks

In the third group of the database system, from the wide scope of work categories those with a high degree of general applicability were selected. The positions contained in the individual characteristics mark technical and organizational pre-requisites for an optimum execution of the tasks. A consideration of the ergonomic specification may avoid particular obstacles in the tasks, however, on its own it does not guarantee that the entire work content in a task does neither poses too much or too little demand on the worker. The following structure of the work tasks in EDS has been applied:

- Putting into and placing out of operation for technical equipment.
- Visual supervision.
- System control.
- Disturbance and emergency treatment.
- Material handling and maintenance.
- Requirements on staff.

All issues listed in above are contained in individual data sheets. Each sheet provides detailed information on the screen (figure 2). In principle, each data sheet is designed in a way that the ergonomic specifications are described qualitatively or with measures and numbers in a list of varying length.

Figure 2: Excerpt from data sheet for room illumination

2.5. Structure of Single Modules in EDS

2.5.1. Consulting module

The consulting module contains data on specific work places as it can be seen in figure 1. The structures correspond to those of the basic module.

2.5.2. Data on body measures, human forces and size of movements

The data on body measures and human forces have an additive character with respect to the previous categories. They provide further information that may be important for the design of technical components. Yet for the application of the data on forces, it is important to note that the values given are isometric maximum forces in all cases. Maximum forces can only be performed for very few seconds.

The tables containing body measures are structured as follows:

- The first and the second column give the number and term for the relevant body measure.
- The subsequent six columns contain the body measures for men and women (mm) each for the 5th, 50th and 95th percentile.
- The last column provides notes on the source in literature.

The material compiled comprises a total of 108 individual measures.

Structure of data sets for body measures (mm):

- Body measures: standing — height.
- Body measures: standing — width.
- Body measures: standing — reach and depth.
- Body measures: standing — circumference.
- Body measures: sitting — height.
- Body measures: sitting — width.
- Body measures: sitting — depth.
- Body measures: head.
- Body measures: hand.
- Body measures: finger.
- Body measures: foot.

The data on body forces give maximum isometric human position forces for occasional activity. They were taken from Rühmann and Schmidtke (1992).

The material compiled contains 46 cases of exerted force. In contrast to the characteristics for body measures, only values for the 5th, 10th and 50th percentile are given, since for the design of work or tools, only information on the lower percentile classes is of importance.

2.5.3. Methods for strain/stress analysis

On the sector of strain/stress analysis three analysis methods are available at the present, this is for prevailing holding tasks (the procedure for the determination of permissible muscle strain in alignment to VDI (German Association of Engineers) and the comparatively restrictive method according to NIOSH) and the procedure presented by Spitzer, Hettinger and Kaminsky for the analysis of energetic stress during dynamic muscle work.

The section strain/stress analysis contains thee different methods:

- Inquiry of muscular strain according to VDI.

WIN-EDS 2000

WIN-EDS 2000 | data search | kind of literature | area of literature | definitions, explanations | abbreviations | info

men modul , anthropometric data of the body

main menu

1 1 0 | Body measures (mm) - standing - height

explanation of importance/consequence (B)

Literature - source

KM1.ppt

Nr:			Benennung	männlich Perzentile weiblich						Quelle
				5	50	95	5	50	95	
1	1	1	body height -streched posture (1)	1715	1825	1965	1575	1679	1787	HdE B-1.2.1/ 1.4.1
1	1	2	body height -streched posture (1)	1662	1761	1863	1538	1632	1738	Lit.Nr. 563 (USA)
1	1	3	body height -streched posture (1)		1679			1558		Lit.Nr. 844 (Korea)
1	1	4	eye height (2)	1595	1690	1794	1472	1566	1664	HdE B-1.2.2/ 1.4.2
1	1	5	eye height (2)	1600	1649	1748	1436	1535	1634	Lit.Nr. 563 (USA)
1	1	6	eye height (2)		1563			1450		Lit.Nr. 844 (Korea)
1	1	7	mouth height (3)	1564	1649	1745	1428	1511	1594	HdE B-1.2.4/ 1.4.4

Figure 3: Excerpt from data sheet for body measures

- Inquiry of muscular strain according to NIOSH.
- Inquiry of muscular stress according to Spitzer *et al.*

The analysis procedure based on the work of Burand and published by VDI is primarily suitable for activities involving the lifting or lowering of material, tools or parts. Although the algorithms for calculation provide recommendations for pure shifting operations on almost identical level, there is no information at the present to which extent these recommendations are scientifically founded.

Similar to the VDI procedure, the NIOSH procedure currently discussed in International Standardisation (draft 1992) is mainly suitable for activities, where material, tools or parts have to be lifted or lowered. As already mentioned, the NIOSH procedure is highly restrictive and provides comparatively low permissible loads.

In contrast, the analysis method according to Spitzer *et al.* is based on experiments and refers to the so-called group evaluation tables published by the authors. In this case two parameters are considered:
- the body position or movement; and
- the type of work.

The analysis method according to Spitzer *et al.* is only applicable for the determination of the energy expenditure during the strain of larger muscle groups of the body. It is not suitable for holding tasks or one-sided strain of smaller muscle groups. Therefore, it has to be emphasised that reliable data on the work dependent energetic stress can only be gained for measurements of the energy expenditure at the work place.

Since, as a rule, a work task is set-up from various elements of activity, requiring different body positions or movements and consequently strain the individual muscle groups to a varying extent, the analysis has to be done separately for each individual element of activity. The computer-aided analysis method summarizes the values of energy transfer determined according to a temporal evaluation to a work transfer value for the entire task.

2.5.4. Module "Computer-aided ergonomic evaluation"

Whereas all modules discussed so far contain ergonomically relevant data, the module "Computer-aided ergonomic evaluation" serves the purpose of creating a defined evaluation task for the evaluation protocol. For this purpose, a decision has to be made, whether test positions have to be taken from the basic or the consulting module (figure 4). For instance, if it is intended to evaluate safety parts on lifting devices from the module "Technical components," by a mouse click on the dialogue button "Technical components," a list of the sections it contains is opened. It includes the section "Equipment for transport, lifting, feeding and storing." A mouse click on the relevant button opens a list of the data sheets included in this field.

According to the evaluation task, the button "safety parts on lifting equipment" has to be clicked. This calls all 17 evaluation positions from this data sheet. The user has to decide by clicking onto the relevant buttons, which individual positions he wants to include in the evaluation protocol.

With the dialogue button "Survey test record" the test protocol thus created is opened. Now the actual values determined on the test object can be entered into the protocol. In addition, it contains

Figure 4. Mask for ergonomic evaluation

information on the type of test method suitable for the analysis and whether the individual positions for health, work safety, performance, comfort, functionality/technical reliability are decisive. The user can set his own priorities in the evaluation.

When the analysis is terminated, a graphical survey of the test result can be called via the button "graphic profile." In addition it is possible to print a list giving all test requirements that are not accomplished. Finally, a conclusive evaluation report can be printed including a graphic depiction of the evaluation, statistics, and individual results.

2.5.5. Module checklist for task analysis and product analysis

If it is intended to do ergonomic tests using the basic or consulting module, this may require modern measurement equipment. If this is not available or if it is sufficient to do a more qualitative task analysis, the module "Checklist" can be selected.

The module Checklist for task analysis is based on the six types of work place contained in the consulting module (figure 1). For each type of work place, a series of areas for analysis have been defined (e.g. environmental impact, tools, controls, physiological strain, etc.). The user then has to decide, which section is important for the test or is insignificant. By clicking onto the relevant buttons, he opens a list of questions. An example using the environmental factor "illumination" may be: "Is the illumination of the work place or room sufficient for all visual tasks?" The user can then answer this question with "yes," "no" or "not applicable" using his own experience or questioning the staff concerned (figure 5).

When all questions have been answered, the user receives an information that he now can proceed to the test result. For this

purpose, he can choose between several depictions. They can print a list of all questions posed sorted according to fields or criteria (e.g. health, performance, comfort, etc.) or only those questions answered with "no" during the test, i.e. where ergonomic requirements have not been accomplished. Here again a sorting according to fields or criteria is possible. Finally, the entire test result can be printed as a test report containing a graphical depiction of the evaluation.

The checklist for product analysis is build up in the same manner and contains up to now the following types of products:

- Motor car.
- Omnibus.
- Truck.
- Earth-moving machinery.

2.5.6. Literary search and definitions of ergonomic terminology

As already mentioned in Section 1 the EDS is completed by a literature compilation, which can be called according to the type of literature and field of application. Naturally, the limited content means that not all relevant information is included. Here a courageous selection and omission had to be done.

Experience has shown that the users wish to obtain more detailed information on some terminology from the field of ergonomics. This need was catered for by offering slightly more than 450 terms around this field alphabetically for selection. When terms have already been defined in standards or regulations these definitions were used.

Figure 5: Cut out from the checklist of process control (areas) and from questions according to 'enviromental factors' on the checklist of process control

3. CONCLUSION

As the description of EDS shows, the system comprises a large amount of ergonomic information that plays a decisive role during the development phase of the product (or work place) as well as its acceptance test.

To meet this demand on an international level, decisions have to be made on the questions, which are the ergonomic core data, which are the differences to the published standards (frequently only covering minimum requirements) and to which extent the core data correspond to international Standardisation and literature.

BULLINGER, H.-J. (ed.), 1994, *Ergonomie; Produkt- und Arbeitsgestaltung* (Stuttgart: Teubner).

LUCZAK, H., 1993, *Arbeitswissenschaft*. (Berlin: Springer).

RÜHMANN, H. and SCHMIDTKE, H. (eds), 1992, *Körperkräfte des Menschen* (Cologne: Schmidt).

SCHMIDTKE, H. (ed.), 1993, *Ergonomie* (3rd ed.)(Munich: Hanser).

VAN COTT, H.P. and KINKADE, R.G. (eds), 1972, *Human Engineering Guide to Equipment Design* (Washington, DC: US Government Printing Office).

WOODSON, W.E., 1981, *Human Factors Design Handbook* (New York: McGraw-Hill).

Ergonomic Methods: Selection Criteria

N. Stanton
Department of Psychology, University of Southampton, SO17 1BJ, UK
C. Baber
Industria l Ergonomics Group, University of Birmingham, B15 2TT, UK

1. INTRODUCTION

The idea of a *usability evaluation* is gaining momentum, and the term *usability* is becoming common parlance in product design. This is a welcome shift in emphasis towards "ease-of-use" in product development. In many respects, the fundamental tenet of usability is that a product should be easy to use. This heightening of interest does not mean that usability (AKA user friendly, ergonomically designed, user-centered design, consumer-oriented product development) is a new concept. Ergonomists have been beating this particular drum for the past fifty years or so. Just over twenty years ago, Ivergard (1976) pointed out that very little research had been published on consumer ergonomics. While that picture has changed somewhat (there is a growing literature on physical aspects of product use), there is still little published on cognitive aspects of product use (Baber and Stanton 1994).

Although one can point to the consequences of not considering usability, there is much debate as to what the term actually means. One of the main problems with the term "usability" is that it means different things to different people. Some may suggest that usability is simply another attempt to introduce "user friendliness" back into product design jargon: usability is simply new wine in old bottles. Others argue that the issues surrounding usability have already been dealt with in "user-centered design". Baber (1993) points out, using the analogy of the soupstone, that the term usability takes on individual meaning to each person involved in evaluation to describe whatever they are doing: the individual adds his or her own ingredients. The trouble with this approach is how can we determine if one product is better than another, or indeed if the product has achieved some acceptable benchmark? Clearly this matter needs to be resolved as a matter of urgency, particularly in the light of recent legislation which makes usability a legal requirement in some products! This is a rather ridiculous situation given the debate and controversy surrounding the concept of usability (Stanton and Baber 1992).

2. WHAT IS USABILITY?

While it is possible to indicate the necessity for usability in product development, as a concept it has proved remarkably resilient to definition; we all know what it is, but have difficulty reaching an agreed, coherent definition (it is our own personal soupstone) which will allow recommendations to be made concerning how best to make something more "usable". This is the first, and perhaps most important, stumbling block in determining methods appropriate to evaluation. If we cannot agree on what usability is, how can we hope to measure it? It is likely that different definitions of the concept will lead people to measure different aspects of product use. This suggests that a usability evaluation may not have a common standard between individuals. If usability

is to be more than an ephemeral concept, we must agree on its constituent ingredients. Stanton and Baber (1992) draw upon a decade of work represented by Shackel (1981), Eason (1984), and Booth (1989) to suggest the factors above serve to shape the concept of usability and define its scope. These are as follows.

1. Learnability: a system should allow users to reach acceptable performance levels within a specified time.
2. Effectiveness: acceptable performance should be achieved by a defined proportion of the user population over a specified range of tasks and in a specified range of environments.
3. Attitude: acceptable performance should be achieved within acceptable human costs in terms of fatigue, stress, frustration, discomfort, and satisfaction.
4. Flexibility: the product should be able to deal with a range of tasks beyond those first specified.
5. The perceived usefulness or utility of the product: Eason (1984) has argued that "the major indicator of usability is whether a ... [product] ... is used". As Booth (1989) points out, it may be possible to design a product which rates high on the LEAF precepts set out above, but which is simply not used.
6. Task match: in addition to the LEAF precepts, a "usable" product should exhibit an acceptable match between the functions provided by the system and the needs and requirements of the user.
7. Task characteristics: these include the frequency with which a task can be performed and the degree to which the ask can be modified, such as in terms of variability of information requirements.
8. User characteristics: another section that should be included in a definition of usability concerns the knowledge, skills, and motivation of the user population.

Whilst we may argue over the relative merits of different ingredients and the labels we give them, this rarely becomes more than a exercise in semantics. ISO 9241 goes some way toward incorporating the above factors, but we feel that it falls short of a comprehensive definition in an important way. From reading ISO 9241 (at the time of writing this was still unreleased), we feel that usability has been defined by what can be measured: usability is what usability evaluations do. This appears to concentrate largely on the LEAF precepts mentioned above (Learnability, Effectiveness, Attitude, and Flexibility). We believe that haste in producing *the definition* of usability should be tempered by rather more circumspect consideration about what is meant by usability.

3. TECHNIQUES FOR EXAMINING USABILITY

The reader will not be surprised to learn that each of the various factors that make up usability has spawned particular approaches to usability evaluation. In this section we present approaches related to aspects of product development. We are particularly concerned that reliance upon one approach exclusively, or a very narrow definition of usability, could lead an individual to perform a limited usability evaluation. This concern should become obvious when we map evaluation methods onto the ingredients of usability.

3.1. Usability in the Design Process

Traditionally, usability assessments have been performed at the end of the design cycle, when a finished product can be evaluated.

However, it has been noted that the resulting changes proposed may be substantial and costly. This has led to a call for usability considerations to be introduced in earlier aspects of the design cycle. The most obvious way in which to collect information about how people perform a task is by watching them do it, or asking them how they go about it.

Observation of human activities may afford the collection of data about the real interaction between the human and the machine. Yet what is observable might not, for example, tell you about the decisions being made, or the alternatives not selected.

Interviews, on the other hand, may enable a quick way to get to the unobservable decision-making process, particularly if the interviewee is describing a recent event and how they dealt with it. However, it has been found that post-event descriptions can vary from what was in fact being done at the time. A solution to this is to get people to "think aloud" (concurrent verbal protocol) as they are dealing with an event — although this may result in them actually changing the way they do things in order to make the process of description easier!

This information can then be analyzed by a variety of means — for example, task analysis, link analysis, time-line analysis and layout analysis. These are described below.

Task analysis has many derivatives: British ergonomists use Hierarchical Task Analysis (HTA), which allows them to describe human behavior in units from a hierarchical presentation, describing behavior in terms of goals, plans and actions. However, HTA often presents only ideal behavior, or an agreed consensus of ideal behavior; the task analysis may then feed into user requirements and system specification. This mapping assumes that existing task practices are ideal, and it may be necessary to abstract further task goals at a general level before proposing desired human activity and specifying system requirements.

Link analysis allows examination of the way in which humans use displays, and the cataloguing of the frequency with which they switch between displays and the time spent at each display. Together with an expert rating of the importance of each display, this enables the presentation of an optimum layout of the displays.

Time-line analysis enables the study of the task sequence and the operations performed. System bottlenecks may become apparent, thus prompting the need for redesign. A new design can then be rerun and reanalyzed to discover if the bottleneck in the system has been cleared.

Layout analysis is a means of examining display and control layouts by four criteria. These are: functional classification, importance of item, sequence of use, and frequency of use. Based on this analysis, the layout may then be subjected to improvement and redesign.

The criteria for acceptance of methods will include the time limit of the project, resources available, skills of the practitioners, and the stage of design. For example, a technique such as link analysis is relatively easy to perform and does not require knowledge of
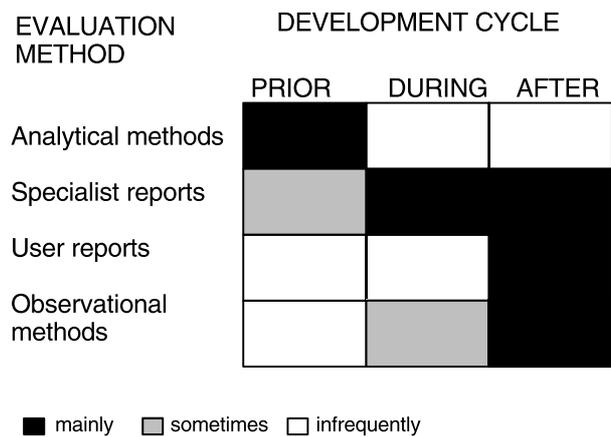
Figure 1.

ergonomics in its use. Hierarchical Task Analysis, however, requires considerable experience if it is to be used effectively.

Usability evaluation and assessment may enter all stages of the design process, from a requirement analysis, through initial design specification and prototyping, up to and including the working product. If we consider the four classes of evaluation method proposed by Whitefield, Wilson, and Dowell (1991) — namely, analytical methods (e.g. Keystroke Level Model: KLM; Task Analysis for Error Identification: TAFEI), specialist reports (e.g. layout analysis), user reports (e.g. interviews) and observational methods (e.g. HTA and link analysis) — each method is appropriate to certain stages of the product development cycle.

As Figure 1 suggests, analytical methods are mainly appropriate prior to product development. At this point, there is a good opportunity to iron out many potential problems at relatively little cost. During product design, specialist reports provide the main input; after product development most of the methods appear to be used. However, this is the most costly point in the design process at which to make changes and therefore underlines the impetus to include usability considerations earlier on. In this way, we may view usability input as a coarse-to-fine design approach. Early on in the design process it is able to offer a range of alternatives, which are successively refined until the product is implemented. This refinement goes hand in hand with the prototyping and evaluation process. Although Figure 1 may represent our notion of current practice, this does not necessarily mean that it is a desirable position. We are particularly keen to involve user evaluations earlier on in the design process.

Rapid prototyping is one means of achieving this. Developing prototypes enables the users to make a contribution toward the design process. In this way identified problems may be reduced well before the final implementation of the product. The contributions of the user do not only take the form of expressed likes and dislikes, but may be more objective in terms of performance testing. This participation may also provide the designers with a greater insight into the needs of the users. This can be particularly important when the users may not be able to express clearly what features they desire of the system, since they are unlikely to be aware of all the possibilities that could be made available. However, participants involved in prototyping should be either "real" end-users or, at least, representative users. Organizations often employ "product specialists" to act as users,

but these specialists will develop expertise which differs from that of "real" users. It would make more sense to use an ad hoc panel of users in the absence of access to "real" users.

In the rapid prototyping process, successive generations of the product are evaluated and the requirement specification finely tuned until the final product emerges. Rapid prototyping is a means of introducing the typically "late" evaluation techniques (illustrated in the "after" section of figure 1 — i.e. human factors methods used after the developmental cycle is complete) earlier on in the development cycle. Thus changes are likely to be more acceptable in terms of cost and therefore more likely to be implemented. Rapid prototyping recognizes that product development should be incremental, and allows usability to be incorporated into a succession of stages. As such, it manages to infiltrate the design process in a subliminal way rather than being a cumbersome add-on. Some organizations conduct prototyping trials in "usability labs". This formalizes the evaluation and provides the opportunity to capture behavior from a variety of sources simultaneously — for example, keystroke, eye movements, verbal protocols, and physical activity. This data is recorded by computer, eye cameras, audio tape recorders, and video recorders, thus enabling the analysis to be conducted at the analyst's leisure, and allowing the users' behavior to be scrutinized in fine detail. However, the mass of data generated will not take into account the fact that much of it is likely to be meaningless in the absence of performance criteria.

3.2. Evaluating Existing Products

Many writers have developed checklists for the evaluation of products in terms of their usability, one of the most detailed and helpful of which was developed by Ravden and Johnson (1989). Of the factors listed above, checklists will be most useful for 2, 3, 6, and 7 (effectiveness, attitude, task match, and task characteristics), and can be used to elicit users' attitudes, or to evaluate the relationship between task performance and computer. However, checklists are not the best way to assess items 1, 4, 5, and 8 (learnability, flexibility, perceived usefulness, and user characteristics). These could be considered either through experimentation (for 1 and 2, learnability and effectiveness) and observation (for 1 and 5, learnability and perceived usefulness), or through detailed specification of performance objectives (4, flexibility), or through development and testing of scientific theories (8, user characteristics). It is these latter approaches which are currently proving difficult to use. The main problem lies in the question of how to define performance criteria and experimental measures, which is mainly due to the fact that performance will be highly context dependent. It will hinge upon the nature of the task being performed and the knowledge that users bring with them to the interaction. In the absence of a coherent theory, one cannot model user characteristics. We are left, therefore, with an engineering, rather than scientific, approach. This means that the most promising way to measure usability effectively is by using real users of the computer system in question — for example when prototyping in system development. Many of the methodologies are similarly heuristic in nature and are not proven nor validated. This means that they should be used with caution. Other approaches available to researchers are verbal protocol, experimentation, task analysis, and simulation; these can be very useful in product evaluation,

but need to be performed after training in order to yield meaningful data.

3.3. Evaluating Early Concepts

While existing products can be evaluated quite simply by asking people to use them and then using a range of techniques to observe and analyze this usage, it is a harder proposition to evaluate conceptual products, i.e. paper-based designs. Yet, it is while the product is in its conceptual stage that the designer will have the best opportunity to incorporate usability into the design. A problem with usability is that, while it is possible to base specifications on products with which the designers are familiar, these products may not in themselves be "usable". As these familiar products are altered through redesign, then the usability specifications will necessarily alter — change the product and the nature of the product's use is changed.

A number of packages exist that allow designers to prototype proposed mock-ups of products. While such prototypes can be used to apply specific guidelines of product design, they cannot be used in an evaluation of usability unless they form part of a rapid prototyping schedule. Furthermore, it is only possible to prototype products when the task has been adequately described. This point is illustrated by Carroll and Rosson's (1991) notion of a task—artifact cycle. Basically, the design of a product will influence its use, which will influence users' goals, which will influence the design of a product. They suggest that scenario analysis or storyboarding as a means of capturing this cycle. This allows designers to maintain a flexible attitude to what and for whom they are designing. The designer's conception of how a task is performed is very different from that of a user, and this means that a gulf may well exist between design and user requirements. Capture techniques for user requirements have been developed, and a number of these which exist in ergonomics, such as task analysis, can be used successfully for this purpose. These can then provide objective information concerning real task requirements, which, in turn, can form the basis of specifications. However, such specifications will only provide "static" product data — i.e. information concerning how the product ought to look, not necessarily how it ought to be used "in anger". Baber and Stanton (1994) present a technique (TAFEI) which they claim can be used to analyze human interaction with products while development is still at the conceptual stage. This technique is potentially very useful to product designers.

4. SELECTION OF METHODS AND TECHNIQUES

To assist in the process of selecting methods and techniques to be employed in the usability evaluation we have devised the following heuristic chart (Figure 2) which could enable product designers to determine which method(s) is/are appropriate for their usability evaluations. This list is not intended to be exhaustive, and we invite readers to add their own. We have tried to highlight the need to consider those demands and constraints being placed on the project that are likely to affect the appropriateness of evaluation methods.

We have reduced the number of questions that the designer needs to ask about the end-user and project. The purpose of these questions is to enable the decision on choice of method to be based upon a considered judgement, rather than an ad hoc selection. We are not necessarily proposing that the procedure
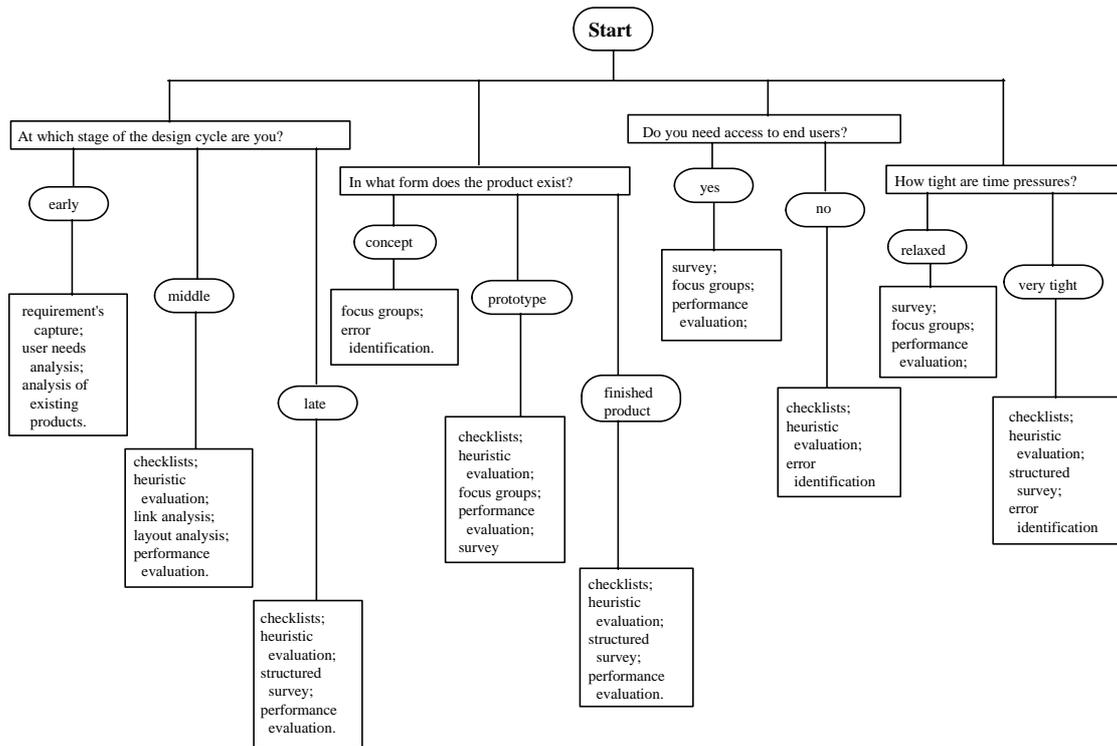
Figure 2.

should be rigidly adhered to, but rather that considerable thought should be given to the choice of any one method in preference to another. We feel that the selection process should be the subject of further research and that the process should be as objective as is possible.

In brief, our approach makes explicit the questions that designers would ask before opting for an evaluation method. The factors to be considered are:

- the stage of the design cycle
- the form of the product
- access to end users
- time pressures

First, the stage that the design cycle is in needs to be ascertained: i.e. early (whilst the product is still in the conceptual stage), middle (when a prototype of the product has been produced) or late (when the final version of product has been produced). Different methods are appropriate at different points in the design process, as we have indicated. The second question asks if the product (or a similar product) needs to be in existence before the method can be used; some methods do not rely upon product existence. The third question relates to the availability of end-users. Again, some methods can be used independent of access to end-users. The forth question determines the time window of the project, as some methods can be performed relatively quickly. At present the checklist method appears to be the most popular in current practice; consideration of Figure 2 makes it clear why this is the case. The use of a checklist in the hands of an experienced designer and ergonomist can make it a very cost-effective evaluation method. Examples of other methods will be found throughout this volume.

CONCLUSIONS

In order to be used, and used effectively, any product should be "usable". This is not simply a truism (although at first glance it may appear so). Products should be designed to conform to basic principles of usability. This article has sought to outline some of these principles, and relate them to methods for usability evaluations. We would caution sole reliance upon outcome measures in usability evaluations, as they do not necessarily guide redesign of products. In usability evaluations we believe that product *use* is of greatest importance, and feel therefore that process measures are of most benefit. This observation is made on the premise that an understanding of how the product is being used will ultimately provide more useful information to the designer than telling them that the product was not usable. Usability can facilitate the design of human-centered products which are not only "user friendly" but also useful. Our interpretation of the concept of usability suggests that it is important to ensure a good task match between products and users. This task-oriented approach — rather than a technological one — is indicated by the methods we propose.

REFERENCES

BABER, C., 1993, Usability is useless. *Interfaces* 21: 24–5.

BABER, C. and STANTON, N.A., 1994, Task analysis for error identification: a methodology for designing error tolerant products. *Ergonomics* 37(11): 1923–41.

CARROLL, J.M. and ROSSEN, M.B., 1991, Deliberated evolution: stalking the view matcher in design space. HCI VI 281–318.

BOOTH, P., 1989, *An Introduction to Human Computer Interaction* (London: LEA).

EASON, K.D., 1984, Towards the experimental study of usability. *Behaviour and Information Technology* 3(2): 133–43.

IVERGARD, T.B.K., 1976, Ergonomics and the Consumer. *Ergonomics* 19 (3): 321–9.

RAVDEN, S. and JOHNSON, G., 1989, Evaluating usability of human computer interfaces: a practical method (Ellis Horwood: Chichester).

SHACKEL, B., 1981, The concept of usability. *Proceedings of the IBM Software and Information Usability Symposium* 1–30 (New York: IBM).

STANTON, N.A. and BABER, C., 1992, Usability and EC Directive 90/ 270. *Displays* 13(3): 151–60.

WHITEFIELD, A., WILSON, F., and DOWELL, J., 1991, A framework for human factors evaluation. *Behaviour and Information Technology* 10(1): 65–80.

Goals Operators Methods and Selection Rules (GOMS)

P. Timmer

Ergonomics and HCI Unit, Department of Psychology, University College London, England

1. INTRODUCTION

One area of growing importance to ergonomists, and an area of specialism, is that of human–computer interaction (HCI). While the commercial trend is toward increasing the role of computing technologies in the workplace, accompanying this trend is an increase in design problems that concern the specification and evaluation of software applications. GOMS, or the Goals Operators Methods and Selection rules method (Card, Moran and Newell 1983), was developed at an early stage in the history of HCI research, and is a method with enduring appeal. When considering ergonomic contributions to the system development process, contributions may be considered along two dimensions; the stage at which the contribution is made, and the nature of the contribution. System development may be characterized in terms of a set of stages: requirements capture; design specification (conceptual and detailed); implementation; and evaluation (including testing and maintenance). The nature of a contribution, at any stage, may be either analytic or empirical. Therefore, when considering the stage of requirements capture, an ergonomist may either survey users to empirically elicit expressed requirements; or generate requirements analytically, following reflection on the domain, tasks, and so forth. Using this scheme of characterization, GOMS may be described as a method for analytic evaluation. To understand the enduring appeal of GOMS, one needs to adopt a slightly more complex view of the stages of system development. While the product under development will necessarily be evaluated, or at the least tested to ensure it works once implemented, specifications (for that product) at earlier stages of development may themselves be evaluated. For example, an ergonomist may generate two interface design specifications and wish to evaluate them both, to know which will be more effective, before the time-consuming and expensive stage of implementation is undertaken. Only by adopting a suitable method for early evaluation can such a design decision be made reliably. Given that at early stages of development device specifications alone exist (the process of implementation comes later), historically analytic methods have been required for early ergonomic contributions. (The recent rise of early prototype construction however, increasingly enables empirical contributions to be made at early stages of development.) GOMS was one of the first methods to support the early evaluation of design specifications, this being possible because GOMS supports the "prediction" of human–computer performance, rather than the "description" of empirically observed interactions. There are few methods in HCI which are predictive, fairly accurate, and which can be carried out quickly on the metaphorical "back of

an envelope". Hence the resilience of the GOMS method within HCI.

2. GOMS OVERVIEW

Models are frequently constructed in HCI design to reason about design decisions. Such models may be of the domain, user behavior, device behavior, user and device (work system) behavior, work system performance, and so forth (Whitefield 1990). The GOMS method outlines a set of components that, when combined, constitute a GOMS model. Such a model is largely of user behavior (both physical and mental). The GOMS method may be considered less a single method and more a family of methods, as models may be constructed at different levels of description, depending upon the purpose behind construction. The starting point for a GOMS model is to construct a hierarchical task description of the sub-tasks that a user must necessarily undertake to complete a task. This "Unit-Task" level of description may then undergo progressive decomposition, to "Functional", "Argument" and finally "Keystroke" levels of description (Card et al. 1983). A GOMS model at the keystroke level of analysis is at once: (1) the most detailed level of description (of the interaction); and (2) the level at which the greatest predictive accuracy and differentiation (from one design to another) are obtained. Before examining the assumptions that underlie the GOMS method, and providing an illustration of model construction, some consideration needs to be given to the scope of the method.

2.1. Scope

Construction of a GOMS model, from a design specification, is appropriate when the task to be modeled is well learned by the user. Card et al. (1983) focused upon the domain of text editing, a domain where a task such as correcting a spelling mistake requires little conscious thought beyond identifying the word to be corrected. An important consequence of scoping models to well-learned tasks is that it is possible to assume that the user does not undertake any complex cognitive behaviors. Instances where a user must reason about their next action, solve a problem prior to action selection, learn something during interaction (that has a bearing on later parts of the interaction), and so forth, do not fall within the scope of an accurate GOMS model. In addition to this "well-learned task" assumption is a further assumption that tasks are performed without error. Predictions of, for example, the time it will take a user to correct a spelling mistake will be accurate (within limits) provided the user makes no slips or mistakes in carrying out the correction task. Finally, when considering a GOMS model's contribution to design, greatest utility is attained when: (a) two interface specifications need early evaluation; and (b) criteria for deciding between specifications largely rest on speed or time to complete different tasks. As will be illustrated, a GOMS model that describes interaction at the level of individual keystrokes can be used to generate predictions of the time to be taken when completing different tasks. The design specification that leads to the fastest time to complete a given task may then be considered the superior design, i.e. the design that should be chosen for implementation. GOMS is therefore a time/speed-oriented method for early evaluation. It offers little assistance to the process of diagnosing, for example, why a dialogue's syntax leads to repeated user failure. The dialogue

should have been well learned by the user prior to modeling, and thereby lead to error-less performance.

2.2. Assumptions

The GOMS method is suitable for modeling user cognition (mental behavior) for well-learned tasks. As for any cognitive task analysis method, a set of assumptions about user cognition are utilized in model construction, assumptions that may be termed a "cognitive architecture" (Anderson 1983). Card et al. (1983) outline the "Model Human Processor" (MHP) architecture as a basis for the GOMS method.

The MHP describes the user in terms of a set of interconnected processors and memories. A single "Perceptual" processor accounts for how the user becomes aware of visual and auditory signals. Drawing upon knowledge from experimental psychology, a uniform time to process a perceptual signal is suggested (100 msec). Processed signals, from each modality, are then transferred to image stores within working memory, visual, and auditory stores respectively. Associated with each image store (and memory) is an assumed time for a signal to decay (if it has not been transferred to Long-Term Memory (LTM)), and a limit on the quantity of information that can be stored (a limit of 5 letters may be stored in the auditory image store for 1500 msec, after which time decay will take place if those letters have not been transferred to LTM). A second "Cognitive" processor is then proposed to operate on all the contents of working memory. Accounting for all cognition in terms of a single cognitive processor is possible because of the well-learned task assumption, separate processors for problem-solving, reasoning, learning, and so forth, are unnecessary as such cognitive behaviors fall outside of the scope of the method. The cognitive processor is assumed to link incoming perceptual information with memories for actions, what is called the "Recognize–Act Cycle". Once a specified action has been established, the third and final "Motor" processor is assumed to convert a mental representation (for an action) into physical behavior itself. These assumptions, embodied within the MHP, inform the structure of particular GOMS models, and are operationalized largely at the Keystroke level of modeling, the level at which times are associated with discrete keystrokes.

3. THE GOMS METHOD

The method itself shall now be considered. Four basic components are assumed: goals; operators; methods and selection rules. A goal is defined as a specification for a desirable state of affairs to be achieved during interaction, and may be decomposed into sub-goals. In text editing, for example, a goal may be to correct spelling mistakes within a document. Here, the top level goal is expressed as "GOAL: EDIT-MANUSCRIPT", and may be decomposed into instances (or "units") of correcting individual spelling mistakes. Each instance of a correction may be expressed as "GOAL: EDIT-UNIT-TASK". Within the MHP, it is assumed that information enters the architecture via the perceptual processor, is processed and action thereby arises. Each unit task is therefore decomposed into "acquisition" and "execution" sub-goals: "GOAL: ACQUIRE-UNIT-TASK"; "GOAL: EXECUTE-UNIT-TASK". The goal structure may thus be represented:

GOAL: EDIT-MANUSCRIPT

- • GOAL: EDIT-UNIT-TASK
- • • GOAL: ACQUIRE-UNIT-TASK
- • • GOAL: EXECUTE-UNIT-TASK

Goals are attained by sequences of operators, which are elementary perceptual, motor or cognitive acts (following the MHP's assumptions). With a goal such as "GOAL: ACQUIRE-UNIT-TASK", a perceptual operator "GET-NEXT TASK" may be employed, which describes the search for the next spelling mistake. Once the mistake has been identified, the execution goal "GOAL: EXECUTE-UNIT-TASK" may be attained by further sub-goals of "HIGHLIGHT-WORD", "MODIFY-WORD", and the mental operator VERIFY-EDIT". Building upon the goal description above, integration of reference to operators takes the following form:

GOAL: EDIT-MANUSCRIPT
- • GOAL: EDIT-UNIT-TASK
- • • GOAL: ACQUIRE-UNIT-TASK
- • • • • GET-NEXT TASK
- • • GOAL: EXECUTE-UNIT-TASK
- • • • GOAL: HIGHLIGHT-WORD
- • • • GOAL: MODIFY-WORD
- • • • • VERIFY-EDIT

Goals may be attained by more than one sequence of operators, each sequence may be described as a method. The goal "HIGHLIGHT-WORD" may be realized using a "MOUSE-METHOD" or a "KEYBOARD-METHOD". The MOUSE-METHOD may be decomposed into: move mouse to word; double click on word. The KEYBOARD-METHOD may likewise be decomposed into: move cursor to word; hold down shift key; use arrow keys to move cursor over word; release shift key. These alternative methods for attaining the goal are then embedded within the expanding task description as follows:

- •••GOAL: EXECUTE-UNIT-TASK
- ••• GOAL: HIGHLIGHT-WORD
- ••• • [select GOAL: USE-MOUSE-METHOD
- ••• • • • MOUSE-TO-WORD
- ••• • • • CLICK ON WORD
- ••• • • GOAL:USE-KEYBOARD-METHOD
- ••• • • • CURSOR-TO-WORD
- ••• • • • HOLD-SHIFT-KEY
- ••• • • • ARROW-OVER-WORD
- ••• • • • RELEASE-SHIFT-KEY

Where more than one method exists for attaining a goal, rules of selection are needed to determine under which circumstances each method is employed. Given that the task is well learned, it is likewise assumed that such rules are well learned and reliably implemented without conscious thought of selection. In the case of the goal "GOAL: HIGHLIGHT-WORD", rules may be:

Rule 1: "USE-MOUSE-METHOD" is the default method
Rule 2: If the cursor is already next to the word to be highlighted, and your hands are on the keyboard, "USE-KEYBOARD-METHOD" should be employed.

Having considered the basic components of the GOMS method, let us now illustrate its use with a simple example.

4. GOMS ILLUSTRATION TO THE KEYSTROKE LEVEL

Let us consider the case of an ergonomist who has designed two competing dialogue specifications for a hardware firm's stock control application. Users of the application are to be well-trained employees of the firm, and the representative task to be considered here is that of specifying the details of a product as a means of querying stock details. Here we shall only be concerned with the
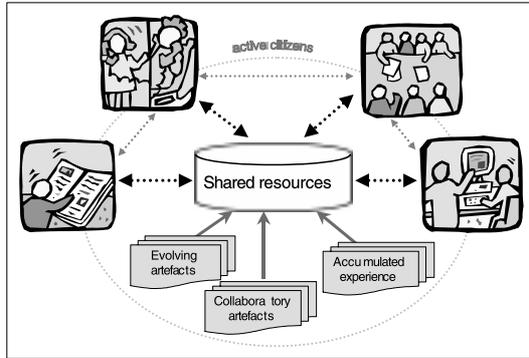
Figure 1. Text-box and command-line interface design specifications

dialogue for specifying the product, thus ignoring how the application presents the results of the query. In particular, we shall consider the dialogue for querying whether or not a 12-inch Stillson wrench is in stock (Newman & Lamming 1995).

Figure 1 shows the two competing dialogue specifications. For the text-box design, two generic data entry fields are presented for the input of details concerning a device and its size. In addition, an acknowledgment button is shown. For the command-line design, the command "cat" specifies that the query refers to a search of the catalogue, "still/d" specifies the first five letters of the name of the device being queried, with the switch "/d" indicating that the preceding letters refer to a device. Finally "12/s" specifies the attribute 12-inches, with the switch "/s" indicating that reference is being made to the size attribute of the device. Using the GOMS method, the following models may be constructed.

GOMS model for text-box specification:

    GOAL: QUERY-DETAILS-STILLSON-12
    •   GOAL: SPECIFY-CATALOGUE-DEVICE
    •   •   ACQUIRE-DEVICE-LABEL
    •   •   HOME-TO-MOUSE
    •   •   MOVE-MOUSE-TO-TEXT-BOX
    •   •   ACTIVATE-TEXT-BOX
    •   •   HOME-TO-KEYBOARD
    •   •   SPECIFY Stillson
    •   GOAL: SPECIFY-CATALOGUE-DEVICE-SIZE
    •   •   ACQUIRE-DEVICE-LABEL
    •   •   PRESS Tab
    •   •   SPECIFY 12
    •   GOAL: TERMINATE-QUERY

    •   •   ACQUIRE OK BUTTON
    •   •   HOME TO MOUSE
    •   •   MOVE-MOUSE-TO-OK-BUTTON
    •   •   PRESS-OK

GOMS model for Command-line specification:
    GOAL: QUERY-DETAILS-STILLSON-12
    •   GOAL: ENTER-CATALOGUE-MODE
    •   •   HOME-TO-KEYBOARD
    •   •   SPECIFY-CATALOGUE-COMMAND
    •   GOAL: SPECIFY-CATALOGUE-DEVICE
    •   •   SPECIFY Still
    •   •   SPECIFY-DEVICE-SWITCH /d
    •   GOAL: SPECIFY-CATALOGUE-DEVICE-SIZE
    •   •   SPECIFY 12
    •   •   SPECIFY-DEVICE-SIZE-SWITCH /s
    •   GOAL: TERMINATE-QUERY
    •   •   PRESS carriage return

Once such models have been constructed, the method provides a set of values for the time it should take a user to execute a range of different low level physical actions with the computer. The basic operators considered are:

time to make a keystroke (K);	
best typist (135 wpm)	0.08s
average skilled typist (55 wpm)	0.20s
average non-secretary typist (40 wpm)	0.28s
worst typist (unfamiliar with keyboard)	1.20s
time to point (P) with a mouse at a target	1.10s
time to home-in (H) on an input device	0.40s
time to mentally (M) prepare to specify an input	1.35s
time for device to respond (Rt) to an input	

(depends on device)

Given these tabulated values, and models that capture the frequency of each basic action with each dialogue specification, prediction of the times to carry out the query task with each design may be calculated. Generating such a prediction involves calculating the total number of each basic operation and multiplying by the time value provided for that operation. Therefore, for the text-box specification, the following prediction is generated (assuming the user is an average non-secretarial typist and system response is instantaneous).

GOAL: QUERY-DETAILS-STILLSON-12	
GOAL: SPECIFY-CATALOGUE-DEVICE	
ACQUIRE-DEVICE-LABEL	
HOME-TO-MOUSE	1H
MOVE-MOUSE-TO-TEXT-BOX	1P
ACTIVATE-TEXT-BOX	1K
HOME-TO-KEYBOARD	1H
SPECIFY Stillson	1M + 8K
GOAL: SPECIFY-CATALOGUE-DEVICE-SIZE	
ACQUIRE-DEVICE-LABEL	
PRESS Tab	1K
SPECIFY 12	1M + 2K
GOAL: TERMINATE-QUERY	
ACQUIRE OK BUTTON	
HOME TO MOUSE	1H

MOVE-MOUSE-TO-OK-BUTTON   1P
PRESS-OK   1M + 1K
Predicted time (Pt) = 3H + 2P + 3M + 13K
        Pt = (3*0.4) + (2*1.1) + (3*1.35) + (13*0.28)
        Pt = 11.09s

For the command-line specification, the following prediction is generated.

GOAL: QUERY-DETAILS-STILLSON-12
GOAL: ENTER-CATALOGUE-MODE
HOME-TO-KEYBOARD   1H
SPECIFY-CATALOGUE-COMMAND   1M + 3K
GOAL: SPECIFY-CATALOGUE-DEVICE
SPECIFY Still   1M + 6K   (a)
SPECIFY-DEVICE-SWITCH /d   2K   (b)
GOAL: SPECIFY-CATALOGUE-DEVICE-SIZE
SPECIFY 12   1M + 3K   (c)
SPECIFY-DEVICE-SIZE-SWITCH /s   2K
 GOAL: TERMINATE-QUERY
 PRESS carriage return   1M + 1K

   Predicted time (Pt) = 1H + 4M + 17K
         Pt = (1*0.4) + (4*1.35) + (17*0.28)
         Pt = 10.56s

In this command-line specification, the keystroke-level model assumes a single mental operator (at (a)) incorporates the specification of both device and switch (at (b)). The same is true at (c). Six keystrokes are shown at (a) to incorporate the space bar keystroke between "cat" and "still/d". It can therefore be seen that the keystroke level model predicts a half-second time saving when interacting with the command-line specification, as compared to the text-box specification. Such a time saving may well constitute the basis for a design decision in favor of the command line interface. An example of one such case, project "Ernestine" (Gray, John, Stuart, Lawrence and Atwood 1995) demonstrates how, when time is an important criteria, and the task is well learned (command language lexicon and syntax), GOMS accurately predicts that it may well be misleading to consider graphical user interface design specifications as a panacea for improved HCI usability.

REFERENCES

ANDERSON, J.R., 1983, *The Architecture of Cognition* (Cambridge MA: Harvard University Press).

CARD, S.K., MORAN, T.P., and NEWELL, A., 1983, *The Psychology of Human-Computer Interaction* (Hillsdale NJ: Lawrence Erlbaum).

GRAY, W.D., JOHN, B.E., STUART, R. LAWRENCE, D., and ATWOOD, M.E., 1995, GOMS meets the phone company: Analytic modelling applied to real-world problems. In R.M. Baecker, J. Grudin, W.A.S. Buxton and S. Greenberg (eds) *Human-Computer Interaction: Towards the Year 2000* (Morgan Kaufmann).

NEWMAN, W.M. and LAMMING, M.G., 1995, *Interactive System Design* (Harlow: Addison-Wesley).

WHITEFIELD, A., 1990, Human-computer interaction models and their roles in the design of interactive systems. In P. Falzon (ed.) *Cognitive Ergonomics: Understanding, Learning and Designing Human-Computer Interaction* (London: Academic Press).

Heart Rate as Strain Index

B. Kapitaniak

Maître de conferences à l'université Pierre et Marie Curie, faculté de
médecine Pitié-Salpêtrière, Paris, France

1. INTRODUCTION

The simplicity and the facility of receiving and constantly
monitoring the ECG signal, without any inconvenience for the
subject and with a low rate of interference, make it very popular
for the studies on work stations. Most often, the ECG is used for
calculating the average heart rate, which is one of the parameters
determining the index figures of the arduousness caused by the
work loads.

The advantages of the heart rate (HR) as a biological parameter
are numerous and undeniable. As far as methodology is
concerned, this parameter may be recorded without any
inconvenience for the subject, and continuously practically
without any time limit.

The heart rate is generated by the autonomous cardiac system
and then modified by an inhibiting action of the parasympathetic
vegetative nervous system or by an activating action of the
sympathetic system. This implies that any action of the vegetative
system has an effect on the heart rate: physical load, emotions,
noise, heat, general mobilization of the organism. All this makes
the heart rate a very interesting parameter for an ergonomist.

2. VALUES OF HEART RATE

In a subject, the instantaneous heart rate (calculated by taking
the inverse of the interval of each successive cardiac cycle)
fluctuates around an average counted over a minute. These
variations correspond to continuous adjustments that follow
breath rate, vessel tonus, thermoregulation, the control of blood
pressure, and, in all likelihood, other function of the vegetative
system. The spectrum of the heart rate includes the frequencies
of several corresponding oscillatory phenomena whose variations
result in this physiological arrhythmia called sinus arrhythmia.

The first records on the sinus arrhythmia (Krough) date back
to the beginning of the twentieth century and refer to the influence
of the breath rate on the heart rate. In 1938 Djourno worked out
a method of measuring the instantaneous heart rate.

The term "sinus arrhythmia" was introduced in the 1950s
to emphasize the physiological nature of that phenomenon. It
describes the arrhythmia occurring without any disturbances
of conductivity, i.e. recorded on perfectly healthy people. The
arrhythmia is noticeable only when observing the
instantaneous heart rate and at rest it can reach up to a 25%
deviation from the average heart rate, which is a relatively
constant value for a given functional state of the body. There
is a good correlation between the acceleration of the heart
rate and the reduction of the arrhythmia. In other words, the
stronger the impact of the adrenergic system on the heart, the
smaller the arrhythmia.

The average heart rate follows a circadian rhythm. It decreases
during sleep by 10 to 15 bpm (beats per min) in the first four
hours, but remains higher in day sleep than in night sleep.

2.1. Value at Rest

The inter-individual differences of heart rate at rest are very wide.
Average values contained between 40 and 100 bpm are considered as
normal, whereas in most cases these values range from 60 to 90 bpm.

The heart rate at rest rises by approximately 5 bpm between
the age of 25 and 60. In pregnant women the heart rate at rest
gradually rises until in the ninth month it reaches values higher
by 30% of the normal rate.

The heart rate may rise by 10 to 12 bpm in the two hours
following a meal. Coffee or tea increase the heart rate even more.

2.1.1. Posture influence

The heart rate is linked to the posture and its variations. The
heart rate is 10% higher when sitting than when lying, and 5 to
15% higher when standing. The shifting of position, both in work
and at rest, is marked by a transitory acceleration or slowing
down determined by the acceleration of the body mass
respectively from head to feet (quick shifting from crouching to
standing) or the opposite.

2.1.2. Environment influence

The environment changes the level of the heart rate at rest. Thus
the exposition to heat leads to an increase of the heart rate in
relation to the increase of the cutaneous blood flow necessary for
thermoregulation.

Exposure to an unknown noise of an intensity over 90 dB
provokes a progressive acceleration of the heart rate reaching 6
to 8 bpm in a few hours. In addition, a rapid transfer to high
altitude provokes an acceleration of the heart rate at rest due to
the decrease of oxygen content in the air. This persists for only a
few days if the altitude does not exceed 2 000 meters, and
disappears after a period of acclimation of about 10 days. For
higher altitudes it persists continuously.

2.2. Values during Physical Work

During physical work the heart rate increases systematically
whatever the task. However the mechanisms of the increase are
different according to the type of effort put in.

2.2.1. Dynamic work

During dynamic work an increase of the cardiac output
proportional to the power of the exercise and to the oxygen
consumption may be observed. It is accompanied by a decrease
of peripheral resistance making blood flow easier and limiting
the increase of blood pressure, which is insignificant during a
moderate dynamic effort and which concerns only systolic blood
pressure during an intense effort.

The increase of the cardiac output first results from a fast
increase of stroke volume at the beginning of the exercise, and
then essentially from cardiac acceleration. For moderate exercises,
therefore, there is an almost linear relation between the heart
rate and the power of the exercise, which allows deduction of
this power from the heart rate and knowledge of the subject's
physical capacity.

Since the heart rate increases more for arm work than leg
work with the same energy consumption, one must be quite
prudent when trying to assess the energy cost of a task
according to the heart rate. This can only be done with a
sufficient degree of certainty if an individual calibration was

realized during a simultaneous recording of the two physiological magnitudes, and moreover during an exercise which involves the same muscle groups as those prompted during the studied task.

2.2.2. Static work
During continuous static work involving a force higher than 20% of the maximal force (maximal voluntary contraction — MVC), the peripheral resistance increases because of the compression of the vessels by the contracted muscles. This increase leads to a significant reduction of the venous return and a defect in cardiac filling. At the same time, the muscles working in hypoxia trigger an adrenergic reflex which results in cardiac stimulation; this, in turn, increases the heart rate and blood pressure. This increase depends on the relative load supported by the muscle in terms of MVC percentage. These reactions are almost independent of the muscular mass involved in the contraction, which bears out their reflex origin.

2.2.2.1. Postural load
The professional work executed in uncomfortable postures generates an increase of the heart rate — standing, leaning forward, standing on tiptoe, kneeling or lying on one's back in restricted spaces, or work carried out with arms outstretched.

A postural constraint which involves levels of force lower than 15% MVC provokes cardiac accelerations between 100 and 120 bpm.

2.2.2.2. Static load
The motionless holding of an external load by muscles in isometric contraction is a very frequent situation in industrial work. At the end of the time limit of static holding, forces between 50% and 60% MVC, values of the heart rate between 120 and 130 bpm, and blood pressure between 150/100 and 170/120 are observed. For higher percentage of the MVC (> 70%) the values recorded at the end of the time limit are even higher.

2.3. Values Involving Emotive Stress
Among the physiological mechanisms involved in emotional reactions, the secretion of suprarenal catecholamines is systematically evoked. One of the consequences that may be easily observed is an acceleration of the heart rate without any relationship to oxygen consumption or even to the increase of cardiac output.

All stress situations lead to a cardiac acceleration: work under time constraint, the subject's anxiety, the occurring of an accident or a danger, increased responsibility, fear or physical pain. Cardiac accelerations up to 115 bpm were observed in dental patients during painful operations.

Conversely, when the cognitive work load is moderate and the work has become monotonous, the heart rate often slows down.

3. HEART RATE MEASUREMENT
3.1. Historical Methods
The first instrumental measurements, confined to laboratory studies, were recorded by Boas as early as 1928. They were based on electrocardiographic detection (ECG) of the cardiac rhythm. But it was Brouha who, in the 1940s, introduced measurements

of heart rate into industry. His field methodology was based on taking the pulse (cardiac pulsation perceptible in the wrist) with no instrument other than the chronometer, and on drawing a cardiac recovery curve established during a three-minute pause, thus constituting only a minimal inconvenience in the operator's work. The Brouha method still remains useful when no modern equipment is available.

In the 1960s and '70s, measurements in the field were carried out thanks to telerecording or the Holter system. Cardiac telerecording consists of a radio carried by the subject transmitting the ECG to a receiver. The result may be recorded on paper, onto a magnetic band and processed later, or on-line.

The "Holter" is a magnetic or hardware recorder which enables long recordings of an ECG signal of up to 48 hours, which can later be processed using various programs. The device is used in cardiology for clinical purposes, but may also be used on work stations. Its main drawback in ergonomics is the cost of the equipment and the training required for the processing of data.

3.2. Current Methods
Today, the heart rate is most often measured with monitors that allow automatic counting, displaying, and recording of the heart rate. The different appliances use either electrodes attached to the skin and connected with cables, or contact electrodes fixed with a chest belt on the thorax.

As the ECG signal picked up by the electrodes is recognized, the device counts the number of heart beats in a predetermined period of time or, depending on the appliance, the time elapsed for a specific number of beats. The choice of the counting time (integration constant) is adjustable between instantaneous heart rate and average rate in 1 minute. The appliance records only the calculated values. It therefore disposes of an autonomy which depends on the choice of the integration constant — the longer it is (for example, 1 minute) the longer the autonomy (24 to 48 hours), whereas for the instantaneous heart rate it seldom exceeds 3 hours.

The advantages of the system are numerous: continuous recording on long duration, simplicity of use, reduced dimensions, insignificant weight, low cost. The main drawback is the lack of control of the quality of the recording (detection of artifacts and parasites), since the original signal is not recorded. The technology of heart rate monitors is evolving rapidly, tending towards miniaturization, increase of recording autonomy, larger choice of time constants.

Currently, there are a dozen heart rate monitors available on the world market. The choice of an appliance must be based on its reliability, its range of choice of the integration constant (minimum required for instantaneous HR: 5 or 15 s and 1 min), its autonomy (minimum required for instantaneous HR: 4 hours and 24 hours with a constant of 15 s), and its cost. The cost of the electrodes must also be included in the running costs when the appliance uses disposable electrodes. This additional cost does not exist in appliances that use permanent contact electrodes.

3.3. Practical Technique of Measurement
The fact that heart rate varies according to numerous psycho-physiological factors makes it compulsory to watch the subject's activity directly during the recording. It is almost impossible to distinguish a cardiac acceleration of metabolic origin from one

of reflex origin using the recorded pattern. A precise description of the subject's activity is therefore necessary in order to interpret the results of his or her heart rate.

For each recording of the heart rate during professional activity one must choose the recording parameters which best correspond to the studied activity. Among these, the choice must consider above all the integration constant and the recording time.

3.3.1. Integration constant

Generally the longer the integration constant, the fewer rapid fluctuations appear. The time constant should be chosen according to the character of the constraint. To study the strain due to the execution of short time rapid gestures (for example, feeding a welding robot with a 24 s cycle), an instantaneous heart rate or a constant of 5 s must be chosen.

When the physical task includes important elements of static load, which lead to cardiac accelerations of reflex origin, the time constant must also be the shortest possible.

For studies concerning a strain due to mental load or stress situations, it is better to record the instantaneous heart rate whereby it is possible later to relate the beginning of an acceleration to a precise event and even to analyze the sinus arrhythmia.

The choice of a time constant implies the choice of a method of observation during the recording of the heart rate. The precision level of the task analysis depends on the task itself. If the content of the task is dynamic, general, and prolonged (for example, manual handling in loading a truck), it is sufficient to observe the physical characteristics of the work accomplished (tonnage, unitary weight of the carried load, walking speed, distance, height of lifting, other events). In such cases, the method best adopted is the recording of the average heart rate for 1 minute.

3.3.2. Recording time

The choice of the recording time depends on the constraint to be studied. The execution of dynamic, general, prolonged and repetitive work with short cycles could be studied with a recording limited to approximately 1–2 hours. The more heterogeneous and complex the task, the longer the recording time must be. The latter must always encompass the time corresponding to the whole set of various operations representing an entire productive set.

4. TREATMENT OF RESULTS

4.1. Absolute Values

The relevant values of the HR are, on the one hand, the average values of the period of work considered, which give an idea of the average workload imposed on the heart, and, on the other hand, the acceleration peaks which reveal the occasional risk sustained by the heart.

The average values may be calculated on complete work cycles provided that the same cycle is repeated continuously and that the profile of the HR in the cycle is not a slope. If, during a cycle or shift, the HR constantly increases, the average value will give an imprecise idea of the workload, in which case the final value is more relevant.

The occasional peak values allow the risk that a workload presents to the operator's health to be measured. Even if such
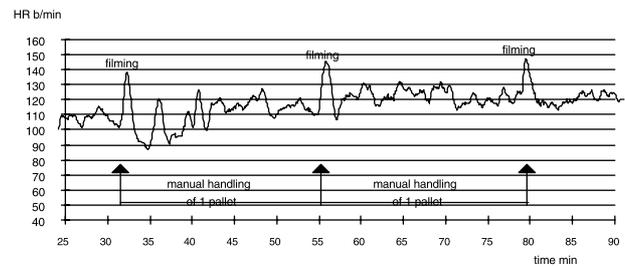
Figure 1. Heart rate during manual handling of small parcels (5 to 10 kg)

peaks are followed by adequate rest periods, they must be limited in value and duration as violent acceleration may be dangerous for the integrity of the cardiovascular system.

Figure 1 illustrates the changes in heart rate during the manual handling of small parcels in the building of pallets for supermarkets. The peaks correspond to the task of filming the pallet with the roll for stabilization.

4.2. Cardiac Cost

Brouha (1963) has introduced in work physiology the concept of *cardiac cost* by analogy with the energy cost of a task. The cardiac cost (CC) is defined as the difference between the HR value during a task and at rest.

4.2.1. Rest reference

The value at rest may be taken in different ways:
- in sitting position in the same condition as those at work, recorded over a period of approximately 15 minutes (Brouha 1963);
- for the study of work stations with recordings of long duration (> 8 h), Malchaire has suggested that the lowest value observed during the recording should be taken as a reference;
- for recordings of 24 h duration, Chamoux has proposed that the value of the fifth centile should be taken as a reference.

Despite all its drawbacks, Brouha's classic reference seems to be the most relevant for calculating the cardiac cost, even if during the shift day it is possible to note negative costs. It takes into account the subject's state at the beginning of the work and allows a better assessment of the effect of the task on the cardiac acceleration. The other references are difficult to obtain for they require recordings over very long periods and their use does not exclude the influence of the variability of the measured magnitude.

4.2.2. Relative cardiac cost

In order to limit the influence of differences in the physical capacities of subjects the relative cardiac cost (Ccr) may be calculated. The Ccr is a cardiac cost expressed in percentage of the maximal theoretical cardiac cost (CCmax) calculated from the maximal theoretical heart rate (HRmax). The proposed equation to calculate HRmax is the simplest and no more imprecise than other more complicated equations.

HRmax = 220 – age
CCmax = HRmax – HRrest
CCr = CC/CCmax * 100%

4.3. Sinus Arrhythmia

Many authors have expressed the opinion that the sinus arrhythmia is a good indicator of stress. A number of ergonomic studies performed on occupations as diverse as air traffic controllers, highway truck drivers, and surgeons confirm the practical usefulness of this parameter as far as the measure of the stress is concerned.

The methods of assessing the sinus arrhythmia applied so far have been mainly designed to test its behavior under the emotional loads during which the average heart rate does not change substantially. A simultaneous analysis of the deviations of the instantaneous heart rate from its average, and the beat-to-beat deviations are usually applied. Using Fourrier's analysis of the heart rate, the following parameters are considered: the frequency of maximal and minimal instantaneous values of the heart rate, the standard deviations from the average values, the values of the differences between two consecutive heart beats, the values of the positive and the negative differences between the consecutive impulses, and their distribution.

5. INTERPRETATION OF RESULTS

The interpretation of the results of recording the heart rate as an index of strain is all the more difficult as the constraints likely to have an effect on the vegetative nervous system are numerous. These difficulties even increase if the intensity of each constraint is moderate. This means that it will be impossible to distinguish the part of the cardiac response which is due to slight physical effort from that due to passing emotion, brief exposure to a noise or moderate heat, with all of them acting simultaneously.

Figure 2 shows the heart rate reached during an exercise on a circus trapeze by an artist revolving on his head. The same physical exercises are executed on a trapeze at 0.5 m and 5 m above the ground during a rehearsal, and at 5 m above the ground during a show. The only explanation for the differences observed during the same physical exercise is the emotional reaction.

When a heart rate occurring in conditions where a subject is exposed to multiple constraints is 40 b/min higher than the value observed at rest, the only observation that can be made is that the strain is excessive, but it is impossible to determine a dominating factor responsible for this excess.

5.1. Basis of Reference Values

The thresholds proposed in the norms or recommendations are first given by way of example and are likely to be modulated. They
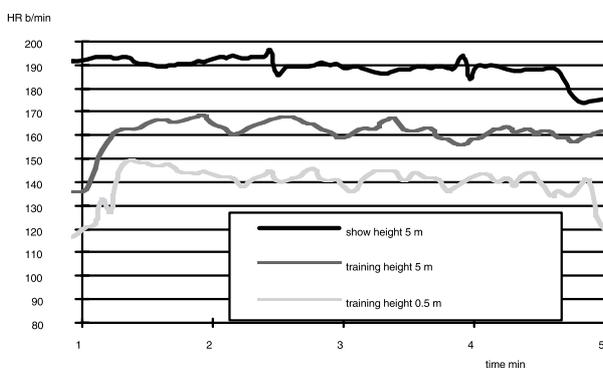
Figure 2. Heart rate during trapeze exercises

must not be considered by the ergonomist, or by those responsible for the production, as thresholds which can be "reached", but rather as limits from which one must try to keep as far as possible.

These thresholds can only be indicated according to the dominating constraint, given the different mechanisms involved by the different constraints. In the case where the constraints are multiple, only a competent ergonomist can interpret the result correctly. This implies of course that all HR recording is accompanied by a direct observation of the activity.

For the cardiac costs observed during the general dynamic exercise, the choice of the reference at rest determines the interpretation of the results. If Brouha's method is used, the average cardiac cost of the shift day thus calculated should not be higher than 30 to 35 bpm. For short periods of work, it is better not to go beyond a cost of 45 to 50 bpm. This is a consensus based on numerous experiences in the field, but with no rigorous experimental demonstration. If the reference of the fifth centile is adopted as the HR of rest, these values must be revised up. The average CCr observed during a shift day should not be higher than 30%, whereas with very short efforts it could go up to 50% provided the period of intense work is followed by an adequate recovery period.

The definition of a "short period" is as much a matter of consensus as the threshold values. In principle, the threshold values proposed for the workshift are normally fixed at 8 hours, and the values for "occasional" work do not exceed 30 minutes per shift (< 6% of work time).

5.2. Reference tables

Table I. Grid for assessing the general dynamic workload from the heart rate for occasional tasks (performed during 8 hours of work) and punctual tasks (performed for less than 30 minutes out of 8 hours of work).

	Classification of cardiac strain					
	Continuous (8 h)			Occasional (< 30 min)		
Class	HR	CC	CCr	HR	CC	CCr
Rest	80	0	0	80	0	0
Slight	100	20	20%	110	30	30%
Moderate	115	35	30%	130	50	50%
High	130	50	45%	150	70	65%
Very high	150	70	60%	170	90	75%

Table II. Limit values of cardiac costs not to be exceeded according to the principal work constraint, either in average values, or in peak values.

Principal constraint	Limit Values			
	Continous		peak	
	CC	CCr	CC	CCr
Dynamic	35	30	50	50
Static	Inapplicable		25	20
Stress	20	15	40	35

BIBLIOGRAPHY

BROUHA, L., 1963, *Physiologie et industrie* (Paris: Gauthier-Villars).

KAPITANIAK, B., 1995, *Choix de méthodes d'études de postes de travail* (Octares éd.), Toulouse, *Performances*, n° hors-série.

LehmaNn, G., 1955, *Physiologie pratique du travail*. Les éditions d'organisation, Paris.

MONOD, H. and KAPITANIAK B., 1999, *Précis d'Ergonomie*. Masson.

Human Body Positioning Analysis

S. A. Lavender and G. B. J. Andersson

Department of Orthopedic Surgery, Rush-Presbyterian-St Luke's Medical Center, 1653 West Congress Parkway, Chicago, IL 60612, USA

1. INTRODUCTION

Positioning analysis refers to the evaluation of the body's posture that occurs during a work task. The analysis of work postures is a fundamental concept in the study of industrial and office ergonomics. Working in "awkward" postures is considered one of the classical risk factors associated with upper extremity cumulative trauma disorders (Putz-Anderson 1988, Sommerich *et al.* 1993). Although it must be remembered that posture is a time-dependent quantity and changes with movement. At some point as movements become faster it makes little sense to talk in terms of limb or body part's (segment) posture or a joint's position and instead focus on the dynamic characteristics of the motions involved as these potentially relate more to the risks inherent in the activity. With slower or even static work activities the posture to a large degree impacts the mechanical loads experienced by the muscles and joints within the body.

The objectives of this chapter are threefold. First, to review why position (body's posture) is important. Second, to review the various ways postures and or joint positions can be described and quantified. And third, to discuss in general terms how postures can be controlled through workplace interventions.

2. WHY IS POSITION (BODY POSTURE) IMPORTANT?

In a biomechanical sense, position data are used to describe the orientation of each individual body part being studied within a two- or three-dimensional space. Operationally, these data may be obtained qualitatively through categorizing observed postures (checklist), or quantitatively using goniometers, or by using video-based, sonic-based or electromagnetic-based motion measurement systems. Qualitative observations and goniometers are used to assess postures directly but in doing so fail to provide information about the body's orientation relative to a world coordinate system. More advanced motion measurement systems provide the orientation of a body part relative to a known global coordinate system. The difference in orientation between a body part and an adjoining body part can be determined, and it is this difference that defines the posture referred to in the orthopedic and ergonomics literature.

2.1. Importance of Global Orientation

In addition to representing the posture, the global orientation becomes important if one wishes to study the load placed on a connecting joint between two adjacent body segments due to gravitational loading. Gravity, depending on the mass of the body part being studied, can be a major contributor to the moment acting on an articulation within the body during a work activity. When an individual is performing a work task in a posture that does not change over time (or changes very slowly) the mechanical loading of the affected joints is in large part due to the 'external' moments created by each body segment's orientation in space relative to gravitational and other applied external forces. The latter might include the weight of a tool in the hands or forces due to acceleration of the body ('G' forces). In order for the posture to be maintained, the net external moment must be neutralized by the combined internal moments stemming from muscles and other passive tissues (ligaments, disc fibers, inactive muscle, etc.). Because of the short lever arm over which muscle forces act, the muscle forces needed to stabilize the body are quite large relative to the applied external forces and therefore are the primary contributors to the mechanical loads placed on the body's joints. Thus, by analyzing the position of body segments relative to the world coordinate system, the external moments due to gravity can be determined. And it is these external moments that are determinants of joint loading.

2.2. How Posture Affects the Muscle Physiology and Biomechanical Relationships

Posture defines many of the biomechanical relationships within the body including the muscle length and the length of a muscle's lever arm. Muscle length determines the overlap that exists in the actin and myosin fibers within the sarcomeres. There is an optimal amount of overlap in these fibers when it comes to developing force within the muscle tissue, thereby making strength a quantity that is dependent upon posture. This variation in muscle strength as a function of muscle length is known as the "length–strength" relationship within muscle. In addition, strength is dependent upon posture because the length of the lever arms over which muscles exert force change with the orientation of one body segment relative to another. For example, the erector spinae muscles have been shown to have a longer lever arm (greater mechanical advantage) when the spine is maintained in a lordotic posture (Tveit *et al.* 1994). Thus, both the muscle length and its mechanical advantage change as a function of body posture.

To some degree this is simplistic in that the length of single joint muscles, for example, the brachiallis can easily be inferred based on a single postural measurement. For muscles that cross multiple joints the relationship is the same in general terms, however, the length is dependent obviously on the orientation of all joints involved. The length of the Biceps is dependent upon forearm rotation, elbow flexion, and shoulder motion (glenohumeral rotation). The lengths of the flexor digitorum profundus muscles are dependent upon the wrist posture and the interphalangeal joints. Each of the joints serves as a pulley for the long tendons extending out to the fingertips. When the pulley system is altered, for example, through carpal tunnel surgery in which the transverse carpal ligament is severed, the relationship between muscle length and joint posture changes such that a worker may be forced to use muscles that are nowhere near their optimal length.

As muscles approach their maximal length more and more of the internal forces are developed through the passive tissues. For example as the trunk is fully flexed Floyd and Silver (1951) reported that the erector spinae muscles became electrically silent. This means that the passive tissues developed the stabilizing internal moment, namely ligaments, passive muscle and tendon stretch, and disc fibers. If repeated often this loading of the passive

tissues may leads to laxity within the structures responsible for stabilizing the vertebral column. Where the passive tissues are not capable of producing the necessary internal moment to stabilize the body part, muscles still must be recruited. Given that the muscles are operating from a weak position due to their stretched position, the risk of overexertion injury is increased near the end range of motion.

2.3. Interplay between Posture and Motion

When motion is involved, the contribution of body segment orientation to the net external moment changes over time. Now the external moment is composed of the forces due to the body segment's linear acceleration ($F = m.a$) and the inertial properties of the body segment as it is being rotated, in addition to the static moments discussed above. Hence, with motion, body positioning while still a contributing factor contributes less to the net moment and the overall anticipated joint loading. The faster the motions the more important the dynamic terms become in terms of mechanical loading and tissue injury (Marras *et al.* 1993, 1995). Alternatively, when motions are very slow the inertial and acceleration components may be negligible, and hence the posture takes precedence in determining the resulting joint loads and the potential injury risk, particularly near the end range of normal motion for an articulation.

3. WHAT ARE "AWKWARD" POSTURES?

Ergonomic texts often refer to "awkward" postures as being risk factors for cumulative trauma disorders. This is distinctly different from awkward motions. For in the latter the posture is a time varying quantity and only held for an instant in time. Therefore, concern regarding awkward postures applies primarily to static work tasks. But what makes a posture "awkward"? Clearly, it takes more than being static for a posture to be awkward. Someone working with no shoulder flexion or abduction can be working in a static (non-changing) posture. However, few would contend that having the arms hanging by the side is an indicator of an ergonomic problem. Conditions often associated with awkward postures are:

1. Joint rotations near the end range of motion.
2. Those that require compound rotations of a joint.
3. Those that require muscle force to maintain.

By themselves each of these factors can be shown to contribute to the awkwardness, however, none define the condition exclusively. For example, we think of extreme wrist flexion or extension as being awkward. However, for the knee or elbow the 'neutral' posture is very near one end of the range of motion. Moreover, we would consider 90° of shoulder abduction as awkward, yet this is not near the end range of motion. Compound rotations, as the name implies, refers to rotations in two or three planes of motion. For example, wrist flexion combined with ulnar deviation while not at the end range of motion would be considered an awkward posture. Postures that result in substantial gravitational loading, as with 90° of shoulder flexion, require muscle force to maintain. Over time the muscles can be expected to fatigue thus leading to discomfort on the part of the worker. Again, however, this condition is not necessary or sufficient by itself to define awkward. The elbow can be flexed 90° while holding a tool or working at a keyboard. Even though there is gravitational loading, and hence, muscle recruitment, this posture

would not be considered awkward. In essence, awkward postures can best be defined as static postures in which one or more of the listed conditions exist.

Awkward postures are cause for concern as these are frequently cited as precursors to injury development in occupational settings. Work postures that require neck flexion, trunk flexion, trunk lateral flexion and twisting, shoulder flexion, shoulder abduction, wrist flexion, ulnar deviation, have all been associated with discomfort or injury reporting in the workplace (Tanzer 1959, Maeda 1977, Armstrong *et al.* 1982, Hagberg 1984, Masear *et al.* 1986, De Krom *et al.* 1990, Punnett *et al.* 1991, Winkel and Westgaard 1992, Armstrong and Lackey 1994). Figure 1 shows a simple model illustrating that the physical characteristics of the work site, the individual employee's characteristics, and the psychosocial work environment all contribute to the work posture attained while performing a particular work task. Thus, ergonomic evaluations of the work tasks need to document the postures observed, preferably in a quantitative fashion, and investigate the workplace, individual and psychosocial factors that may be responsible for the observed work postures.

3.1. How Long are Awkward Postures Held?

Two key parameters of the work posture that affect an employee's ability to perform the work task(s) are the duration of the posture and the rest or recovery time between the work postures. Figure 2 shows the hypothetical interaction between the rate of fatigue, the degree of postural deviation, the duration of a work task, and the recovery time between work tasks. The longer a posture must be held (duration) the faster and individual will fatigue. The rate of fatigue is also dependent upon the degree of postural deviation, with greater deviation from a neutral relaxed posture leading to an increased onset rate of fatigue. This is primarily due to the increased muscle force that is most often associated with greater postural deviation. The time between work tasks allows muscles responsible for maintaining the body posture to rest and recover. The greater the recovery time between work tasks the slower the onset of fatigue. Therefore, any analysis of postures observed in the workplace is not complete unless these two parameters are also recorded. While the link between muscular fatigue and injury is not necessarily known at this time, the link between muscular fatigue and discomfort is pretty
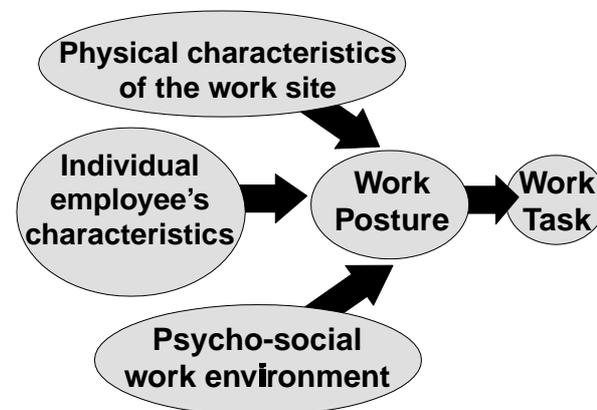
Figure 1. Potential factors accounting for the body postures observed in task performance.
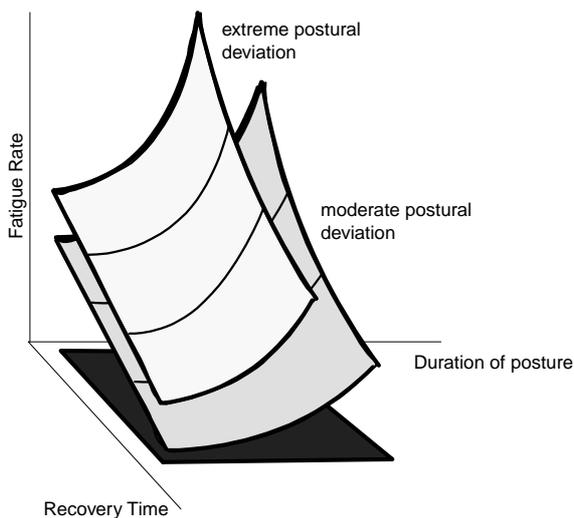
Figure 2. A conceptual model showing the theorized interaction between postural deviation, posture duration and recovery time.

straightforward. Fatigue has generally been associated with reduced productivity, work quality, and attentiveness, in addition to an increased accident potential.

4. DESCRIBING AND QUANTIFYING POSTURES AND OR JOINT POSITIONS

Postural descriptions usually form the basis of an ergonomic analysis. Often work postures, or ranges of motion in dynamic activities, serve as the flag that captures the ergonomist's eye. This is because, in general, the more a posture deviates from that of a relaxed standing posture with the arms at the side, the more loading experienced by the joints, tissue compression, and or force required in the muscles. Thus, when an ergonomist tours a facility and sees individuals working with the hands above shoulder level, the back bent or twisted, or the wrist extremely flexed, immediately there is cause for concern. This is especially true if the postures are held for any duration. If the postures are transitory as part of an overall motion, then it is the end range of the motion that will capture the ergonomist's eye. Once observed, the work postures need to be described so that risk assessments can be made and the observations communicated to others.

Several techniques are available for describing and quantifying postures. They range from simple gross descriptions, for example pinching or pressing (Armstrong *et al.* 1979), to categorical descriptions joint angles (Armstrong *et al.* 1982, McAtamney and Corlett 1993), to full quantitative measurement of three-dimensional angles. If measuring angles, even coarsely, the first and foremost requirement is the selection of a measurement system. In ergonomics, postures are most often described in terms of the orthopedic angles as defined by the American Academy of Orthopedic Surgeons (1965). This system defines each joint's posture by evaluating the angle between adjacent body parts. Other systems quantify postures by determining each body segment's orientation relative to a global coordinate system. This type of postural description is required to use the 2D and 3D Static Strength Prediction Models developed by the University of Michigan (Chaffin 1970, Chaffin *et al.* 1977).

The second major issue is the degree of accuracy that is required. In an ideal world we would always collect extremely accurate postural data. However, in reality increased accuracy comes with an increased price. The cost of measurement equipment, data collection and processing time, and training on the part of the individuals assessing the workplace may make that added accuracy gained by more precise measures not cost effective. In general, however, the better you can describe the problem that exists the better chance that your will solve the problem. Thus, the more quantitative the description the better the reading of the true problem.

In deciding on the appropriate workplace assessment method the evaluator should keep in mind how the assessment will be used. Qualitative descriptions of postures may be adequate when the assessments do not have to be communicated to others, and budgets for workplace interventions do not have to be justified. But in most cases one's observations need to be summarized and communicated to management in a form that allows all to understand the seriousness of the problem. Reporting that a task needs to be modified because the operator "works in a posture with flexed wrists" may not adequately communicate the problem and the need to invest the money in correcting the problem. By quantifying, the same report could have communicated that the operators, on average, exhibited 50° of wrist flexion. Not only does this accurately describe the nature of the problem, but it also provides a benchmark by which future intervention efforts can be compared. Goals can be developed for the intervention process, for example, that the average wrist flexion be reduced to < 15°. Mock-up workstations can be created before the resources are invested to ascertain whether the postures were controlled by the proposed intervention by comparing pre and post measurements.

When measuring orthopedic angles a goniometer can be used to quantify postures. However, it is imperative that the measurements be obtained without interfering with the work process. Therefore, most of the time postural measurements are made from videotapes rather than from the observed individuals directly. This introduces a number of errors into the process however. Ideally the video camera is oriented in a plane orthogonal to the plane of measurement and adjusted such that the center of the image corresponds to the articulation of interest. With two-dimensional motions, for example lifting tasks within the mid-sagittal plane, locating the camera in an orthogonal plane is a possibility. High contrast adhesive markers attached to the limbs and the torso facilitate the extraction of measurements from videotape by providing consistent reference marks on the limbs and joint center locations. In reality most studies need to quantify the posture of multiple body parts at the same time thereby leading to wider camera angles and increased error in measuring postures further from the center of the video-camera lens. One of the challenges facing the ergonomist is where to position the camera such that all the articulations of interest are visible at the same time and are not blocked by machinery or other equipment in the workplace.

With tasks that create postural deviations in three dimensions it becomes very difficult to accurately measure postures from the screen directly. Because we can extract the necessary information from a two-dimensional image on the video monitor to simulate the observed posture. The ergonomist is then better off measuring the simulated posture than attempting to measure from the video

monitor directly. The use of multiple video cameras increases the accuracy of three-dimensional postural descriptions. Wherever possible a synchronization signal is used to insure that video data are describing the same motion or posture. In a laboratory setting when an individual performs a task within a calibrated region of space, the data from two or more video cameras can be obtained and integrated to resolve the three-dimensional coordinates of the body segments. But these systems are very difficult to use in normal work environments. More recently magnetic based systems have become available that allow three-dimensional posture and motion measurement of multiple body segments without relying on video data. Unfortunately, because these systems are susceptible to interference introduced by metallic objects in the work environment, they are of limited use outside the laboratory.

When the study is limited to specific body parts, for example the spine or the wrist, electrical goniometers or inclinometer may be used. The simplest version of an electrogoniometer consists of two rigid tangs that can be affixed to the adjacent body segments

Figure 3. Electrical goniometers used to quantify postures and motions of the wrist (a) when performing a keyboarding task (photo courtesy Richard Marklin) and on the spine (b) when performing a beverage delivery job.

hinged by a potentiometer. These devices are usually strapped or taped to the skin such that the axis of rotation aligns with the joint's axis of rotation (Figure 3a). For the elbow an electrical goniometer designed to measure flexion should be should be positioned such that it is on the lateral side of the upper and lower arm with the potentiometer centered over the joint between the humerus and the radius. A single axis potentiometer works well if the joint can be conceptualized as a hinge joint. With the knee, for example, this assumption may not be valid as there can be three-dimensional motion in addition to a shifting center of rotation as the femur tibial contact point changes with knee flexion. Thus, for extremely accurate measurements a more complex set of goniometers may be needed (Chao 1976).

Trunk postures and motions have been measured using various types of instrumentation. Nordin *et al.* (1984) reported a flexion analyzer that was essentially an inclinometer strapped on the back. This device provides data indicating the duration of the working time that was spent in five flexed postures consisting of 18° intervals between 0 and 90°. This device is useful for obtaining an overall picture of the trunk postures required in a job across many tasks and work cycles. The device clearly distinguished the difference between dentists, who flexed forward only moderately, and warehouse workers who worked in deeply flexed postures. The same device was later used by Magnusson *et al.* (1990) to study the trunk postures in assembly line workers performing highly repetitive work. Snijders *et al.* (1987) reported on the development of a device that can be worn beneath the clothing and can collect three-dimensional trunk postural data throughout a workshift of eight hours or more. The data are stored on a multichannel tape recorder and can be analyzed later. Marras *et al.* (1992) have developed the Lumbar Motion Monitor, which in addition to measuring the instantaneous posture of the lumbar and lower thoracic spine in three dimensions, provides the angular velocities and accelerations associated with the movements. The LMM gives very precise information regarding the postures and motions from each individual activity sampled. If one is looking at primarily static task then the postural data are the primary focus of the analysis. If one is looking at a dynamic material handling task then the motion parameters such as range of motion, velocities, and accelerations become the focus of the analysis. This device is now commercially available (Figure 3b) as are other devices that perform a similar function. The key to these devices is the software used to collect and analyze the workplace data. It needs to allow the observations to be grouped according to the tasks and subtasks observed and allow the user to enter relevant workplace data, for example reach distances, weights handled, and working heights, with each observation recorded. This insures that the motions and the workplace parameters responsible for those postures or motions are linked within the database. This is essential when analyzing the data to determine the best type of workplace interventions.

5. SUMMARY

Analysis of work postures, whether qualitatively or quantitatively performed, is a crucial step in the identification of ergonomic problems. The work postures define the strength demands and the degree of muscular loading experienced by the affected articulations and are frequently associated with discomfort or injury in the workplace. Various means are available for

performing postural analysis. The more quantitative the description and the more employees studied the better the chance that the physical characteristics of the work site responsible for the posture will be identified. Once work posture problems are identified several options are available to control work postures, or at least, the total duration of an employee's work experience in the identified postures.

REFERENCES

AMERICAN ACADEMY OF ORTHOPEDIC SURGEONS, 1965, *Joint Motion — Method of Measuring and Recording* (Chicago: American Academy of Orthopedic Surgeons).

ARMSTRONG, T.J., CHAFFIN, D.B. and FOULKE, J.A., 1979, A methodology for documenting hand positions and forces during manual work. *Journal of Biomechanics*, 12, 131–3.

ARMSTRONG, T.J. and LACKEY, E.A., 1994, *An Ergonomics Guide to Cumulative Trauma Disorders of the Hand and Wrist* (Fairfax: American Industrial Hygiene Association).

ARMSTRONG, T.J., FOULKE, J.A., JOSEPH, B.S. and GOLDSTEIN, S.A., 1982, Investigation of cumulative trauma disorders in a poultry processing plant. *American Industrial Hygiene Association Journal*, 43, 103–16.

CHAFFIN, D.B. and BAKER, W.H., 1970, A biomechanical model for analysis of symmetric sagittal plane lifting. *AIIE Transactions*, 2, 16–27.

CHAFFIN, D.B., HERRIN, G.D., KEYSERLING, W.M. and GARG, A., 1977, A method for evaluating the biomechanical stresses resulting from manual materials handling jobs. *American Industrial Hygiene Association Journal*, 38, 662–75.

CHAO, E.Y., 1976, *Experimental Methods for Biomechanical Measurements of Joint Kinematics* (Cleveland: CRC Press), 385–411.

DE KROM, M.C.T.F.M., KESTER, A.D.M., KNIPSCHILD, P.G. and SPAANS, F., 1990, Risk factors for carpal tunnel syndrome. *American Journal of Epidemiology*, 132, 1102–10.

FLOYD, W.F. and SILVER, P.H.S., 1951, Function of erectores spinae in flexion of the trunk. *Lancet*, 1, 133–4.

HAGBERG, M., 1984, Occupational musculoskeletal stress and disorders of the neck and shoulder: a review of possible pathophysiology. *International Archives of Occupational and Environmental Health*, 53, 269–78.

MAEDA, K., 1977, Occupational cervicobrachial disorder and its causative factors. *Journal of Human Ergology*, 6, 193–202.

MAGNUSSON, M., GRANQVIST, M., JONSON, R., LINDELL, V., LUNDBERG, U., WALLIN, L. and HANSSON, T., 1990, The loads on the lumbar spine during work at an assembly line: the risks for fatigue injuries of vertebral bodies. *Spine*, 15, 774–79.

MARRAS, W.S., FATHALLAH, F.A., MILLER, R.J., DAVIS, S.W. and MIRKA, G.A., 1992, Accuracy of a three-dimensional lumbar motion monitor for recording dynamic trunk motion characteristics. *International Journal of Industrial Ergonomics*, 9, 75–87.

MARRAS, W.S., LAVENDER, S.A., LEURGANS, S.E., RAJULU, S.L., ALLREAD, W.G., FATHALLAH, R.A. and FERGUSON, S.A., 1993, The role of dynamic three-dimensional trunk motion in occupationally related low back disorders. *Spine*, 18, 617–28.

MARRAS, W.S., MARKLIN, R.W., GREENSPAN, G.J. and LEHMAN, K.R., 1995, Quantification of wrist motions during scanning. *Human Factors*, 37, 412–23.

MASEAR, V.R., HAYES, J.M. and HYDE, A.G., 1986, An industrial cause of carpal tunnel syndrome. *Journal of Hand Surgery*, 11a, 222–7.

MCATAMNEY, L. and CORLETT, E.N., 1993, RULA: a survey method for the investigation of work-related upper limb disorders. *Applied Ergonomics*, 24, 91–9.

NORDIN, M., ORTENGREN, R. and ANDERSSON, G.B.J., 1984, Measurement of trunk movements during work. *Spine*, 9, 465–469.

PUNNETT, L., FINE, L.J., KEYSERLING, W.M., HERRIN, G.D. and CHAFFIN, D.B., 1991, Back disorders and nonneutral trunk postures of automobile assembly workers. *Scandinavian Journal of Work and Environmental Health*, 17, 337–46.

PUTZ-ANDERSSON, V., 1988, *Cumulative Trauma Disorders: A Manual for Musculoskeletal Diseases of the Upper Limbs* (London: Taylor & Francis).

SNIJDERS, C.J., VAN RIEL, M.P.J.M. and NORDIN, M., 1987, Continuous measurements of spine movements in normal working situations over periods of 8 hours or more. *Ergonomics*, 30, 639–53.

SOMMERICH, C.M., MCGLOTHLIN, J.D. and MARRAS, W.S., 1993, Occupational risk factors associated with soft tissue disorders of the shoulder: a review of recent investigations in the literature. *Ergonomics*, 36, 697–717.

TANZER, R.C., 1959, The carpal-tunnel syndrome. *Journal of Bone and Joint Surgery*, 41A, 626–34.

TVEIT, P., DAGGFELDT, K., HETLAND, S. and THORSTENSSON, A., 1994, Erector spinae lever arm length variations with changes in spinal posture. *Spine*, 19, 199–204.

WINKEL, J. and WESTGAARD, R., 1992, Occupational and individual risk factors for shoulder–neck compliants: Part II — The scientific basis (literature review) for the guide. *International Journal of Industrial Ergonomics*, 10, 85–104.

Human Factors: Reliability and Risk Assessment

F.-J. Yu†, S.-L. Hwang§ and Y.-H. Huang‡

†Chung Shan Institute of Science and Technology, ‡Institute of Nuclear Energy Research, CAEC and §Department of Industrial Engineering and Engineering Management, National Tsing-Hua University Hsin-Chu, Taiwan

1. INTRODUCTION

In most of production systems, such as assembly lines of a semiconductor plant, manufacturing processes of a chemical factory or process controls in power plants, the human operators follow the standard operation procedure (SOP) to perform their tasks. However, some errors many occur and result in product deficiency or system breakdown. Comparing the equipment errors and human errors, the former is more obvious whereas most of the human errors except few significant ones that cause an accident are hidden in the operating process. Such latent human errors have to be clarified by some special techniques, otherwise the exist in the process and may gradually damage the system.

Each human task may involve several potential error modes, but only some of the error modes have high occurrence probability that cause severe consequence. If a task contains human error modes with high error probability or severe effects, this task is called the critical human task. From a quality control point of view, the more efficient strategy to improve product quality or system reliability is by identifying the potentially critical human tasks so that the corresponding errors can be reduced.

Here, a newly developed method on human error analysis, so-called the "Human Error Criticality Analysis (HECA)" is introduced. HECA is a practical method to identify the potentially human tasks and their associated error modes, as well as showing the relationships among these error modes.

2. BACKGROUND OF THE HECA

Methods have been developed for criticality analysis. The most popular method used in the military is Failure Modes, Effects and Criticality Analysis (FMECA), which identifies all possibilities of catastrophic and critical failure so that they can be eliminated or minimized through re-design at the earliest life cycle time. The FMECA method has many advantages since it is easy to understand, well accepted, a standardized approach, non-mathematical, etc. Nevertheless, FMECA is a hardware-oriented method and it analyzes hardware reliability while assuming the software and human to be perfect, which may lead the analyzers to neglect the risk caused by human error. The fundamental of FMECA, such as basing on SOP, has been applied to the development of HECA, which is used systematically to evaluate the human-related failures.

Another widely used method from the human perspective is human reliability analysis (HRA) methods used to identify human error path by event or fault trees. The HRA methods have some databases of human error to support the technique, but they have some limitations on evaluating the critical human tasks. The most noticeable deficiency is that the probability of error occurrence is calculated based on the event tree, and thus the degree to which the error affects the risk cannot be fully determined until the outcome is integrated into another consequence analysis technique. One of the steps in HECA is estimating human error probability (HEP) such as that in the HRA method, but for some improvement the risk assessment and human reliability can be analyzed simultaneously by HECA.

3. IMPLEMENT OF THE HECA

3.1. Main Steps

The analysis process of HECA is based on the SOP. Using the criticality matrix and HECA event tree, the HECA report can be completed to produce outputs such as critical human tasks list, critical human error mode list, reliability information and HECA event tree. Thus, HECA consists of four main steps: task analysis, HECA event tree construction, HEP estimation and HECA worksheet analysis.

Step 1: Determining SOP and performing task analysis:
- Coding and describing each human task in HECA worksheet.
- Identifying human error modes of each human task in HECA worksheet.
- Predicting the error effect of the error mode in HECA worksheet.
- Classifying the severity of error effect in HECA worksheet.

Step 2: Constructing the draft HECA event tree.

Step 3: Estimating the HEP of human error modes (λ_{ij}), probability of hardware failure (γ_{ij}) and error–effect probability (β_{ij}):
- Filling the HEP of human error modes (λ_{ij}) in HECA worksheet.
- Filling the probability of hardware failure (γ_{ij}) in HECA worksheet.
- Filling the error–effect probability (β_{ij}) in HECA worksheet.
- Calculating the human tasks (λ_{ij}) via HECA event tree.
- Calculating the human reliability of tasks (R_i) and the human reliability of mission (R_m).

Step 4: Completing HECA worksheet:
- Calculating the criticality index of error modes (CI_{ij}).
- Calculating the criticality index of tasks (CI_i).
- Drawing and analyzing criticality matrix.
- Listing the critical human tasks, critical human error modes, reliability information and showing the HECA event tree.

3.2. HECA Worksheet

The HECA worksheet has to be completed through deep interviews with managers or specialist of the operating system. The HECA worksheet (Table 1) is composed of 13 items, defined as follows:
- Human task identification number: number of human task.
- Human tasks: describe all the tasks operated by person(s) in a selected mission.
- Human error modes: state of error committed by person(s). This mode may either be an omission or commission. All human error modes should be listed and analyzed whether they may cause immediate or latent failures of the system.
- Probability of hardware or software failure (γ_{ij}) includes the failure probability of material, equipment, tool and even software. The failure rate of hardware can be generated from the database, e.g. MIL-HDBK-217F (1992) or by the judgment of an experienced engineer.

Table 1   Human Error Criticality Analysis (HECA) worksheet (a sample)

human task identification number (ID)	human task	human error modes	probability of hardware failure(γ_{ij})	HEP of human error mode(λ_{ij})	HEP of human task (λ_i)	error effect	severity classification	error-effect probability (β_{ij})	criticality index of human error modes(CI$_{ij}$)	criticality index of human task (CI$_i$)	human reliability (R$_i$,R$_m$)	remarks
1...	Inspecting before assembly Recording part #											
1.1		1.1.1 No record	☐	0.005		Part's quality	C4		1			
:	:	1.1.2 wrong record	☐	0.005		part's quality	C4		1			
:	:					:						

- HEP of human error mode (λ_{ij}): probability that ith human task is performed and jth human error mode happened. This mode may just be a simple error (i.e. human error only) or a compound error (i.e. human error leading to hardware failure has occurred or simply a prior human error has occurred). If the human error mode is a compound error, this HEP of human error mode is called CHEP (conditional human error probability). The joint HEP of jth human error mode (λ_{ij}) multiplied by the failure probability of the hardware (γ_{ij}) (assume the human error and hardware failure are independent), i.e. joint HEP = $\lambda_{ij} \times \gamma_{ij}$. There are many ways to estimate the HEP of human error mode (λ_{ij}); some are statistical, and some are judgmental. But it is important to differentiate whether the HEP is a simple HEP or a CHEP to avoid misinterpretation.
- HEP of human task (λ_{ij}): probability when ith human task is performed and an error mode occur. Table 2 shows the levels and description of different λ_i. This is calculated by:

$\lambda_i = 1 - \Pi_j(1 - \lambda_{ij} \times \gamma_{ij})$, if all of the human error modes are simple errors and each of them are independent.

$\lambda_i = 1 - \Pi_j(1 - \lambda_{ij} \times \gamma_{ij})$, if all of the human error modes are compound errors and not only is each of them are independent, but also the human error and hardware failure are independent; or

$\lambda_i = 1 - [(1 - \lambda_{i1}) \times (1 - \lambda_{i2}) \times \dots (1 - \lambda_{ij} \times \gamma_{ij}) \times (1 - \lambda_{in} \times \gamma_{in})]$, if some of the human error modes are compound errors and the others are simple errors.

- Error effect: include the error impacts to hardware system, system function, mission objectives, personnel safety, etc.
- Severity classification: impact of the human error mode on the system level. Can be considered from the standpoint of safety, reliability, maintainability, quality, cost, etc. Table 3 is an example of severity classification of cost.
- Error–effect probability (β_{ij}): conditional probability that the error effect will result in the identified severity classification given that the jth human error mode of ith human task has occurred. Table 4 describes the failure effect of β_{ij} in a case.
- Criticality index of error modes (CL$_{ij}$): a criticality number standing for the occurrence probability of ith human task attributable to jth human error mode which results in the identified severity classification. Criticality is equal to the HEP of human error mode (λ_{ij}) multiplied by the probability of error effect (β_{ij}), i.e. CI$_{ij}$ = $\lambda_{ij} \times \beta_{ij}$ (simple errors) or CI$_{ij}$ = $\lambda_{ij} \times \gamma_{ij} \times \beta_{ij}$ (compound errors).
- Criticality index of human task (CI$_i$): occurrence probability

of ith human error that results from one or more human error modes that will induce the identified severity classification. Can be calculated using CI$_i = 1 - \Pi_j(1 - \lambda_{ij} \times \beta_{ij})$ (simple errors) or CI$_i = 1 - \Pi_j(1 - \lambda_{ij} \times \gamma_{ij} \times \beta_{ij})$ (compound errors).

- Human reliability (R_i or R_m) is the probability that a person correctly performs ith human task in a required period and in a specified environment. Human reliability of mission (R_m) is the probability of successfully performing all the tasks of the mission. Based on the data of HEP of human error mode (λ_{ij}), HEP of human task (λ_i) and HECA event tree, one can calculate for human reliability (R_i or R_m). If all the sequential human tasks are independent of each other, the human reliability of ith task and human reliability of mission can be calculated by:

$R_i = 1 - F_i = 1 - CI_i = 1 - \Pi_j(1 - \lambda_{ij} \times \beta_{ij})$ (if all the errors are simple error), where F_i is the failure probability of ith task, $R_m = \Pi_i R_i$ (series system) or $R_m = 1 - \Pi_i(1 - R_i)$ (parallel system); or

$R_i = 1 - F_i = 1 - CI_i = 1 - \Pi_j(1 - \lambda_{ij} \times \gamma_{ij} \times \beta_{ij})$ (if all the errors are compound error).

According to the HECA worksheet data, we can construct the criticality matrix. Then, the final HECA event tree can be presented subsequently. Figure 1 shows a partial HECA event tree in a case.

Figure1. A partial HECA event tree.

Table 2. HEP of human error modes χ_{ij}) of a case study.

Level	Description	λ_i vlaue		Average
A	Frequent	1.0	0.1	0.5
B	Probable	0.1	0.01	0.05
C	Occasional	0.001	0.0001	0.005
D	Remote	0.001	0.0001	0.0005
E	Improbable	0.0001	0.0	0.00005

Table 3. Severity classification of cost.

Level	cost loss percentage	description
C1	100	hardware system will result in 100 cost loss given the error effect happen
C2	75	hardware system will result in 75 cost loss given the error effect happen
C3	50	hardware system will result in 50 cost loss given the error effect happen
C4	25	hardware system will result in 25 cost loss given the error effect happen
C5	0	hardware system will result in 0 cost loss given the error effect happen

Source: Adapted from MIL-STD-882C

Table 4. Error-effect probability β_{ij}) of a case study.

Level	Failure effect	β_{ij} value	average
B1	actual loss	$\beta_{ij} = 1.0$	1.0
B2	probable loss	$0.1 < \beta_{ij} < 1.0$	0.5
B3	possible loss	$0.0 < \beta_{ij} < 0.1$	0.05
B4	no effect	$\beta_{ij} = 0.0$	0.0

Source: Adapted from MIL-STD-1629A

3.3. Criticality Matrix

The criticality matrix provides a method of identifying and comparing the error occurrence probability among human tasks (or among human error modes) with respect to the error effects which consider safety, cost, and reliability of the system. The matrix is constructed by plotting human tasks (or human error mode) identification numbers against the severity classification category which can be represented by either the error occurrence probability level or the criticality index of human tasks CI_i (or the criticality index of human error modes CI_{ij}). Figure 2 is an example of criticality matrix in a case study.

The completed matrix shows the distribution of criticality of human tasks relevant to severity classification, and provides a criticality order for assigning correction priorities. The further away the diagonal appears from the origin, the greater is the
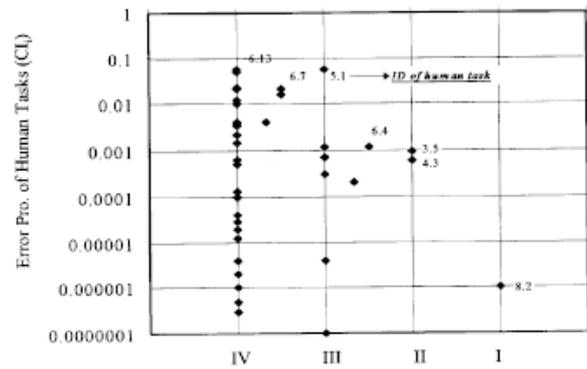
Figure 2. An example of Criticality Matrix.

criticality of the related human task, and the more urgent it is to implement a correction action or design change. For instance, task 8.2 in figure 2 is the critical one.

4. CONCLUSIONS

HECA has been applied to some real cases. For instance, the igniter operating tasks (Yu *et al.* 1997) and the initiator assembly tasks of the missile system (Yu *et al.* 1998). Based on the results of analysis in initiator assembly tasks and the constraints of cost and time factors, some corrective actions are taken and their effectiveness of improvement have been evaluated. It reveals that not only the human assembly reliability has been enhanced by 9%, but also that the risk occurrence probability and safety severity decrease significantly.

The current limitation of HECA method is that the processes are getting complex as applying to a very large system, and this may be solved by computerization of HECA in the future.

REFERENCES

MIL-STD-1629A, 1980, *Military Standard-Procedures for Performing a Failure Mode, Effects and Criticality Analysis* (Washington, DC: US Department of Defense).

MIL-HDBK-217F, 1992, *Military Handbook-Reliability Prediction of Electronic Equipment* (Washington, DC: US Department of Defense).

MIL-STD-882C, 1993, *Military Standard-System Safety Program Requirements* (Washington, DC: US Department of Defense).

YU, F.-J., HWANG, S.-L. and HUANG, Y.-H., 1997, The system safety analysis of igniter assembly behavior by standard operation procedure. *Journal of Explosive and Propellants, R.O.C.*, 13, 23–6.

YU, F.-J., HWANG, S.-L. and HUANG, Y.-H., 1999, Task analysis for industrial work process – from aspects of human reliability and system safety. *Risk Analysis: An International Journal*, 19(3), 401–15.

Human Factors Design Tools for Consumer– Product Interfaces

J. V.H. Bonner

Teeside University, Institute of Design, School of Law Arts and Humanities, Middlesbrough, UK

1. INTRODUCTION

As the use of computer-embedded technology in consumer products increases and advances are made in novel display and control devices, so the design considerations of consumer–product interfaces draw closer to the design of computer interfaces. Conventional consumer products such as video cassette recorders (VCR), microwave ovens, pagers and heating control systems, along with more state-of-the-art products, such as personal digital assistants (PDA) and digital television, now adopt many interface characteristics similar to desk-top computers including touch sensitive displays, cursor controls, icons and menus. It is unwise, however, to assume that computer interface design knowledge can easily migrate to consumer–product interfaces.

There are distinct differences between the two types of products particularly within the context that they are used. Unlike computers, consumer products usually have very specific purposes or functions. Many products have been a part of domestic life for longer and therefore have a much deeper historical and cultural role in the environments in which they are used. The functions of an oven, for example, have been established over many years, leading to deep-rooted expectations about its purpose in the home. Interaction activity between a user and a consumer product is usually less time consuming than computing tasks and more intermittent. Furthermore, users generally lack willingness to understand or use new features. Commonality between different product interfaces is rare and users often do not expect similar controls and displays between one product and another. It is also far more common for a user to also be the purchaser of a consumer product than is the case with commercial computer users. This can affect user attitudes towards usability. Many consumer products are used by more than one individual, creating collective responsibility to care for and maintain consumer products. The challenge for product interface designers is to understand this complex web of user–product interaction.

Over the years, the computing industry has embraced and accommodated interface design issues with the development of new disciplines such as software design (Winograd 1996) and human computer interaction (HCI) (Laurel 1990, Preece *et al.* 1994). In contrast, most consumer products still rely upon the skills of manufacturing and production engineers along with design engineers and industrial designers. There has been little need for a strong emphasis on interface design issues beyond conventional "knobs and dials" human factors knowledge, although this is changing. Jordan *et al.* (1996), for example, offer a useful range of informal usability methods which can be applied to consumer products. However, with the onset of "convergence technology" (the merging of telecommunication, computer and consumer products), new consumer products are beginning to emerge that have few design precedents. The design of product interfaces is becoming less defined by hardware constraints and more defined by the imagination of designers and the needs of potential users.

This has created a demand for designers and design tools capable of creating, building and testing speculative and conceptual proposals in which the interaction between the user and the product is as important as the product itself. Many consumer product manufacturers now employ interaction designers whose specific role is to address the interaction aspects of the "softer," less defined and often more complex product interfaces. Interaction design is not, however, entirely synonymous with human factors. The heart of interaction design is fundamentally a creative process where pragmatic consideration of good product interfaces must be balanced with marketing and aesthetic demands as well as usability issues.

This chapter briefly describes a human factors-based design toolset which begins to fulfil these requirements by allowing human factors issues to be addressed very early in the design process but also allowing creative, incomplete and speculative interface design solutions to be proposed.

2. RATIONALE FOR THE DESIGN TOOLS

Many traditional human factors methods require the formalization or visualization of an interface before any serious usability study can take place. Even with an iterative, software-based rapid prototyping approach, design solutions need to be quite "mature" before evaluation studies can occur. With this in mind, these design tools were developed to support the interface design process prior to the use of software-based prototypes.

In developing the design tools, much of the theoretical underpinning has been drawn from the field of HCI, reflecting the recent shift from a cognitive engineering paradigm to theories considering the importance of context such as activity theory (Nardi 1996). The design tools described here are still in their infancy and many of the techniques adopted should be familiar to practicing HCI specialists. In summary, the design tools are based on the following criteria.

They should support the interface design process at very early conceptual stages especially where few "role models" exist or where novel display and control devices are being considered:

- users and designers should be involved in both the design and evaluation of interfaces;
- the context in which the product interface could be used should be considered;
- the use of "transitional objects" (concepts, ideas and suggestions placed in real world situations) to allow users and designers to discuss and suggest design proposals; and
- design methods already familiar to designers should be used.

3. OVERVIEW OF THE DESIGN TOOLS FOR NOVEL CONSUMER–PRODUCT INTERFACES

The toolset comprises of four related and inter-linked design tools where iterations can occur between and within the design tools. They are:

- contextual enquiry;
- card sorting;

- scenario design; and
- inspection methods.

All these occur before the building of software-based prototypes.

3.1. Contextual Enquiry

The first design tool is used to understand the context within which the product will be used. This requires conventional observational and data gathering exercises such as interviews, video, focus groups and so on. Designers are solution rather than problem focused and in many cases the tendency is to observe until a potential solution is found. For many designers the approach used in this tool is unconventional because broader, more "peripheral" data is gathered. The key to this design tool is to encourage designers to move the focus of their observation work so that more than the product in question is observed. It is also important that designers suspend their natural "solution generating" process as long as possible and allow the complexity and subtlety of the contextual frame to unfold. From the contextual enquiry work, designers propose two "outputs" for the next design tool.

First, a typical event or scenario that they feel will capture important interaction events that will test the usability and acceptance of an interface, for example preparing a new meal from a cookbook using a microwave and conventional oven.

Second, a range of possible existing and proposed control and display solutions that support existing or new product functions which are appropriate for the chosen scenario. These can be incomplete or even ambiguous proposals. An example could be a new recipe planner for a microwave interface.

Complete or well-defined solutions are not a necessary prerequisite to progressing to the next design tool.

3.2. User Requirements Capture through Card Sorting

The purpose of this design tool is two fold: first, to describe elements of the task or scenario and to design product functions using cards; second, to allow designers and potential users to discuss, arrange and sort these cards allowing a set of user requirements to be produced. The cards are composed by the designers and can describe concepts, design features or control and display proposals both in text or images. Cards should appear "rough and ready" to suggest to the participants that they represent tentative solutions and are amendable. The scenario cards should reflect not just tasks but also other events that may affect the use or non-use of the product. The function cards describe different interface features, for example alternative methods of selecting and altering the power level of a microwave.

Once the cards are completed by the designers, workshops are set up where representative users can discuss, place and sort the cards. Participants' first sort and place the scenario cards that best describes, by consensus, the structure and order of the scenario (Figure 1). It is important that the group plan by consensus to allow users to articulate differences in approach and understanding of the tasks within the scenario. These differences in approach provide useful contextual information.

Once this stage of the card sorting is complete, the designers have a physical representation of the scenario. The card "map" should be photographed to allow further analysis of the participants' placement and inclusion of cards. These "maps" can also

Figure 1. Participants sorting and placing cards

Figure 2. Making notes on consensus statements

be compared with results of other workshops. The next stage requires the participants to go through each function card and agree on its importance to the scenario. The cards are sorted by placing them into labeled bins such as "really useful" and "really useful but not for this scenario." Again this has to done by consensus to promote discussion.

The favored cards are then placed into the scenario plan providing important indicators to the designers of where and how proposed ideas would be adopted within the scenario. Participants should be actively encouraged to introduce new cards or change existing cards. During the card sorting activities one of the designers should record significant events (Figure 2) that have implications for interface design. Within the design tools, guidance is provided for the designer on how to do this.

Analysis takes place through debriefing exercises with designers and participants and also between the design team, through video analysis and reviews of the scenario maps and selection strategies. By setting different scenarios through workshops and exercises, designers normally get a very good feel for what type of innovative proposals can be implemented.

3.3. Interface Design Development through Scenario Design

The next design tool refines and details the interface elements into a storyboard and produces a list of all the proposed inter-
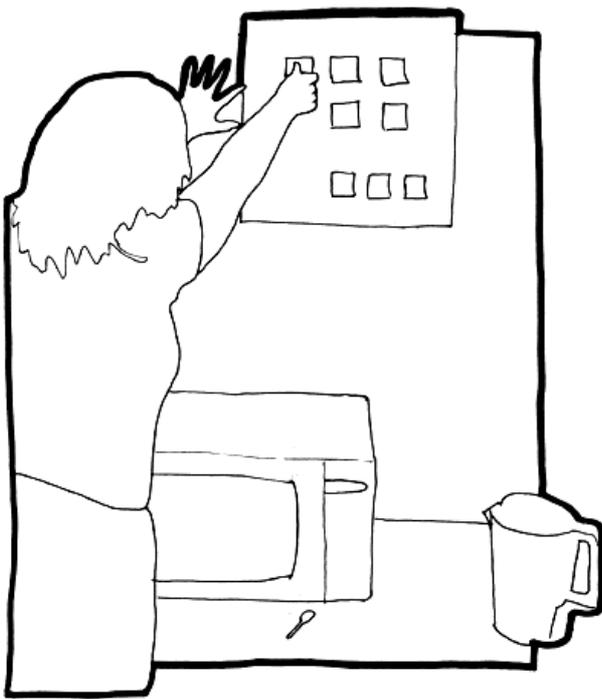
Figure 3. Taking tabs off the board before placing on a mock-up product
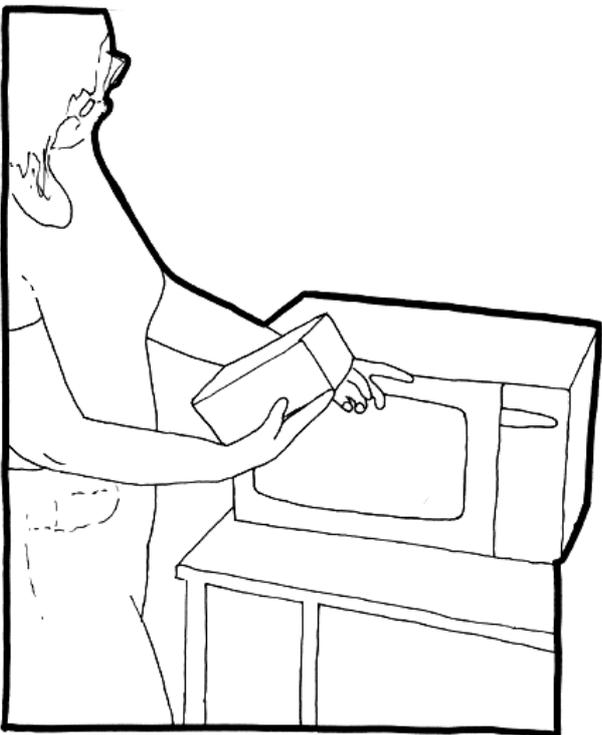
Figure 4. Use of a mock-up in the scenario design tool

face control and display function descriptions. These function controls and displays should then be converted into "tabs" (Figure 3). These will be more detailed and coherent descriptors based on the function cards used in the card-sorting tool. It is important that fundamental concepts and ideas are still perceived as being open to critical judgement and alteration.

The scenario for this design tool should be as realistic as pos-sible. For example, if one of the scenarios proposed was to cook a meal for four, this should be done in practice with real food and cooking implements. During the task, participants take off each tab from a board when they are required in the task and place them on a mock product. The mock-up does not have to be par-ticularly realistic apart from dimensional requirements. The mock-up becomes a very important vehicle and point of discussion between the participants and the designers (Figure 4).

For example:

Participant: "at this stage I want to tell the microwave that I want this pie to be defrosted for 5 minutes." Selects a preferred "defrost" tab from the board."[N]ow I need a way of telling it that I want this done for 5 minutes." Looks a tab board for possible options.

Designer: "from the different ways of doing this on the board, which do you think would be the best way at this stage?"

If the tasks involve the use of a working product (such as a microwave to actually cook the meal) then, as far as possible, the mock-up and tab board should be referred to before using a working product. This prevents the existing interface from driving the flow of interaction activity. The results from the design tool emerge through video analysis, user statements and opinions and mock-up designs.

3.4. Inspection Methods for Designers

The final design tool is an inspection method approach that does not require involvement with participants. It provides a systematic way of considering many of the high-level human factors issues that affect the usability of product interfaces. One of the advan-tages of using this design tool is that it can be used with concep-tual ideas that have little low-level detail and include novel interaction styles. Other types of inspection methods exist but require the interface design to be more complete or assume more common computer-based controls and displays.

The inspection methods include design heuristics and check-lists that deal mainly with issues related to the anticipated con-text in which the product will be used and places broader attention on users, tasks and the environment beyond the scenarios that have been selected in the previous design tools. They also deal more specifically with interaction design issues and how users typically process information. The intention of this design tool is to re-examine the proposed interface design and take a more objective and broader critical review of usability issues. In going through the inspection methods, it should be possible to examine interaction issues more methodically.

Where possible, solutions are provided when a usability prob-lem is identified. Revised solutions should be evaluated by using more focused scenario design exercises. Some problems are, how-ever, best resolved by moving forward to the next design phase, prototyping, particularly where "time-based" interaction activi-ties are concerned. Software-based prototypes can now be developed using more traditional evaluation methods that have been successfully applied in the development of many computer and consumer-based products.

4. CONCLUSIONS

A substantial number of design tools and methods have been developed in the past, but few have been adopted by designers.

Designers complain that they do not offer sufficient benefits to justify the implementation costs. The design tools outlined here, which have been based on HCI principles and theories have been adapted, developed and refined by involving designers (the end users) in the development process and have been through many iterations. The tools have to fit the requirements of designers, meaning that they have to be easy to learn, quick to implement, require little formal training and integrate into existing design cultures.

The key to the success of the design tools has been the use of "transitional objects" (cards, scenarios, foam models being used as points of discussion) overcoming many of the communication problems that can occur between users and designers. None of these concepts or approaches are radical or new to designers but traditionally have not been used so explicitly in this way and so early on in the design process. A further benefit is that designers do not have to resort to lengthy user studies or users having to be heavily briefed before being effective.

Improvements to the design tools are still being made and further evaluation studies are planned. Nevertheless they appear to satisfy two often-conflicting demands: appropriateness within a product design culture supported with sound theoretical underpinning.

REFERENCES

JORDAN, P.W., THOMAS, B., WEERDMEESTER, B.A. and McCLELLAND, I., 1996, *Usability Evaluation in Industry* (London: Taylor & Francis).

LAUREL, B., 1990, *The Art of Human–Computer Interface Design* (Reading, MA: Addison-Wesley).

NARDI, B. (ed), 1996, *Context and Consciousness: Activity Theory and Human Computer Interaction* (Cambridge, MA: MIT Press).

PREECE, J., ROGERS, Y., SHARP, H., BENYON, D., HOLLAND, S. and CAREY, T., 1994, *Human–Computer Interaction* (Harlow: Addison Wesley).

WINOGRAD, T. (ed), 1996, *Bringing Software to Design* (Reading, MA: Addison-Wesley).

Human Sensory Measurement Application Technology (HSMAT)

K. Suzuki

Research Institute of Human Engineering for Quality Life, Osaka, Japan

1. INTRODUCTION

New technologies became widespread at the same time as the rapid economic growth that started in Japan in the mid-1970s, making our lifestyles very convenient and comfortable. Technological advances in manufacturing and information sciences have made it possible to create customized products and environments in response to various human needs. However, these new technologies or products often failed to provide true satisfaction for their users. More and more people are now experiencing technology-related stress and fatigue, and the number of illnesses caused by these problems is increasing.

At the end of the 1980s when this phenomenon emerged, industry started paying more attention to human well-being. Technologies, products and environments needed to be designed to satisfy human psycho-physiological (including sensory) needs in order to reduce problems related to the human factor at home as well as the workplace, and to improve the overall quality of life. Aiming to create a sense of well-being in fields such as health, safety, convenience and comfort, human-oriented manufacturing forms the basis of achieving a more comfortable and rewarding lifestyle. Human-oriented manufacturing requires objective comprehension of the physiological and psychological state of humans, and also the influence of products and environments.

On this basis, "Human Sensory Measurement Application Technology" (HSMAT) was established in 1990 as a national 9-year research project to compile physiological and psychological data to form the core of the analysis, and to develop standard technologies for measuring human senses. This chapter describes the objectives and major results of the HSMAT project.

2. PROJECT OVERVIEW

The HSMAT project has a total budget of ~$140 million and is part of the Industrial Science and Technology Frontier Program overseen by the Agency of Industrial Science and Technology of the Ministry of International Trade and Industry (MITI). The project was commissioned through the New Energy and Industrial Technology Development Organization (NEDO). The Research Institute of Human Engineering for Quality Life (HQL) based in Osaka, Japan, undertook this project in cooperation with six national research institutes, the laboratories of 25 major corporations and six universities. The main objectives of the HSMAT project were to establish technologies that quantitatively measured and evaluated human sensations, and apply these technologies to the design of user-friendly products and reduced-stress living and working environments meeting various individual needs.

The HSMAT project consisted of two stages: I (1990–94) and II (1995–98). During Stage I, individual and basic research was conducted to establish human sensory measurement technologies.

During Stage II, the focus of the research was shifted to the integration and application of these technologies, especially toward developing human sensory indices that reflected human sensations.

2.1. Stage I (1990–94): Fundamental Research

Stage I of the project included the development of technologies for generating and controlling the external stimuli that influence human senses, and simple, non-invasive, quantitative determination of physiological values resulting from external stimuli. These technologies have made it possible to estimate sensory measures based on physiological measures. Particular developments included non-invasive devices that made it easier to make objective and practical measurements of variations in human senses and in bodily conditions caused by external stimuli. These measurement technologies involve measuring metabolic functions, the central nervous system, the autonomic nervous system, and the ocular function.

Quantitative physiological data was recorded using physiological measurement devices, and the relationship between physiological responses and external stimuli was studied. This has allowed an evaluation of human senses, i.e. the levels of fatigue, stress, alertness, and comfort with regard to the human-friendliness of products and environments.

The systems also created synthetic environments by controlling different external stimuli including heat, noise, vibration, lighting and composition, and were able to make assessments under different conditions without using human subjects.

2.2. Stage II (1995–98): Development of Indices and their Applications

The primary objectives of Stage II were to adopt the human sensory measurement technologies developed during the first stage in order to establish technologies for developing human sensory indices and to develop application technologies that incorporated human sensory indices into the design of industrial products or environments. These technologies evaluate the human-friendliness of products, and also of living and working environments.

2.2.1. Human sensory indices

Human sensory indices are technologies for measuring and evaluating stress, fatigue and alertness induced by physical environments or products, and for evaluating the compatibility of environments and products to humans in terms of comfort, familiarity and ease of operation. These technologies were developed by refining and systematizing the human sensory measurement technologies developed during Stage I.

2.2.2. Human sensory indices application technologies – applied case studies and the development of evaluation simulators

The validity of human sensory indices was demonstrated in close to real-life conditions, and statistical data were collected during the demonstration to fine-tune the indices. A series of applied case studies was conducted in newly constructed laboratories to confirm the robustness of each human sensory index, and to integrate the related indices. Each laboratory was equipped with a simulated environment presentation system that created desired

working or living conditions by adjusting environmental factors, such as lighting, room temperature, noise, air flow, humidity, and color temperature.

Evaluation simulators that evaluated products and environments in place of human subjects were also developed.

3. OUTCOMES

Subsequent sections describe the results of the research undertaken in the HSMAT project. These sections are based on a written report published by the Human Sensory Measurement Application Technology Evaluation Committee and the Evaluation Group of the Industrial Technology Council.

3.1. Human Sensory Indices

3.1.1. Indices for measuring stress, fatigue and alertness

These indices estimate the mental conditions (stress levels) of humans by quantitatively measuring physiological conditions such as blood pressure, heart rate and skin temperature. Other indices evaluate work adaptability by humans by estimating the level of physical activity (from sleep to active) determined by physiological factors such as EEG, blood pressure and core temperature. The indices can be applied to evaluate the adaptability of monitoring work required at plants and airport control towers to prevent serious accidents.

3.1.1.1. Development of technology and a method for measuring, analyzing and evaluating stress

A high-performance stress hormone evaluation system was developed on the basis of evaluating human stress levels by monitoring changes in the volume of stress hormones found in saliva (and urine). The device makes it possible for companies to improve employee health care without professional medical help. If worker stress levels are examined on a regular basis using the device, stress-related diseases caused by overwork can be prevented.

3.1.1.2. Telemeter-style non-restrained physiological measurement devices

These three portable devices measure human physiological conditions such as systolic blood pressure, electrodermal activity and blinking. The net weight of the device for measuring systolic blood pressure is 200 g, the wrist-watch style electrodermal activity measurement device only 60 g and the headband-like device for measuring blinking is 70 g. These newly developed devices are much smaller and lighter than other devices with similar functions. They are superior because they are also very easy-to-use, less restrained, and are suitable for practical use. Stress levels can be evaluated by using these three devices together.

3.1.1.3. Non-contact skin temperature measurement system

This system is a world-first for automatic and non-invasive measurement of facial skin temperatures using infrared and color images. Stress levels are estimated from the images, and expressed numerically.

This system can be also applied to workplace stress management and to evaluating the development process for stress-reducing products.

3.1.1.4. Visual sensory fatigue measurement device

This is a portable and lightweight measurement device for simultaneously measuring three major visual functions (focusing, pupil diameter and eye movements). It weighs 350 g, one-fifth of the weight of conventional models.

3.1.1.5. Monitoring system for estimating alertness levels

This monitoring system is a device that estimates human alertness levels by measuring heart rate and core (eardrum) temperature. By monitoring the alertness of workers who monitor work at plants and airport control towers, the system can prevent accidents caused by human error or declines in work efficiency.

3.1.2. Indices for measuring and evaluating environmental adaptability

Using physical environmental quantities as parameters, these indices measure and evaluate the adaptability of environments to humans to design comfortable work and living environments in tune with human sensory perceptions. The indices are determined from major environmental factors, namely vision, sound and oscillation, and heat.

3.1.2.1. Technology for designing the visual environment of offices

This model evaluates the spatial impressions of offices and other spaces, including perceived comfort levels and the sense of openness. It evaluates any given space based on its dimensions, coloring and lighting. A visual environment adaptability evaluation index was created to provide guidelines for designing working and living environments.

3.1.2.2. Color presentation technology

This technology quantitatively evaluates chromatic adaptation in actual situations and accurately predicts corresponding color images in various states of illumination.

After the validity of the method for predicting corresponding color images was confirmed, a newly developed chromatic adaptation transform received international recognition after being presented to the International Commission on Illumination (CIE). In 1994, the proposal was published as a CIE technical report entitled, "A Method of Predicting Corresponding Colors under Different Chromatic and Illuminance Adaptation."

3.1.2.3. Sensitivity to vibration while sitting

This research resulted in vibration stimuli presentation technology based on waveform control, and an evaluation method for sensitivity to vibrations. Differences in the curvature of the spine are considered, and the degree of discomfort estimated when the whole body is subjected to vertical vibrations while sitting.

It will be possible to standardize the sensitivity level for different postures or for when different parts of the body are subjected to vibrations. This novel method considers changes in posture responding to actual situations.

3.1.2.4. Human thermal model and thermal sensation prediction model

Physique and other individual differences were considered when constructing the human thermal model, which estimates the skin temperature and perspiration of different parts of the human body.

The thermal sensation prediction model accurately estimates skin temperature and thermal sensation, even in dynamic and uneven thermal environments (errors within ± 0.5%). The effects of thermal environments in real-life settings can be evaluated using the latter model, which predicts the changes in thermal sensation when a person moves from one room to another or leaves a house. The model is a potential new thermal comfort index.

3.1.3. Indices for evaluating product adaptability

Actual consumer products such as clothes, a chair and a bed were used to develop practical indices for evaluating product adaptability in terms of human body size, shape and movement. Other indices measuring product usability via human–product interaction have been developed to evaluate equipment operability in specific working conditions, including monitoring cranes and plants.

3.1.3.1. *Practical postural measurement and evaluation system*

This system evaluates the adaptability of chairs based on the posture of a worker sitting and performing a task over a long period. It evaluates the shape of the sitting surface and the distribution of body pressure. The system can be used as a design support tool, easily evaluating design ideas and creating chair models in the product development process, and also for data acquisition and data processing when necessary. Companies can use the system to collect basic information when establishing design or evaluation criteria.

3.1.3.2. *Platform system for supporting the design process*

This is a prototype design support system to allow designers to easily find data relevant to human life engineering. An advanced system has been developed on the basis of the prototype. The platform system allows designers to create high-quality product designs using common data and evaluation criteria in a relatively short time. A variety of products can be designed and evaluated by running computerized human models (computer manikins) on this system.

3.2. Human Sensory Indices Application Technology

3.2.1. Applied case studies

The combined effects of multiple environmental factors have been clarified using experiments in various conditions created with a simulated environment presentation system that adjusts different environmental factors, including lighting, room temperature, noise, air flow, humidity and color temperature. Careful investigation of the effectiveness and limitations of human sensory indices in use should facilitate the development of systems or devices, which can measure and evaluate products and environment comfort in real-life settings.

A database containing various types of human data obtained through the HSMAT project, and a handbook providing information on advanced technologies for gauging biological, behavioral and subjective measures will be produced after refining the indices and other resulting technologies. In the near future, comprehensive guidelines indicating proper application of human sensory indices will be constructed by integrating and systematizing the results of the HSMAT project.

3.2.2. Evaluation simulators

3.2.2.1. *Thermal sweating manikin*

This thermal sweating manikin simulates human skin temperatures and perspiration. Unlike conventional thermal manikins, it can simulate skin temperature and perspiration changes over time, and can also be made to perspire profusely. Micro-climates (humidity) within clothing can also be simulated. The manikin can make different poses including, standing, sitting and lying down, and can be used to design and evaluate products such as bedclothes, chairs, and comfortable clothing.

3.2.2.2. *Human comfort meter*

This is a small manikin with movable joints and adjustable height that is equipped with 12 thermal comfort sensors. The manikin allows accurate evaluation of the adaptability of products to humans, and expresses thermal sensations as numerical values.

4. CONCLUSION

The HSMAT project was the first national research project to systematically study humans in order to develop technologies directly related to our lifestyles. Although this project has only outlined some human characteristics, the results are expected to contribute greatly to improving product design and working and living environments. Some of the results will be applied to develop technologies for scientific analysis of human behavior for a new national project starting in 1999. The HSMAT project results and the new developments will further accelerate technological development in human life engineering.

REFERENCES

SCIENCE AND TECHNOLOGY FRONTIER PROGRAM, 1998, Human Sensory Measurement Application Technology (Phase Ii), Pre-Final Evaluation Report, March (Evaluation Group of the Industrial Technology Council, Human Sensory Measurement Application Technology Evaluation Committee).

TSUCHIYA, S., IIDA, T., MORIKUNI, I., HANE, T., MIYATA, K., KOGI, H., INUKAI, Y., FUKUI, Y. and YOSHIOKA, M., 1998, Industrial science and technology frontier program — human sensory measurement application technology. *Industrial Science and Technology*, 39, 1–17. [in Japanese]

Human Systems Engineering Process and Methods

J. A.B. McKneely†, D. F. Wallace†, A. A. Perry‡ and J. J. Winters§

†Naval Surface Warfare Center, Dahlgren Division
‡Synetics, Inc.
§Basic Commerce and Industries, Inc.

1. INTRODUCTION

Traditionally, human factors engineering issues are not well integrated within system design efforts. Where human factors engineers *are* called upon, it is usually late in the design process or even post-design. Therefore, human factors engineering solutions are often not implemented due to the increased cost of late design changes. System design deficiencies become operations problems and require highly skilled users to overcome these deficiencies. These skill requirements drive increased training demands and potential user availability problems. Exacerbating this problem is the rate of technological change that levies significant constraints on system development time. Program managers are pressured to demonstrate development progress frequently resulting in premature allocation of system functions when the requirements are not yet defined. These factors substantially reduce or eliminate time for late changes.

Systems engineers are not always familiar with the issues of human–system integration. Further, human factors engineers and systems engineers speak different technical languages. To address both issues, a human systems engineering process has been defined using systems engineering terminology. For this chapter, 'Human Systems Engineering is defined as the activities involved throughout the system life cycle that address the human element of system design' (IEEE 1220-1994, 1998). These activities include the definition and synthesis of manpower, personnel, training, health hazards, safety issues, human factors and any other issues addressing human activity within the system. This concept of the human systems engineering process while consistent with traditional definitions of human factors and human systems integration, fully imbeds human engineering into the systems engineering process and employs common terminology. The concepts and process presented in this chapter influenced and are integrated into IEEE 1220-1998, *Standard for the Application and Management of the Systems Engineering Process* (1998), a commercial systems engineering standard.

This chapter will describe the human systems engineering process, identify examples of methods that support the process and provide examples of design efforts that applied this process.

2. SYSTEMS ENGINEERING

Systems engineering is an iterative and interdisciplinary process through which system requirements are transformed into a system solution. Systems engineering addresses not only the physical design of the system, but also all phases of the system lifecycle. The process begins with definition and analysis of the requirements, translates those requirements into a functional definition of the system and creates from those functions a detailed design. Systems engineering is iterative and begins with concept-level requirements; develops and decomposes more detailed requirements, functions and designs; and integrates system components to form a coherent definition of the system. Solutions
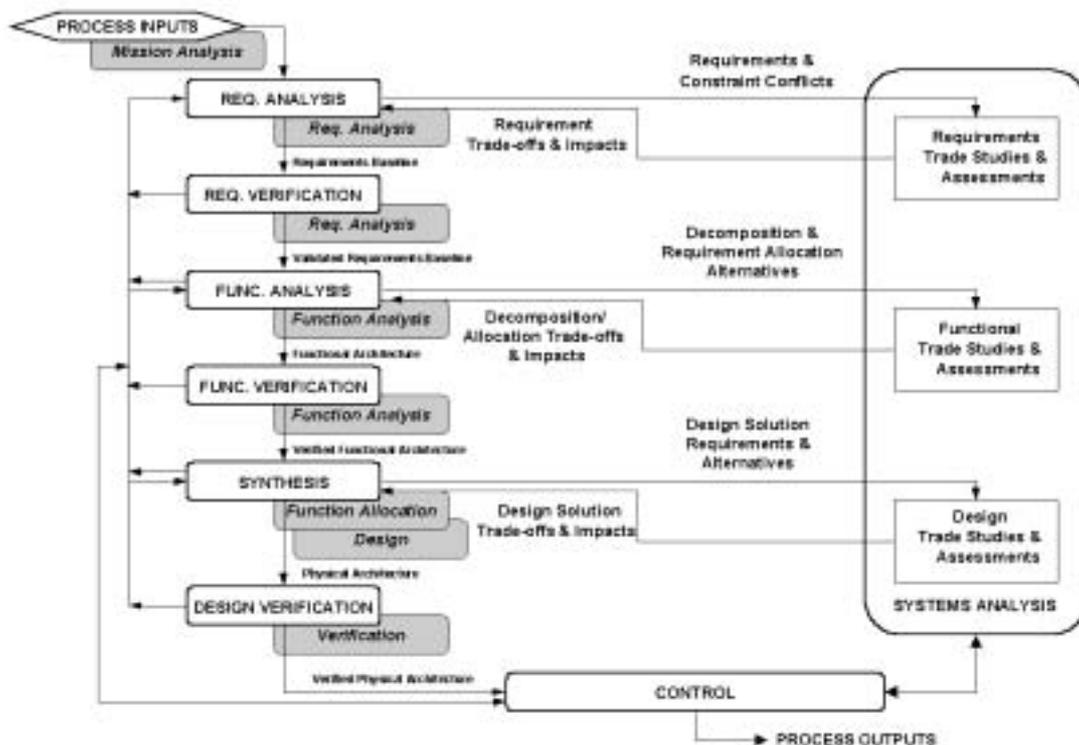
Figure 1. Human systems engineering process and systems engineering process of IEEE 1220–1994 (1994)

are continually revised, improved, updated or modified until system design factors are optimized. Systems engineering is also interdisciplinary, addressing hardware, software and human issues in addition to cost, manufacturing processes, logistics and lifecycle issues. To have a measurable impact on the final system solution, human engineers must use the terminology of systems engineers and operate within the framework of the systems engineering process.

3. HUMAN SYSTEMS ENGINEERING

To achieve the desired impact, human systems engineering must be fully integrated with systems engineering. Human systems engineering products must be the result of a disciplined engineering effort that is conducive to overall system analysis. To meet these needs, the human systems engineering framework was modeled after the systems engineering framework developed for the evolving IEEE 1220 specification. Figure 1 depicts how human systems engineering (shaded blocks) integrates with the systems engineering process.

In addition to consistency with the systems engineering process, the human systems engineering process must be consistent with its heritage in human factors engineering. To meet this goal, multiple sources of human factors process descriptions were evaluated. Key documents included the ShipSHAPE process (Malone 1997), the NATO panel research group's report on *Analysis Techniques for Man–Machine System Design* (Beevis *et al.* 1994), *Human Factors in Systems Engineering* (Chapanis 1996) and an HFES Conference Workshop on human factors methods (Chapanis and Shafer 1995). Leveraging process descriptions of human factors engineers ensured that the process defined in the systems engineering framework would be true to the problems and issues it is intended to solve.

As depicted in Figure 1, the human systems engineering process follows the basic top-down systems engineering framework. The six major steps of the human systems engineering process are mission analysis, requirements analysis, function analysis, function allocation, design and verification. The following sections define and describe the major steps of this human systems engineering process and provide examples of methods that are applicable to that design step.

3.1. Mission Analysis

The start of a system engineering effort is mission analysis, where the overall system purposes, objectives, capabilities, operational environments and basic functions are determined. Mission scenarios are identified or created. The focus in this step is on the definition of the system boundaries, treating the system as a 'black box' and defining inputs, outputs, environments and other constraints.

To articulate the mission, narrative mission descriptions and graphic mission profiles can be employed. Narrative mission descriptions detail the mission in terms of the characteristics, sequences and times of mission events, mission constraints and environmental conditions. Details of the description facilitate identification of top-level functions required of the system. Graphic mission profiles illustrate the sequence of events that will determine the performance requirements of the system.

3.2. Requirements Analysis

Requirements analysis is performed to identify the characteristics of the system necessary to meet mission requirements. Intended users and maintainers of the system and activity-related needs of users are determined. The feasibility and internal compatibility of the system requirements are assessed. Measures of effectiveness and measures of performance of the mission, human and job/task requirements are defined. In this step, the role of the human, and manning, training and cost guidelines are also developed.

3.3. Function Analysis

Function analysis output supports development of the system's functional architecture — the sequence of operations or events that turn inputs into desired outputs — and compares design alternatives. Although the system may be broken into functions, tasks and subtasks to be performed, they are not allocated to any particular system component. This step and the three following steps are initially performed at a high system level with little function decomposition, but are iterated to greater levels of detail.

Function flow diagrams and behavior graphs are methods to analyze system functions. Function flow diagrams depict the sequential relationship of the system functions. This method logically inventories the system functions and structure. Behavior diagrams describe system behavior of control and information flow and scenario modeling, providing a complete behavioral model of the system.

3.4. Function Allocation

A critical step of the human systems engineering process is determining what functions are to be performed by humans. Function allocation distributes the defined functions among available resources (humans, hardware, software or combinations). The allocation of some functions will be mandatory and predetermined by constraints established in the Mission Analysis or Requirements Analysis steps of design. Allocation decisions are determined by a comparison of performance among humans, hardware, and software; cost factors; anticipated operator workload; and cognitive support for the operators. Allocation decisions should be made to maximize total system performance and effectiveness. Allocation of a function may require redefinition of its component subfunctions. Function allocation will be guided by information requirements and decisions required to initiate, sustain and otherwise support mission functions. How these allocation decisions impact total system performance is determined and adherence to defined requirements assessed.

Function allocation has been aided with guidelines to assess human capabilities and function requirements, such as 'Fitts' lists' and evaluation matrices and while these guidelines are helpful, they can be dated and must be used judiciously. Today's technology affords us far more automation potential than in the past, but it comes with a cost. With automation comes a certain degree of cognitive distancing between the operator and the task; therefore, this automation must be applied carefully. Methods to support this step rely heavily on the engineer and designer and are not well automated themselves. This is an area ripe for more research and development of standardized methods.

3.5. Design

This step is where the bulk of human factors and ergonomics methods have traditionally focused and where the majority of human system issues are resolved. A time-based description of the allocated system architecture must now be defined, and interactions between tasks, and between humans and equipment examined. The flow of information and objects between components of the system is defined. The architecture is analyzed and functions and tasks are redefined as necessary. Determination is made as to whether the specified levels of activity (physical, cognitive, temporal) for humans and equipment can be met with projected or currently available resources. Once the functional architecture meets mission and system requirements, operator interfaces may be specified and designed. Changes made to the functional architecture at this step will require a return to earlier steps to ensure that all system/mission requirements will be met. The following substeps support the design step.

3.5.1. Task design and analysis

Human tasks required to ensure successful completion of the function are developed in terms of cues required to alert the human that a decision/action needs to be taken, the decision/action to be made, the information required to support the decision and mechanisms to implement the results of the decision/action. The critical characteristics and interactions are also articulated. In addition to supporting other steps of the human systems engineering process, Operational Sequence Diagrams (OSD) and Critical Task Analysis (CTA) methods are excellent tools for task definition and analysis.

3.5.2. Human interface design

Interfaces between humans and hardware, software and other humans are designed such that physical and procedural interfaces are considered. Link analysis and design option decision trees facilitate this substep of design.

3.5.3. Performance, workload and manning level estimation

Physical (perceptual, psychomotor, physiological, etc.) and cognitive workload levels for individuals and teams are estimated to provide approximate system performance. Workload stressors and their effects on human performance, operator coping strategies and the effects of task neglect/delay are defined. Workload and the resultant manning and training requirements should be optimized to meet required performance levels. Task network models and simulation provide mechanisms for early assessment of system performance with the defined human role. Subjective workload ratings during prototype testing also facilitate performance evaluation.

3.6. Verification

The potential performance of the system with respect to its ability to achieve required levels of operation is assessed. Verification may be carried out either during conceptual stages using analytical or executable system models or after a physical prototype or mock-up has been constructed using human-in-the-loop simulations. Verification of some system components may be concurrent with design of other components. If the system under design is unable to achieve the required levels of performance

and operation, then either the requirements must be altered or the design must be improved through re-allocation of functions or selection of an alternate design. Storyboarding with subject-matter experts, followed by prototype usability testing, will help verify the system meets its requirements. Early in system design, modeling can be employed to assess the design; however, it is critical to conduct human-in-the-loop testing of the system to fully verify the design.

4. APPLICATIONS

Human systems engineering has been effectively applied to system design efforts. The following are examples of projects that have employed the steps described in this chapter and supported the evolution of the human systems engineering process.

The NATO Seasparrow Surface Missile System (NSSMS) used this approach in the development of the graphical user interface (GUI) of the system (Wallace *et al.* 1995). In this program, mission and operational requirements were identified and OSD developed. The OSD articulated tasks that supported required functions and were used to define information requirements, displays and critical system events which impact decisions. This task information was then used for system design and assessment. Models of the design and simulations of human performance measuring workload were run. This design data was then used to develop prototypes, which were analyzed with human-in-the-loop testing. By applying this top-down approach, the human systems engineering team worked with the systems engineering team to define an optimized NSSMS GUI display.

A top-down approach facilitates design of all systems, not just GUI designs. A similar approach was applied in the redesign of the Naval Space Operations Center at the Naval Space Command (Bohan and Wallace 1997). This project focused on the re-design of a collaborative work environment where team interaction was critical. At project commencement, missions were defined and required functions identified. Tasks were then articulated with OSD, providing the framework for link analysis. A rapid prototyping tool modeled the layout and simulated communication. Based on this analysis, optimal layouts were prototyped for human-in-the-loop evaluations. This evaluation method provided two benefits to the project. First, the users provided insight about the design that engineers alone do not have; and second, the users fully supported the design as they were part of its development from cradle to grave. The newly designed center was unveiled in May 1997 and since proven successful.

Once the human systems engineering process was defined and documented for the Navy's SC-21 Science and Technology Manning Affordability Initiative and was instantiated within IEEE 1220-1998, the process was used to develop the web site (www.manningaffordability.com/hes) which provides a mechanism to 'teach' systems designers about human systems issues. Greater levels of detail of the human systems engineering process are included here.

5. CONCLUSIONS

The rapid and continual advancement of technology makes the human more likely to be the limiting factor in system design and performance, making it increasingly important that the human factors and ergonomics communities work together with

systems engineers. Communication and integration are central to *successful* design development and there are many mechanisms that facilitate both across systems design teams. Involvement in the professional groups and societies of the designers and specialty engineers with which we interact is an excellent mechanism for understanding the systems design process more fully and becoming familiar with specialized terminology. Advanced systems engineering environments that support the full systems engineering process can facilitate communication and integration while reducing design costs and errors by removing the need to translate system design data between disciplines, automatically flagging design changes and notifying the appropriate personnel.

Great gains in improved design will be realized by the use of a systems engineering terminology to articulate a human systems engineering process which feeds the systems engineering process directly. The human systems engineering process is built on the foundation laid by ergonomics and human factors communities over past decades. It is differentiated from more traditional approaches in that it is fully imbedded into the systems engineering process in practice and terminology. Use of a rigorous process, such as that defined here, is critical to ensuring enhanced system and human performance.

REFERENCES

BEEVIS, D., BOST, J.R., DORING, B., NORDO, E., PAPIN, J.P, SCHUFFEL, I.H. and STREETS, D., 1994, *Analysis Techniques for Man–Machine System Design*. NATO Technical Report AC/243 (Panel 8)TR/7, vol. 1, issue 2 (Brussels: NATO Headquarters, Defence Research Station).

BOHAN, J.A. and WALLACE, D.F., 1997, Team ergonomics and human engineering methods for the design of collaborative work environments: a case study. In *Proceedings of the Human Factors Society's 41st Annual Meeting*, Albuquerque, NM (Human Factors and Ergonomics Society), 1066–70.

CHAPANIS, A., 1996, *Human Factors in Systems Engineering* (New York: Wiley).

CHAPANIS, A. and SHAFER, J.B., 1995, Human Factors Methods. Workshop presented at the 39th Annual Meeting of the Human Factors and Ergonomics Society, 9–13 October, San Diego, CA.

IEEE 1220-1994, 1998, *Standard for Application and Management of the System Engineering Process* (New York: IEEE).

MALONE, T.B., 1997, *Ship/System HSI For Affordability and Performance Engineering*. Technical Report (Falls Church, VA: Carlow International, Inc.).

WALLACE, D.F., DAWSON, J.A. and BLAYLOCK, C. J., 1995, Tactical information GUI engineering and requirements specification (TIGERS): a top-down HMI engineering process. In *Proceedings of the Human Factors Society's 39th Annual Meeting*. San Diego, CA (Human Factors and Ergonomics Society), 1184–9.

Job Analysis and Ergonomic Assessment after Injury

L. S. Strickland† and G. S. Rash‡

†Spalding University, Louisville, KY 40203, USA

‡BIO/MED Research & Consulting, Louisville, KY 40241, USA

1. INTRODUCTION

Returning an injured worker to the work environment at the highest possible level of independent functioning is the ultimate goal of any rehabilitation effort. Assuring the long-term success of the individual within the worker role will require a well-stated problem encompassing the individual's "design" and the design of the workplace. For the ultimate goal of returning a worker to the workforce, the job site evaluator must assess the worker, the work site and the actual work being performed.

2. JOB SITE ANALYSIS

2.1. Need for Job Site Analysis

When working with an injured worker there is need to perform a job site analysis to:

- focus the rehabilitation effort and determine the specific end goals that are meaningful, relevant and measurable;
- determine work site modification needs if the individual will not be able to attain his/her prior level of physical or cognitive functioning;
- ensure an appropriate match between the injured worker and his/her previous job to reduce potential future injuries; and
- assist the employer in making the job easier, safer and less tiring.

2.2. Physical Demands

Job analysis is a process of obtaining an accurate description of all job requirements. This evaluation is used in conjunction with other industrial therapy services in the effort to return an injured worker to the job. Detailed information in all the following job demands should be included as part of the job analysis:

- Walking.
- Standing.
- Sitting.
- Climbing.
- Crouching.
- Bending.
- Lifting.
- Carrying.
- Pushing/pulling.
- Reaching.
- Handling and fingering.

Other job demands such as sensory, and cognitive and psychological stress must also be taken into consideration to return successfully an injured worker to the job. The job site analysis incorporates the worker and their abilities along with assumptions of potential or future exposure to unsafe or harmful work practices.

2.3. Recommendations

After conducting the analysis, recommendations should be made about environmental modifications, tool and equipment modifications, individual work practices and the workers health maintenance. Environmental modifications, such as changes in the work station so that the worker has a sit/stand option, or changes in the lighting may be appropriate. Incorporating the use of shock-absorbing gloves, cutting or remove ledges under tables, etc. may be recommended as tool or equipment modifications. Additionally, modifications to the individual's work practices and techniques such as reducing reaching above shoulder level or reducing forward bending may be included. Last, the worker's health maintenance and general wellness may need addressing. These recommendations may consist of programs involving ongoing exercise programs to reinforce good body mechanics and posture, or simple 1–2-minute neck and upper back exercise breaks every 2 h.

3. TRAINING

To make extensive ergonomic recommendations of the work site, the practitioner should have received specific training in ergonomic studies and be recognized as a qualified ergonomics practitioner (BCPE 1995). Several government publications can be used as references in the development of the practitioner's knowledge of job analysis and include *Dictionary of Occupational Titles*, 4th edn (1991), *Handbook for Analyzing Jobs* and *A Guide to Job Analysis* (Superintendent of Documents). To assure a well-stated problem that includes recommendations, practitioners working in industrial rehabilitation and ergonomics must have education in, and be familiar with, the following areas:

- Workplace design.
- Anatomy and physiology.
- Kinesiology.
- Biomechanics.
- Anthropometry (the science of the variation in body size among individual, genders and races).
- Sensory nervous system.
- Musculoskeletal system.
- Normal and abnormal psychology.

4. TOOLS OF THE TRADE

In preparing for a job site analysis, the practitioner must gather the appropriate tools to be used in the assessment process. These items may include (but are not limited to) a written job description provided by the employer, tape measure, stopwatch, scale, push–pull gauge, chalk, goniometer, plumb line, Polaroid (or digital) camera, photo release forms, laptop computer and a video camera. The employer should provide prior approval of any pictures and/or videotaping of a work site and/or work production practices. Sample job site analysis forms are given in Tables 1 and 2.

The exact form a practitioner uses within their practice will be based on many factors. The type of work environment and jobs being evaluated, the needs of the worker, employer, and other interested parties such as insurance case managers. The format should be developed in response to feedback from referral sources. Many employers and case managers may want recommendations and concise summaries on the front page of the evaluation with the detailed analysis after the summary. The analysis forms provided do not take into account reasonable

Table 1.

JOB SITE ANALYSIS

Job Title: _____ Employee: _____

Location: _____

Work Hours: _____ to _____ Days _____ Overtime Per Wk. _____

The Job

General Job Description:

Equipment, Tools, Work aids:

Essential Job Functions:

Time Spent:

Standing _____ hrs. Sitting _____ hrs. Inside Bld. _____ hrs. Out of door _____ hrs.

Activity	Yes	No	Repetitive (Frequency)	Continuous (# Hours)	Comments
Stoop/Bending					
Squat/Crouch					
Kneeling					
Climbing					
Reaching Overhead					
Manipulate/Finger					
Grasping/holding					

Lift/Carry	Hrs/Reps/Position	Lift/Carry	Hrs/Reps/Position
0 – 10 lbs.		11- 20 lbs.	
21 – 50 lbs.		Over 50 lbs.	

Environmental Conditions:

Cold _____ Heat _____

Lighting:_____ Vibration: _____

Dampness _____ Dust/Fumes _____ Noise _____

Other:

---

accommodations that need to considered when determining if an individual can return to a particular job. These accommodations will be unique to the situation based on the employer's resources and the abilities of the worker. If appropriate, the reasonable accommodations can be included under the "other comments" section of the evaluation.

4.1. Accounting for Demographics

Workforce demographics have been changing over the past two decades. The face of the workforce is more female and older in age than the workers in the past. Large increases in female and older workers have an impact on workplace safety and injury management. Age and gender have been correlated with dynamic balance skills, ability to climb and the ability to lift. Older workers, female workers and older female workers may be at high risk for injury and, once injured, may face more challenges in returning to work. These individuals may also be more susceptible to re-injury. The 1991 revised National Institute of Occupational Safety and Health (NIOSH) lifting equation defines lifting capacity of young, healthy workers; therefore, the job site evaluator must

The Person

Relevance & Meaningfulness of job:

Values & Interests:

For each performance component, determine the level of challenge required to perform job	Level of Performance			Comments
	Min	Mod	Max	
Physical Components • Sensation/ Sensory Integration				
• Visual Perception				
• Range of Motion				
• Strength & Muscle Tone				
• Coordination				
• Balance & Posture				
• Endurance				
Mental ▪ **Cognitive:**				
▪ **Decision making:** ▪ **Concentration requirements:**				
• **Multiple tasking/Shifting attention:**				
• Mood & Affect				
• **Behavior**				
• **Thought content**				
• **Emotions**				
• **Body Image**				
• **Volition**				
Sociocultural • Cultural values, beliefs, language				
• Interpersonal skills				
• Management of social roles				

The essential functions of a job are those duties that must be performed unaided or with responsible accommodations

Min= is needed for job, but not for essential job functions, occasional, required< $1/3^{rd}$ of the time

Mod= is needed to carry out essential job functions, but is not the major component(s) for essential function Frequently, required > $1/3^{rd}$ < $2/3^{rd}$ of the time

Max= is needed to carry out essential job duties -- is the major component(s) needed for essential job function, constant, required > $2/3^{rd}$ of the time

consider the context of the work as it relates to the individual worker's abilities to assure success.

4.2. Workers Physical Conditioning

An individual's physical conditioning may have a profound effect on work capacity. Poor conditioning and unhealthy lifestyles can lead to an increase in risk of injury. Post-injury, the worker should be involved in a structured health and fitness program after formal medical interventions are completed to progress cardiovascular conditioning, heighten flexibility and strength to help control

fatigue and improve body posture and positioning during work tasks.

5. CASE EXAMPLE

A previous case will be used to illustrate the process. A 42-year-old female warehouse employee who had worked for the company for 4 years sustained an acute lumbar strain 1 week before while working the first shift in a snack food warehouse. The employee is a union employee. Soon after the onset of pain, 3 h into an 8-hour shift, the employee informed her supervisor of the injury.

```
┌─────────────────────────────────────────────────────────────────┐
│ Recommendations:                                                  │
│                                                                   │
│        Environmental Modifications                                │
│                                                                   │
│                                                                   │
│                                                                   │
│                                                                   │
│        Tool/Equipment Modifications                               │
│                                                                   │
│                                                                   │
│                                                                   │
│                                                                   │
│        Modification of Work Techniques/Approach                   │
│                                                                   │
│                                                                   │
│                                                                   │
│                                                                   │
│        Health Maintenance/Wellness Recommendations:               │
│                                                                   │
│                                                                   │
│                                                                   │
│                                                                   │
│ Other Comments:                                                   │
│                                                                   │
│                                                                   │
│                                                                   │
└─────────────────────────────────────────────────────────────────┘
```

_____   _____
Consultant   Date

(Adapted from Burt & Smith, 1996 Watson 1997)

The employee has received a muscloskeletal assessment from a physical therapist and has received therapeutic modalities to modulate symptoms and promote soft tissue relaxation and healing. Exercises for mobilization and spinal restriction correction were performed for 5 days after the incident.

5.1. Injured Worker's Job Site Analysis

Job analysis is the foundation for providing information to assist in returning the injured person to the workforce. The analysis should include information pertaining to job tasks and/or specific services for an injured employee. Identification of stressful postures and positions, ergonomic recommendations and reasonable accommodations to return the worker quickly to the same or alternate work tasks are the goal. Table 2 shows the job site analysis forms completed for the worker in question.

From the job site analysis, it can be seen that the significant items are that the worker is on her feet for the entire work shift. She lifts items weighing 10–12 lb throughout the majority of her work day and she has routinely to stoop and bend many times a minute continuously for durations of ~2 h. Additionally, she must routinely reach overhead, which extends her trunk.

5.2. Injured Worker's Job Analysis Recommendations

Most return-to-work recommendations typically include generic suggestions, which would be applicable to any worker doing the

Table 2.

JOB SITE ANALYSIS

Job Title: _Warehouse loader_____ Employee: _____Jane Worker_____
Location: _____Warehouse area C_____
Work Hours: __7:00AM_ to _____4:30PM_ Days ___5/wk___ Overtime Per Wk. _≈ 2 hrs./day 3-5
days/wk_mandatory 12 hrs. per week during part of year, often volunteers for overtime__

The Job

General Job Description:
1. Picking of products from warehouse shelves to supply orders.
2. Placing items in 24"X 32" box cases.
3. Tapping cases closed with strapping
4. Loading 10-12.5 # cases of snack food products into trailer at a loading dock.
5. Checking off orders and signing out

Equipment, Tools, Work aids:
6 foot Ladder to reach snack food products on overhead shelves which are 8 and10 feet off ground

Essential Job Functions:
Loading 10 –12.5 pound cases of snack food product into trailers, Picking items off shelves, placing into cases.

Time Spent:
Standing _8-10_hrs. Sitting ____-0-_ hrs.   Inside Bld. ___8-10___ hrs.   Out of door ____-0-_ hrs.

Activity	Yes	No	Repetitive (Frequency)	Continuous (# Hours)	Comments
Stoop/Bending	X		5-8 times/min.	1-1 ½ hrs	Performs this task when picking products off shelves. Performs this task for 2 hours, then switches to loading. Bends over box while strapping box
Squat/Crouch	x		1-2 times/min		When loading trucks, squats to lift 24"x32" boxes
Kneeling		x			
Climbing	X		3X/5min		Climbs ladders to retrieve some items needed for loading. Most items are located from waist to 5" overhead height
Reaching Overhead	X		7-8/min		To pick products off shelves, last 4 cases loaded on truck require overhead lifting of cases
Manipulate/Finger	X		20/min		Tripod pinch & release light weight 8-12 oz bags of snack food product
Grasping/holding	x				Holds bags of product for ≈ 3 seconds Holds boxes of orders for 1 minute while loading trucks

Lift/Carry	Hrs/Reps/Position	Lift/Carry	Hrs/Reps/Position
0 – 10 lbs.	6 hrs./8 hrs shift 140 times per hr.	11- 20 lbs.	8-10 times per hour
21 – 50 lbs.	Rarely	Over 50 lbs.	Never

Environmental Conditions:
Cold ___unheated warehouse_____ Heat_____unaircond. warehouse_____
Lighting:__45 foot candles_____ Vibration: _____--_____
Dampness _____--_____Dust/Fumes _____--_____ Noise _____--_____
Other:
Concrete floors – standing throughout the day

---

specific job. This is because more times than not the first job analysis performed by a professional is after a worker has been injured. This was the case in the present example. Simple items like anti-stress mats in front of the product shelves for reducing muscle fatigue, using a picking tool for the lightweight 8–12 oz bags of snack foods to reduce the hyperextension of the trunk, and lifestyle changes to incorporate exercising would be helpful for most all workers.

Specific to the present worker is the use of the picking tool for the lightweight snack food bags. This would reduce the number of times the worker flexes the trunk and goes into extreme extension. Additionally, she needs to stretch her neck and upper back a minimum of once an hour and she needs more alternating of job task. She should not remain in one position for > 10 min and should alternate between task more routinely. With respect to lifting, she needs to avoid frequent lifts of > 10 lb and restrict repeated trunk bending and rotating. This worker will also need an exercise program designed by a rehabilitation therapist to address her specific needs and a lifestyle change, which includes routine exercise.

The Person

Relevance & Meaningfulness of job: Worker is a single mother. Position at warehouse provides sole income for family of four. Worker reports satisfaction in her role within the organization.
Injury: Acute Lumbar strain at L3-5.

For each performance component, determine the level of challenge required to perform job	Level of Performance			Comments Worker has been using a lumbar support belt periodically during loading tasks.
	Min	Mod	Max	
Physical Components • Sensation/ Sensory Integration		X		
• Visual Perception		X		
• Range of Motion			X	**Movement in trunk extension, R sidebend is restricted. Job requires overhead reaching. Worker does no routine fitness program outside of work.**
• Strength & Muscle Tone			X	**Job requires lifting, bending and carrying throughout day. 10 –12.5 # lifted No routine conditioning program**
• Coordination		X		**Tripod pinch used during picking of product from shelves**
• Balance & Posture			X	**Tasks of climbing, reaching overhead and carrying. Worker has rounded shoulders, no routine stretching program. Poor body mechanics demonstrated during lift & carry**
• Endurance			X	**Job requires lifting carrying and climbing through out shift.**
Mental ▪ **Cognitive:**		X		
▪ **Decision making:** ▪ **Concentration requirements:**	X			
• **Multiple tasking/Shifting attention:**	X			
• Mood & Affect	X			
• **Behavior**	X			
• **Thought content**	X			
• **Emotions**	X			
• **Body Image**	X			
• **Volition**	X			
Sociocultural • Cultural values, beliefs, language	x			
• Interpersonal skills	X			
• Management of social roles	X			

The essential functions of a job are those duties that must be performed unaided or with responsible accommodations
Min= is needed for job, but not for essential job functions, occasional, required< 1/3rd of the time
Mod= is needed to carry out essential job functions, but is not the major component(s) for essential function Frequently, required > 1/3rd < 2/3rd of the time
Max= is needed to carry out essential job duties – is the major component(s) needed for essential job function, constant, required > 2/3rd of the time

6. CONCLUSION

From the example given, one can see that except for the added task of accounting for the workers specific injury limitations, a post-injury job site assessment is really no different than any other job site assessment. However, this is a significant exception. As stated in the training section, the knowledge of workplace design, anatomy and physiology, kinesiology, biomechanics, anthropometry, the nervous system, the musculoskeletal system and normal/abnormal psychology are needed. Many of us lack the knowledge of how the pathology of the condition affects the other components. This can be learned over time, or one can solicit the aid of a medical/rehabilitation specialist to deliver a team approach. The later approach being the preference for these authors.

Recommendations:

Environmental Modifications:

Flooring: Anti-stress mats located in front of the product shelves would reduce muscle fatigue

Tool/Equipment Modifications
Reaching overhead & bending: The weight of the product being "picked" off shelves weights 8-12 oz. Worker could use a lightweight reacher when obtaining product off high and low shelves in warehouse to avoid extreme truck extension and flexion.

Modification of Work Techniques/Approach
Breaks: Provide two, 1-2 minute neck and upper back exercise breaks every two hours

Alternate job tasks: Frequent alternation between sit, stand and walking should be performed. The worker should not remain in one position for longer than 10 minutes and should alternate between tasks every two hours.
Lifting: Avoid frequent lifts above 10#, restrict repeated trunk bend and trunk twist

Health Maintenance/Wellness Recommendations:

Stretching and Flexibility program as designed by rehabilitative therapist. Lifestyle modifications to incorporate flexibility, strength and endurance activities into daily activity configuration.

Other Comments:

REFERENCES

AJA, D., 1996, Finding a niche in job-site analysis. *Occupational Therapy Practice*, 1, 36–41.

BLAZ, J., 1998, Fit for duty, how OT-managed health and fitness programs can prevent work-site injuries and reinjuries. *OT Week*, 12, 12–13.

BOARD OF CERTIFICATION IN PROFESSIONAL ERGONOMICS, 1995, *Information on Certification, Policies, Practices, and Procedures* (Bellingham: BCPE).

BURT, C.M. and SMITH, P., 1996, Work evaluation and work hardening. In I. L. W. Pedretti (ed.), *Occupational Therapy Practice Skills for Physical Dysfunction* (St Louis: Mosby), 556.

CALLAGHAN, J. and SWITZER-MCINTYRE, S., 1994, Job analysis. In G. Key (ed.), *Industrial Therapy* (St Louis: Mosby), 86.

ISERHAGEN, S.J., 1988, *Work Injury: Management and Prevention* (Rockville: Aspen).

MILLER, D.M., 1998, Industrial therapy: diverse and dynamic. *Occupational Therapy Practice*, 3, 36–40.

SUPERINTENDENT OF DOCUMENTS, 20402 (Washington, DC: US Government Printing Office).

WATERS, T. R., PUTZ-ANDERSON, V. and GARG, A., 1994, *Applications Manual for the Revised NIOSH Lifting Equation* (Cincinnati: National Institute for Occupational Safety and Health).

WATSON, D.E., 1997, *Task Analysis: An Occupational Performance Approach* (Bethesda: American Occupational Therapy Association).

Job Load and Hazard Analysis

M.K. Mattila

Occupational Safety Engineering, Tampere University of Technology, PO Box 541, FIN-33101 Tampere, Finland

1. INTRODUCTION

Job Load and Hazard Analysis (JLHA) is a simple and standardized method for workplace investigations for occupational healthcare and for occupational safety management. The method is best suited for screening of the risks at jobs, preventing health and safety hazards at the workplace, and for determining the contents of the occupational healthcare program and the safety prevention program.

The method was developed in Finland after the Finnish Occupational Health Care Act came into force in 1983. The Occupational Health Care Act requires workplace investigations for the analysis of hazards inherent in the work and working conditions. Workplace investigations as the basis of occupational healthcare is a new concept first defined in the Act. It soon became evident that workplace inspections were a central problem hindering enforcement of the Act (Mattila 1985). The aims of the method are:

- identification of the health hazards to which workers are exposed in their work and by their working conditions;
- to help determine the contents of the occupational safety and health program according to the real needs that arise from jobs and working conditions; and
- to support the prevention of hazards and to facilitate the selection of appropriate preventive measures.

2. BASIS OF JOB LOAD AND HAZARD ANALYSIS

The JLHA comprises the identification of health hazards, their assessment, conclusions and proposals concerning measures to be taken, and follow up. The method is based on the principles of job analysis, self-evaluation, participatory ergonomics, risk assessment and group work. At the same time the conduct of the analysis result is based on the expertise of safety and health professionals.

Information is collected during the visit to the workplace using forms of analysis where the number of items is kept as small as possible. Different types of loads and hazards are considered:

- chemical hazards;
- physical hazards;
- physical workload;
- mental stress factors; and
- injury (accident) risk.

Each item is assessed on a three-point ordinal scale according to occurrence and relevance to workers' health. Written guidelines for assessment are provided (Mattila 1989).

The information is gathered, and the assessment is made in two ways: (1) the occupational healthcare and safety personnel interviews workers and makes observations at the workplace during an inspection tour, whereafter they make their assessments;

and (2) the workers complete a questionnaire form (in which five items are illustrated with several typical examples). It is suggested that a group carries out the analysis at the workplace where, for example, the physician, the nurse, the safety supervisor and worker representatives take part.

3. SAFETY AND HEALTH IMPROVEMENT PROGRAM BY COOPERATION

The cooperation team formed for group problem-solving discusses the findings. The team can comprise, for example, the safety specialists who participated in the safety inspection and, whenever possible, a member of line management. Working together, the team comes up with a comprehensive assessment of risks of jobs and makes ideas for preventive measures.

Next, the occupational health professionals independently evaluate the jobs on the basis of all data. These conclusions, together with proposals for correction or prevention, are then presented to the management of the company. As a result the occupational health professionals present a proposal for the occupational health program.

In this way the JLHA method organizes the cooperation between different levels of the company and yields information about job safety. The following connections and effects are considered particularly important:

- The occupational health program is based on the results of the risk assessment of the workplace.
- The occupational safety personnel receive information about hazards at work while they are involved surveying the workplace.
- Occupational health professionals have access to the occupational safety personnel's information pertinent to workplace hazards, e.g. accident data or the result of industrial hygiene measurements.
- Management obtains the information it needs about operational disturbances, hazards and workers' experiences, and the hazard survey of the workplace can lead to improvements in the production system and to planning for preventive measures.
- The risk assessment of the workplace becomes part of the company's safety information system and constitutes a database on work conditions.

4. RECOMMENDATIONS

The JLHA method has the following characteristics:

1. Job analysis makes it possible to analyze all working conditions. This approach has proved practical in linking information about hazards and stress to the contents of occupational healthcare.
2. The cooperative approach is realized through a cooperation team that evaluates the gathered information. The cooperation team contributes the advantages of group problem-solving. The results become more comprehensive, different personnel groups are concerned in the process and have the change to express themselves, the necessary cooperation takes place, and the acceptability of and confidence in the assessment improves.
3. Workers' active role in hazard screening guarantees that all the problems recognized at the workplace come up for closer

consideration. Workers have proved themselves ready to play an essential part in identifying hazards.

4. The independent role of occupational healthcare personnel is further strengthened when they make assessments independently, and when their conclusions and proposals are reflected in the contents of the occupational healthcare program.

5. The JLHA has numerous indirect, special advantages: proposals for measures to eliminate unsatisfactory working conditions are more numerous and of better quality, cooperation between occupational healthcare and occupational safety receive an efficient forum, workers' confidence in the analysis increase, and the analysis has an invigorating effect on all occupational safety activities at the site and obviously have a positive effect on its safety climate.

6. The method offers numerous possibilities for the utilization of information. Most of the utilization occurs in occupational healthcare, where the information is useful in the initial medical examinations of new workers, in advising workers about safe performance, and in studies on the relations between working conditions and workers' health status. But the information has also other users, e.g. in preventing hazards during planning, in production design and in management.

7. Systematic workplace investigations may be one part of the information system of organization and may help the management in guiding the system, thereby contributing to the realization of safety objectives.

REFERENCES

MATTILA, M., 1985, Job load and hazard analysis: a method for the analysis of workplace conditions for occupational healthcare. *British Journal of Industrial Medicine*, 42, 656–66.

MATTILA, M., 1989, Improvement in the occupational health program in a Finnish construction company by means of systematic workplace investigation of job load and hazard analysis. *American Journal of Industrial Medicine*, 15, 61–72.

MATTILA, M. and KIVI, P., 1991, Hazard screening and proposals for prevention by occupational health service: an experiment with job load and hazard analysis at a Finnish construction company. *Journal of the Society of Occupational Medicine*, 41, 17–22.

Medical Equipment Usability Testing: An Introduction

V. J. Rice

Ammed Center & School, Ft Sam Houston, TX 78234–6131, USA

1. USABILITY TESTING INTRODUCTION

The proof of the pudding is in the eating.

Usability testing is conducted in order to develop a product that, when placed in the hands of users, will perform as intended. Conducted iteratively throughout the product development process, usability professionals seek to prevent production of a poorly designed product that will compromise the safety and productivity of the individuals who use it.

McClelland has stated that usability testing is the systematic evaluation of the "interaction between people and the products, equipment, environments, and services they use" and "is the fundamental principle that underpins all ergonomics" (McClelland 1990, p. 218). However, what is included in the "systematic evaluation" of the "interaction" often depends on the training of the individual conducting the usability testing. Usability testing is not merely establishing user preference regarding what users like or dislike about a product, nor is it concerned with merely handing out equipment to physicians or other healthcare workers and asking for their open-ended, informal feedback. Stanton and Baber (1996) suggest usability testing consists of the eight factors seen in table 1, while Booher (1990) suggests five ingredients to evaluate equipment including manpower, personnel, training, human factors, system safety, and health hazards (table 2). Combining the efforts of these two authors with personal experience reveals the authors' expectations of usability testing (table 3). Usability testing is a comprehensive, structured approach intended to answer questions about whether the end users of equipment can use it easily, safely, effectively, and efficiently.

Table 1. Components of Usability Testing (Stanton and Baber, 1996).

Components	Definition
Attitude	Acceptable performance within acceptable human costs (fatigue, stress, frustration, discomfort, satisfaction).
Effectiveness	Acceptable performance should be achieved by a defined proportion of the user population over a specified range of tasks and enviroments
Learnability	Allow users to reach acceptable performance within specified time frame.
Flexibility	The product should be able to deal with a range of tasks, beyond those first specified
Task match	The needs and requirements of the user should match the functions provided by the system
Task characteristics	The frequency with which a task can be performed and the degree of modification according to variability of information requirements
Product utility	How frequently the product is used
User characteristics	Knowledge, skills, and motivation of the users

Table 2. Components of MANPRINT (Booher, 1990).

Domain	Description
Manpower	Number of human resources required and available to operate and maintain equipment/system
Personnel	Aptitudes, experience, and other characteristics necessary to achieve optimal system performance
Training	Requisite knowledge, skills, and abilities needed by personnel to operate and maintain equipment/system
Human factors Engineering	Comprehensive integration of human characteristics into system development, design, and evaluation to optimize human-machine interface and performance
System safety	Ability of the system to be used, operated, and maintained without accidental injury to personnel
Health hazards	Inherent conditions in the use of the equipment /system that can cause death, injury, illness, disability, or reduce job performance of personnel

1.1. Process/Procedure

To be most effective, both in terms of ensuring a quality product and keeping costs low, usability testing should be incorporated throughout the design process. The process typically incorporates the nine steps shown in figure 1, which are reiterated during each phase of product development: prototype/initial development, laboratory-based efficacy testing (static, mock-up, simulator), and field-testing. (For a detailed description of the steps involved in usability testing, see Rice 1995.) During the first steps of the process, the human factors professional meets with product developers, management, purchasers, salespersons, subject matter experts, and users (in medical applications, the user may be the medical technician or care provider, while the "purchaser" may be the patient who is "purchasing" the service or treatment). Using a variety of techniques (interview, questionnaire, focus groups, observation, task/function analysis, etc.), the design objectives for each group are identified according to their desires for the final outcome features of the product. Usability evaluations should identify both positive and negative aspects of a product, from the points of view of all interested parties. By identifying what constitutes a successful product for all, the results of user testing can help to generate a product that is desired by and does not harm the end user, and creates financial gain for the manufacturer. The critical success factors for each group are then incorporated into measurable performance criteria for experimental testing. The success factors are prioritized, to determine which are critical and which may used as tradeoffs between groups, should that become necessary.

After the nine steps in the usability assessment are completed (table 3), the results are used to reassess the product design, and the process is started again, in a consequent stage of the development process (product development, laboratory testing, field testing). Although usability testing is most effective when conducted in an iterative fashion throughout the design process, other procedures are used. For example, instead of repeating the entire nine steps, one or two steps may be repeated, if the desired changes so warrant.

Table 3. Components of Usability Testing (Rice, 1999).

Component	Description
Task match/ Effectiveness	1) Establishment of critical success factors for all interested parties (manufacturer, developer, purchaser, user) and assessment of the products in terms of meeting those factors. 2) Flexibility to meet changing needs.
Human factors	1) Match between known human characteristics (abilities and limitations of physical, cognitive, and emotional domains) and requirements to operate equipment. Educational level required. 2) Detailed description of the user populations, along with match or mismatch with equipment requirements 3) Selection criteria, if necessary: aptitude, experience and other characteristics necessary to safely operate equipment.
User performance	1) physical 2) cognitive (decision making, etc.)
User attitudes	Users subjective ratings including ease of use, preference, perception of safety, comfort, etc.
Training	Amount of training, type of training (written, oral, performance, training for stressful enviroments), depth of training, user knowledge and skills necessary to benefit from training, need for over-riding previous learning, and efficacy of training.
Manpower	Number of personnel to operate and maintain equipment and/or system.
System safety	1) Ability of the system to be used, operated, and maintained without accidental injury to personnel, and within acceptable human costs (fatigue, stress, frustration, discomfort, satisfaction). 2) Comprehension of and efficacy of warnings.
Human error	Evaluation of potential human error and establishment of failsafe designs or error tolerant systems.
Health hazards	Inherent conditions in the use of the equipment/ system that can cause death, injury, illness, disability, or reduce job performance of personnel.
Integration	Integration of the equipment or operational system within the existing procedures and organizational enviroment.
Cost	Costs listed in other categories (manpower, training) are combined with other costs (production, supplies) to achieve and estimated cost comparison or cost/ benefit ratio.

1.2. Fidelity

If the field testing of the equipment is accomplished in a simulated environment, such as in an integrated anesthesia training environment, then a final assessment conducted during or immediately following actual use is suggested to ensure ecological validity (Bogner 1998). Difficulties with using the equipment may be identified during or immediately following actual use that would not be identified during simulations. This was the case in research on decision-making under stress in which Mackenzie and his colleagues (1996) videotaped actual emergency room cases in which difficulty of package removal and inaccessible data via a video display terminal were revealed.

The proof of the pudding is in the eating. This means the product can only be determined as being safe, effective, efficient, and acceptable when it is put to use by the user population. This is the basis of usability testing. However, information should also

1 SME / USER / INVESTIGATOR

3 ESTABLISH DESIGN OBJECTIVES

5 DEVELOP PERFORMANCE CRITERIA

6 ESTABLISH MEASUREMENT TECHNIQUES

7 WALK-THROUGH / TRAINING

8 ASSESSMENT (conduct testing, evaluate data, identify results)

9 PRODUCT (assess results in terms of product design)
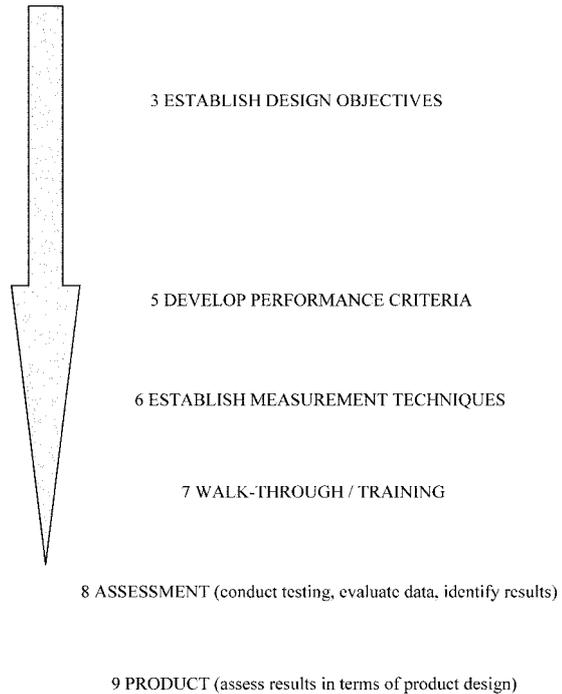
Figure 1. Usability Testing Process., SME = subject matter expert)

be gathered immediately after new products go on the market. If this is not done, direct evidence of success may be missed as people tend to complain about a product that is difficult to use or causes pain immediately after its use; but rarely voice their pleasure. Immediate follow-up with new users results in quick retrieval information on the product.

To reiterate, usability testing of medical equipment should yield a design, which achieves the critical goals of all persons who develop, manufacture, sell, purchase, prescribe, and use the equipment.

2. UNIQUE CONSIDERATIONS FOR MEDICAL EQUIPMENT USABILITY TESTING

The unique considerations for usability testing of medical equipment include the user population (aptitude, capabilities, and limitations), the context in which the equipment will be used, the complexity of the equipment, and the integration of the equipment into medical or rehabilitation based practice.

2.1. User Population

The *user population* includes medical personnel (physicians, nurses, technicians, therapists), patients, and non-professional healthcare providers, such as friends and family members. Therefore, the patient's disease process, the residual capabilities of the patient, and the skills and limitations of each user group must be carefully considered during the design process. Obviously, the educational background, experience, and personal characteristics of these groups are diverse.

The healthcare system is increasingly moving toward the use of technical level personnel. The physician may not be the sole user of complicated medical equipment. Instead, physicians

assistants, nurses, and a variety of specialty technicians may share in the responsibility. A physician typically has seven to ten years of education post bacheloreate degree, while many medical technicians attend a two-year trade school following high school. The design of all aspects of the equipment and/or system should be aimed at the lowest educational level of potential users.

Increased longevity of the normal aging population, of rehabilitation patients (for example, persons experiencing spinal cord injuries), and of those with newly identified diseases (such as autoimmune disorders) has increased the population requiring home-based rehabilitation equipment. The individuals needing care may be providing it themselves and the symptoms of their disease or disability may interfere with their ability to correctly use the equipment. Care may be furnished by lay providers, who may be retired or infirm themselves. For example, lay providers may be of high school or college age and need to furnish care before their working parents return home. Increased longevity and the increase in home-based care has created even greater diversity in the capabilities of user groups (Bogner 1998). The strengths and capabilities of these individuals (users) are not within the typical human factors' database used to match task requirements with human capabilities. The need for identification of user characteristics and evaluation of user groups are therefore even more important to prevent human error in the operation of home-based equipment.

The aptitude, skills and knowledge necessary for operating the equipment, as well as for understanding training and warning labels of the users, must be known. Equipment use, instructions, training programs, and warnings should be specifically designed for the user populations. However, this is not always the case, as shown in research conducted on readability of patient education literature. Merrit, Gates, and Skiba (1993) found that pamphlets designed for use by persons experiencing hypercholesterolemia required college-level reading skills according to three different methods for analyzing literacy levels. Thus, most of the individuals in the United States who could benefit from these pamphlets would be unable to comprehend and utilize information important to their health and well-being. Techniques such as editorial reviews and readability scoring should be part of usability testing for training, instructions, and warnings.

Systems designers and managers sometimes use training to make up for design limitations; however, this can result in unexpected and excessive costs for training courses, facilities, and devices, while not necessarily assuring an appropriate level of performance. An even worse prospect, poorly designed training, can induce unanticipated error which can potentially harm care-givers or their patients. Training which is specifically designed to accomplish learning within a particular context must be carefully designed. For example, if the individual receiving training will be functioning in a highly stressful environment, then specific training strategies designed to ensure fidelity of training should be used. Although the manufacturer cannot always follow their product to ascertain whether training has been conducted, it can (and perhaps should) become part of the sales package. In a study focused on the evaluation of mechanical aids for moving patients, trained nurses were used as subjects. Although the nurses were responsible for moving patients using mechanical devices as part of their job, 61% had received only informal on-the-job training. The researchers suggested specific,

practical training in assessing situations, selecting the correct sling, and correctly using mechanical aids for nursing personnel (McGuire, Moody, and Hanson 1996). In the introduction of new equipment and/or training, it is important to remember that the new learning may be superseded/interrupted by prior knowledge. By considering training during the equipment development process, it can be carefully developed as part of the total equipment package, rather than as a backup for poor design.

2.2. Context

The *context* in which the item is to be used is of particular consequence. As pointed out by Bogner (1998), without consideration of the context of use, it is impossible to identify factors (including equipment design) that may contribute to human error. Some of the contextual factors within medical-based facilities include high stress, time pressures, substantial workload, sleep deprivation, and staffing shortages. It is well known that decision-making and cognition are altered by stress levels, sleep deprivation, and fatigue. These potential contributors to human error should be considered and evaluated as part of usability testing, either through use of known information (databases) or real-time research.

One method of attaining ecological validity is through "naturalistic" research. This may involve computer controlled simulations or real-world video taping to delve into the medical contextual environment. This method is suggested for evaluating the medical environment, which is highly dynamic and uncertain, having ill-defined and/or changing goals, and entailing considerable time.

Other aspects of the medical arena, tasks, and cognitive processes of users must also be assessed and utilized in the design process. Cognitive research using a concurrent verbal ("think aloud") protocol has proved successful for assessing physician's informational needs in medical computing, and for use in iterative system development. Using this technique yields information that appears unattainable through interview, questionnaires, and simple observation. According to the authors, this technique helps the designer to comprehend the context, thought processes, and decision-making in healthcare operations so that medical information systems and decision support systems can be designed accordingly.

The context for rehabilitation or assistive devices differs from that of hospital-based equipment. Home-based products can potentially be used in the patient's home, at the beach, in a car, at work, or on a sports field. Context also can include users being relaxed and undistracted or panicked, frenzied, and sleep deprived. The weather could be sunny and bright, foggy, windy or rainy. Fail-safe designs, with several safety features may seem like "over-kill"; but such designs may ultimately prove themselves cost effective both in terms of safety and user preference. At the very least, tolerant designs should be implemented for cases in which dire consequences could result.

At times, usability testing is either not included, or it is not accomplished during the development process or in the environment in which the product will be used. Although not optimal, evaluations of already existing products can be completed in order to recommend design changes to the manufacturer or to give recommendations to those who purchase or prescribe equipment. Most patients and family members do not select their

own equipment without input from healthcare personnel, and many pieces of equipment are prescribed by healthcare personnel. One example of post-production usability testing is that conducted on the effectiveness of mechanical aids for moving patients from their beds to chairs, from their chairs to the bath and shower, and vice versa. McGuire and colleagues (1996) evaluated operator and patient safety, subjective evaluation (confidence, comfort, effort), performance, and advice needed during testing. Although areas for potential improvement of the equipment were noted, the crux of the article centered on the general guidelines delineating selection criteria for those purchasing and prescribing mechanical aids for their patients.

As noted in the opening section of this chapter, evaluating equipment during or immediately following actual use can be helpful and can address contextual issues. Questions can be answered as to whether people are using the product, its effectiveness, and their level of satisfaction with the product. Survey techniques can be used. In one such study, persons with spinal cord injuries were surveyed about the use of bowel care/shower chairs at home. Over 66% of those who regularly used these chairs felt their safety was compromised during their use and less than 54% were satisfied with the design and usability of the chairs. Caregivers also reported being dissatisfied with the chairs. Reports of patient falls (35%) and pressure ulcers (24%) stated they were specifically related to chair use. Other issues raised were lack of hand access to perianal area for digital stimulation, difficulty rolling and turning the chair, difficulty keeping the chair clean, and lack of durability of the seat and brakes on the chair (Nelson, Malassigne, Amerson, Saltzstein, and Binard 1993).

Without assessing the equipment in the environment in which it will be used, the information gained will be incomplete, the equipment may or may not be as effective as it could be, and potential harm to users may be missed. Usability testing of medical equipment should be part of a good quality assurance and safety program for any manufacturer.

2.3. Complexity of Medical Equipment

For *complex medical equipment*, the design considerations go beyond those for simple pieces of equipment. Considerations include physical and cognitive requirements, the human–computer interface, hardware and software, monitoring and input devices, as well as visual/auditory/written instructions, training and warnings. The contained knowledge-base and access to that data can also be part of the total assessment (such as information on diagnostic flow charts). The physical and cognitive requirements demanded by the equipment may involve relatively conventional information regarding human characteristics, as well as storage of detailed anatomical, physiological, or medical information. For example, traditional human factors information includes anthropometric dimensions such as reach and body size. An evaluation could involve keyboards used with medical equipment. Reduced size keyboards have been integrated into some designs, however the reduction cannot go below 16.7 mm center-to-center key space (standard is 19.05 mm) without performance degradation. Medical information might be provided by virtue of written guidance, available on an expert system. Usability testing could involve testing of the use of an expert system and whether the system guided the practitioner quickly

and easily to the correct information. The former involves retrieving data from known sources, while the latter requires evaluation of patient-oriented thought processes by healthcare personnel, in order to design an effective system. Another item of conventional knowledge is the rule of thumb for discrimination of auditory or visual information ($\pm$ 7). The warning signals for monitoring anesthesia during an operation exceed that capability. In a typical operating room, an alarm sounds every 4.5 minutes, with 75% of the alarms being false (they do not signal a change in patient status requiring intervention). Also, alarms may require a corrective response, alert the anesthetist to an abnormal, but intended state, signal an artifact that requires no action, or be a reminder for other behaviors. The point being that the complexity of the equipment and the situation yields unique situations to be evaluated.

2.4. Product Integration

A product will not be used if it is not *integrated* into the system into which it is introduced. Knowledge of the current system and methods of that systems use, how and why information is gathered by healthcare professionals, and how tasks are conducted are essential to encouraging integration of a product. For example, in conducting a pilot study of a medical imaging system for data management, objective cognitive-based performance was added to the usual subjective preference testing. After introduction of the new system, it was found improvement was masked by users' previous strategies and their consequently incorrect interactions with the new technology (Wu, Smith, and Swan 1996). A lack of education, training, practice, and delineating the differences in the two systems, and the strategies needed for each, almost prevented the adoption of a superior system. Users' attitudes and motivation can also prevent acceptance of a new product or system.

Using quality management techniques, such as continuous assessment protocols, can help to ensure the product achieves what it is designed to achieve, and that it is integrated into the daily operations of the healthcare system. This process is initiated at the start of the project and at each successive stage to re-evaluate whether the prototype conforms to the expectations of the developers, designers, and various user groups. Traditional approaches of modeling behavioral and cognitive aspects based on a users' requirements analysis and use of an observational task analysis have not been found to reveal the true picture, and have resulted in the failure of the integration of new information management tools. For example, the missing ingredient in one such system was information on how the recorded data was to be used. Once that was known and understood by the designers, it was possible to redesign the system for more than simply recording of information.

Physicians and other healthcare professionals often do several things at once, such as asking questions of their patients while they are conducting a physical or functional evaluation. Therefore, computer-based data recording and retrieval may not be the best solution in healthcare, as it interrupts the patient care process. Speech recognition has been investigated, as it gives users freedom of movement and the ability to carry on concurrent tasks, while simultaneously entering or retrieving data. Since healthcare workers tend to record a brief description of what may be a lengthy evaluation, use of pen technology also provides greater

mobility compared with being tied to the computer keyboard and terminal. It also makes use of their already learned responses and methods of providing patient treatment. Both speech recognition and pen technology have the propensity to improve the interaction between the physician and the patient by increasing eye contact, and improving the relationship between the patient and the care provider.

Each of these examples illustrates the point: integration of a new product within the cultural and contextual system in which it is to be used vastly improves chances of success.

3. RATIONALE FOR CONDUCTING USABILITY TESTING OF MEDICAL EQUIPMENT

Who bears the cost of poorly designed medical equipment? The patient who may receive less than optimal treatment, or be injured or killed; the healthcare provider who struggles with appropriate use of the equipment, thus wasting time and energy; the purchaser and prescriber of equipment who must conduct informal usability tests and/or closely study equipment-based publications to determine the best equipment for their use; the manufacturer whose product does not sell, and neither the sales personnel or management understand why; the manufacturers, insurance companies, and patients who must pay extra costs to cover lawsuits, which have resulted from wrongful injury or death lawsuits. All of us pay, in one way or another. Some pay with their lives or quality of life, and some pay financially; but we all pay.

Usability tests should include cost estimates for informed decision-making regarding production and/or purchase. It has been suggested that adding human factors results in an increase of total design costs from 0% to 8%, with 8% being high unless it includes costs associated with integrated logistics support (Booher 1990). This is perhaps a small price to pay for the safety and productivity improvements which are generated through usability testing.

Criticisms of usability testing state that too many people must be included in assessments (with such a variety of users), they take too much time, and they are too expensive. However, research conducted by Lewis (1994) demonstrated that observing five participants allowed discovery of 80% of a product's usability problems (as long as the average likelihood of problem detection ranges between 0.32 and 0.42) , and that observing additional participants revealed fewer and fewer new usability problems.

Managers use systematic methods for product evaluation including demonstrations, clinical scenario simulations, and conceptual models with an interdisciplinary team and tape-recording the joint sessions for later analysis. In essence, managers have devised informal usability testing strategies, with an emphasis on application integration. It is not known how much time this takes of vendors, or if potential purchases turn down equipment should vendors not comply with the informal testing, continue to provide unsolicited material, or insist on using their own marketing strategies. However, if independent usability research has been conducted, this information could be provided early in the process and should be an integral part of any marketing strategy.

Investment on the front end will yield both immediate and long-term productivity and cost avoidance advantages. Costs of retrofitting expensive equipment or re-developing equipment based on human factors deficiencies is extremely costly, but, it seems, industries rediscover this fact continuously.

Unfortunately, unless legal action is taken or a report is filed with the Food and Drug Administration, the manufacturer may bear no "visible" consequence of their poor design. However, they still pay financially with loss of consumer confidence.

All interested parties want safe, usable, and affordable products. What will influence manufacturers to institute testing to help deliver such a product? Top management, who design and structure the reward system for their employees can influence the organizational structure that will ensure total quality management of their products. The key to their hearts may be assuring them of a competitive edge in a global market, cost savings, their experiencing a loved one whose suffering is perceived as having been avoidable, or legally mandated compliance. Will such a mandate occur? At this point, the answer to that question is unknown. What is known is that usability testing can reduce waste, unnecessary operations, and maintenance costs, while increasing productivity and avoiding injury.

REFERENCES

BOGNER, M.S., 1998, An Introduction to design, evaluation, and usability testing. In V.J. Rice (ed.) *Ergonomics in Health Care and Rehabilitation* (Boston: Butterworth-Heinemann), pp. 231–47.

BOOHER, R.E., 1990, MANPRINT implications for product design and manufacture, *International Journal of Industrial Ergonomics*.

LEWIS, J.R., 1994, *Human Factors,* 36(2), 368–78.

McCLELLAND, I., 1990, Product assessment and user trials. In J.R. Wilson and E.N. Corlett (eds), *Evaluation of Human Work: A Practical Ergonomics Methodology* (New York: Taylor & Francis).

McGUIRE, T., MOODY, J., and HANSON, M., 1996, An evaluation of mechanical aids used within the NHS. *Nursing Standard,* 11(6), 33–8.

MACKENZIE, C.F., JEFFERIES, N.J., HUNTER, A., *et al.*, 1996, Identifying systems problems in trauma patient anesthesia: An analysis of performance deficiencies in airway management as captured by video and self-reports. *Human Factors,* 38, 636–45.

MERRITT, S.L., GATES, M.A., and SKIBA, K., 1993, Readability levels of selected hypercholesterolemia patient education literature. *Heart and Lung,* 22(5), 415–20.

NELSON, A., MALASSIGNE, P., AMERSON, T., SALTZSTEIN, R., and BINARD, J., 1993, Descriptive study of bowel care practices and equipment in spinal cord injury. *Sci Nursing,* 10(2), 65–7.

RICE, V.J., 1995, Human factors in medical rehabilitation equipment: Product development and usability testing. In K. Jacobs and C.M. Bettencourt (eds) *Ergonomics for Therapists* (Boston: Butterworth-Heinemann) pp. 77–93,.

STANTON, N. and BABER, C., 1996, Factors affecting the selection of methods and techniques prior to conducting a usability evaluation. In P.W. Jordan, B. Thomas, B. Weerdmeester, and I. McClelland (eds) *Usability Evaluation in Industry* (New York: Taylor & Francis) pp. 39–48 .

WU, S.C., SMITH, J.W., and SWAN, J.E., 1996, Pilot study on the effects of a computer-based medical image system. *Proceedings/AMIA Annual Fall Symposium*, 674–8.

The MUSE Method for Usability Engineering

K. Y. Lim

Nanyang, Technological University, School of Mechanical and Production
Engineering, Singapore 639798

1. INTRODUCTION

MUSE (M̲ethod for US̲ability E̲ngineering), a human factors
structured analysis and design method, was developed between
1987 and 1992 to advance the methodological foundation of
human factors design. The method has evolved through a number
of versions since its early conception in 1988 (see Walsh, Lim,
Long and Carver 1988, Lim, Long and Silcock 1992). MUSE is
the first and arguably the most complete structured human factors
method to be developed.

MUSE is the result of a project commissioned specifically to
solve the "too-little-too-late" problem of human factors input to
interactive systems development. Human factors inputs may be
"too little" since frequently only advice is offered. Inputs may be
"too late" because evaluation is often the main (or worse the only)!
human factors contribution to interactive systems development.

As a structured analysis and design method, MUSE addresses
the problem by specifying explicitly the human factors design
inputs and processes required at each stage of the design cycle.
By making the scope, format and timing of its design contributions
explicit, MUSE enables a more timely and effective intersection
of human factors and other system design inputs. In this way,
MUSE facilitates cross-disciplinary integration of methods, such
as with structured software engineering methods like the Jackson
System Development (JSD) method, Structured Systems Analysis
& Design Method (SSADM), etc. Thus, an efficient system
development process is ensured, which in turn enables a more
effective uptake of human factors design contributions.

MUSE provides design support throughout the system
development life-cycle, including explicit stages and procedures
for user requirements analysis, task analysis and u!ser interface
design. Thus, the method advances human factors design
specification (as opposed to the traditional focus of human factors
on design evaluation only), by taking system development from
the description, analysis and synthesis of user tasks to a user
interface design specification. It should be noted that MUSE
advocates iterative design, and facilitates prototyping by specifying
explicitly the design descriptions to be derived at each stage of
system development. The MUSE method also exploits the
potential of a common notation for cross disciplinary design
communication, by recruiting the well developed structured
diagram notations offered by the Jackson System Development
(JSD) method, a software engineering structured analysis and
design method. The notations exploited by MUSE include basic
constructs such as sequence, selection, iteration and hierarchy;
as well as more powerful constructs to describe concurrent,
interleaved and uncertain events. These constructs enable a
human factors designer to provide a more specific description of

user tasks and a concise way of specifying dynamic aspects of a
user interface design. Examples of the latter human factors
specification may include relating and timing software functional
supports and screen actuations to user tasks, and the definition
of contexts for triggering particular error and dialogue message
displays.

It can not be over-emphasized that the MUSE method need
not be applied in its entirety. Although the method includes very
detailed products and procedures, a human factors designer may
choose to apply MUSE at its top level only; that is to use its
stage-wise design framework to guide human factors input to
system development and to identity cross disciplinary contact
points to facilitate design collaboration. In this respect, MUSE
can be configured for integration with any explicit system
development method, since its conception of the human factors
design process is generic. In another scenario of method
application, a human factors designer may choose to apply only
the early stages of MUSE to support requirements and task
analysis, followed later by task synthesis. In this case, the designer
may also wish to exploit its structured diagram notations to enable
more explicit and specific user task description (see Lim 1996).
Naturally, the MUSE method may be applied in its entirety, for
instance in large systems development projects or for long life
systems, both of which would require comprehensive design
process and documentation to support systematic specification
and later system maintenance. Case-studies of all of these
scenarios of application of MUSE exist.

A brief account of the MUSE method follows. For a detailed
description, the reader is referred to Lim and Long (1994).

2. MUSE: A STRUCTURED HUMAN FACTORS M̲ETHOD FOR US̲ABILITY E̲NGINEERING

MUSE involves a number of design stages grouped into three
phases as follows:

i   Information Elicitation and Analysis Phase. Background
    information to support design is derived and processed at
    this phase including human factors analyses (see Figure 1).
ii Design Synthesis Phase. This phase is concerned with the
    initial definition of the system and its sub-systems. One or
    more conceptual designs of the target/new system is then
    synthesised.
iii Design Specification Phase. The three stages in this phase
    address the detailed design of a new system, including human
    factors specifications of a user interface design.

A brief de!scription of the design stages of MUSE follows.

2.1. Extant Systems Analysis Stage

At this stage, background information to support system
development is derived from users and other sources; e.g. critical
user needs and problems, salient characteristics of current tasks,
features and rationale of the existing user interface design, etc.
Related extant systems may also be analysed as appropriate. The
objective of the analysis is two-fold. First, information is gathered
to set new design requirements against the context of the current
system. Second, extant system designs are assessed with respect to
new system requirements. On the basis of the information, the
consequential implications of porting (or non-porting) of particular
design features from the extant system to the new system, may be
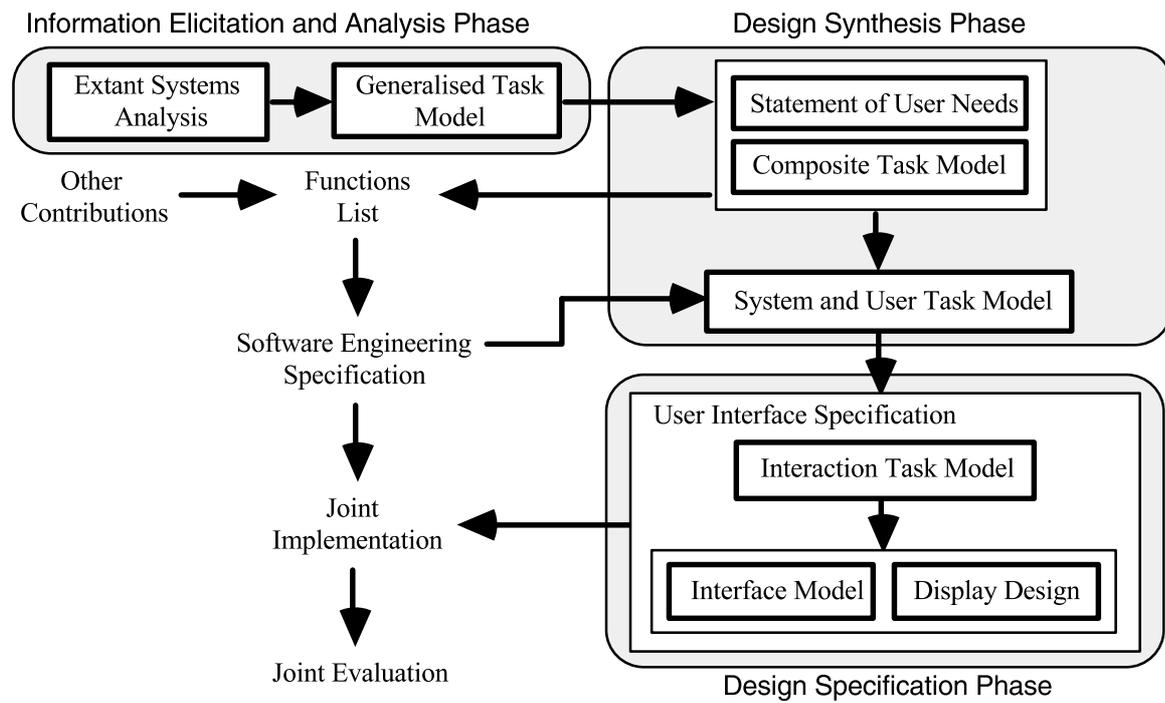
Figure 1. A simple representation of! MUSE: a structured human factors method

considered. For instance, assessments of possible transfer of learning effects (both positive and negative effects) would be supported.

2.2. Generalised Task Model Stage

This design stage is concerned with processing the information gathered at the preceding stage. The objective is to generate (predominantly) device independent descriptions to support analytic mapping between specific extant design features and new system requirements. To this end, generalised task models are derived to characterise the conceptual requirements of the new system and of each extant system analysed. By comparing the models, appropriate extant design features may be identified and selected for porting to the new system. Similarly, new design features that conflict with the current system may be avoided.

2.3. Statement of User Needs Stage

Conclusions drawn from extant system analysis are collated and summarised with respect to the user. The statements derived constitute an explicit set of human factors requirements for the new system. The information documented should include problems experience!d by users; design requirements, constraints, and rationale of existing and new design features; performance criteria; domain semantics; etc. A user-centered design basis is thus established.

2.4. Composite Task Model Stage

A conceptual design of the new system is generated at this stage to support function allocation between the human and computer. To this end, a composite task model is synthesised from compatible sub-sets of the generalised task models derived earlier. On-line and off-line tasks are then designated in the composite task model.

2.5. System and User Task Model Stage

This stage addresses the decomposition of on-line and off-line tasks (see sub-section 2.4 above). Specifically, on-line task components of the composite task model are decomposed to generate a system task model to define the cycles of human and computer actions entailed by the interactive task. Similarly, off-line task components are decomposed to gen!erate a user task model to describe tasks not supported by the computer. During decomposition, design features of extant systems that are consistent with the statement of user needs, may be incorporated.

2.6. Interaction Task Model Stage

At this stage, an interaction task model is derived by decomposing the human action components of the system task model. A device level description of user inputs required by the interactive task is derived. Points for triggering and removing screen displays may then be defined by locating the start- and end-points of coherent groups of user inputs (corresponding to interactive task actions). The resulting specification provides a user-oriented perspective of the error-free execution of the interactive task (errors are addressed later in sub-section 2.8). Note that the development of training programmes and user manuals would be supported by such a task model.

Generally, two rules of thumb guide the derivation !of an appropriate interaction task model. First, the model should be decomposed to a level understood easily by design team members and end-users. Second, the terms used in its description should be consistent with the user interface primitives (e.g. screen objects and actions) of the chosen implementation environment and hardware (e.g. basic keystrokes). On the basis of an appropriate interaction task model, design activities of the remaining stages of the method may then be undertaken.

2.7. Interface Model Stage

At this stage, a set of human factors descriptions (termed interface models) of the behaviour of user interface objects, is derived by decomposing computer action components of the system task model. The models specify how appearances and behaviours of user interface objects would change in response to user inputs, and to changes in attribute states of corresponding real world domain entities. The specification derived provides a computer-orie!nted perspective of the interactions required to execute the on-line task.

2.8. Display Design Stage

This design stage addresses three aspects of human factors design. First, potential user errors are identified by examining each input prescribed by the interaction task model. Design modifications may then be considered to remove the causes of error. Problems that can not be solved may be accommodated by a more tolerant or forgiving design, and as the last resort, by a greater emphasis on user training and selection (if applicable). To address these concerns, design iterations may be required with the two preceding stages described above.

Second, static characteristics of computer displays are specified; for instance the composition and layout of screen displays and computer support functions, and the design of error, feedback and help messages. Finally, the dynamic characteristics of screen displays are then specified; for insta!nce the context for actuating particular displays to support human-computer dialogue and the interactive task. The set of human factors design descriptions thus derived specify how, when and what computer functions and messages are presented to support each stage of the user's task.

It should be emphasised that the MUSE method provides detailed design procedures and notations for each of the design stages described above. Space constraints, however, preclude an account of these procedures here. For a more detailed account of MUSE, the reader should refer to Lim and Long (1994).

Note also that MUSE has been integrated with JSD (a structured software engineering method) in 1991, to support more effective cross-disciplinary design collaboration. The integrated method derived is named MUSE•JSD. For a more detailed account of the latter, the reader should refer to a related entry in this encyclopedia and to Lim and Long (1994). Since the completion of MUSE development in 1992 and its official introduction in 1994, follow-up work on the method has been undertaken at several locations and across different countries. In particular, MUSE has been applied in several system development cases by researchers as well as designers from large commercial companies that participated in a European Community project. Derivative versions of MUSE have also emerged to tailor the method more specifically for particular application domains (e.g. safety critical process control), and for cross-disciplinary integration with particular object oriented design methods and structured software engineering methods. The results of some of these projects have been published and reported at conferences. Other projects are ongoing including the development of basic tools to support method application.

REFERENCES

LIM, K. Y., 1996, Structured Task Analysis: An Instantiation in the MUSE Meth!od for Usability Engineering. *Interacting with Computers,* 8(1), (Holland: Elsevier Science), pp. 31-50.

LIM, K.Y. and LONG, J.B., 1994, *The MUSE Method for Usability Engineering.* Cambridge University Press (UK), 350pp.

LIM, K. Y., LONG, J. B. and SILCOCK, N., 1992, Integrating Human Factors with the Jackson System Development Method: An Illustrated Overview, In: Barber, P. and Laws, J. (eds.), *Ergonomics Special Issue on Methodological Issues in Cognitive Ergonomics III,* 33(12), (London: Taylor and Francis).

WALSH, P., LIM, K. Y., LONG, J. B. and CARVER, M. K., 1988, Integrating Human Factors with System Development. In: Heaton, N. and Sinclair, M. (eds.) *Designing End-User Interfaces. Pergamon Infotech State of the Art Reports 15:8,* (England: Pergamon Infotech Limited), pp. 111-120.

MUSE–JSD: Structured Integration of Human Factors and Software Engineering Methods

K.Y. Lim

Nanyang Technological University, School of Mechanical & Production Engineering, Nanyang Avenue, Singapore 639798

1. MOTIVATION AND RATIONALE FOR DEVELOPING MUSE–JSD

The typical involvement of human factors only at the evaluation stage of system development often results in poor realization of its design contributions. Contributions may be "too-late" because human factors activities are not presently basic to design specification. The lateness of human factors involvement makes it more difficult to formulate contributions, as information on the decisions made during conceptual design are frequently poorly documented. As a result, human factors recommendations may be incorporated ineffectively during system development. Worse, no action may follow human factors recommendations because the desired modifications might be too far reaching, and hence too difficult and expensive, to implement at later stages of system development. Further, the possibilities for modification may become restricted due to interlocking design dependencies. Developed designs may thus become "frozen" and more resistant to change. Consequently, late human factors involvement may result in contributions which are little more than advice, i.e. "too-little". The overall outcome is often an inefficient system design cycle skewed heavily towards design maintenance (which includes design modification).

This "too-little-too-late" problem of human factors input to system development prompted a growing awareness that early and continued human factors involvement in the system development process is vital for effective uptake of its contributions. Early involvement is more effective since appropriate user requirements may be established to constrain design at later stages of system development. An early definition of user requirements is particularly important because human-related design issues predominate at early stages of system development. Continuous human factors involvement then supports the appropriate interpretation and translation of user requirements into system design specifications. Continuous involvement also ensures that human factors inputs are timely and contextually relevant to system development concerns that evolve through the design cycle. In this way, human factors contributions to system development may be optimized.

A basic requirement for ensuring comprehensive human factors input is to define and locate explicitly its design concerns against existing conceptions of the system development process. To this end, system development methods that provide life-cycle support and well-developed methodological characteristics would

provide an explicit design framework against which human factors inputs could be located and timed appropriately. These requirements are met by a class of methods named structured analysis and design methods. These methods are characterized by systematic design stages defined explicitly in terms of their scope, process, product and notation. Further, the framework provided by structured analysis and design methods could also guide the development of a similarly explicit, complete and structured human factors method. Indeed, a structured human factors method named MUSE was developed in this way (Lim and Long 1994).

By developing a structured human factors method, cross-disciplinary design concerns may be coordinated more effectively by integrating human factors and software engineering methods. The benefits that would accrue from a structured integration of human factors and software engineering methods are:

- A structured integration of cross-disciplinary design processes would facilitate explicit representation of human factors design in the overall agenda for system development. Project resources may then be budgeted appropriately to accommodate human factors design. Design schedules and deliverables may also be specified better. By making explicit at project outset the resource requirements and allocation for human factors design, the frequently reported encroachment of its resources may be avoided. In addition to improved project planning and resource management, incorporating an explicit human factors design process would facilitate the recruitment of existing standards and guidelines. Similarly, an explicit design process would help define specific contexts for recruiting particular declarative human factors knowledge for application at each stage of system development.

- A structured integration of cross-disciplinary design processes could potentially support the use of a common notation for describing human factors and software engineering design specifications. Thus, human factors design inputs may be communicated more effectively to software engineers. The use of a common notation is indeed realizable, since existing structured analysis and design methods often include well-developed notations that are also appropriate for human factors descriptions.

- As the methodological characteristics of structured methods are well-defined, structured integration of design processes would enable cross-disciplinary reference/contact points to be defined. In this way, collaboration between human factors designers and software engineers would be facilitated, leading to improvements in the uptake of human factors contributions.

These benefits motivated the selection of structured analysis and design methods as substrates for integrating human factors and software engineering design processes and methods. Presently, a number of conceptions of structured integration of cross-disciplinary design methods have been developed. A leading example of such an integration is MUSE–JSD, an integration of MUSE with the Jackson System Development method (a structured human factors and software engineering method respectively). Preliminary conceptions of the MUSE–JSD method were published by Lim *et al.* (1990, 1992). MUSE–JSD is the first and arguably the most explicit and complete conception of

a structured integration of software engineering and human factors methods.

Generally, MUSE–JSD comprises parallel streams of human factors and software engineering design contributions to interactive systems development. The design streams are elaborated at a lower level of description to characterize their design stages in terms of their scope (and associated products); process (and associated procedures); and notation (and associated description schemes). In addition, design interdependencies between the two streams are specified to check any design drift that may arise. Specifically, identifying locations at which human factors and software engineering design intersect specifies obligatory contact points. At these points, human factors designers and software engineers are required to agree on the design scope, assumptions and information to be used in the development of the system. In other words, the design basis is defined and adhered to at subsequent stages of system development. A more efficient convergence of cross-disciplinary design outputs is thus ensured by the MUSE–JSD method.

A brief account of the integrated method follows. Before reading this account, it is best that the reader refers first to the MUSE chapter in this volume. For a detailed account of procedures of the MUSE–JSD method, refer to Lim and Long (1994).

2. OVERVIEW OF THE MUSE–JSD METHOD

Figure 1 shows a schematic representation of the integrated method. Some characteristics of MUSE–JSD are immediately discernible. For instance, it can be seen that the integrated method comprises parallel streams for software engineering and human factors design (supported by JSD and MUSE respectively). These

streams are structured into design stages, which are interlinked at several locations.

Figure 1 should not be interpreted as indicating MUSE to be more complex than JSD. The present manner of representation is only intended to show MUSE in greater detail, since the JSD method is already well reported and is maintained largely unchanged in the integrated method. Other characteristics of MUSE–JSD are:

1. The structured human factors method, namely MUSE, comprises three design phases that address information elicitation and analysis (such as user requirements and task analysis); followed by design basis definition and conceptual design synthesis; and finally user interface design specification. For a detailed account of the MUSE method, see the related chapter in this volume and to Lim and Long (1994).

2. The design scope (and associated products), process (and associated procedures) and notation (and associated description schemes) for all stages of the integrated method (i.e. MUSE–JSD) are defined explicitly. Thus, design collaboration between software engineers and human factors designers is facilitated. In particular, the well-defined characteristics of the method facilitate the specification of cross-disciplinary requirements for collaboration in respect of the scope, format, granularity, timing and communication of intermediate design inputs.

3. The methodological structure of MUSE–JSD supports the application of accepted design principles such as:
   - avoiding premature design commitment by emphasizing conceptual design before detailed design;
   - conducting of early evaluation, either analytically by the
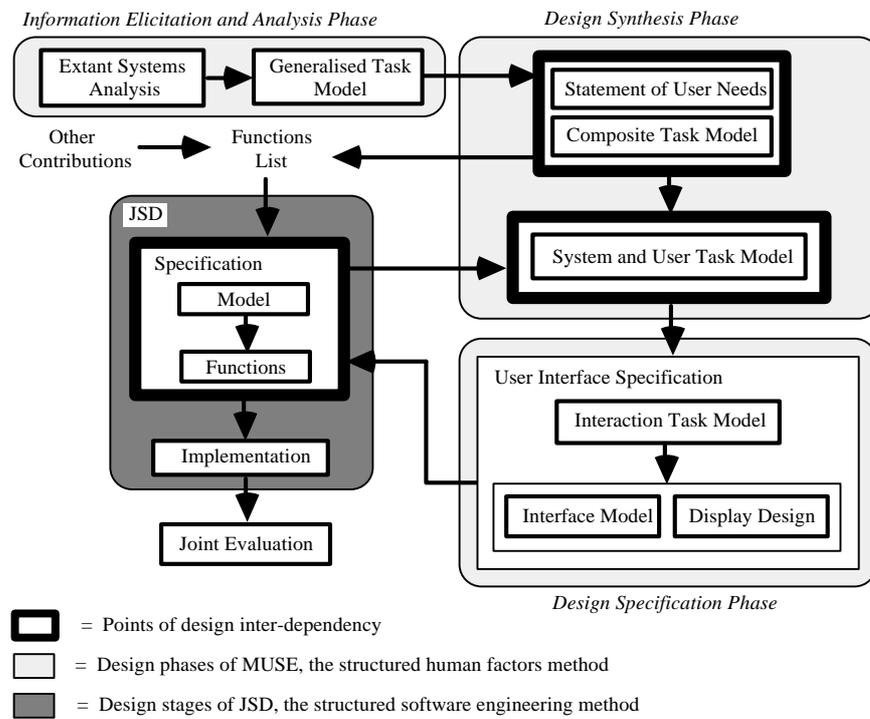
Figure 1. Representation of the MUSE–JSD method.

designer(s) or empirically through prototyping. Both types of evaluations are facilitated by design descriptions to be derived at each stage of system development as defined explicitly by the method; and

- iterative design which is facilitated by the well documented design rationale and descriptions generated by the method;

4. To exploit positive transfer of learning by designers (hence facilitate uptake of the method), appropriate existing human factors methods and techniques have been incorporated into MUSE. For the same reason, the software engineering method chosen for integration (namely JSD) has been maintained largely unchanged in MUSE–JSD.

5. Design interdependencies have been identified explicitly in MUSE–JSD by comparing the information requirements and design concerns addressed at each stage of software engineering and human factors design (as defined by the JSD and MUSE methods respectively). To account for these design interdependencies, *obligatory* contact points have been specified to ensure convergence in the scope of system development advanced in parallel in the individual design streams. In this regard it should not be construed that informal contacts between the two groups of designers are precluded. Such contact points are not specified explicitly in the method since they tend to be situation-specific. For a generally applicable method it would be more appropriate to leave informal contact points as *discretionary*.

The two *obligatory* contact points specified in MUSE–JSD occur between the Design Synthesis Phase of MUSE and the Modeling and Functions Stages of the JSD method. The nature of the contact points is described below.

2.1. Contact Point 1 between the Modeling Stage (JSD Method), and the Composite Task Model and Statement of User Needs Stages (MUSE Method)

At this contact point the human factors descriptions derived are discussed with software engineers. The objective is to augment the design information to be used later to construct a JSD Model to define the subject-matter of the system. Generally, the design information discussed comprises object and action attributes, user needs and problems, event lists, user task semantics, user task support requirements, etc. Following the discussions, a functions list is drawn up to characterize the scope of system development. The functions list is then carried forward to constrain design activities undertaken in succeeding stages of both streams of the integrated method.

2.2. Contact Point 2 between the Functions Stage (JSD Method), and the Composite Task Model and Statement of User Needs Stages (MUSE Method)

At this contact point JSD analysts would address the specification of information flows among JSD Function processes, JSD Model processes, and the external world. These software engineering concerns intersect human factors concerns addressed at the System Task Model Stage of MUSE. Specifically, intersections occur between the functional sequences described in the system task model (design product of MUSE), and JSD specifications of information functions and the input subsystem. Since both streams of the integrated method are presently concerned with conceptual design extension and functional definition, close contact is particularly important to ensure a convergent scope of system development. It should be emphasized that inadequate contact at this stage may incur additional design iterations, when human factors and software engineering specifications are integrated. The result is thus less efficient design.

For a more detailed account of MUSE–JSD and its cross-disciplinary design interdependencies, see Lim and Long (1994).

REFERENCES

LIM, K.Y. and LONG, J.B., 1994, *The MUSE Method for Usability Engineering* (Cambridge: Cambridge University Press).

LIM, K.Y., LONG, J.B. and SILCOCK, N., 1990, Integrating human factors with structured analysis and design methods: an enhanced conception of the extended Jackson system development method. In D. Diaper *et al*. (eds), *Proceedings of the 3rd IFIP Conference on Human–Computer Interaction (INTERACT '90)* (Amsterdam: Elsevier Science), 225–30.

LIM, K.Y., LONG, J.B. and SILCOCK, N., 1992, Integrating human factors with the Jackson system development method: an illustrated overview. *Ergonomics* [special issue on Methodological Issues in Cognitive Ergonomics III], 33.

Noise: Metrics and Equipment

S.E. Lee, G.S. Robinson and J.G. Casali

Virginia Tech Center for Transportation Research, Blacksburg, VA 24061, USA

1. INTRODUCTION

Over time, exposure to noise above a certain level (generally accepted to be in the range of 75–85 dBA) will cause hearing damage in humans. For this reason, most developed countries have established limits on noise exposure for industrial workers. In the USA these limits are found in the Occupational Safety and Health Administration's (OSHA) Occupational Noise Exposure Standard and Hearing Conservation Amendment (Code of Federal Regulations 1983), as well as in other regulations for mining, construction and the military. The application of these standards requires a basic knowledge of the physics of sound, the common metrics for noise measurement and the equipment to be used. Other considerations include: measurement of high-level, impulse noises (such as explosions) which can cause instantaneous hearing damage; the frequency characteristics of the noise in question; and the correct settings for the equipment used.

2. SOUND PRESSURE LEVEL

Sound is defined as oscillations in pressure in a medium (i.e. gas, liquid or solid) which has both mass and elasticity. Sound is caused by a disturbance from an equilibrium condition (e.g. in air, the pulsating airstream from rotating fan blades). In air, the equilibrium condition is the ambient air pressure. Sound is also sometimes defined as the human auditory sensation evoked by these oscillations in pressure (from the human perspective rather than the physical perspective). *Noise* is usually defined as unwanted sound and, more technically, as sound of a random nature.

The most common unit of measurement for sound and noise is sound pressure level (SPL), measured in micropascals (Ostergaard 1986). Because common sounds measured in micropascals (mPa) have a range of ~1 000 000, a compressed log scale called the decibel (dB) scale is almost always used. Correspondingly, common sounds measured in dB have a range of ~120, which is much more convenient for calculations. The formula for SPL in dB is:

$$SPL = 20 \log_{10}(P/P_0), \qquad (1)$$

where P is the maximum measured acoustic pressure and i_0 is the reference pressure, usually 20 mPa.

3. BROADBAND VERSUS NARROW BAND VERSUS TONAL NOISE

In describing a noise, its frequency characteristics are also important. The oscillations in ambient pressure produced by a noise source travel outward from the source in waves, and the number of waves emanating from the source per time unit defines the frequency of the noise. The usual unit for frequency is Hz or cycles per s. For example, for a pure tone (i.e. single frequency) sound of 200 Hz, the sound source is vibrating 200 times per s. Sounds do not have to be pure tones, and in fact, most noises are complex; that is, they are composed of many frequencies occurring simultaneously. These are also known as broadband noises. Still other noises may be narrow band, or composed of a few tones near a certain frequency, so that the overall effect is very close to tonal. An example of a common pure-tone sound is the sound made by striking a tuning fork. The noise of a vacuum cleaner is a broadband sound, while a car horn is narrow band in nature. The young, healthy human ear is capable of detecting sounds with frequencies ranging from ~20 to 20 000 Hz, but is most sensitive from ~1000–5000 Hz. Frequency correlates roughly with pitch, so low-frequency noises are thought of as deep or bass (such as a foghorn), while high-frequency noises are thought of as high-pitched or treble (a whistling tea kettle, for example).

In measuring sound pressure levels, it is often possible to
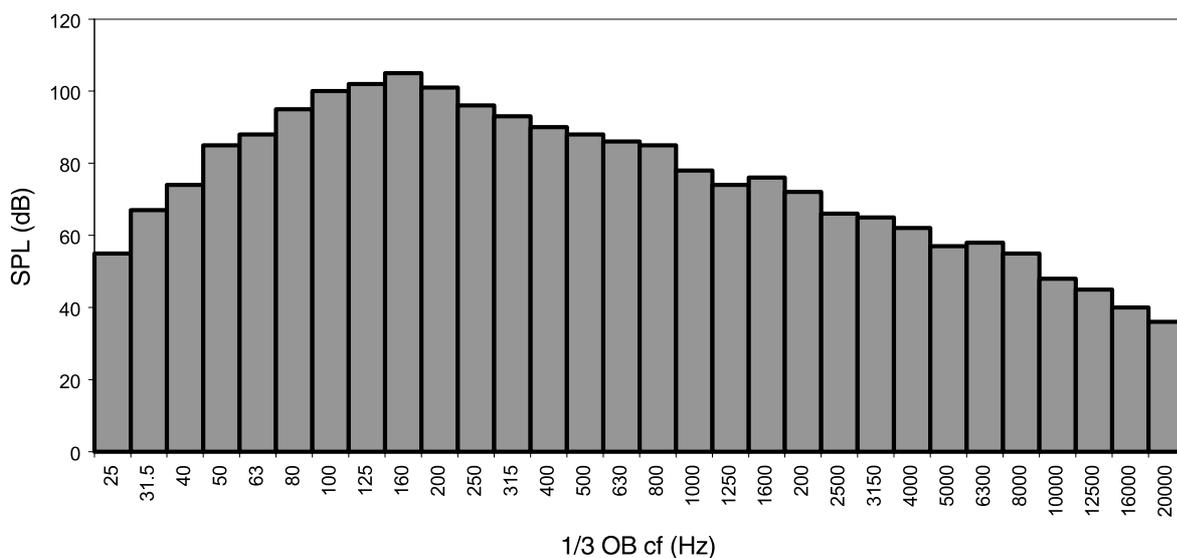
Figure 1. Example of low frequency industrial noise measured in one-third octave bands.

use electronic filtering to measure the different frequency bands of the broadband noise in question. For instruments that will provide this type of measurement (see Section 7 for details on instrumentation), the most common filter sets are either octave band and one-third octave band, where an octave represents a doubling of frequency. When this type of measurement is employed, the result is often portrayed graphically, with amplitude in dB on the vertical axis and the octave or one-third octave band center frequencies along the horizontal axis (Figure 1). By noting at which frequencies the amplitude is greatest, one can characterize the broadband noise as high or low frequency biased (and in some cases, the noise will be nearly flat across frequencies). In figure 1, the noise is biased toward the lower frequencies, a common finding in measuring industrial noise.

4. A- AND C-WEIGHTINGS

The human ear is not equally sensitive to all frequencies, and in particular, it is more sensitive in the range of ~1000–5000 Hz (Ward 1986) due to several anatomical mechanisms which amplify these frequencies. As a result, the ear is not only made more sensitive, but it is also more susceptible to hearing loss resulting from noise exposure in this frequency range. In order to account for the differential sensitivity and damage risk at these frequencies, the A-weighting scale was developed. A-weighted noise measurements (usually expressed as dBA) take the basic SPL measurement and weight the different frequencies in a manner similar to how the human ear processes the sounds; in other words, sounds from 1000 to 5000 Hz are given extra weight, but sounds outside this range are de-emphasized (Earshen 1986). The resulting metric is thought to be a more accurate predictor of the risk of hearing damage, and also closer to how the human auditory system processes low SPL. The dBA metric is so commonly used that some people mistakenly think that any dB measurement is implicitly A-weighted. The correct practice, however, is to make the weighting scheme explicit, and for the remainder of this paper the two are distinguished, where dB is the same as dB(linear), or unweighted, and dBA indicates an A-weighted measurement.

Another (slightly less common) weighting scheme is the C-weighted, or dBC, scale. This weighting scheme is also loosely related to the response of the human auditory system, but at fairly high SPL. At these high levels, the auditory system is still more sensitive to sounds in the 1000–5000 Hz range, but the curve is not as pronounced. In fact, the C-weighting curve, when drawn, is nearly flat, and thus very similar to dB(linear). One useful application of dBC measurements is that when compared with dBA measurements of the same source, a higher dBC reading indicates that the sound source has significant low frequency components, thus providing a very rough form of frequency analysis.

4. TIME CONSTANTS

The time constant is an important consideration in measuring sound pressure levels. There are three commonly used time constants: fast, slow and impulse. Basically, the sound level meter dynamics are such that when an input (sound) is introduced, the system's response rises exponentially towards the maximum, and reaches 63% of the final, steady-state reading within one time constant. For "fast," it reaches this point in 0.125 s, for

"slow," in 1.0 s and for "impulse," in 0.035 s. When a sound level meter is set to "slow," for example, the needle or digital display is providing a snapshot of the noise that has entered the instrument in the last second.

5. COMMON METRICS

In the field of acoustics and hearing conservation, several types of measurements are commonly available from the various types of sound measurement equipment. In this section the most common types of measurements will be discussed, along with the appropriate usage for each type. Some of these measurements are instantaneous (or over very short periods) while others average the measurement over an extended period.

5.1. Instantaneous Metrics

5.1.1. L_{max}

L_{max} is the maximum sound pressure level (dB) measured over a stated period, using an exponential time constant (fast, slow or impulse). When the exponential time constant is not provided, it is assumed that the fast response has been used. The L_{max} measurement can only be obtained with sound-measuring equipment with a "maximum-hold" feature, which serves to capture and store the L_{max} data. This metric is useful in exploring the maximum levels encountered in events which occur quickly, such as the slamming of a door or the noise emitted by a passing truck (Yeager and Marsh 1991).

5.1.2. Peak

By contrast, the peak sound pressure level is given by:

$$PEAK_{SPL} = 20 \log_{10}(P_{max\text{-}inst}/P_0), \qquad (2)$$

where $P_{max\text{-}inst}$ is the maximum *instantaneous* sound pressure that occurs during a given time period and P_0 is the reference pressure (20 mPa). Peak measurement thus truly captures the highest instantaneous SPL occurring during a specified time period. No time constant is used, unlike L_{max}, for which the measurement is usually taken over 0.125 s. For this reason, peak measurements can exceed L_{max} measurements of the same noise by as much as 20 dB. As with the L_{max} metric, a "peak-hold" feature can be found on many types of sound-measuring equipment (Yeager and Marsh 1991). Peak measurements can be useful for describing high-level, instantaneous sounds, such as explosions or gunshots, which cause instantaneous hearing loss.

5.1.3. L_{min}

Both PEAK and L_{max} are useful in describing noise in industrial and leisure environments, but it is also occasionally useful to know the minimum noise levels. The L_{min} metric is measured similarly to the L_{max} metric, including the need for a special "minimum-hold" feature (Johnson *et al.* 1991). This setting is most often used in laboratories, when it is sometimes desirable to conduct experiments in a very low noise environment.

5.2. Average Metrics

5.2.1. L_{eq}

In most cases, some sort of an averaging metric is needed; the metrics described above do not indicate any sort of average noise level, but instead only provide very brief snapshots of the extremes of the noise in question. Because of the logarithmic nature of

decibels, ordinary arithmetic operations (such as addition and division) cannot be used. Perhaps the most common averaging metric is L_{eq}, or equivalent continuous sound level, which can be thought of as the equivalent steady-state (unvarying) sound pressure level (usually A-weighted) which has the same energy as the time-varying, averaged sound. In other words, the L_{eq} provides an average sound pressure level for a noise which varies over time, and which possesses the same acoustical energy as another unvarying noise. The formula for L_{eq} follows:

$$L_{eq} = 10 \log_{10}\{1/T[S(10^{Li/10}t_i)]\}, \qquad (3)$$

where i is the ith interval, T is the total time, t is the duration of the ith interval, and L_i is the SPL during t_i (L_i is usually A-weighted). The L_{eq} metric is used frequently, and may be encountered in determining the average noise level within a work area, community noise issues, and some government regulations concerning noise and hearing protectors (Earshen 1986).

5.2.2. TWA

Although some government regulations use the L_{eq} metric, OSHA has developed its own metrics for measuring noise exposure in the workplace. One of the more commonly encountered OSHA metrics is the time-weighted average, or TWA. Since OSHA is concerned about the noise exposure of workers in a standard, 8-h day, the TWA is always referenced to 8 h (in other words, 8 h is the time weighting applied to the average noise levels in the workplace). The formula for TWA is:

$$TWA = 16.61 \log_{10}\{1/8[S(10^{LAi/16.61}t_i)]\}, \qquad (4)$$

where i is the ith interval, t is the duration of the ith interval in h and L_{Ai} is the A-weighted SPL during t_i. People working in industrial hearing conservation are most likely to encounter the TWA metric. Dosimeters, or personal noise exposure monitoring devices, provide the TWA as one of the main data outputs. Note that regardless of the actual duration of exposure, it is always weighted against an 8-h standard, so take care *not* to replace the eight with the actual exposure duration (Earshen 1986).

5.2.3. Dose

OSHA, as well as regulatory agencies from other industrialized countries, have established standardized limits on the amount of noise exposure a worker can receive in any given shift (usually taken to be 8 h). This can either be expressed as a TWA, discussed above, or noise dose. Basically, once a regulatory agency sets a limit for noise exposure (in the USA, current OSHA regulations limit noise exposure to an 8-h TWA = 90 dBA), this limit is considered to be a 100% dose, and a mathematical formula can then be used to derive doses above and below the 100% benchmark. An important concept in dose calculations is the exchange rate, which is the number of decibels which is considered to result in a doubling or halving of noise dose. For example, OSHA uses a 5 dB exchange rate (most other countries use 3 dB); this means that while a worker can be exposed to 8 h of 90 dBA noise, if the noise level goes up by 5 dB to 95 dBA, he or she could only be exposed for 4 h. Conversely, the same worker could be exposed to 85 dB noise for 16 h. The formula for dose is fairly simple, as it is simply the sum of the ratios of the actual noise exposures to the allowed noise exposures, where the allowed noise exposures may be obtained from a table or calculated; the

resultant sum is then multiplied by 100 to put it into percentage form. See Casali and Robinson (1999) for further detail on noise dose calculations. If the noise dose is known, then the TWA can be calculated using a simpler formula:

$$TWA = [16.61 \log_{10}(D/100)] + 90, \qquad (5)$$

where D is in percent.

5.2.4. SEL

Sound exposure level, or SEL, is a metric that has been advocated for use in the regulation of hearing protection (Casali and Robinson 1999). It basically takes the SPL formula (equation 2) and multiplies P by the number of seconds exposed and P_0 by a reference duration of 1 s. If L_{eq} is known, SEL can be calculated directly by:

$$SEL = L_{eq} + 10 \log_{10}(T/T_0), \qquad (6)$$

where T is the exposure duration in seconds and T_0 is the reference duration of 1 s. SEL thus provides an indication of the level of a 1-s duration sound that is equivalent to the level obtained by integrating a time-varying sound over the exposure duration (Casali and Robinson 1999).

6. EQUIPMENT

6.1. Sound Level Meters

Sound level meters, or SLM, are perhaps the most common piece of sound-measuring equipment. In their most basic form, they provide a simple sound pressure level reading, without regard to L_{eq}, noise dose, or TWA, although the measurements obtained can then be used in such calculations. Most SLM on the market today have several measurement options that can be employed. The fast/slow/impulse time constant is a common option. For steady-state noise, either the fast or slow setting will provide the same reading. For rapidly fluctuating noise, the slow response will provide a long time average reading, and also has the benefit of making the display more readable (on the fast setting, an analog display will produce a rapidly fluctuating needle in the presence of fluctuating noise, and a digital display will update rapidly, making it difficult to read; Earshen 1986). Most SLM also have the option of dBA/dBC readings; the choice is often dictated by the purpose of the measurement (for example, most OSHA measurements should be taken using dBA, slow response).

The typical SLM has an integrated microphone, with the result that the unit is relatively compact and extremely portable. Some SLM incorporate filter sets so that octave band or one-third octave band measurements can be taken; however, the filter sets found in most SLM operate one at a time, and must be accessed sequentially, which can be time-consuming when this type of measurement is required. There are ANSI and ISO standards for sound-measuring equipment according to its stated degree of accuracy and intended use; it is important that the equipment purchased meets the relevant standards.

6.2. Dosimeters

With the advent of the OSHA Noise Standard in 1971, the need arose for a small, wearable instrument to measure the individual noise dose received by a worker as he or she moved around in the workplace, and the personal noise dosimeter resulted. With microelectronics, it became possible to incorporate the circuitry

for such a device into a small package, not much larger than a deck of cards, and fairly lightweight, that could be comfortably worn by the worker through a workshift. Current advances in microchip design have reduced the size of some of these devices to just a little more than an identification badge. In some models the microphone is built in, and the dosimeter is thus worn on the clothing as close as possible to the ear, while in other devices, the microphone is external, and the dosimeter itself is worn on a belt, with the microphone clipped to the shoulder of the shirt near the ear.

Dosimeters are set electronically (usually through a software interface) to incorporate the correct exchange rate, exposure limit, and time constant. The worker then wears the dosimeter throughout the workshift. The dosimeter often has a protective covering to prevent the worker from changing the settings during the workshift, and the microphone usually has a windscreen covering to alleviate the effect of someone blowing on the microphone and to provide some protection from physical damage. After the workshift is complete, the dosimeter is removed and stopped, and the data are downloaded into a personal computer via a software interface. The data that can be obtained from dosimeters typically include noise dose, TWA, peak sound level, actual running time, L_{eq}, L_{max}, L_{min}, SEL, and various histograms of sound pressure level versus time.

6.3. Spectrum Analyzers

As mentioned previously, it is often desirable to have a breakdown of the noise in question by frequency, and this is often done with real-time spectrum analyzers. The most commonly encountered filter sets are the octave band and one-third octave band. In these devices, the filter sets run in parallel, in real time, so that both instantaneous and long-term average measurements can be obtained. (Mathematically based analyzers, called fast Fourier transform or FFT analyzers, are also available, and these provide analysis down to pure-tone resolution.) The typical spectrum analyzer is quite expensive and includes measurement options such as all of those discussed for SLM. Spectrum analyzers typically allow measurements to be viewed in either tabular (alphanumeric) format or graphically (Figure 1). Many analyzers include computer interfaces so that data can easily be downloaded. Also, many spectrum analyzers have a memory function that allows large amounts of data to be stored so that the computer download can be delayed until a convenient time (especially useful when taking industrial measurements). The data obtained from spectrum analyzers can be used to assess hearing damage risk, to pinpoint the source of a noise, and to perform various calculations that require data in this format (e.g. to determine the masked threshold in the field of warning signal design, or to determine the materials need for industrial noise control).

6.4. Microphones

Microphones are an integral part of all sound-measuring equipment. They may be incorporated into the equipment or attached to the device using a cable to allow remote equipment location. When a cable is used, care must be taken that the device is calibrated correctly, as longer cables may result in "cable resistance drop," an increasing attenuation of measured sound pressure levels as the signal is carried through the cable. Most

microphones have an in-line preamplifier to provide a boost to the signal on its way to the measuring device, as the actual sound pressure levels are quite small. Microphones should not be placed too close to the noise source; a good rule of thumb is that they should be placed no closer than _ wavelength away from all reflecting surfaces, including the sound source itself. For a 100 Hz sound, for example, this rule means that the microphone would be placed at least 76 cm away from the sound source and all reflecting surfaces (Earshen 1986). OSHA regulations also state that many of the required noise measurements must be taken in the vicinity of the operator's head position. In addition, microphones are not equally responsive at all angles of incidence to the sound being measured, and some are designed to be used only in certain environments. Two of the most common types of microphones are briefly discussed below.

6.4.1. Free-field

Free-field microphones are meant to be used in free-field environments (those with no meaningful reflections of sound). These microphones have directional properties and must be aimed toward the noise in question. This type of microphone is required to be used by European noise measurement standards, and so is commonly found in instruments designed for use in European countries (Earshen 1986).

6.4.2. Random-incidence

Random incidence microphones, on the other hand, have a response that is independent of the angle of arrival of a sound wave, so that they need not be pointed directly at the sound source. These microphones are meant to be used in reverberant or semi-reverberant spaces (spaces with significant reflection of sound waves; Johnson *et al.* 1991). Most USA (e.g. ANSI) standards specify the use of random-incidence microphones, but it is essential to check any relevant standards before beginning measurements to make sure the correct microphone is being used. This is especially important when using equipment made or purchased in another country.

7. CALIBRATORS

Although calibration has been left to the end of this paper, it is perhaps the most important element in obtaining correct and valid noise measurements. The sound measurement equipment itself is fairly stable, and should be returned to the factory or service center for bench calibration every year (every 2 years for some types of equipment), or if a problem is detected. The weakest link in the sound measurement chain is the microphone, as it is a delicate component susceptible to damage or skewed readings due to atmospheric conditions, rough handling, moisture, or large electromagnetic fields. It is therefore essential that the microphone be calibrated to the device with which it is being used at *both* the *beginning* and *end* of every measurement session. If the device is turned off, the microphone must be recalibrated. Likewise, if the microphone is removed and attached to another device, it must be recalibrated to that device. The calibrator is usually a cylindrical device into which the microphone is inserted and tightly coupled. The calibrator emits a pure tone at a specific sound pressure level (most commonly 1000 Hz at 94 dB), and the measuring equipment is read and adjusted if necessary to match this level. The calibrator usually runs on batteries, and care should be taken

that the batteries are fresh. In addition, the calibrator itself should be sent to the service center to be checked and calibrated on a schedule prescribed by its manufacturer.

8. CONCLUDING REMARKS

This has been only a brief overview of noise measurement, metrics and equipment. As can be seen, this is a complex topic that requires specialized knowledge. Some important factors to review before beginning any noise measurement are:

- Are there any standards or laws that apply to this type of noise measurement?
- What type of measurement is appropriate for the intended purpose?
- What is the appropriate piece of equipment to obtain this type of measurement?
- Which equipment settings should be used for this measurement?
- What type of microphone is appropriate?
- What is the appropriate calibration protocol for the measuring equipment, especially the microphone?
- What data records are necessary for the application?

In-depth answers to these questions can be found in the relevant standards as well as in the cited references.

REFERENCES

CASALI, J.G. and ROBINSON, G.S., 1999, Noise in industry: auditory effects, measurement, regulations, and management. In W. Karwowski and Marras (eds) *The Occupational Ergonomics Handbook* (Boca Raton: CRC Press), 1661–92.

CODE OF FEDERAL REGULATIONS, 1983, *Department of Labor Occupational Noise Standard*, Title 29, Chapter XVII, Part 1910, Subpart G, 36FR 10466, 29 May 1971, amended 48FR 9776-9785, 8 March 1983 (Washington, DC: US Department of Labor, Occupational Safety and Health Administration).

EARSHEN, J.J., 1986, Sound measurement: instrumentation and noise descriptors. In Berger, Ward, Morrill and Royster (eds), *Noise and Hearing Conservation Manual* (Akron: American Industrial Hygiene Association), 37–95.

JOHNSON, D.L., MARSH, A.H. and HARRIS, C.M., 1991, Acoustical measurement instruments. In C.M. Harris (ed.) *Handbook of Acoustical Measurements and Noise Control* (New York: McGraw-Hill), 5.1–21.

OSTERGAARD, P.B., 1986, Physics of sound. In Berger, Ward, Morrill and Royster (eds), *Noise and Hearing Conservation Manual* (Akron: American Industrial Hygiene Association), 19–36.

WARD, W.D., 1986, Anatomy and physiology of the ear: normal and damaged hearing. In Berger, Ward, Morrill and Royster (eds), *Noise and Hearing Conservation Manual* (Akron: American Industrial Hygiene Association), 177–95.

YEAGER, D.M. and MARSH, A.H., 1991, Sound levels and their measurement. In C.M. Harris (ed.) *Handbook of Acoustical Measurements and Noise Control* (New York: McGraw-Hill), 11.1–18.

The OCRA Method: Assessment of Exposure to Occupational Repetitive Actions of the Upper Limbs

E. Occhipinti and D. Colombini

Research Unit "Ergonomics of Posture and Movement" EPM – Milano, I-20145 Milan, Italy

1. GENERAL ANALYTICAL MODEL AND DEFINITIONS

The description and assessment model of tasks that imply potential biomechanical overload for the upper limbs is aimed at identifying and quantifying the following four main risk factors: repetitiveness (frequency), force, awkward postures and movements, lack of recovery periods. These factors, when taken together, characterize work-related exposure as related to time pattern (duration).

Additional risk factors that are to be considered as enhancers of the actual risk should be added to these. Each risk factor is to be properly described and classified (that is assessed, even if roughly). This means on the one hand to detail every single working action and on the other to consider all the factors contributing to the overall "exposure" in a general and integrated frame. The definitions regarding tests, cycles and technical actions reported in Table 1 are important to this end.

Table 1 lists the main terms used in this section, together with the definitions that best fit the author's operational choices for exposure assessment and for the description of the main risk factors.

The suggested procedure for assessing the single risk factors and the overall exposure follows the general phases listed hereunder:

- Pinpointing the typical tasks of any job, and — among them — those that take place in repetitive and similar cycles for significant lengths of time.
- Finding the sequence of technical actions in the representative cycles of each task.
- Describing and quantifying the risk factors within each cycle (frequency, force, posture, additional factors).
- Reassembling of the data concerning the cycles in each task during the whole work shifts, taking into consideration the duration and sequences of the different tasks and of the recovery periods.
- Brief and structured assessment of the risk factors for the job as a whole (exposure index).

1.1. Organizational Analysis

Organizational analysis should come before the analysis of the four main risk factors and of additional factors. It is essential to focus on the real *task duration of repetitive tasks* and on the existence and distribution of *recovery periods*.

The first phase of the analysis is finding the distribution of work times and pauses within the work shift/s.

If the task is characterized by cycles with mechanical actions

Table 1. Main definitions of recurring terminology in exposure assessment.

ORGANISED WORK:
the organised grouping of work activities that are carried out within a single working shift; it may be composed of one or more tasks.

TASK:
specific working activity whose objective is the attainment of a specific operational result. The following are identified:
REPETITIVE TASKS: characterized by repeated cycles with mechanical actions.
NON–REPETITIVE TASKS: characterized by the presence of non-cyclical mechanical actions.

CYCLE:
a sequence of technical, mainly mechanical, actions of relatively short duration, that repeats itself over and over, always the same.

TECHNICAL ACTION (mechanical):
an action that implies a mechanical activity; not necessarily to be identified with the single joint movement, but rather with the complex movements of one or more body regions that enable the completion of an elementary operation.

MAIN RISK FACTORS

RECOVERY:
period of time within the workin/g shift or within a cycle, during which no repetitive mechanical actions are carried out. It consists of relatively long pauses after a period of mechanical actions, during which the metabolic and mechanical recovery of the muscle can take place. Lack of recovery is the relating risk factor.

REPETITIVENESS:
the presence of events (cycles technical actions) that are repeated in time, always the same.

FREQUENCY:
number of technical (mechanical) actions per given time units (no. of actions per minute).

FORCE:
the physical effort required by the worker for the execution of the technical actions.

POSTURE:
the ensemble of postures and movements used by each of the main joints of the upper limbs in order to execute the sequence of technical actions that characterize a cycle. Awkward posture: risky postures for the main joints of the upper limbs.

ADDITIONAL RISK FACTORS:
the additional risk factors may be present in repetitive tasks but are not either necessarily or always present. Their type, intensity and duration leads to an increased level of overall exposure.

it will be defined as a *repetitive task*. If it is characterized by control operations, and therefore without mechanical actions, it will be defined as a recovery for the upper limbs. The tasks with non-repetitive mechanical actions remain and shall be defined as *non-repetitive tasks*.

The number of foreseen cycles within a *repetitive task*, and the net duration of each cycle, must be counted at this point. The number of cycles often coincides with the number of pieces to be worked in each shift.

1.2. Frequency of Actions (Repetitiveness)

The characterization of repetitiveness can be used to discriminate in general the tasks that must be assessed. To this end, the presence of a repetitive task for the upper limbs can be defined as the consecutive activity, lasting at least 1 h, in which the subject carries

out work cycles similar to each other and of relatively brief duration.

Once repetitive tasks have been submitted to analysis, there is the more important problem of quantifying and assessing the level of repetitiveness.

A frequency measurement proposal that is applicable in the field is the analytical counting of technical actions (mechanical) as defined above. After that, an analysis of postures and movements will make it possible to estimate the duration and frequency of even single joint movements (the type of qualitative and quantitative joint involvement, the static or dynamic component of a movement).

A description of the technical actions requires often the filming of the job, which must then be reviewed in slow motion. Often, the company already has records available in which the task is described and numbered, and the elements constituting successive technical actions are timed (methods–time measurements) .

From the technical action description it is possible to obtain: the number of actions per cycle, the action frequency in a given time unit: number of actions per minute; the overall number of actions performed by the upper limbs (right; left) during the task/tasks, and consequently in the shift.

In this model, the contents of individual or grouped technical actions (by contents we intend the force and the postures associated with technical actions) are described qualitatively and quantitatively when the risk factors "force" and "posture and movements" are studied.

1.3. Force

Force more directly represents the biomechanical involvement necessary to carry out a given action — or sequence of actions. Force may be intended as being external — applied force — or internal — tension developed in the muscle, tendon and joint tissues. The need to develop force during work-related actions may be related to the moving or the keeping still of tools and objects, or to keep a part of the body in a given position. The use of force may be related to static actions (contractions) or to dynamic actions (contractions).

Force quantification in real implementation contexts is a problem. Some authors use a semi-quantitative estimation of external force via the weight of the objects being handled. In other cases, it has been suggested to use mechanical or electronic dynamometers. The quantification of internal force is suggested by means of surface electromyography techniques. All of these methods present implementation difficulties. In this proposal we suggest to use a specific Scale developed by Borg (Category Scale for the Rating of Perceived Exertion). It can describe muscular effort as subjectively perceived for any body region. The results of the implementation of CR-10 RPE Borg's Scale, when used for an adequate number of workers, have turned out to roughly be comparable to those obtained with surface electromyography.

The actions that require minimal muscle involvement could be identified as 0.5 in Borg's Scale; then the involvement description procedure could only describe those actions, or groups of actions, that require more force than the minimal amount, always by using Borg's Scale. Once this procedure has been carried out, the average weighted score for the whole of the cycle must be calculated.

1.4. Posture and Types of Movements

Upper limb postures and movements during repetitive tasks are of basic importance in contributing towards the risk of various musculoskeletal disorders. A definite agreement is found in literature as to the potential damage coming from extreme postures and movements of each joint, from postures maintained for a long time (even if not extreme), and from specific, highly repetitive movements of the various segments.

On the other hand, the description of postures and movements of each segment of upper limbs during technical actions of one cycle completes the description of "the repetitiveness" risk factor.

The analysis of postures and movements shall be concerned with each single segment of upper limbs (hand, wrist, elbow, shoulder): it is aimed at checking the presence and time pattern in the cycle (frequency, duration) of static postures and dynamic movements involving each segment/joint considered.

The description may be more or less analytical but has to be able to appreciate at least the following items:

- Technical actions requiring postures or movements of a single segment beyond a critical level of angular excursion. The angular excursion critical level can be determined according to criteria available in the literature.
- Technical actions involving static postures and/or movements which, also in acceptable angular excursion, are maintained or repeated in the same way.
- The duration expressed as a fraction of cycle/task time of each condition reported above.

Joint combination of such description factors (posture/time) will provide the classification of posture effort for each segment considered.

1.5. Additional Factors

There are other factors that are considered important, apart from those which have already been discussed here. They always have their origin in work, and must be taken into consideration whenever assessing exposure. They have been described as additional here, not because they are of secondary importance, but because each of them can be either present or absent in the various occupational contexts which are examined and assessed.

The list of these factors — albeit not exhaustive — includes the following: (1) the use of vibrating tools, even if only for some of the actions; (2) the requirement for extreme precision (tolerance of ~1 mm in positioning an object, for instance); (3) localized compressions on anatomical structures of the hand or of the forearm, due to tools, objects or fixtures on the workplace; (4) exposure to cold; (5) the use of inadequate gloves for the task at hand; (6) objects with slippery surfaces; (7) rapid or sudden wrenching movements are required; and (8) gestures implying return shock (such as hammering hard surfaces).

This model only considers factors of physical or mechanical origin. It does not include psychosocial factors because they are not easy to quantify.

1.6. Distribution and Duration of Recovery Periods

A recovery period is the period during which one or more of the muscle groups which are usually involved in the working tasks are basically inactive. The following may be considered as recovery periods:

- work breaks, including the lunch break when taken;
- those periods during which tasks that do not involve the usual muscle groups are carried out; and
- those periods within a cycle, during which actions implying the total rest of the usually active working groups are carried out; to be significant, these periods must be at least 10–20 s consecutively.

Hence, the analysis of the recovery periods is first and foremost a check of their presence, duration and distribution within the cycle, and a macroscopic examination of their presence, duration and frequency within the whole shift.

In tasks implying high action frequency of the upper limbs (e.g. 40–50 actions per min), recovery periods within the cycle are basically non-existent. Analysis must therefore be targeted to quantifying the duration, distribution and frequency of "light" work periods and pauses. Apart from the partial exception as represented by recovery periods for actions implying protracted static contractions, the description and assessment of recovery periods should be based on the following:

- a description of the actual sequences of tasks implying an overload of the upper limbs, of "light" non-repetitive tasks, and of the pauses;
- frequency of the recovery periods with reference to the actual number of working hours per day; and
- ratio between the "total recovery time" and the "total repetitive working time".

No definite univocal criteria for the assessment of recovery periods exist in the literature.

For practical purposes one can refer to an Australian Health and Safety Commission document which established that work periods with repetitive movements and without recovery periods, extending > 60 min, cannot be acceptable. This document also supplies general criteria supporting the ratio of 5:1 between working time with repetitive movements and recovery periods. These two guidelines seem wholly acceptable as far as current knowledge goes. Their *critical* use offers an interpretation of the descriptive data collected on sequence, duration and frequency of recovery periods relating to cycles with a prevalence of dynamic actions.

2. OCRA: A CONCISE INDEX FOR THE ASSESSMENT OF EXPOSURE TO REPETITIVE ACTIONS OF THE UPPER LIMBS

2.1. Introduction

A procedure is here reported for calculating a concise index of exposure to the risks of musculoskeletal disorders associated with Occupational Repetitive Actions (OCRA) of the upper limbs. The report is based on the quantification figures for the various risk factors proposed in the previous pages.

The OCRA index is based on three premises:

- There is a need for an integrated assessment of the contribution of the main occupational risk factors (i.e. repetitiveness, force, posture, lack of recovery time, additional factors).
- Interest has been displayed in the development of a "model" for a concise index along similar lines to the one proposed by Waters *et al.* for the assessment of manual lifting tasks.
- In the present proposal the technical action is identified as

the specific characteristic variable relevant to repetitive movements of the upper extremities. The technical action is factored by its relative frequency during a given unit of time. The "exposure index" (OCRA) is the ratio of the number of technical actions (derived from tasks featuring repetitive movements) effectively performed during the shift to the number of recommended technical actions. In practice:

OCRA =
(total number of technical actions actually performed during the shift)/(total number of recommended technical actions during the shift).

The following general formula is used to calculate the total number of recommended technical actions to be performed during the shift:
Number of recommended technical actions =
$$\sum_{1}^{n} \times [CF \times (Ff_j \times Fp_j \times Fa_j) \times D_j] \times Fr \qquad [1]$$

in which

1,n = task(s) featuring repetitive movements of the upper limbs performed during the shift;
CF = frequency constant of technical actions per minute, used as a reference;
Ff; Fp; Fa = multiplier factors, with scores ranging between 0 and 1, selected according to the behavior of the "force" (Ff), "posture" (Fp) an "additional elements" (Fa) risk factors, in each (j) of the (n) tasks
D = duration of each (j) repetitive task in minutes;
Fr = multiplier factor, with scores ranging between 0 and 1, selected according to the behavior of the "lack of recovery period" risk factor, during the entire shift.

When the exposure index is less or equal to 1, exposure can hypothetically be assumed to be non-significant, or at least acceptable.

Exposure becomes significant when the exposure index is > 1. The higher the index, the greater the exposure.

Since the values of all the variables included in the equation for calculating the index are still hypotheses awaiting validation, for practical purposes it is advisable to adopt a prudential classification system of the results of the exposure index, based on the "traffic light" approach (green/yellow/red).

In practice, given the current status of experiences in the use of the index and the related results of WMSD surveys in real contexts, the following statements may be made:

- OCRA index scores < 0.75 indicate that the condition examined is fully acceptable (green area);
- OCRA index scores from 0.75 to 4.00 (yellow area) are borderline (uncertain). However, though exposure is not substantial, it may be significant and therefore careful monitoring for induced health effects should be introduced (health surveillance); and
- OCRA index scores > 4.00 (red area) are definitely significant, and the higher the value the higher the risk. Actions should be undertaken to improve working conditions (for which the analytical data will help determine priorities), as well as close monitoring for induced effects.

2.2. Procedures for Calculating the Total Number of Recommended Actions in the OCRA Index

The action frequency constant (CF) has been set at 30 actions per min under optimal conditions (i.e. other risk factors not significant); this solution is based on a critical review of literature suggestions and on practical considerations. However, it is to be underlined that the criteria suggested for the interpretation of the results of the index, with an extensive borderline zone, are primarily due to the uncertainties deriving in the choice of the constant value. The elements for determining the multiplier factors for force, posture, complementary factors, lack of recovery periods are reported in Table 2. For the analysis of the different risk factors, reference is to be made to description and quantification procedures reported by Occhipinti (1998) and Colombini (1998).

Annex A reports a data sheet for calculating the OCRA Index.

2.3. Final Remarks

The concise index of exposure proposed in this report represents a preliminary endeavor to organize the data obtained from the descriptive analysis of the various mechanical risk factors. However, it must be emphasized that the proposal is entirely experimental. At this juncture, its value lies in its ability to: (1) classify or at least group together the various scenarios which might give rise to different degrees of exposure to the various significant risk factors, and thus steer priorities of preventive actions; and (2) identify situations which do not constitute a problem, at least as far as is currently known. The design of the index is based on indications contained in the literature. The proposed index cannot under any circumstances be used in its present form as a standard, or as the expression of threshold values.

The OCRA index naturally still requires validation, particularly by means of parallel studies of induced effects in groups of exposed workers. To this aim a preliminary study was performed in 8 different work contexts in which 462 workers were exposed to different conditions of repetitive movements of upper limbs. The preliminary study showed, among others, a good agreement ($r^2 = 0.88$; $p < 0.001$) between the OCRA index

and the overall occurrence of WMSD of the upper limbs as calculated in the different work contexts.

The linear regression line best expresses this agreement:

$$y = 0.614 + 0.858x,$$

where $y = \log_{10}$ of the overall frequence (%) of WMSD and $x = \log_{10}$ of the OCRA index.

The exhaustive description of the risk factors associated with a given repetitive task, the quantification of consequent exposure in a concise, albeit approximate, index and the need to perform parallel studies on the clinical effects on exposed workers, all represent both an opportunity and a commitment to carry out further researches and investigations.

ANNEX A

Data sheet for calculating exposure risk

The first part of the data sheet lists the main items characterizing the repetitive tasks analyzed, followed by the second part which serves more specifically to calculate the OCRA index.

In particular, the first part of the data sheet identifies and quantifies the following:

- production department or line and type of work performed by the exposed workers;
- items characterizing each repetitive task (up to a maximum of four repetitive tasks per shift) such as average cycle duration (s); average action frequency (number of actions per min); total duration of each task (min);
- total number of actions performed in each repetitive task and during the entire shift;
- breaks and non-repetitive tasks that could be regarded as recovery periods;
- sequence of tasks and breaks as they occur during the shift; and
- number of hours spent in the shift without recovery periods.

The second part of the data sheet is used directly to calculate the desired index. For each task analyzed the calculation starts from the frequency constant (CF) of 30 actions per min. This constant is multiplied by the perceived effort factor (Ff), as obtained from the relevant conversion table, for each task.

Now another multiplier is calculated, this time for the posture

Table 2: Elements for determining the multiplier factors for Force (Ff), Posture (Fp), Additional Factors (Fc) and Recovery Periods (Fr).

Force Factor	R.P.E. (Borg scale)	≥0,5	1	1,5	2	2,5	3	3,5	4	4,5	5
	mean effort ≥5 (% of m.c.v.)	10	15	20	25	30	35	40	45	50	
Multiplier Factor (Ff)		1	0,85	0,75	0,65	0,55	0,45	0,35	0,2	0,1	0,01
Postural involvement Index score		0-3	4-7	8-11	12-15	16					
Multiplier Factor (Ff)		1	0,70	0,60	0,50	0,3					
Additional Element index score		0	4	8	12						
Multiplier Factor (Ff)		1	0,95	0,90	0,80						
No. of hours without adequate recovery		0	1	2	3	4	5	6	7	8	
Multiplier Factor (Ff)		1	0,90	0,80	0,70	0,60	0,45	0,25	0,10	0	

PART A – Summary of data for calculating index of exposure to repetitive movements of the upper limbs.

| Department or line........................ | Station or task .... | Shift........................ |

Characterization of repetitive tasks performed during shift

	A	B	C	D
• duration of task in shift (min)				
• mean cycle duration (sec)				
• action frequency (no. of actions/min)				
• total actions in task				

• total actions in shift (sum of A, B, C, D) ☐ Ae

Characterization of non-repetitive tasks performed during shift

	X	Y	Z
• duration (min)			
• comparable to recovery			
• not comparable to recovery			

Total no. of minutes of non-repetitive task comparable to recovery ☐ min ☐

Characterization of breaks during shift

• duration of meal break (min)
• other breaks
• total duration of other breaks (min)

Time-wise distribution of tasks and breaks in shift
(describe exact sequence of tasks and breaks, and their relative duration in minutes)

1 hour

No. of hours in shift featuring lack of recovery times N.=_____

	A	B	C	D	
• minutes spent with previous adequate recovery periods					Dre
• minutes spent without previous adequate recovery periods					Dso

PART B   Calculation of the OCRA index for the right or left upper limb.

	A	B	C	D	tasks
• action frequency constant (no. of actions/min.)	30	30	30	30	CF

• **force factor (perceived effort)**

BORG	0,5	1	1,5	2	2,5	3	3,5	4	4,5	%
FACTOR	1	0,85	0,75	0,65	0,55	0,45	0,35	0,2	0,1	0,01

☐☐☐☐ FF

• **posture factor**

SCORE	0-3	4-7	8-11	12-15	16
FACTOR	1	0,70	0,60	0,50	0,33

SH [ ]	SH [ ]	SH [ ]	SH [ ]	(*)select
EL [ ]	EL [ ]	EL [ ]	EL [ ]	lowest factor
WR [ ]	W [ ]	WR [ ]	WR [ ]	among elbow,
HA [ ]	HA [ ]	HA [ ]	HA [ ]	wrist
				and hand
(*)	(*)	(*)	(*)	FP

• **additional items factor**

SCORE	0	4	8	12
FACTOR	1	0,95	0,90	0,80

☐ FA

• **duration of repetitive task (min.)**

☐

= ☐

*No. of recommended actions per repetitive task and totals (partial result without recovery factor)

					π
α	β	γ	δ	(α+β+γ+δ)

• **factor for lack of recovery time**
(No. of hours without adequate recovery)

N. HOURS	0	1	2	3	4	5	6	7	8
FACTOR	1	0,90	0,80	0,70	0,60	0,45	0,25	0,10	0

☐ × ☐ = ☐
FR   π   Ar

$$\text{OCRA Index} = \frac{\text{Total no. of actions observed in the repetitive tasks}}{\text{Total no. of recommended actions}} = \frac{Ae}{Ar} = \boxed{}$$

---

factor (Fp). Here too, the factor is chosen on the basis of a conversion table that matches descriptive values with multiplier factors. The values for the factors pertaining to the four segments of the upper limb at greatest risk for each limb (i.e. hand, wrist, elbow, shoulder) must be entered into the appropriate spaces. It is advisable to select the lowest multiplier factor for elbow, wrist and hand (since the index is designed specifically for repetitive actions performed by these segments).

The next multiplier to be calculated is for the additional items factor (Fa).

The result of these three multipliers (not indicated in the data sheet) represents the frequency constant per minute weighted by the factors for force, posture and additional elements. The result, multiplied by the duration (min) of each task analyzed, is used for calculating the number of recommended actions for each individual task and, when added together, for the entire shift (p).

At this stage, it is necessary to weight the total number of recommended actions (p) obtained in the partial result indicated above, using the factor relative to the presence and distribution of recovery periods. This is achieved by obtaining the factor for recovery periods (Fr) from the conversion table. This factor is then used as a multiplier of the figure resulting from the previous equation (p).

In this way, the total number of recommended actions for the shift can be obtained (Ar).

This factor is the denominator of a fraction in which the numerator is the total number of actions effectively performed during the shift (Ae) calculated in the first part of the data sheet. The ratio represents the index of exposure to repetitive movements (OCRA).

REFERENCES

ANDERSON, P., 1998, *Cumulative Trauma Disorders: A Manual for Musculoskeletal Diseases of the Upper Limbs* (London: Taylor & Francis).

COLOMBINI, D., 1998, An observational method for classifying exposure to repetitive movements of the upper limbs. *Ergonomics*, 41, 1226–89.

HAGBERG, M., SILVERSTEIN, B., WELLS, R., SMITH, M.S., HENDRICK, H.W., CARAYON, P. and PERUSSE, M., 1995, *Work-related Musculoskeletal Disorders. A Reference Book for Prevention* (London: Taylor & Francis).

OCCHIPINTI, E., 1998, OCRA a concise index for the assessment of exposure to repetitive movements of the upper limbs. *Ergonomics*, 41, 1290–1311.

WATERS, T. R., PUTZ ANDERSON, V., GARG, A. and FINE, L. J., 1993, Revised NIOSH equation for the design and evaluation of manual lifting tasks. *Ergonomics*, 36, 749–776.

OWAS – A Method for Analysis of Working Postures

M.K. Mattila

Occupational Safety Engineering, Tampere University of Technology, FIN-33101 Finland

1. INTRODUCTION

OWAS method (the Ovako Working Posture Analyzing System) is one of the simpler observation methods for postural analysis (Karhu *et al*. 1977). The OWAS method has proved to function well in practice on the plant level, and it has proved to be fruitful in achieving improvements in the work system and in preventing health problems (Karhu *et al*. 1981).

The OWAS method has proved to be easily adaptable to daily workplace analysis and is capable of evaluating numerous postures at a variety of workplaces. The OWAS method can be used for the following purposes (Mattila *et al*. 1993):

i   standardized ergonomics evaluation of the postural load
ii   improvements and planning of workplaces, work methods, tools and machines
iii   use by the occupational health services in planning of work for disabled individuals
iv   scientific research to be used with other field methods.

The OWAS method is widely used as the ergonomics analysis method because (Kivi and Mattila 1991):

i   it is intended to serve as a practical tool for daily analysis at workplace level
ii   it is oriented to correcting measures, not only to problem identification
iii   it has been developed to analyse a wide range of different postures, as in the case in building construction
iv   it has proved to function as a tool for fruitful cooperation between different specialties in the company — e.g. occupational safety and health professionals, managers and engineers
v   it is an observational technique well suited to the current methods of occupational health care.

2. FUNDAMENTALS OF THE OWAS METHOD

The theoretical background of the OWAS method is in job and task analysis, in observational techniques, and in risk assessment. When using the OWAS method in job analysis the job is usually divided into tasks. By analysing the working postures at all the tasks the whole posture analysis of a job will be carried out. The OWAS method is based on observations done by trained specialists and is therefore aimed to be an objective method for analysis. High reliability coefficiencies have been reported in several studies.

The OWAS method includes the idea of risk assessment also when the idea of risk acceptance and the emergency of risk reduction by work improvements is integrated in the interpretation of results.

3. THE STRUCTURE OF THE OWAS METHOD

The OWAS method is based on the idea of making momentary observations of postures at certain intervals of the jobs studied and, after receiving sufficient observations, reliable material is available to describe the postures of the job as a whole. It requires only a few seconds to perform an observation and to record it.

The original OWAS recommendation is to make the observations at 30 second intervals. In some studies however shorter intervals, e.g. 15 seconds, have been used. Recently the OWAS method has also been used based on videotaped information of the task and even then 5-second intervals have been used for tasks with short time-spans. When using videotape recordings for the OWAS analysis, special attention has to be paid to the validity, reliability, and correctness of the work-posture information.

The OWAS method identifies the most common postures
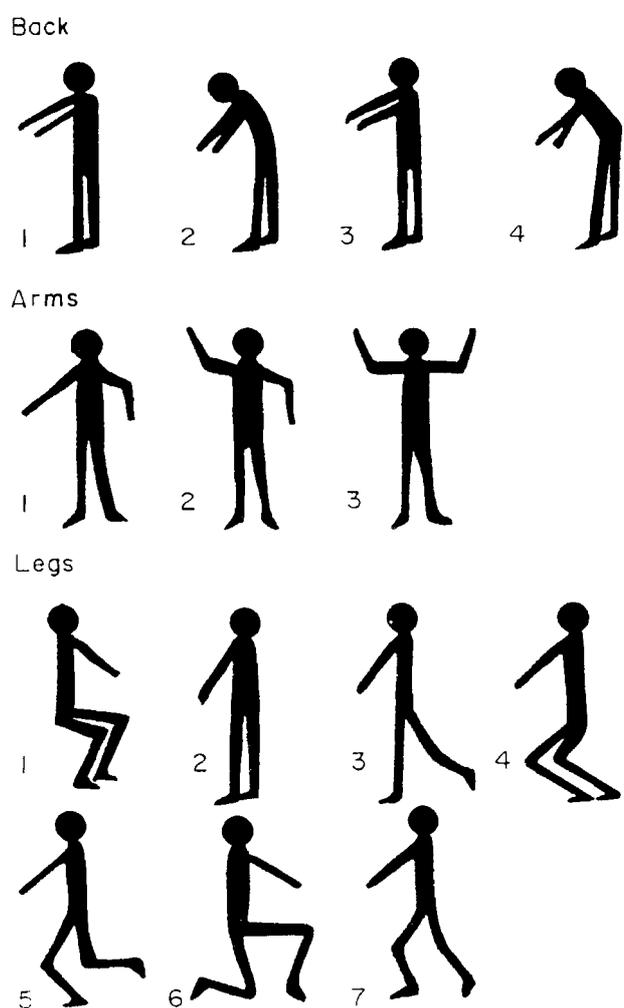
Figure 1. OWAS codes for different body parts (Karhu et al. 1977)
Back code: 1, straight; 2 , bent; 3, twisted; 4, bent and twisted
Arms code: 1, both below shoulder level; 2, one above shoulder level; 3, both above shoulder level
Legs code: 1, sitting; 2, both straight; 3, one straight;4, both bent; 5, one bent; 6, kneeling; 7, walking
Load effort code: 1, ≤10 kg; 2, ≤20 kg; 3, >20 kg

for the back, arms, and legs, and estimates the weight of the load handled or the (extent of the) hand force muscle (effort). The OWAS method utilizes a four-digit code to describe various postures of the back, upper limbs, lower limbs, and the force needed (figure 1).

In the analysis of posture observations the amount of different posture codes will be calculated and their relative distribution presented. The presentation of the result shows the percentage distribution of observations according to observation criteria in all four posture determinants: back, arms, legs, load/effort (figure 2).
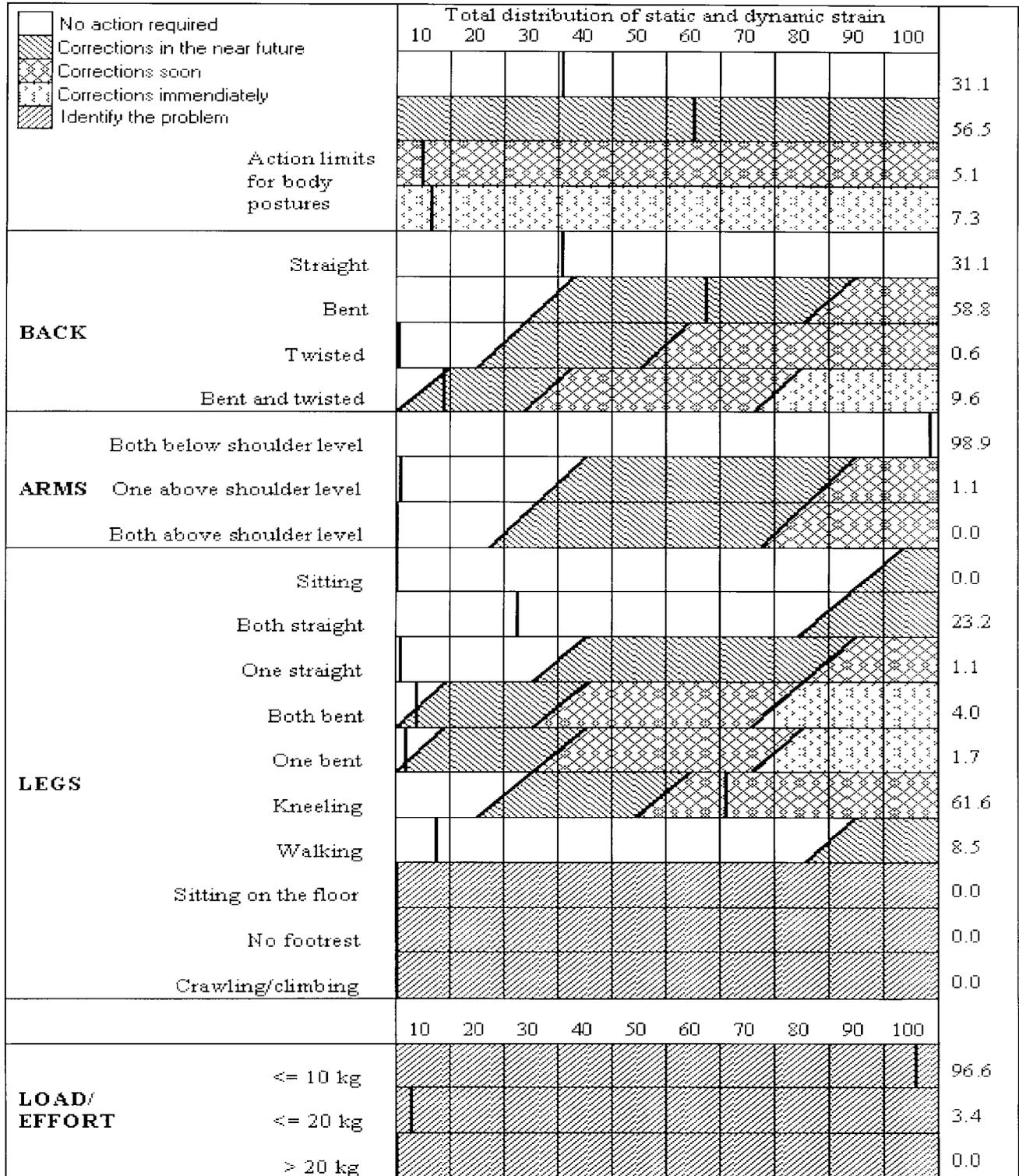
Figure 2. Distribution of the postures observed in the task analysed, with recommendations for corrective actions (Mattila *et al.* 1993)

Table 1. The OWAS action categories for evaluation of working postures

OWAS category	Description
Action category I	Work postures are considered usually with no particular harmful effect on musculoskeletal system. No actions are needed to change work postures.
Action category II	Work postures have some harmful effect on the musculoskeletal system. Light stress, no immediate action is necessary, but changes should be considered in future planning.
Action category III	Work postures have a distinctly harmful effect on the musculoskeletal system. The working methods involved should be changed as soon as possible.
Action category IV	Work postures with an extremely harmful effect on the musculoskeletal system. Immediate solutions should be found to change these postures.

The observed posture combinations are classified according to the OWAS method into four ordinal scale action categories. This classification of postures (see table 1) is based on experts´ estimates of the health hazards of each work posture or posture combination on the musculoskeletal system.

4. COMPUTER SYSTEM

Originally the coding and analysing procedures of OWAS were carried out using handwritten forms, which required a lot of paperwork. ADP applications have therefore been developed for OWAS, and it has been shown that these offer new possibilities for data analysis. Knowledge of the activity carried out during different postures has been mentioned as being useful for prevention, and this has been added to the traditional OWAS record forms (Kivi and Mattila 1991).

The computer system which has been developed consists of three elements: (1) posture coding in the field, (2) data transfer to the PC, and (3) data analysis and presentation through the PC.

The OWAS program developed for data analysis and presentation consists of the following functions:

i   printing the summary form for all observations according to the classification of measures needed

ii   distribution of postures into evaluation classes according to postures and according to body parts

iii printing the form for the job or the task, including analysis of the acceptability of postures

iv   the most critical postures of subtasks

v   analysis of jobs and tasks according to the percentage of unacceptable work postures.

Based on the computer-aided analysis, it is possible to identify the tasks causing the highest risk estimations. It is also possible to discover which really are the riskiest posture codes and the sub-tasks at which they occur.

5. THE PARTICIPATIVE PROCEDURE

The participative procedure has proved to be fruitful in introducing the improvements based on OWAS analysis. The results of the OWAS analysis have been discussed by a cooperative team, an approach which has proved to be fruitful in improving

prevention. In one experiment in the construction industry, the occupational safety manager (who also served as the head of the production unit), the workers´ safety representative, the physician, the nurse, and the researchers took part in a meeting of the cooperation team. The team discussed the possibilities for preventive measures, which were then presented in the form of written suggestions to the employer.

6. RECOMMENDATIONS

i   The OWAS method has proved to be successful in safety-oriented risk assessment based on workplace settings where working posture is one of the main problems.

ii   The OWAS method has proved to be suitable in situations where the previous hazard screening has shown that work postures are a major risk factor for the health of personnel.

iii The computer-aided application of the OWAS method has enlarged the possibilities for the analysis of OWAS results, and it has made it possible to organize participative meetings of work groups very soon after the analysis is completed.

iv   Good documentation of the OWAS analysis, supplemented with videotapes and slides, has offered a powerful and reliable basis for teamwork in developing corrective measures.

v   The OWAS documentation simplifies the transfer of knowledge obtained by ergonomics analysis to the designers of production methods and to machinery and product designers (Vilkki *et al.* 1993).

vi   The OWAS documentation provides practical training material for new employees and for retraining experienced workers. This training material is considered especially effective as it is based on a company's own production technique.

vii The OWAS documentation, accompanied where possible by videotapes, supports the participative approach in occupational health and safety for the development and introduction of preventive measures and improvements.

viii Worker involvement is crucial in order that everyone´s work interests are recognized and the aims and principles of the analysis are understood.

ix   Team work and group problem-solving have proved to be requirements for successful intervention. The group should include representatives of management, employees, safety experts and health professionals.

x   Feedback to employees concerning the results and follow-up steps is necessary to motivate a continuous process in developing the work environment.

xi   The implementation of the corrective measures must be performed quickly by the line management in order to manifest the management's interest in improving the work environment, whether in changes in working methods, planning, or delivery.

REFERENCES

KARHU, O., KANSI, P., and KUORINKA, I., 1977, Correcting working postures in industry: a practical method for analysis. *Ergonomics*, 8(4), 199–201.

KARHU, O., HÄRKÖNEN, R., SORVALI, P., and VEPSÄLÄINEN, P., 1981, Observing working postures in industry: examples of OWAS application. *Applied Ergonomics*, 12(1), 13–17.

KIVI, P. and MATTILA, M., 1991, Analysis and improvement of work postures in the building industry: application of the computerized OWAS method. *Applied Ergonomics*, 22(1), 43–8.

MATTILA, M., KARWOWSKI, W., and VILKKI, M., 1993, Analysis of working postures in hammering tasks on building construction sites using the computerized OWAS method. *Applied Ergonomics*, 24(6), 405–12.

MATTILA, M., VILKKI, M., and TIILIKAINEN, I., 1992, A computerized OWAS analysis of work postures in the papermill industry. In Mattila and Karwowski (eds), *Computer Applications in Ergonomics,* *Occupational Safety and Health* (Amsterdam: Elsevier Science), pp. 365–72.

VILKKI, M., MATTILA, M., and SIUKO, M., 1993, Improving work postures and manual materials handling tasks in manufacturing: a case study. In Marras, Karwowski, Smith, and Pacholski (eds), *The Ergonomics of Manual Work* (London: Taylor & Francis), pp. 273–6.

Prevention of Work Injury

A. Garrigou

Département Hygiène Sécurité and Environnement, IUT A, Université de Bordeaux 1, France.

A. Weill–Fassina

Laboratoire d'Ergonomie Physiologique et Cognitive, Paris, France.

J.–P. Brun

Faculté des Sciences de l'Administration, Chaire en gestion de la santé et de la sécurité du travail dans les organisations, Université Laval, Québec, Canada.

F. Six

Groupement d'Ergonomie de la Région Nord, Lille, France.

M. Chesnais

Equipe Psychologie Cognitive et Ergonomique, Université Paris, France.

D. Cru

Agence Nationale pour l'Amélioration des Conditions de Travail, Lyon, France

1. INTRODUCTION

The object of this article is to discuss the specificities of the preventionist's job based on an ergonomics point of view of their practice. The prevention function has developed within enterprises and control organizations, both private or public, to take into account factors such as the increase of industrial activities, new risks for workers, the nearby population, and the environment. This has led to an increase in the number of the participants involved in these matters (doctors, ergonomists, toxicologists, health and safety officers, etc.). In this article we will focus on health and safety preventionists operating within enterprises in France and Québec and highlight some of the difficulties encountered in their daily lives. Although managers, designers, and workers have a tendency to restrict the work of preventionists to health and safety matters, we will indicate the additional functions they could fulfill within organizations.

2. PREVENTION: ITS MULTIFOLD ACTORS AND OBJECTS

The prevention function that has developed in enterprises directly involves many participants, such as health and safety officers or technicians, occupational physicians, nurses, ergonomists, health and safety representatives (Comité d'Hygiène et Sécurité Condition de Travail in France, Comité de Santé Sécurité in Québec), and union and management representatives.

In their daily work, these men and women have a vital role to play within an enterprise so that it can develop efficiently while, at the same time, guaranteeing a reliability of operation and working conditions that is compatible with the health of the workers and nearby populations, together with a healthy environment.

Some recent studies (Brun *et al.* 1998) have revealed a great variety in the tasks and objectives dealt with by preventionists, which represent the specificities of their jobs and which can have

more, or less, importance according to the enterprises or countries involved:

- Respect of health and safety and/or environmental laws and regulations that are constantly evolving
- The management of budgets
- The follow-up of design projects, sites, the checking of facilities and devices, etc.
- Ensuring that equipment meets requirements
- The setting up of prevention policies and programs as defined by the enterprise concerned
- The drafting of safety instructions and toxicology data sheets
- The organization of training sessions for operators, foremen, designers
- Ensuring obedience to safety instructions and regulations by shop-floor workers
- Pre-planning of crises management, fire-fighting intervention, evacuation of personnel, etc.
- The completion of reports on accident investigations or on occupational diseases
- The follow-up and treatment of accident statistics
- Involvement in meetings (health and safety groups, project meetings, etc.)
- Meeting people from other control organizations and agencies; any follow-up arising from these meetings such as notices of improvements and/or changes required

These activities and contacts with the various personnel concerned with prevention are time-consuming, especially in the establishment of relationships and/or completion of documents, and this can sometimes hinder prevention actions in the field. The numerous subjects that must be dealt with within the constraints and limits of time and resources available can present a risk of fragmentation and dispersion for the preventionist. The necessary response to hazards in an emergency can prevent in-depth analyses of some situations.

3. MODELS EXISTING IN ORGANIZATIONS

To carry out their various tasks, preventionists have to adapt to work or risk models, which can be more or less developed, explicit or implicit. They guide the managers to decisions about investments, organizational change, choices of technologies, and health and safety policies.

3.1. The Paradigm of Technical Rationality

Many of the decisions taken by the participants within an enterprise, especially those of industrial designers, are strongly influenced by the paradigm of technical rationality. In this current of thought, there is a gap between human and technical factors. According to Pahl and Beitz (1984): "The engineer's main task is to apply his scientific knowledge to the solution of technical problems and then to optimize within the given material, technological and economical constraints."

Often anchored in industrial business, this paradigm gives the managers and designers the illusion that to carry out a project successfully one needs only to work from hard sciences. This approach, therefore, tends to define the objectives of a project by favoring technical aspects rather than the future work organization or the training of the operators. The latter will only be defined at a later stage without taking into account the human interaction likely to occur in work situations. Thus, many designers ignore

the fact that, when developing technological activities, they are also, without realizing it, designing the frame for organizations and hazards that will appear in the new work process. The general principles of prevention and precaution are often overlooked in the industrial design process.

3.2. The Person at Work and Hazards Models

In many industrial or service businesses, the human activity model remains focused, consciously or unconsciously, on a physical or physiological dimension, in which the person is perceived as a system of energy transformation. Often, it is also considered that people are exposed only to concrete and visible hazards which endanger the body (machine operation, falls, posture, noise, etc.); risks that could be called "virtual" (too much information to handle, work rhythm, etc.) are seldom taken into account by preventionists.

This mechanical vision accompanies a stereotype opposing manual and intellectual activities; perhaps it is a legacy of work representation from nineteenth- or early twentieth-century industries and factories, or perhaps it remains because it is the more visible, if not the most spectacular, aspect of work. Such representations of people at work and the hazards involved result in only the physical injuries, whether due to occupational diseases or accidents, being accounted as damage to health.

The cognitive dimensions relevant to any activity are thus greatly underestimated, despite awareness that the excessive demands of having too much information to explore or consider, together with the pressures resulting from having too little time in which to see through these activities, could be the cause not only of various efficiency and reliability malfunctions of the system, but also of long-term effects on physical and mental health.

3.3. From the Statute of Prescriptions to the Role of Experts

The Taylorian tradition of work organization separates design from execution. Even if more discretionary forms of work develop, giving place to the initiative of operators, due to certification procedures, the statute of prescriptions and work regulations are strengthened in other cases because of economical and legal requirements.

In order to design the procedures, the preventionist must foresee a certain number of risky situations and combine different levels of safety standards drawn from general regulations, and apply them to particular risks or to the envisaged task. The various states in which a system can exist are most frequently predefined within the framework of nominal situations, the hypothesis being that the operator has only to carry out and strictly obey the recommendations in order to meet safety, reliability, and efficiency conditions.

However, although the prescriptions and recommendations are indispensable tools for work regulation, they are not enough (Trinquet 1996). Indeed, the difficulties are manifold and can be found in the implicit aspect of recommendations which do not set out the full job specification (de Terssac 1992) since, because of variable situations, it is impossible to foresee every facet of coordination necessary between safety and production activities.

3.4. Approaches Focused on the Person

Whenever a breach of regulations prejudicial to the performance of the system or to health is discovered, the legal logic for compensation or damages resulting from work accidents, as well as the general line of thinking and social representations, is often regarded as the responsibility of the individual expert or skilled person because he or she was careless or not sufficiently alert. Prevention or training measures are then centered on the acquisition of individual safety behaviors, and little account is taken of the organization of the work, management decisions, the interaction of operators within a collective vertical job in a team, as opposed to the horizontal work of their hierarchy, the exchange of information and coordination activities (Carballeda *et al.* 1994; Reason 1990).

3.5. Accidents and Their Causes

The individual management of prevention, added to psychological defense and mourning mechanisms have a tendency, in case of accident, to look for a culprit, which interferes with the understanding of the accident mechanism in order to improve prevention modes. This causal attribution mechanism (Kouabénan 1998) is particularly clear when different groups present different "explanations" of the accident. In fact, to blame someone else for an accident, or the non-obedience to a regulation being its cause, is a psychic protective reaction (Brun 1992). This lead to neglecting the complexity of work situations, the intricacy of the conditions or actions that enabled the unwanted events to occur and that changed them into "normal accidents" (Perrow 1984).

The limits of such uni-causal approaches have led researchers in the 1980s to admit the multi-causality of accidents (Leplat and Cuny 1979; Hale and Glandon 1987). In addition, prevention organizations (such as INRS in France) have developed appropriate analysis tools that were widespread throughout enterprises and professional training courses (Chesnais 1993; Hendrix and Benner 1987).

4. THE PRACTICE OF PREVENTIONISTS: THE DIFFERENT MODELS UNDER STRESS AND THEIR CONTRADICTIONS

According to their professional background and their academic training, preventionists will interiorize and feed these models, or at least come across them. In their practice they will have to face the contradictions generated by these models because of the fragmentation of the representation of the activities suggested and the limits mentioned earlier; the preventionist then risks finding it difficult to account for the complexity of situations and assess the impact of the preventive measures decided for the activities of the people. In addition to which the lack of time and resources available leads the preventionist to relate work accidents to simple causes, i.e. the operator, rather than to complex causes revealed by more accurate analyses that often require more time and assistance from some kind of expertise that he does not necessarily have.

The preventionist then risks being caught between the mechanistic model of behavior and the necessity of integrating a strong relational dimension inherent in his functions, which supposes his daily positioning with regard to all the personnel.

One of the difficulties he then encounters is how much advice he will give rather than coercion activities, how much will he play the role of an expert rather than a mere animator.

The establishment of trusted relationships, the guarantee of his credibility rest on a gamble, which recurs with each interaction situation and which can be challenged whenever there is a conflict. New regulations insist on this advisory part and it seems that awareness of the ergonomics approach is appreciated to face new missions (Chesnais *et al.* 1995).

An important part of a preventionist's activity is based on the classification of the different actions to be carried out, according to their degree of urgency. They require the elaboration of a compromise between heterogeneous demands and logistics: safety and reliability, the will of partners, social relations within joint institutions (management, workers representatives for example with the CHSCT or CSS), but also efficiency and return after investment, etc.

5. NEW APPROACHES TO PREVENTION

The questions raised direct us towards new approaches to prevention practices. The viewpoint presented here belongs to the tradition of activity ergonomics, originally developed within the French-speaking trend of ergonomics (de Keyser *et al.* 1988; Montmollin 1992; Wisner 1994). It also rejoins the debate opened within the English-speaking ergonomics on safety and security trend (De Greene 1991; Embrey 1991; Dwyert 1996).

5.1. The Human Models pertinent for Prevention

The limits of the models that have been discussed earlier make it necessary to complete (Garrigou *et al.* 1994) by taking into account a *cognitive dimension* (looking for information, treating them, decision making, memorization, etc.), a *psychic dimension* (accounting for the identity and personality construction process through the work activity) and a *social dimension* (enabling us to account for the belonging to social groups that will convey values, as well as the coordination of actions in collective work). It is only then that the preventionist will be able to describe and explain the difficulties encountered by operators and thus contribute to changing situations.

5.2. Industrial Variabilities, Incorporated Know-how and Regulations

A great many studies in ergonomics (Faverge 1972; Laville and Teiger 1972; Wisner *et al.* 1986) have emphasized the lack of stability of work process, particularly due to industrial variabilities (sensitiveness of raw materials and processes, variations in the environment: temperature, humidity, vibrations, acidity, etc.) and of the plants working under degraded mode (which is more frequent than thought). Under such conditions the operators are brought to develop know-how and individual and collective strategies to collect information, indications, of the current operations, in order to anticipate incidents or accidents (Sanderson 1989). Such strategies contribute to various sensorial modalities that will gave rise to embodied knowledge (Garrigou *et al.*, 1998a; Aubert, to be published); we must point out that because of their characteristics, they are very often difficult to put into words. This knowledge plays a crucial part in the

efficiency and reliability of the functioning of organizations and industrial systems.

In such a context, when industrial variability cannot be reduced, work regulations or recommendations are unable to foresee everything because of the great diversity of situations, and if they are so numerous they can no longer be managed. Then it is necessary to consider regulations as supports for action and assistance in decision-making; they must be articulated around the knowledge developed by the operators: work regulations are necessary but not sufficient.

5.3. Re-thinking of the Definitions of Health

If originally the role of preventionists consisted in carrying out different actions (training, animation, change of equipment, etc.) so that certain situations could not harm people's health, the very evolution of work (in Western countries: technologies of information, more discretionary organizations, the aging of populations) makes it necessary to think about a more pertinent definition of health. Dejours (1995) emphasized that the traditional definition of health is centered on the static state of good or bad health. He defends the idea that health is a dynamic construction by or from the people; current research (Laville 1998; Antunes Lima 1995; Avila-Assunçao 1998) shows that the process of safeguarding health may require a collective construction. To achieve the objectives of the working system, tiredness and pain can be avoided or lessened to those most affected by a redistribution of tasks to younger workers, while older workers carry out non-formal functions as advisers or trainers to the younger ones. Organizational conditions which allow for individual or collective maneuvering margins (care for others, mutual assistance, etc.) thus enables health protection to become *a vital aspect of the preventionist's action* within the framework of organizational changes.

5.4. Reviewing the Operators' Risks Management

The works of Cru (1995) and de la Garza (1995) point out that operators who have to face everyday-risk situations do not wait passively for possible unwanted events to occur. The management of these risks is often integrated into their activity. This requires an individual mobilization in the way of attention, a certain form of discipline or permanent personal organization, when dressing or setting up PPES (personal protection equipment), achievement of production or maintenance operations, etc. Beyond a physical activity that can be important, this mobilization and attention requires the intake of a lot of information.

This individual mobilization is articulated around more or less developed forms of collective mobilization (Garrigou *et al.* 1998b). Thus, important research on sites where asbestos has been removed has revealed that when workers put on personal protective equipment or when they worked in risk areas, this collective mobilization is noticeable in the way that care is shown towards one another: each one making sure that his neighbor feels well, that his protection suit is not damaged, and that everything is all right for the "young workers" just starting, etc. Collective mobilization, as far as care is concerned, then, is the main guarantee that work on the site will be safe, and it reinforces individual mobilization. It is essential, therefore, that preventionists to understand the conditions better (technical,

organizational, or individual) that improve such processes and, on the other hand, those who hinder or prevent them.

From this perspective, it may be necessary for operators to show that they are in control of risks by exhibiting professional skills, and questioning hierarchy or safety arrangements (Dodier 1995). The establishment of these preventive competencies, therefore, both helps to achieve production targets and give the operator the opportunity to display his ability at controlling hazards. In this case, the relationship with danger, safety regulations, and the activity of the preventionists is reversed; the operator is no longer governed by hazard and safety recommendations, but he himself determines his own safety and destiny. This is an aspect with which preventionists are not very familiar but one which they will have to learn to work with.

5.5. To Think instead of Others or with Them?

Expertise in health and safety is indeed necessary, particularly when the preventionist intervenes early in design projects and when projects are highly confidential. However, such an approach remains unsatisfactory because the preventionist is doing the thinking instead of the workers. Decisions could be made without the operators being able to assess their advantages or drawbacks and any new constraints which might be generated in their activities. Participative approaches (Wilson 1991; Noro and Imada 1991) that have been praised in the 1980s have attempted to mobilize the workers and their competencies; the results of such experiments were often mitigated. In order that participative approaches of future users in design projects, that are presently being re-used in new design organizations (for example in concurrent engineering), have a chance of being efficient, at least three conditions seem to be necessary: (1) the recognition of what the users' representatives or the users themselves can input according to their professional experience; (2) a commitment of the deciders about objectives and means granted to such practices; (3) the existence of methodological approaches that sustain such a process. If this is not the case, these practices risk, for different reasons, facing resistance from users and designers. More particularly, according to the *"good professional's ideology"*, he should have no problems since he knows how to solve them; in this context, which is fairly frequently found not only with designers but also with others professions, anyone who admits to having problems takes the risk of being judged by his peers or seniors as someone not competent; such mechanisms can hinder feedback and cause suffering for some people.

One should expect from the participative approach a recognition of the operators' cautiousness, their professional language, codes of practice that rule their collective work, equipment better adapted to specific situations and the foreseeing of any malfunctioning in respect of organization, techniques, activities, and the overcoming of malfunctions (Cru 1995). As far as the preventionist's activity is concerned, this requires the setting up of a social construction process and rules of participation among the various actors involved in an enterprise.

6. CONCLUSION

The discussion proposed here concerning the activity of preventionists has consisted in characterizing its practice by confronting it to other modes of knowledge at work that stem from sociology, psychodynamics and, especially, ergonomics.

More questions have been raised than answers given. We think it wise, therefore, to improve understanding of what the activities of preventionists are in order to help the development of new prevention practices that would match their new applications (Brun *et al.* 1998).

A reflexive approach on their practice (Schön 1983) would enable better understanding of the complexity of the constraints and stakes where they have to find compromises. Different aspects could be tackled, such as:

- the establishment between safety, reliability, production, and criteria of rentability at shorter or long term;
- the levels of reduction of the complexity of working situations necessary for prevention;
- the possibility of conciliating the role of an expert with the recognition of the operator's competency, so that they have the opportunity to share their experience;
- efficient strategies and actions to change, at least in part, the reference models of the various actors in the enterprise (described in section 3);
- strategies enabling the preventionist to increase his credibility so that he is accepted very early on in design projects, from the technical or organizational viewpoint;
- the fact that preventionists themselves have to protect themselves from the various types of risks related to their own responsibility can breed fear and anxiety, and lead them to develop various psychical defenses to deal with these problems.

Such reflection on the activity of preventionists is vital, both for the teaching and training of preventionist's courses, the developing of methodological approaches adapted to needs of preventionists and, later on, the improvement of safety conditions, reliability, and efficient working of installations.

REFERENCES

ANTUNES LIMA, F. de P., 1995, Les contraintes au travail et la dimension éthique de l'activité, les valeurs dans l'activité de travail et leur signification pour la transformation ergonomique. Thèse de doctorat d'Ergonomie. Paris, Laboratoire d'Ergonomie du CNAM.

AUBERT, S., Embodied knowledge, to be published in International Encyclopedia of Ergonomics and Human Factors.

AVILA-ASSUNCAO, A., 1998 ,*De la déficience à la gestion collective du travail: les troubles musculosquelettiques dans la restauration collective.* Thèse de doctorat en ergonomie. Paris: EPHE.

BRUN, J., 1992 *Les hommes de lignes: analyse des phénomènes sociaux et subjectifs dans l'activité de travail des monteurs de lignes électriques.* Thèse de doctorat en ergonomie, École Pratique des Hautes Études, Paris.

BRUN, J.P., LOISELLE, C., GAUTHIER, G., et BÉGIN, C. 1998, *Le métier de préventionnistes: entre l'arbre et l'écorce*, éd. Héritage et Transecta.

CARBALLEDA, G., GARRIGOU, A. and DANIELLOU, F. 1994, Organizational changes: organization stabilization vs. workers destabilization: the case of a high-risk process control plant. In G.E. BRADLEY and H.W. HENDRICK (eds) *Human Factors in Organizational Design and Management - IV* (Amsterdam: North-Holland), pp. 161–6.

CHESNAIS M., 1993, *Enseigner la prévention des risques professionnels: l'arbre des causes* (Paris: INRS).

CHESNAIS,C. and ROCHER M., 1995,*Utilisation de la démarche ergonomique dans les pratiques de prévention des CRAM. Résultats d'entretiens avec 40 préventeurs de CRAM ayant suivi un Stage*

d'ergonomie. Rapport technique Contrat INRS CRAM CNRS A.1/1.008.

CRU, D., 1995, *Règles de Métier , langue de Métier: dimension symbolique au travail et démarche participative de Prévention Le cas du Bâtiment et des travaux publics*, Diplôme de l'École Pratique Des Hautes Études

De GREENE, K., B., 1991, Emergent complexity and person-machine systems, *International Journal of Man–Machine Studies,* 35, 219–34.

De KEYSER, V., DE CORTIS, F., and VAN DAELE, A., 1988, The approach of Francophone Ergonomics: studying New Technologies. In V. DE KEYSER, T. QUALE, B. WILPERT and S. A. RUIZ QUINTILLA (eds), *The Meaning of Work and Technological Options* (London: Wiley), pp. 147–63.

De La GARZA, C., 1995, *Gestions individuelles et collectives du risque et du danger dans la maintenance d'infrastructure ferroviaires.* Thèse de doctorat d'Ergonomie. Paris: LEPC-EPHE.

DEJOURS, C., 1995, Comment formuler une problématique de la santé en ergonomie et en médecine du travail, *Le Travail Humain*, 58(1), 1–16

DODIER, N., 1995, *Les hommes et les machines: la conscience collective dans les Sociétés technicisées* (Paris: Metailié).

DWYER,T., 1996, A wager on the future: a sociological paradigm for occupational safety and health. In Abstracts from the workshop on Occupational Health Strategies, Dublin, October 1996.

EMBREY, E.D., 1991, Conception et mise en oeuvre de programmes visant à la réduction de l'erreur dans l'industrie: considérations techniques, organisationnelles et culturelles. In M. NEBOIT and E. FADIER (eds), *Facteurs Humains de la fiabilité et de la sécurité des systèmes complexes* (Nancy: INRS), pp 53–9.

FAVERGE, J.-M., 1972, *L'analyse du Travail.* In M. Reuchlin (ed), Traité de Psychologie Aplliquée. Vol. 3, (Paris: PUF).

GARRIGOU, A., CARBALLEDA, G., and DANIELLOU, F., 1994, The contribution of activity analysis to the understanding of maintenance difficulties in high-risk process control plant. In F. AGHAZADEH, *Advances in Industrial Ergonomics and Safety, VI* (London: Taylor & Francis), pp. 199–206.

GARRIGOU, A., CARBALLEDA, G., and DANIELLOU, F. (1998a) Know-how in maintenance activities and reliability, in a high-risk process control plant. In *Applied Ergonomics*, 29(2), pp. 127–32.

GARRIGOU, A., MOHAMMED-BRAHIM, M. and DANIELLOU, F., 1998b, *Une étude ergonomique sur les chantiers de déflocage d'amiante.* Rapport OPPBTP (Paris: OPPBTP).

HALE, A.R. and GLENDON, A.I., 1987, *Individual Behavior in the control of danger* (Amsterdam: Elsevier).

HENDRICK, K.. and BENNER, L.,1987, *Investigating Accidents with STEP* (New York: Marcel Dekker).

KOUABENAN, D.R., 1998, Beliefs and the perception of risks and accident, *Risk Analysing,* 18(3).

LAVILLE , A., 1998, Les silences de l'ergonomie vis à vis de la santé, Actes des deuxièmes journées *Recherche* and *Ergonomie*, Toulouse, February.

LAVILLE, A. and TEIGER, C., 1972, Nature et variations de l'activité mentale dans les tâches répétitives: essai d'évaluation de la tâche de travail. *Travail Humain*, 35(1–2), 99–116.

LEPLAT, J. and CUNY, X., 1974, *Les accidents du travail* (Paris: PUF).

MONTMOLLIN, M. de, 1992, The future of ergonomics; hotchpotch or new foundations? *Le travail Humain ,* 55(2/1991), 171–81.

NORO, K . and IMADA, A ., 1991, *Participatory Ergonomics* (London: Taylor & Francis).

PAHL, G. and BEITZ, W., 1984, *Engineering Design* (London: Design Council).

PERROW, C., 1984, *Normal accidents: Living with High-risk Technology* (New York: Basic Books).

REASON, J., 1990, *Human error* (Cambridge: Cambridge University Press).

SANDERSON, P.M., 1989, The human planning and scheduling role in advanced manufacturing systems: an emerging human factors domain. *Human Factors*, 31(6), 635–66.

SCHÖN, D.A., 1983, *The reflexive practitioner: how professionals think in action* (New York: Basic Books).

TERSSAC (de) G., 1992, *Autonomie dans le travail* (Paris: PUF).

TRINQUET, P., 1996, *Maîtriser les risques du travail: Le cas du bâtiment et de l'industrie française* (Paris: PUF).

WILSON, J., 1991, Participation: a framework and a foundation for ergonomics? *Journal of Occupational Psychology,* 64, 67–80.

WISNER, A., 1994, Situated cognition and action: implication for ergonomics work analysis and anthropotechnology. In *Proceedings of the 12th congress of IEA,* Toronto, August, vol. 1, pp. 1–11.

WISNER, A., DANIELLOU, F., and DEJOURS, C., 1986, Uncertainty and anxiety in continuous process industries. In K. NORO (ed.), *Occupational Health in an Automated Factory* (London: Taylor & Francis).

.

Psychophysiological Methods

W. Boucsein

Physiological Psychology, University of Wuppertal, 42119 Wuppertal, Germany

1. INTRODUCTION

The aim of this chapter is to introduce ergonomists and human factors researchers to the unique benefits of the psychophysiological approach. Although collecting physiological data is not unfamiliar to this field, psychophysiology has more to offer than physiological recording. One of its major goals is the integration of physiological, subjective and performance/behavioral data. Furthermore, psychophysiological theories emphasize the importance of the nervous system and explain how the neural modulation of physiological systems promotes or limits the range of work related behavior. Therefore, researchers and those who apply ergonomics in the field to create better work environments should consider complementing their methodology with relevant physiological measures. In addition, the choice of psychophysiological measures should be based on neurophysiological models providing a rationale for the use of particular measures. The present article furnishes a brief introduction to theory, methodology and application of psychophysiology in the field of ergonomy and human factors. More detailed information is provided in Backs and Boucsein (1999) on "engineering psychophysiology."

1.1. Aims and Methods of Psychophysiology

Psychophysiology studies the relation between psychological processes and physiological activity by *non-invasive recording* (e.g. Andreassi 1995). Psychophysiological techniques provide unobtrusive measures of central nervous system activity taken from the electroencephalogram (EEG), of autonomic nervous system activity such as heart rate (HR), blood pressure and electrodermal activity (EDA), of somatomotor activity such as electromyogram (EMG) or eye movements, of body temperature and of endocrine measures such as catecholamine or cortisol excretion (Table 1). Psychophysiology refers also to a certain approach to the mind–body problem, assuming that physiological response patterns reflect cognitive, emotional, behavioral and social events. This implies modeling of brain–behavior relationships that may help to generate hypotheses on the use of appropriate physiological measures for the psychological phenomenon under investigation (Figure 1). Theoretical concepts that are commonly investigated by means of psychophysiology comprehend arousal and sleep, orienting and habituation, information processing and learning, emotion and motivation, or stress and stress-related diseases. The progress in psychophysiological techniques has been fostered by international societies such as the Society for Psychophysiological Research (SPR) or the Federation of European Psychophysiology Societies.

Physiological recording techniques have been employed by work physiologists to study bodily effects of work place features and environmental factors such as noise or heat. Beyond these, psychophysiological measures have proved to be valuable tools

Table 1. Psychophysiological parameters of different categories of strain. The variables are grouped according to their respective physiological systems (central nervous, cardiorespiratory, electrodermal, somatomotor, thermoregulatory, endocrine). '≠' means that the values of the parameter in question increase with increasing strain, and 'Ø' that they decrease. More evidence is available for double arrows than for single ones. Adapted from Boucsein, and Backs (1999). Used by permission of the publisher.

Physiological measure	Category of strain		
	Physical	Mental	Emotional
EEG alpha activity (8-12 Hz)		↓↓	
EEG theta activity (4-7 Hz)		↑↑	
P3 amplitude		↑↑	
P3 latency		↑	
CNV amplitude		↑	
Heart rate	↑↑	↑	↑
Heart rate variability		↓↓	
Respiration rate	↑	↑	
Finger pulse volume amplitude		↓	↓
Systolic blood pressure	↑↑	↑	
Diastolic blood pressure	↑	↑	
EDR amplitude		↑	↑
EDR recovery time		↑	
NS.EDR frequency		↑	↑↑
Eye blink rate		↑↑	↑
Saccadic eye movement frequency		↑	
Pupil diameter		↑	↑
Electromyographic activity	↑	↑	↑
Muscle tremor	↑↑		
Core temperature	↑↑		
Finger temperature			↑
Epinephrine		↑↑	↑
Norepinephrine	↑↑		↑
Cortisol		↑	↑↑

to gain information about the worker's psyche that complements performance and subjective measures. Recently, the availability of lightweight portable recorders for multiple physiological measures has brought about an important progress for the use of psychophysiology in the field of ergonomics (Fahrenberg and Wientjes 1999). Field settings that make use of *ambulatory monitoring* comprise both work place and leisure time activities. Because of the lack of control in real life settings, field studies should be combined with experiments in the laboratory or in simulated environments. Here, psychological factors that may determine physiological responses can be manipulated in order to establish unambiguous psychophysiological relationships.

In addition to the *physiological* measures, *subjective* and *behavioral measures* constitute a genuine part of the psychophysiological approach. Subjective arousal, well-being, ratings of bodily symptoms and other stress related variables can be obtained discontinuously during working pauses by means of rating scales. As behavioral indicators, various performance measures such as speed, accuracy and amount of work can be
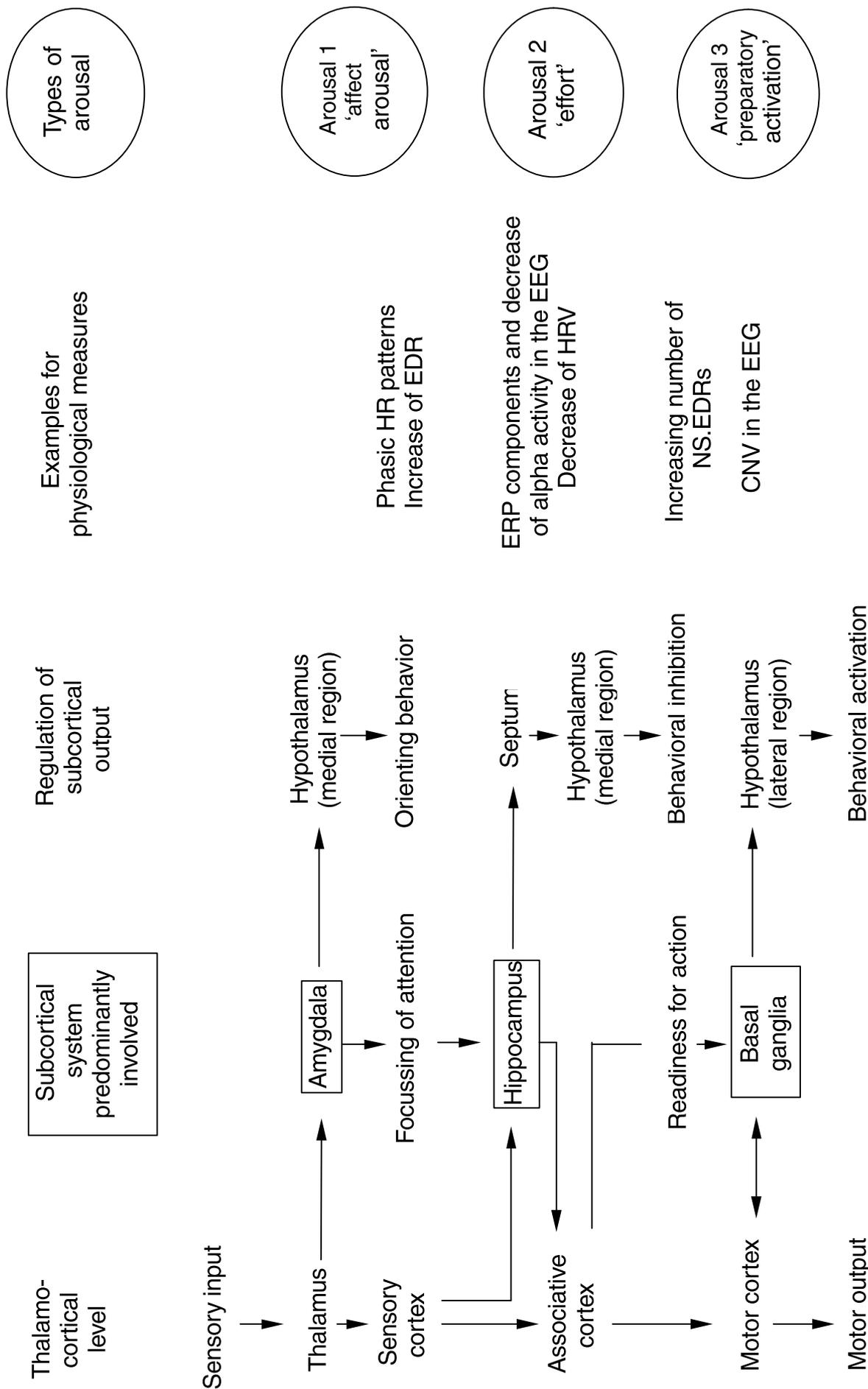
Figure 1. A three-arousal model for the use of psychophysiology in ergonomics. CNV, contingent negative variation; EEG, electroencephalogram; ERP, event related potential; HR, heart rate; HRV, heart rate variability; NS.EDR, non-stimulus specific electrodermal responses. Adapted from Boucsein and Backs (1999). Used by permission of the publisher.

continuously obtained from the task itself. In addition, behavior may be videotaped, including facial expressions and gross body movements. However, the analysis of such data is time consuming and is therefore not often performed by psychopsychologists.

1.2. A Brief History of Psychophysiology in Ergonomics

Early in the 1910s the German psychologist Hugo Münsterberg suggested the use of scientific psychology to improve every day life. Collectively, he referred to the application of psychology to address problems in education medicine, law, business and industry as *psychotechnics*. Its major application started to become military personnel selection and training during World War I.

Consistent with this tradition, psychologists in the USA and UK made valuable contributions to the design of weapons systems and personnel protection equipment during World War 2. Psychologists had been involved in personnel selection and training for many years, but their involvement in ergonomics was something new. Descriptors of these activities that appeared in the USA were terms such as *engineering psychology*, *human factors* engineering or just human engineering, along with the term *ergonomics* in the UK. All these terms are still in use and often used interchangeably. Perhaps the beginning of the use of psychophysiology in ergonomics was in 1939 with a study of psychophysiological "tension" in flight using what was probably the first airborne polygraph. For the next two decades, psychophysiological research in the field was limited by both the technology available and the *Zeitgeist* of this period and tended to emphasize muscular measures during skilled psychomotor tasks, including eye movements. An example is the classical study of Fitts and colleagues of pilot's eye movements during instrument landings in 1950 that helped to determine the arrangement of the basic flight instruments to what is called a standard "T" arrangement that is still generally used today (for more details, see Boucsein and Backs 1999).

In the 1970s the use of psychophysiology in human factors flourished in the USA, in part because of technological advancements in data acquisition and analysis equipment, and in part because of the Department of Defense funds for "biocybernetic" research. In particular, research on EEG measures and the recently discovered event-related brain potentials was supported because of the common belief that physiological correlates of psychological events should be measured as close to the brain as possible. Unfortunately, the applicability of EEG derived measures in the field did not meet the probably unrealistically high expectations. This may have lead to some disillusion with psychophysiological methods in general. As a consequence, psychophysiological research in human factors in the USA waned during the 1980s.

Whereas the use of psychophysiology in ergonomics in the USA continued its traditional focus on defense, especially aviation, applications during the 1970s and 1980s, research in Europe had a different focus. European research during this period was concerned with the psychophysiological effects of increasing workplace automation and adopted a *stress-strain* theoretical approach, unlike the information processing approach of most of the human engineering psychophysiology in the USA. More recently, the ongoing computerization of work places has shifted the attention of European researchers from physical stress to the

issues of mental workload and emotional strain (Gaillard and Kramer 1999). Following this line, psychophysiological research continues to gain influence on the organization of work settings, as well as on the identification and prevention of stress related disorders.

In 1993, an international scientific organization "Psychophysiology in Ergonomics" (PIE) has been formed by the present author and others to promote and advance the understanding of psychophysiological methods and their application to ergonomic environments.

2. METHODOLOGICAL ISSUES

2.1. Theoretical Concepts

Despite the great technical advances in recording and data evaluation that have been made in psychophysiology during the past decades, only limited progress has been made in the development of psychophysiological theory. Very seldom psychophysiologists have made attempts to integrate their measures into theory. Instead, rather simple theoretical frameworks such as the one-dimensional *arousal* concept from the 1950s are still used in the field (Boucsein and Backs 1999). Arousal has, however, proved to be a complex and *multi-dimensional* phenomenon that can be best taken care of by modeling different neurophysiological systems for different kinds of arousal as proposed by Pribram and McGuinness (1975).

Figure 1 depicts a simplified model of the three major kinds of arousal and their neurophysiological origins (for more details, see Boucsein 1992, Figure 48). Arrows indicate the principal information flow within the system.

Arousal System 1 is centered around the amygdala and is labeled the "affect arousal system." If is responsible for focusing attention and for generating hypothalamic reaction patterns. For example, if a novel or unexpected event occurs in the environment, an orienting response is generated. Such a response can be characterized by a phasic HR deceleration followed by an acceleration and/or by an electrodermal response (EDR), sometimes labeled "skin conductance response," due to a phasic increase in sweat gland activity. Attention is shifted towards the new event that may be supported by somatomotor responses such as head- or eye-movements.

Arousal System 2 is centered around the hippocampus and is labeled the "effort system." It has the property to connect or disconnect input and output. The physiological patterns generated by this system can be regarded as concomitants of central information processing. For example, the positivity in the event related potential (ERP) labeled "P3" that peaks in the EEG ~300 ms after presentation of a stimulus increases with the increasing amount of processing capacity assigned to that particular stimulus. Furthermore, diminishing EEG activity in the alpha-band (8–12 Hz) and decreasing heart rate variability (HRV) are regarded as indicators of stimulation or situation that requires greater information processing.

Arousal System 3 is centered around the basal ganglia and is labeled the "preparatory activation system." Its activation results in an increased readiness of brain areas involved in somatomotor actions such as grasping or other body movements. Its physiological concomitants are the contingent negative variation (CNV), an increase in cortical negativity that can be observed before the execution of a motor response, and a tonic increase in HR.

The rather straightforward chain of input–output relationship indicated by arrows in the left part of Figure 1 can be modified by situational contexts such as additional emotional load caused by imminent danger or time pressure. In such a case, the "effort system" disconnects Arousal Systems 1 and 3 to prevent an immediate action and facilitate a deliberate analysis of the situation. This is performed by certain cortical–subcortical brain areas including the so-called Papez Circuit. In such a case, an increase of spontaneous, non-stimulus specific electrodermal responses (NS.EDR) will indicate the emotional load that comes with the situation (Boucsein 1992).

Although far from covering the whole range of psychophysiological relationships in the field of ergonomics and human factors, the model depicted in Figure 1 provides a framework for generating refined hypotheses for the use of psychophysiological measures in the field. Boucsein and Backs (1999) have exemplified this with respect to workload. From a psychological point of view, the issue of workload may be embedded in the broader concept of stress-strain processes. Although not very familiar to most North Americans, the use of the term "strain" for the impact of stress is gradually spreading over from Europe. A common distinction has been made between *physical* and mental workload or strain. Mental load may be further split into *mental* and *emotional strain*. Mental strain in this narrower sense refers to the concept of mental effort, while emotional strain refers to the excess mental effort that comes from anxiety-evoking cognitive aspects of the task such as the imposition of deadlines, worries about one's capacities to perform the job, etc. Table 1 provides hypotheses concerning the sensitivity of physiological measures for the three different kinds of strain as derived from a literature analysis (for more details, see Boucsein and Backs 1999, Tables 2–7).

The rows in Table 1 are grouped according to the various physiological systems. If there is evidence from the literature that a particular physiological measure is associated with a specific kind of strain (i.e. physical, mental or emotional), an upward or downward pointing arrow is inserted in the appropriate column to indicate that the measure increases or decreases with an increase in that kind of strain. Two arrows indicate greater evidence for the relation.

Researchers who want to apply psychophysiology in the field may use the theoretical model depicted in Figure 1 together with the empirical framework provided in Table 1 to determine the *specific validity* of their designated physiological measures. Stimulus- or situation-related emotional strain will mainly affect Arousal System 1 (i.e. "affect arousal"), while mental strain will very likely be a result of an increased activity in Arousal System 2 (i.e. "effort") or in Arousal System 3 (i.e. "preparatory activation"). If emotional load is going to persist, certain measures related to System 2 such as spontaneous EDR will be affected. Thus, if the situation under investigation is thought to exert some emotional load on the subject, care should be taken to include measures that reflect both the "affect arousal" and the "effort" system. If the physiological measures are chosen accordingly, the sensitivity and diagnosticity of the psychophysiological approach to ergonomics can be considerably improved. Furthermore, hypotheses about what kind of physiological changes might be present in the work setting can be formed more accurately.

2.2. Psychophysiological Recording and Evaluation

The majority of psychophysiological responses are measured by means of sensors placed on the body surface. Some oculomotoric measures can be recorded by video monitoring. Endocrine measures are obtained by collecting samples of body fluids such as urine or saliva. Recorded signals may be distorted by noise and other *artifacts* resulting from electromagnetic fields, body movements, etc. in particular under real life conditions. Signal amplification, filtering processes and analog–digital conversion (e.g. sampling rate) may exert additional influences on the recorded signal (Luczak and Göbel 1999). Therefore, any application of psychophysiological methods requires training in both methodology and evaluation of data.

As a rule of thumb, sampling frequencies of > 100 Hz are required for each physiological measure. Hence, a huge amount of data will result from a long-term recording that is typically applied in the field. Pre-processing during recording can, however, considerably reduce the amount of data to be collected (e.g. recording HR instead of the raw electrocardiogram). Averaging over a certain period of time will further reduce the amount of data. However, valuable information can be lost during such a procedure that may be needed for further evaluations. For example, the calculation of HRV requires the length of each interbeat interval. Those would have been lost during an early averaging procedure.

Most physiological measures require *parametrization* before being ready for statistical evaluation. In some cases, standard procedures such as frequency analysis (e.g. fast Fourier transformation) or averaging over a certain time span may be sufficient. However, it is more likely that custom software has to be applied, e.g. for obtaining certain parameters of an EDR from electrodermal recording such as amplitude or recovery time (see Boucsein 1992, Appendix). Integration or low-pass filtering and rectification is required to determine the total muscle tension from the electrical spikes recorded in the surface EMG and an additional high-pass filter may be applied to attenuate motion artifacts (Luczak and Göbel 1999). In any case, data sets should be scrutinized for all kinds of distorted or missing data that requires an interactive evaluation.

Unfortunately, the use of certain psychophysiological measures by a particular researcher or research group is frequently determined by the availability of appropriate equipment and past experience. However, the investigation of complex human organisms in complex environments requires a *multi-level* (i.e. physiological, behavioral and subjective), *multivariate* approach (i.e. recordings from different systems with specific sensitivity such as demonstrated in Table 1).

3. ERGONOMIC APPLICATIONS

The following three sections provide information about selected applications of psychophysiological methods in ergonomics and human factors. A representative selection of studies can be found in two special issues of *Ergonomics* (Gundel and Wilson 1993, Boucsein *et al.* 1998) and in a special issue of *Biological Psychology* (Gaillard *et al.* 1996).

3.1. Psychophysiology and Traditional Work

Both HR and EMG have been successfully used as continuous psychophysiological measures for energy expenditure and physical strain in traditional work. Long lasting cardiovascular and muscular strain without appropriate recovery during rest breaks and/or leisure time after work may result in so-called psychosomatic disorders over the long run. Traditionally, psychophysiological research has primarily focused upon *cardiovascular disorders* such as coronary heart disease (Gaillard and Kramer 1999). Unfortunately, one-dimensional arousal concepts still prevail in this kind of research. As a consequence, successful stress management has almost become identical with the reduction of general arousal and the activity of the sympathetic nervous system. However, some types of arousal may have adverse effects for health while others do not. For example, "effort" and "preparatory activation" are necessary for an adequate job performance. They can be accumulated with low health risk while a large amount of "affect arousal" may constitute a considerable risk. Factors such as decision authority and social support may exert their stress-reducing properties by preventing excess "affect arousal." Psychophysiological indices of orienting and defensive behavior seen in cardiovascular and electrodermal patterns can be used to determine the effectiveness of stress-reducing factors (Figure 1). Defensive responses being temporarily evoked by an increased mental workload are characterized by increases in HR and blood pressure and a decrease in HRV and blood pressure variability (Mulder *et al.* 1999). The decrease of finger pulse volume has proved to be a sensitive measure of defensiveness for laboratory studies and may be applied in that context as well.

The development of *musculoskeletal disorders* is also an important issue in long-term consequences of stress during traditional work. Lack of relaxation of certain muscle groups may constitute a health risk in both repetitive muscular work and in modern computer work (see section 3.2) and can be continuously monitored by means of EMG. Muscle tremor constitutes a specific indicator of physical strain and fatigue. Highly demanding and repetitive work (such as performed by cash register operators) emerges a combination of physical and mental strain. Elevated systolic and diastolic blood pressure, increased trapezius muscle tension as well as epinephrine and norepinephrine levels constitute the typical psychophysiological pattern during and after this kind of work. A high percentage of such workers develops neck–shoulder problems in the long run.

Lundberg and Johansson (1999) summarized the effects of *assembly work* and other types of repetitive work on psychophysiological stress indicators. Traditional assembly work is associated with physical constraint, machine pacing, piece-rate remuneration and often with under utilization of human resources. A comparison of such repetitive work at an assembly line in a car factory with a more flexible type of assembly work in autonomous groups (six to eight people who had a considerable amount of freedom and variety in their job including responsibility for work organization and quality control) yielded about the same mean urine epinephrine and norepinephrine levels in both groups. However, psychophysiological stress responses during work in the assembly line showed an increase that was not seen in the more flexible type of group work. The autonomous group kept a moderate and relatively stable levels of epinephrine and norepinephrine throughout the shift. In addition, their urine

epinephrine levels were significantly lower after the shift as compared with the traditional assembly line workers. The inability to recover from psychophysiological strain after work in the assembly group was even more pronounced for female than for male workers. This may be partly due to the multiple roles of women that do not allow for recovery after work.

The amount of cortisol that can be reliably obtained from saliva probes now is not merely an indicator of short-term emotional strain (Table 1). Moreover, cortisol excretion constitutes a major endocrine indicator of long term stress. This relation has been demonstrated in highly demanding jobs such as air traffic control.

3.2. Psychophysiology and Automated Work

Major changes in various work places occurred during the last three or four decades as a consequence of increased automation. Supervisory monitoring without considerable physical load constitutes most of the work in process industries such as chemical plants, power stations, etc. In particular, the shift from man–machine interaction to a mere man–computer interaction has diminished physical workload but increased mental workload. *Office automation* is probably the most obvious example of pervasive spread of human–computer interaction (HCI). However, the transition process from industrial to information-age office designs has been slow and incomplete in most large institutions, since elements of computer technology were superimposed on relatively unchanged work structures. Work has become more mentally challenging in some kinds of jobs, whereas boredom and dissatisfaction prevail where the human role has been diminished to mere monitoring or data entry tasks.

Psychophysiology has been successfully used to determine the amount of mental workload, to organize an optimal workflow and to establish appropriate schedules for rest breaks during HCI (for a summary, see Boucsein 1999). Elevated epinephrine levels have been found during and after work in administrative workers with a high amount of HCI compared with those performing mainly secretary work. Increases in EEG beta activity (> 13 Hz) and decreases in HRV were observed as consequences of increased task complexity and informational load. Inadequate temporal structures of HCI such as delays in the work flow caused by *system response times* have been a major source of stress since the introduction of time-sharing computer systems in the late 1960s. The adverse effects of involuntary delays in HCI still persist even though computers have become incredibly faster. Hardware speed is often jeopardized by bulky software packages, powerful input and output devices (e.g. graphic tablets), network functions (both intra- and internet) and huge data banks with simultaneous access for many users. Furthermore, the possibility of multi-tasking provided by Windows-based systems raises the additional problem of adequate work flow scheduling with respect to the different durations of tasks being performed in the background.

Psychophysiological investigations have stressed the need for an *optimal scheduling* of the temporal workflow in HCI. Cardiorespiratory measures such as HR, blood pressure and respiration rate show an increase if scheduling becomes too tight, together with an increase in working speed, higher error rates and task related bodily symptoms. On the other hand, if non-optimal scheduling results in temporal delays, an increase in EDA indicates emotional tension (Boucsein 1999). The adverse

influence of both kinds of suboptimal scheduling accumulate if additional cognitive demands such as time pressure are imposed on the operator. The increase of cardiorespiratory activity in the case of a too-tight scheduling indicates an overstressed "preparatory activation" system. On the other hand, the higher amount of NS.EDR found during involuntary temporal delays points to an increased effort that can not result in behavioral activation. Because behavior is inhibited, emotional strain is experienced (Figure 1).

Optimal scheduling is also important for *rest breaks* during work. As in any kind of repetitive work, time is required for the recovery from psychophysiological stress such as muscle tension and elevated HR that may accumulate during data entry tasks. Appropriate break schedules provide the opportunity for such a recovery and range from microbreaks (i.e. several seconds after several minutes of work) to 10 min every hour. However, such a predetermined work/rest schedule may not be suitable for more complex HCI tasks. In a field study performed in a rather complex computerized environment, Boucsein and Thum (1997) found an advantage for more frequently interspersed short breaks prior to the early afternoon as indicated by an increase in HRV and diminishing EDA during the break. Conversely, long rest breaks after a longer working period yielded a better recovery in the afternoon as seen in the same psychophysiological measures. Neck-EMG showed a considerable increase during breakdowns of the computer system as opposed to a decrease during regular rest breaks.

Psychophysiological measures can also be used for triggering mode changes in *adaptive automation* systems. In contrast to operator performance, that may be used as criterion as well, they do not require overt behavior and can be obtained continuously, thus allowing for a rapid response to changes in the operator's state. Both simple threshold-level indicators based on EEG, EOG and head movements, and more complex systems based on multivariate analysis of several physiological measures have been proposed for the use in process control and in long haul transport (see Section 3.3). Power band ratios taken from the EEG such as the operations beta/(alpha + theta) index have been successfully used for determining changes in task engagement in flight management simulation. Autonomic nervous system parameters such as HRV may indicate the operator's engagement in voluntary or compensatory effort. In addition, the absence of NS.EDR for several minutes has been used as a trigger for an auditory alarm in order to regain alertness (for an overview, see Byrne and Parasuraman 1996). However, this technique has not yet moved out of the laboratory. Both complex arousal theories and human information processing resource theories are used as theoretical framework within this area.

3.3. Psychophysiology and Extended Work

Working at odd hours, especially during night and after a prolonged period of wakefulness, is very common in our modern society. *Night shiftwork* or the operation of extended transmeridian flights are typical examples for situations in which a co-variation of performance decrement and increased sleepiness is observed. Sleepiness occurs as a consequence of several factors: the displacement of work to the circadian phase, an extended duration of wakefulness and the loss of sleep (Akerstedt 1995). In addition, work-induced fatigue, vigilance decrement and monotony develop with extended time-on-task.

The amount of sleepiness is commonly determined on different levels: by subjective estimates such as visual analogue scales and by using objective measures for the sleep tendency. The latter may be measured during repeated controlled opportunities to fall asleep such as the Multiple Sleep Latency Test. As physiological measures, both oculomotoric (Sirevaag and Stern 1999) and EEG recordings (Gundel *et al*. 1999) can be used to determine the amount of sleepiness or fatigue during work at odd hours such as night shiftwork or *long-haul transport operations*. During sleepiness, eye blink frequency and duration increase as well as the amount of slow eye movements. Alpha-band activity dominates in the EEG while phases of micro-sleep are characterized by the desynchronization or disappearance of alpha activity and the occurrence of theta waves (4–8 Hz). Eye blink rate as well as eye closure duration increase with time-on-task. Both frequency and velocity of saccadic eye movements decline during extended periods of work.

A multivariate study with seven oculomotoric measures performed by Morris and Miller (1996) revealed that 36% of the performance decrement in partially sleep deprived pilots during a 4.5-h flight simulator mission could be predicted by a combination of blink amplitude, blink closure duration and long closure duration blink rate. This is of particular interest for long-term monitoring because oculomotoric measures can be obtained by using unobtrusive techniques such as video monitoring. Pupil diameter can be also used for monitoring fatigue. Large and stable pupils are associated with alertness, while pupil dilatation indicates mental/emotional challenge or expectancy. Pupil constriction and pupillary oscillations are associated with sleepiness but appear to be impervious to time-on-task effects (Sirevaag and Stern 1999).

The neurophysiological mechanism underlying sleepiness is relatively well known. A pacemaker in the suprachiasmatic nuclei of the hypothalamus determines the *circadian rhythm*, bringing about a minimum of sleepiness in the late afternoon and a maximum in the early morning (~5:00 h). Body core temperature follows the same circadian pattern and is therefore frequently used as a marker for changes in the circadian rhythm. Melatonin shows the reverse pattern and seems closely related to both sleepiness and body temperature (Akerstedt 1995). The production of melatonin is triggered by light exposure, especially during the afternoon. Both sleep loss and work-induced fatigue exert their arousal-reducing influence on the "effort" system (Figure 1). Focusing of attention is considerably reduced and EEG alpha and theta activity are facilitated. As a sign of fatigue, HRV is increased. The breakdown of the "effort" system corresponds to what cognitive psychology has called the failure of an "upper" mechanism in monitoring and altering the parameters of a "lower" mechanism (Mulder *et al*. 1999). As a consequence of this failure, automated processes performed by the "lower" mechanism are subjected to deterioration by fatigue. Performance may be still sufficient which is, however, achieved by additional energy expenditure. This is shown by an increase in cardiovascular activity, EMG activity and epinephrine excretion.

Various situational and personal factors may contribute to the ability to stay alert during prolonged operational hours. Furthermore, countermeasures may be taken such as allowing for prophylactic naps, the application of caffeine, or the improvement of sleep quality by means of hypnotics in situations

where sleep is impaired by situational factors. The effectiveness of such countermeasures for minimizing the impact of sleep loss can be determined by the combination of subjective, behavioral and physiological measures.

4. CONCLUSIONS

Psychophysiology provides a unique approach to study the human being not only in the laboratory or in simulated environments but also at real work places. The ongoing micro-computerization and substantially reduced costs of commercially available data acquisition and analysis equipment facilitate the application of psychophysiological measures in the field. Two major goals can be achieved by applying the psychophysiological approach in the field of ergonomics: to create better work places and to prevent the development of work-related diseases.

Physiological measures can be recorded continuously without interrupting the workflow. They provide information that is not readily available from performance measures and subjective ratings. They have the advantage not being subjected to faking as psychological scales are very often. However, several bio-electrical and biochemical measures mentioned in table 1 are not easy to obtain at real work places. For example, the use of EEG parameters can be limited by recording artifacts in the field. Blood samples of epinephrine and norepinephrine require special medical care and compliance. It is generally recommended to test psychophysiological measures in a simulated environment first, before they are going to be applied in the field. Such a combined laboratory-field approach can be easily performed for most automated work places where the man–machine interaction makes use of a computer system.

However, a psychophysiological approach not only refers to the use of physiological measures in conjunction with behavioral and subjective data. It is also highly desirable to ascribe to a neurophysiological model such as depicted in Figure 1 as a theoretical background. This not only provides a better understanding of what is recorded by the various measures, it also avoids being trapped in oversimplified hypotheses derived from one-dimensional arousal or cognitive theories assuming only sequential information processing.

REFERENCES

AKERSTEDT, T., 1995, Work hours, sleepiness and the underlying mechanisms. *Journal of Sleep Research*, 4 (suppl. 2), 15–22.

ANDREASSI, J. L., 1995, *Psychophysiology: Human Behavior and Physiological Response* (3rd ed.) (Hillsdale: Lawrence Erlbaum Associates).

BACKS, R.W. and BOUCSEIN, W. (eds), 1999, *Engineering Psychophysiology* (Mahwah: Lawrence Erlbaum Associates).

BOUCSEIN, W., 1992, *Electrodermal Activity* (New York: Plenum).

BOUCSEIN, W., 1999, The use of psychophysiology for evaluating stress-strain processes in human computer interaction. In R.W. Backs and W. Boucsein (eds) *Engineering Psychophysiology* (Mahwah: Lawrence Erlbaum Associates), 231–52.

BOUCSEIN, W. and BACKS, R.W., 1999, Engineering psychophysiology as a discipline: historical and theoretical aspects. In R.W. Backs and W. Boucsein (eds) *Engineering Psychophysiology* (Mahwah: Lawrence Erlbaum Associates), 8–25.

BOUCSEIN, W. and THUM, M., 1997, Design of work/rest schedules for computer work based on psychophysiological recovery measures. *International Journal of Industrial Ergonomics*, 20, 51–7.

BOUCSEIN, W., HENNING, R.A. and PORGES, S.W., 1998, "Psychophysiology in ergonomics." *Ergonomics*, 41(5).

BROUGHTON, R.J., 1992, Qualitatively different states of sleepiness. In R.J. Broughton and Ogilvie (eds) *Sleep, Arousal, and Performance* (Boston: Birkhäuser), 45–59.

BYRNE, E. and PARASURAMAN, R., 1996, Psychophysiology and adaptive automation. *Biological Psychology*, 42, 249–68.

FAHRENBERG, J. and WIENTJES, C.J.E., 1999, Recording methods in applied environments. In R.W. Backs and W. Boucsein (eds) *Engineering Psychophysiology* (Mahwah: Lawrence Erlbaum Associates), 71–95.

GAILLARD, A.W.K. and Kramer, A.F., 1999, Theoretical and methodological issues in psychophysiological research. In R.W. Backs and W. Boucsein (eds) *Engineering Psychophysiology* (Mahwah: Lawrence Erlbaum Associates), 26–40.

GAILLARD, A.W.K., BOUCSEIN, W. and STERN, J. (eds), 1996, "Psychophysiology of workload." *Biological Psychology*, 42(3).

GUNDEL, A. and WILSON, G.F., 1993, "Psychophysiological measures in transport operations." *Ergonomics*, 36(9).

GUNDEL, A., DRESCHER, J. and TUROWSKI, J., 1999, Alertness in airline pilots during night flights: assessment of alertness using EEG measures. In R.W. Backs and W. Boucsein (eds) *Engineering Psychophysiology* (Mahwah: Lawrence Erlbaum Associates), 130–43.

LUCZAK, H. and GÖBEL, M., 1999, Signal processing and analysis in application. In R.W. Backs and W. Boucsein (eds) *Engineering Psychophysiology* (Mahwah: Lawrence Erlbaum Associates), 55–70.

LUNDBERG, U. and JOHANSSON, G., 1999, Stress and health risks in repetitive work and supervisory monitoring work. In R.W. Backs and W. Boucsein (eds) *Engineering Psychophysiology* (Mahwah: Lawrence Erlbaum Associates), 271–93.

MORRIS, T. L. and MILLER, J.C., 1996, Electrooculographic and performance indices of fatigue during simulated flight. *Biological Psychology*, 42, 343–60.

MULDER, G., MULDER, L.J.M., MEIJMAN, T.F., VELDMAN, J.B.P. and VAN ROON, A.M., 1999, A psychophysiological approach to working conditions. In R.W. Backs and W. Boucsein (eds) *Engineering Psychophysiology* (Mahwah: Lawrence Erlbaum Associates), 101–25.

PRIBRAM, K.H. and MCGUINNESS, D., 1975, Arousal, activation, and effort in the control of attention. *Psychological Review*, 82, 116–49.

SIREVAAG, E.J. and STERN, J.A., 1999, Application of oculomotor measures to fatigue and cognitive factors. In R.W. Backs and W. Boucsein (eds) *Engineering Psychophysiology* (Mahwah: Lawrence Erlbaum Associates), 220–30.

Scientific Management Influences on Ergonomic Analysis Techniques

P. G. Dempsey

Liberty Mutual Research Center for Safety & Health, Hopkinton, MA 01748, USA

1. INTRODUCTION

Ergonomics is a discipline that "borrows" from several related branches of science including, but not limited to, psychology, anatomy, physiology and engineering. Activities occurring around the time of World War II have been traditionally described as the beginning of ergonomics (Europe) and human factors engineering (USA). Comparatively, ergonomics is quite young if one uses these chronological landmarks; however, there is a fairly rich body of literature going back to even the previous century that clearly describes investigations of manual work with the purpose of enhancing human performance.

The industrial revolution had a dramatic effect on the design, analysis and management of manual work. Adam Smith's principles of the "Division of Labor" in *An Inquiry into the Causes and Nature of the Wealth of Nations* epitomized the new philosophy towards designing manual work required by the developing industrial base. Previously, craftsmen would be responsible for producing an entire product. Under the division of labor, an individual would be responsible for only a small contribution to a product, repetitively performing a narrow task or small set of tasks throughout the day. The division of labor was not as simple as arbitrarily dividing the work into equal units, but became a studied process with the goal of optimizing production. This epic change in work design required new approaches to the analysis and management of work. These new approaches provided direction to the design of work and eventually provided a feedback loop for existing designs. New products and a desire continually to increase production efficiency created the need for specialists, early on called "efficiency experts," to perform these functions. The Scientific Management movement formed the core of the early efficiency experts in the USA.

Today, ergonomic analysis of manual work plays an important role in the design and re-design of manual work for the purpose of enhancing human performance. The aspects of human performance considered by modern ergonomists extend beyond those of the early efficiency experts, but these differences may be explained to a considerable extent by the advancement in knowledge of the human responses to work. Different systems of compensation and accommodation for disabled workers and even different social values may considerably affect the focus of work analyses. Operationally, the analysis approaches used in the scientific management of work early this century are fundamentally the same as some techniques used today by ergonomists to study manual work.

An understanding of the evolution of ergonomics techniques provides more than a pedagogical history lesson. It provides insight into the goals and philosophic reasoning for analyzing work the way we do. We can then use this insight to support or question the appropriateness of the tools that we use today and develop in the future for analyzing manual work. This entry provides an overview of the contributions of Scientific Management to ergonomics.

2. SCIENTIFIC MANAGEMENT

The industrial revolution in the UK was followed by an industrial revolution in the USA. By the beginning of the 20th century, there was a substantial industrial base in the USA, particularly in the north-eastern area of the country. At that time, the Scientific Management movement was beginning to foster. The works of Frederick W. Taylor and Frank and Lillian Gilbreth formed the essence of this movement. There was a strong focus on maximizing productivity of workers engaged in industrial production, although Scientific Management would eventually find its way to non-industrial forms of work such as the surgical operating room. Similar lines of work were also being pursued in Europe and elsewhere. For an excellent example and references of work in Europe, see Amar (1920) which may also be the first ergonomics textbook.

Scientific management was an approach to optimizing the output of a facility, and was applied to almost every aspect of a production facility including accounting, line management, facilities design, as well as the production floor. The discussion here will be restricted to activities involving the design and analysis of manual work.

Two significant components of Scientific Management were time study and motion study, with the former more closely associated with Taylor and the latter with the Gilbreths. Time study predated Taylor, and was closely associated with the industrial revolution (for examples of early time studies, see Babbage 1835). Time study was used to determine work standards for task elements, whereas motion study examined task elements with the goal of optimizing efficiency through workplace design and the elimination of wasteful motion. Ideally, both techniques are used to analyze work. It should be noted that the distinction between time study and motion study was not always made, and that confusion is still made to this day.

The work of Taylor and the Gilbreths (1909) went far beyond time study and motion study, and included a systemic view of work design. As will be illustrated below, the primary goal of Scientific Management investigations of manual work focused on the effect of specific aspects of the work system design on fatigue and output. The Gilbreths' work, in particular, was focused on enhancing productivity through fatigue elimination.

Taylor (1911) described experiments to determine the optimal shovel sizes for shoveling loads of materials with varying densities (e.g. rice coal versus iron ore). The result was the provision of eight-to-ten different shovels for use at the Bethlehem Steel Company. Taylor (1911) summarized the effort as follows (comparing shoveling rice coal to shoveling iron ore): "In the one case, he was so underloaded that it was impossible for him to do a full day's work, and in the other case he was so ridiculously overloaded that it was manifestly impossible to even approximate a day's work" (p. 67 [page number refers to the 1967 edition.]) Taylor (1911) understood that either underloading or overloading the muscular system is inefficient, and focused on hand tool design in this case. The experiment provided early empirical evidence for "providing the right tool for the job," which

is a contemporary hand tool guideline. Although the goal of the experiments was to maximize tonnage shoveled per day, it appears as if Taylor (1911) at least superficially recognized the physiological responses to muscle loading, even if the recognition was a result of output concerns.

From a modern ergonomics perspective, the pig iron experiments described by Taylor (1911) have the perception of being the most "barbaric." The goal of the experiments was to maximize the weight of pig iron loaded onto rail cars. In this case, the temporal design of the work (i.e. work and rest schedules) was deemed the critical system component to manipulate. The main result of the study was the development of an equation to calculate tonnages handled. The acceptable daily tonnages were a function of distance carried (task variable), with the restriction on the grade of the plank going to the rail car (workplace variables) when applying the results. It was recognized that the limits would have to be reduced under high ambient temperatures (environment variable), and that only a small percentage of workers was capable of performing the job (administrative control). Modern ergonomic approaches to manual handling task analysis, such as psychophysical tables and equations used to predict acceptable loads, use the same approach. The difference is that contemporary methods provide loads aimed at protecting the majority of the population from injury, not maximizing the productivity of a minority of the population.

The work of the Gilbreths (1909) is even more compatible with contemporary ergonomics. Gilbreth's (1911) approach to analyzing manual work represents a modern systemic view of ergonomics: "A careful study of the anatomy of the worker will enable one to adapt his work, surroundings, equipment, and tools to him. This will decrease the number of motions he must make, and make the necessary motions shorter and less fatiguing" (p. 10). Gilbreth's (1909) study of bricklaying serves as a fine example of the ergonomic benefits of early motion studies. The results of the job design changes clearly reduced postural loading of the low back. Perhaps more important is the detailed manner in which Gilbreth broke down job elements to perform a task analysis. For instance, the act of laying one brick the "wrong way" was broken down into 18 task elements. The primary goal of the analysis was to enhance efficiency with a strong emphasis on reducing needless motion and fatigue.

These examples are by no means exhaustive, but illustrate the systemic approach to work design used by the founders of the Scientific Management movement. The approach typically involved studying tasks, or task elements, with the goal of modifying task, environment, machine and tool variables to enhance productivity, often times through the reduction of fatigue. The approach differed vastly from ergonomics in that worker selection was used extensively, whereas contemporary ergonomics promotes designing work that can be performed safely by the majority of the population, eliminating the need for selection. It should be noted that worker selection was not unique to the Scientific Management movement. Münsterberg's (1913) extensive use of psychological tests to select workers for various occupations provides an excellent example of the pervasive use of such testing, as well as a nice example of early experimental psychology research that certainly falls into what today would be considered in the realm of ergonomics.

3. TASK ANALYTIC APPROACHES TO ANALYZING MANUAL WORK

Task analysis, one of the most widely used techniques in ergonomics, had perhaps the most substantial influence of the Scientific Management techniques on modern ergonomics practice. Task analysis is a robust and powerful tool that enables ergonomists to study manual work (and many other activities) in a systematic manner. The components of task analysis include (Stammers and Shepherd 1995):
- data collection stage;
- task description stage; and
- analysis stage.

The task description contains the task requirements whereas the task analysis is the actual *analysis* of the behavioral implications of the task identified in the description (Stammers and Shepherd 1995). The analysis used in scientific management tended to focus on efficiency of activities (e.g. reduction of wasteful motion) and elimination of fatigue. Today, the analysis phase may focus on a broad range of measures that influence human performance, depending on the task and its context.

The origins of task analysis often include mention of the landmark report by Miller (1953). The purpose of the methodology reported was to provide a framework for specifying training requirements. The examples provided by Miller (1953) show the focus on breaking work into small units, just as was done by Taylor and the Gilbreths. Miller's (1953) methodology was more involved and far more advanced from the standpoint of comparing human capacity to task demands, but the general approach to describing work tasks is similar. The notion of performing a task analysis for delineating training requirements is functionally identical to Taylor's instruction cards for industrial tasks. In some regards, the task description methods form the core commonalties across time, and the analysis progressed and evolved as the knowledge of the human responses to work evolved.

4. DISCUSSION

Today, many ergonomic tools incorporate the division of manual work into tasks or task elements. This division provides a structured means of determining which tasks or task elements may result in task demands exceeding human capabilities. The control of work-related musculoskeletal disorders of the upper extremity and low back is currently a priority in many industrial organizations. Numerous analysis techniques used are indeed task analytic approaches, many of which are founded in fundamental time study and motion study techniques.

During the Scientific Management era, the task analytic approach was widely used as a means of enhancing productivity through the redesign of the workplace or the elimination of "wasteful" elements. The reduction of fatigue was a primary goal, with the assumption that fatigue was a detriment to productivity. The analysis was focused on the influence of specific aspects of the work system on efficiency and fatigue. However, today the primary goal of ergonomic analyses of manual work appears to be the prevention of disability, including the reduction of incidences and absenteeism. This shift in philosophy may be due to several factors, including better recognition of the deleterious effects of manual work on the human body, an increase in the

level of safety and health at work that the public finds acceptable, and a better understanding of the effect of workplace injuries and illnesses on organizational effectiveness and economic competitiveness.

Regardless of the reasons for the change in outcome measures focused on, we should consider the consequences of retaining essentially the same methodologies while focusing on different outcome measures. Often times, factors such as motion and productivity are easily measured, and the influence of changing a task parameter can usually be readily observed. Particularly in the case of productivity, the effect of design changes can be measured by throughput. However, disability and absenteeism are tremendously more complex, and the influence of task parameters on these measures is not readily observed or immediately realized. In fact, our knowledge of these causal relationships is lacking in many areas, in no small part because of the difficulty defining and measuring them for empirical study. As we look forward, it should remember that understanding the determinants of disability and absenteeism may require more complex analyses at the job, organizational and perhaps even societal level.

REFERENCES

AMAR, J., 1920, *The Human Motor* (London: George Routledge & Sons).

BABBAGE, C., 1835, *On the Economy of Machinery and Manufactures* (London: Charles Knight).

GILBRETH, F.B., 1909, *Bricklaying System* (New York: Myron C. Clark Co.).

GILBRETH, F.B., 1911, *Motion Study* (New York: Van Nostrand Co.).

MILLER, R.B., 1953, *A Method for Man–Machine Task Analysis*. WADC Technical Report 53-137 (Patterson: Wright-Patterson AFB, Wright Air Development Center).

MÜNSTERBERG, H., 1913, *Psychology and Industrial Efficiency* (Cambridge, MA: Riverside).

STAMMERS, R.B. and SHEPHERD, A., 1995, Task analysis. In Wilson, J.R. and Corlett, E.N. (eds) *Evaluation of Human Work: A Practical Ergonomics Methodology* (2nd ed.) (London: Taylor & Francis), pp. 144–68.

TAYLOR, F.W., 1911, *The Principles of Scientific Management* (New York: Harper & Bros; repr. W.W. Norton & Co., 1967).

Survey Design

P. Carayon

Ecole des Mines de Nancy, 54042 Nancy Cedex , France
and University of Wisconsin-Madison, Department of Industrial
Engineering, Madison, WI 53706, USA

P. Hoonakker

P2 Research and Advice, 3572 RK Utrecht, The Netherlands

1. INTRODUCTION

Surveys or questionnaires are used to collect information from people concerning a range of ergonomics or human factors variables (Salvendy and Carayon 1997). Questionnaires can be used to evaluate human–machine system performance, e.g. usability evaluations of products or human–computer interfaces, and to evaluate demands on people, e.g. workload, fatigue, stress, and attitudes. They can also be a useful tool in implementing ergonomic programs as a diagnostic tool and in the evaluation phase. They allow the researcher to obtain large amounts of data from large numbers of people at relatively low cost and relatively quickly (Sinclair 1995). Particular attention should be paid to the development of questionnaires. The methods used to develop, implement, and use questionnaires are very important for the quality and usefulness of the data collected. Before developing a questionnaire survey, it is important to clearly specify the objective of the questionnaire: what will the questionnaire be used for? Figure 1 summarizes the issues related to survey design discussed in this article. Two issues related to measurement are highlighted: (1) objectivity/subjectivity, and (2) reliability/validity. Issues in the process of developing and conducting a survey are also discussed.

2. OBJECTIVITY/SUBJECTIVITY

Questionnaires are sometimes perceived as non-scientific because of their subjectivity. It is important to recognize that any data can be placed on a continuum of objectivity–subjectivity. Objectivity has multiple meanings and levels in the literature. According to Kasl (1987), objective data is one not supplied by the selfsame respondent who is also describing his distress, strain,

```
OBJECTIVE                         SUBJECTIVE
low in dependency         high in dependency
on cognitive and             on cognitive and
emotional processing     emotional processing
```

Figure 1: Issues in survey design
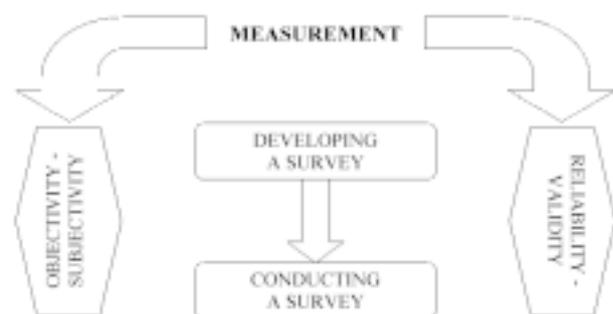
Figure 2: Objectivity/subjectivity continuum

or discomfort. This embraces the outsider perspective. With regard to measurement of psychosocial work factors, Kasl (1987) feels that measures of psychosocial factors can be less subjective when the main source of information is the employee, but that this self-reported exposure is devoid of evaluation and reaction. Similarly, Frese and Zapf (1988) conceptualize and operationalize "objective stressors" as not being influenced by an individual's cognitive and emotional processing. Therefore, it is more appropriate to conceptualize a continuum of objectivity and subjectivity (see Figure 2). The level of objectivity/subjectivity of a measure will then depend on the degree of influence of cognitive and emotional processing. For example, ratings of work factors by an observer cannot be considered as purely objective because of the potential influence of the observer's cognitive and emotional processing. However, ratings of work factors by an outside observer can be considered as more objective than an evaluative question answered by an employee about his/her work environment (e.g. "how stressful is your work environment?"). On the other hand, ratings of the same object by two or more experienced observers do not always achieve a good level of inter-rater reliability.

It is important to recognize that the wording of a question will determine somewhat the degree of objectivity–subjectivity. For instance, self-reported measures of psychosocial work factors can be more objective when devoid of evaluation and reaction. Any kind of data can be placed somewhere on this continuum from "low in dependency on cognitive and emotional processing" (e.g. objective) to "high in dependency on cognitive and emotional processing" (e.g. subjective). Some degree of subjectivity is also involved in so-called objective, quantitative measures. For instance, physical measurement involves decisions regarding, for instance, the timing and frequency of measures: these decisions are made by human observers, and therefore involve some degree of subjectivity. Questionnaires tend to be on the right side of the continuum; however, the degree of subjectivity/objectivity depends highly on the content and form of the questions.

3. DEVELOPING A SURVEY

The development of a questionnaire survey is important to ensure that the data collected is useful and of high quality. Fowler (1995) lists seven principles for designing good survey instruments:

1. The strength of survey research is asking people about their firsthand experiences: what they have done, their current situations, their feelings, and perceptions.
2. Ask one question at a time.
3. A survey question should be worded so that every respondent is answering the same question.
4. If a survey is to be interviewer-administered, wording of the questions must constitute a complete and adequate script such that, when interviewers read the question as worded, respondents will be fully prepared to answer the question.
5. Clearly communicate to all respondents the kind of answer that constitutes an adequate answer to a question.
6. Design survey instruments to make the task of reading questions, following instructions, and recording answers as easy as possible for interviewers and respondents.
7. Measurement will be better to the extent that people answering questions are oriented to the task in a consistent way.

The process of developing a questionnaire survey involves the following steps.

- Step 1 – Conceptualization
- Step 2 – Operationalization
- Step 3 – Sources of questionnaire
- Step 4 – Constructing the questionnaire
- Step 5 – Pre-testing of the questionnaire

3.1 Step 1

Before engaging in the process of writing questions or putting questions together, one should define precisely the concepts to be measured. This process of conceptualization is especially important when dealing with complex phenomena. Concepts often take the form of variables with a collection of related variables. For instance, the concept of musculoskeletal discomfort can be measured by the level of pain involved in various body parts: the attributes of the variable "pain experienced in various body parts" could represent different levels of pain, e.g. low–medium–high.

3.2 Step 2

The step of operationalization has been defined as the process by which one specifies empirical observations that can be taken as indicators of the attributes contained within a given concept (Babbie 1990). The operationalization process starts with a listing of the different sub-dimensions of the variable. At this stage, it is important to review previous research and work on the topic. Then, one should decide all the things that the variable represents, as well as what the variable does not represent. This avoids creating overlap between the variable of interest and some other variables. It is also important to pay special attention to the opposite of the variable.

3.3 Step 3

Before developing new questions, one should examine existing questionnaires which have been validated and used in previous studies. At this step, an extensive review of the literature is necessary. When reviewing existing questionnaires, one should examine the content of the questionnaire (what is measured?), as well as the context for which the questionnaire was developed. The content of the questionnaire should match the conceptualization and operationalization of the variables under study. It is also important to examine the context for which the questionnaire was intended. Using standardized questionnaires can be useful when one is interested in comparing data of a particular sample to already-collected data on a range of jobs, occupations, and organizations. However, whereas standardized questionnaires are interesting when one wants to make comparisons, they may not fit the objective of the questionnaire survey. In this case, the researcher needs to develop a tailor-made questionnaire.

3.4 Step 4

The construction of questionnaire items is a delicate process that requires experience and knowledge. The past experience of survey researchers and designers provides a wealth of guidelines useful in the writing of good questions and the generation of data useful to analysis. Converse and Presser (1986) provide guidelines for writing questions. First, one should decide on the type of questions:

- Rating scales: degree of agreement–disagreement, degree of satisfaction–dissatisfaction, frequency, quantity. Examples: "All in all, how satisfied would you say you are with your job: very much satisfied, somewhat satisfied, not too satisfied, not at all satisfied?"
- Semantic differential scales: simultaneous ratings across multiple dimensions. Examples: "Please indicate your reactions to the computer as a means of work. How do you find the computer interface:
  UNDERSTANDABLE 1 2 3 4 5 6 7 CONFUSING?"
- Multiple choice questions or dichotomous questions: respondents select one category. "Do activities at work make your back problem: better, worse, no change?"
- Open-ended questions.

The next stage in constructing the questionnaire involves the selection of the response scale and categories. For instance, musculoskeletal discomfort can be measured in terms of severity, frequency and duration. See Fowler (1995), Converse and Presser (1986), and Sheatsley (1983) for various response-scale categories. The balance, polarity, and number of categories on the scale are important considerations (Charlton 1996). Balanced scales with equal numbers of positive and negative alternatives are preferred. Whether or not to provide a neutral midpoint is another important decision. The number of categories on the response scale depends on the degree of discrimination required.

The following stage in constructing the questionnaire is the assembly of the questionnaire items and instructions. Dillman (1978) created "The Total Design Method" — that is, a set of guidelines for formatting a questionnaire. The format of a questionnaire can be as important as the nature and wording of the questions asked. A poorly organized and laid out questionnaire can lead respondents to miss questions, confuse them, and, finally, result in poor data quality. The order in which questions are asked can also affect the responses. If the order of questions is an especially important and sensitive issue in a given study, pre-testing of different versions of questionnaires with different ordering should be performed.

In addition, every questionnaire should contain clear instructions and introductory comments. Respondents should be informed about the purpose of the survey and procedures for filling out the questionnaire. In most instances, it is important from an ethical point of view to provide summary data to the study participants. Specific instructions may be necessary for some questions. For instance, rank-ordering a list of items is often difficult for respondents. The same information can be obtained by having respondents rate the items. However, if one finds it important to have respondents rank-order a list of items, the list should be made as short as possible and clear instructions should be given.

3.5 Step 5

The final step in the development of a questionnaire survey is the pre-test. The questionnaire should be reviewed for grammatical errors and typos, and also for content and clarity. The pre-test can involve review by experts and/or administration of the questionnaire to a small number of subjects. The latter can be performed using either focus groups or respondents filling out the questionnaire individually. Whatever method is used, several issues should be examined:

- questions: clarity, relevance, response categories
- questionnaire format: instructions, layout
- time needed for completing the questionnaire

4. CONDUCTING A SURVEY

Conducting a questionnaire survey involves a sequence of steps and a range of decisions. The opportunities for errors in the survey process are multiple. Sinclair (1995) emphasized the potential communication problems in different stages of questionnaire design and survey. At the survey stage, an important communication problem is the inaccuracy and misinterpretation of the information recorded on the questionnaire. Once a questionnaire has been designed, there are still many opportunities for error.

The data-collection method itself can lower the quality of the data collected. The researcher should pay attention to the timing of data collection. Data collected at the "wrong" time may be of lower quality. One should also think about the conditions in which the questionnaire will be filled out: on-the-job, at home, etc. If questionnaires are completed at the workplace, respondent should be given privacy and a quiet place to perform this task. Study participants should be explained the purpose of the survey, as well as the methods used by the researcher to handle the questionnaires. In particular, if respondents are asked to put their name (or an identifier) on the questionnaire, the researcher should clearly explain the way questionnaires will be handled (and by whom) in order to ensure confidentiality.

Once questionnaire data has been collected, various activities should be performed before the data can actually be analyzed:

- questionnaire "cleaning". At this stage, the researcher assigns ID numbers to the questionnaire and gets the questionnaires ready for data entry. This may involve, for instance, recoding of questions: a question on the job experience of a person ("10 years and 6 months") may need to be recoded ("10.5 years") in order to facilitate the data entry. Open questions (e.g. job title) should also be recoded and/or categorized if the data is to be entered in the database.
- setting up the database. The database should be set up to facilitate the data entry process, and to allow the researcher to perform all the necessary data analyses subsequently. This requires the researcher to know in advance the bulk of the data analyses to be performed.
- data entry and verification. The data entry process should be made as easy as possible for the data entry person. Appropriate physical working conditions should be provided to the data entry person. Full verification of data entry is rarely done, except when the sample size is small. However, verification of a sample of questionnaires can be done to evaluate the accuracy of the data entered.
- data cleaning. Once data has been entered into the database, a range of analyses should be performed to verify the accuracy of the data. For instance, frequency tables should be computed for all variables to examine the range of data and to check for outliers and extreme variables. Cross-tabulations between supposedly related variables can also be performed to verify the accuracy of the data.
- data analysis. Once the researcher feels confident about the accuracy of the data entered into the database, data analysis can start. The first stage in the data analysis involves the evaluation of the data quality (reliability and validity, see section below).

5. RELIABILITY AND VALIDITY

When constructing a survey or a questionnaire, the quality of the measurement method plays a crucial role. The quality of measurement is assessed by examining reliability and validity. Reliability refers to the issue of measurement repeatability: when we measure something at two different occasions, the two measures should result in the same outcomes. Validity refers to the content of measurement: are we measuring what we think we are measuring? There are several methods to evaluate reliability and validity (Carmines and Zeller 1990).

We can evaluate reliability by measuring a concept at two different times (test–retest–reliability), by looking at the internal consistency (of the answers) of questions that are supposed to measure the same concept, and by comparing with other methods of measurement of equal or higher level, for example standardized (and validated) questionnaires. A measure that is often used to evaluate internal consistency is the Cronbach's Alpha: it is a measure of the homogeneity of a group of items in a survey or questionnaire.

Three forms of validity can be distinguished (Nunnaly 1978): predictive validity, content validity and construct validity.

- *Predictive validity* (also known as criterion validity) refers to the comparison of the measurement under evaluation to another variable that lays outside the domain of the concept being measured. This variable is known as the criterion or "the golden standard". For example, experienced health as being measured with a survey can be compared with a medical examination. In this example, the medical examination is the criterion to which the questionnaire data (i.e. experienced health) is compared.
- The *content validity* of a measurement instrument can be established by examining the content of the questions very carefully. Are the items (in a concept) well chosen? Do they measure what we think they measure? Do they represent the entire domain of the concept? An often-used method for establishing the content validity of a questionnaire is to ask experts about clarity and completeness of a questionnaire (McDowell and Newell 1987). In addition, one can find out if the questions used to measure a particular concept are well understood by the target population (and in "their own language") by interviewing them (see discussion on pre-testing questionnaires above).
- *Construct validity* is important when measuring abstract concepts (e.g. job satisfaction, mental workload, stress). The abstract concept (the construct) is operationalized by several questions. When results of statistical analyses show that the questionnaire items show a high degree of internal consistency, one can conclude that the different questions do indeed refer to one (underlying) construct. The following statistical analyses can be used to evaluate construct validity: reliability analysis, correlation analysis and factor analysis.

6. CONCLUSION

Questionnaires can provide useful information on various ergonomic variables (Salvendy and Carayon 1997). Questionnaires are often perceived as not scientific because of

their subjectivity. However, it is important to recognize that any data can be placed on a subjectivity/objectivity continuum. In this chapter, we have reviewed principles, guidelines and methods for ensuring the high quality of data collected via questionnaire surveys. Before designing a questionnaire, it is important to clearly specify the objective of the questionnaire survey.

REFERENCES

BABBIE, E., 1990, *Survey Research Methods.* (2nd ed.) (Belmont, CA: Wadsworth).

CARMINES, E. G., and ZELLER, R. A., 1990, *Reliability and validity assessment* (Beverly Hills, CA: Sage).

CHARLTON, S. G., 1996, Questionnaire techniques for test and evaluation. In T. G. O'Brien, and S. G. Charlton, (eds) *Handbook of Human Factors Testing and Evaluation.* (Mahwah, NJ: Lawrence Erlbaum), pp. 81–99.

CONVERSE, J. M., and PRESSER, S., 1986, *Survey Questions: Handcrafting the Standardized Questionnaire* (Beverly Hills, CA: Sage).

DILLMAN, D. A., 1978, *Mail and Telephone Surveys: The Total Design Method* (Chicago, IL: University of Chicago Press).

FOWLER, F. J. J., 1995, *Improving Survey Questions - Design and Evaluation.* Applied Social Research Methods Series, 38 (Thousand Oaks, CA: Sage).

FRESE, M., and ZAPF, D., 1988, Methodological issues in the study of work stress. In L. Cooper, R. Payne, (eds) *Causes, coping and consequences of stress at work.* (Chichester: Wiley).

KASL, S. V., 1987, Methodologies in stress and health: Past difficulties, present dilemmas, future directions. In S. V. Kasl, C. L. Cooper, (eds) *Stress and health: Issues in research and methodology* (Chichester: Wiley), pp. 307–18.

MCDOWELL, I., and NEWELL, C., 1987, *Measuring Health: A Guide to Rating Scales and Questionnaires* (Oxford, UK: Oxford University Press).

NUNNALY, J. C., 1978, *Psychometric theory* (2nd ed.) (California: McGraw Hill).

SALVENDY, G., and CARAYON, P., 1997, Data-collection and evaluation of outcome measures. In G. Salvendy, (ed.) *Handbook of Human Factors and Ergonomics* (2nd ed.) (New York: Wiley), pp. 1451–70.

SHEATSLEY, P.B., 1983, Questionnaire construction and item writing. In P.H. Rossi, J.D. Wright and A.B. Anderson, (eds) *Handbook of Survey Research* (Orlando, FL: Academic Press), pp. 195–230.

SINCLAIR, M. A., 1995, Subjective assessment. In J. R. Wilson, E. N. Corlett, (eds) *Evaluation of Human Work: A Practical Ergonomics Methodology* (2nd ed.) (London: Taylor & Francis), pp. 68–100.

KEY REFERENCES

CARMINES, E. G., and ZELLER, R. A. 1990. *Reliability and validity assessment* (California: Sage).

CONVERSE, J. M., and PRESSER, S. 1986. *Survey Questions: Handcrafting the Standardized Questionnaire* (Beverly Hills, CA: Sage).

FOWLER, F. J. J. 1995. *Improving Survey Questions: Design and Evaluation.* Applied Social Research Methods Series, 38 (Thousand Oaks, CA: Sage).

NUNNALY, J. C. 1978. *Psychometric theory* (2nd ed.) (California: McGraw Hill).

SALVENDY, G., and CARAYON, P. 1997. Data-collection and evaluation of outcome measures. In G. Salvendy (ed.) *Handbook of Human Factors and Ergonomics* (2nd ed.) (New York: Wiley), pp. 1451–70.

A Survey of Ergonomics Methods

N.A. Stanton and M.S. Young

Department of Psychology, University of Southampton, Highfield, Southampton, SO17 1BJ, UK

1. TEXTS ON ERGONOMICS METHODS

There appears to be a growing number of texts in recent years describing, illustrating and espousing a plethora of ergonomics methods (Wilson & Corlett 1995; Salvendy 1997; Stanton 1998). The rise in the number of texts reporting on ergonomics methods may be seen as a response to the requirement for more inventive approaches for assessing the user and their requirements. In many ways this may be taken to mean that the call for user-centred design has been taken seriously by designers. However, this success has forced the ergonomics community to develop methods to assist the design of products and devices. This demand seems to have resulted in the pragmatic development of methods having priority over scientific rigour. In a recent review of ergonomics methods, Stanton & Young (1998) identified over 60 methods available to the ergonomist. The abundance of methods might be confusing for the ergonomist, Wilson (1995) goes as far as to suggests that a

> . . . method which one researcher or practitioner is an invaluable aid to all their work may to another be vague or insubstantial in concept, difficult to use and variable in its outcome. (p. 21)

This quote highlights the fact that most methods are used by their inventors only. Despite the proliferation of methods, there are few clues in the literature available for ergonomists to enable them to discriminate between methods. The purpose of this article is to address some of these issues and focus attention on usability measures and methods that examine user activity and behaviour. Some researchers have attempted to address the problem. For example, an overview of the methods is presented in a number of books (Wilson & Corlett 1995; Salvendy 1997; Stanton 1998). Stanton & Baber (1996) argue that it is not suprising that the checklist is the ubiquitous method, as it is the only method that is independent of these factors.

2. SURVEY OF ERGONOMICS METHODS

In order to evaluate ergonomists' practices in the use of methods, it was necessary to conduct a survey of professional ergonomists. In a recent survey of 6 organisations, conducted by Baber & Mirza (1998), a vrey limited range of methods were reportedly used in product evaluation. The methods typically used consisted of questionnaires, interviews, observations, checklists and heuristics. Baber & Mirza (1998) report that the frequency with which a method is used is highly dependent upon its ease of use, and most respondents reported combining three or four methods to obtain an overall picture of the product under evaluation. Stanton & Young (1998) followed up this survey with eighteen self selected respondents from a pool of 163 members of The Ergonomics Society listed on the professional register, who were asked to comment on their experience with ergonomics methods.

Stanton & Young (1998) were particularly interested in which methodologies are used, what they may be applied to, and whether such techniques are useful.

2.1 Results and Discussion of Survey

Details of 27 methods used by ergonomists are presented in Table 1. Of the 27 reported methods, only 11 were reported by 6 or more respondents (i.e. from Mock-ups to Walkthroughs in Table 1) and only these methods were treated to statistical analysis. In general terms, the results of the survey seem to confirm the analyses undertaken thus far. First, there were no references by the respondents to reported evidence for reliabilty or validity in the literature which concords with our earlier investigations (Stanton & Young, 1998). Second, the respondent's evaluations of the techniques is consistent with our own experience.

The statistical analysis of the respondent's ratings of the techniques revealed three interesting and statistically significant results. First, some methods were rated as easier to use than others (Chi-square, corrected for ties = 33.0595; $p < 0.0005$). Checklists were rated as significantly easier to use than simulation (Z, corrected for ties, = -3.3994; $p < 0.001$), guidelines were rated as significantly easier to use than prototyping (Z, corrected for ties, = -2,578; $p < 0.01$) and interviews were rated as significantly easier to use than mock-ups (Z, corrected for ties, = -2.1381; $p < 0.05$). We noted earlier that only a limited range of methods are used but there is no guarantee that these are the most appropriate. Baber & Mirza (1998) report that product designers tend to restrict their methods to interviews, observation and checklists (which was confirmed in our study). Similar to the report by Baber & Mirza (1998), our finding suggests that this is likely to be due to the space of applying the methods. Second, the reported ease with which methods are applied depends upon whether software support is used. The results show that where no software support is used, the method is rated as easier to use than where software support is provided (Z, corrected for ties, = -2.6597; $p < 0.01$). Although perhaps this is a counter-intuitive finding, our own experience suggests that software can make even a relatively easy method quite complex and cumbersome. With some irony we would suggest that developers of ergonomics software cannot afford to ignore ergonomics in the design of their product! However, it is our experience that software can make the ergonomist's activities more efficient in the long term. Finally, the data suggest that users of ergonomics methods perceive differences in ease of use of methods depending upon the level of training they have received (Chi-square, corrected for ties, = 6.0639; $p < 0.05$). Those who have received no training rate the methods as easier to use than those who have received formal training (Z, corrected for ties, = -1.9919; $p < 0.05$). We suggest that this result is probably due some misconception regarding how to use the method and would recommend formal training in any approach used.

Table 2 shows the reported links of each method with other methods. As can be seen there are strong links between questionnaires and interviews as a combined approach. Interviews are also linked with observation, HTA and prototyping. Observation is linked with questionnaires and HTA. Finally, HTA is linked with HEART. Some of these links with other methods are by necessity. For example, HTA requires observation and inter-

Table 1. Summary of survey responses (variables "Frequency" through "Time" are medians of respondents' ratings; "Software" and " Training" variables are frequency counts; "Years" is mean reported use)

	Mock-ups	TLA	HTA	Q'naires	Interviews	Observations	Prototyping	Checklists	Guidelines	Simulation	Walkthrough	Link	Layout
n	9	6	10	15	17	12	6	8	7	8	6	5	3
Frequency	2	2	2.5	2	4	3.5	3.5	3	3	2	2.5	2	2
Ease	2	3.5	3	4	4	4	3	4	4	2	3	4	5
Cost/benefit	3	3.5	4	3	4	3.5	5	4	4	4	4.5	4	4
Time	2	3	2.5	3	3	3	2.5	4	3	1.5	3	3	4
Software	4	3	3	6	2	3	5	1	2	6	1	1	0
No software	5	3	7	9	15	9	1	7	5	2	5	4	3
No Training	3	1	1	2	4	3	2	4	5	2	2	1	1
Informal	6	4	1	4	8	5	1	3	2	4	2	3	1
Formal	0	1	8	9	5	4	3	1	0	2	2	1	1
Years	10.57	16.25	3.33	11.69	12.94	11.91	10.67	11.6	10.5	6.29	8.6	7.75	10.5

	Event trees	Fault trees	HEART	Focus group	Rep grids	ARA	Protocols	TAKD	Studio	JACK	Activity	THERP	Expert	THEA
n	4	3	3	3	3	2	2	1	1	1	1	1	1	1
Frequency	2.5	2	4	3	1	1.5	3	5	5	2	4	5	5	1
Ease	3	3	4	3	3	2.5	3.5	5	4	2	5	2	3	2
Cost/benefit	4	3	2.5	4	2	3	4	3	4	4	5	4	4	3
Time	3	3	3	3	3	2	2.5	3	3	3	4	3	2	2
Software	3	2	2	1	1	0	0	0	1	1	1	0	0	0
No software	1	1	1	2	2	2	2	1	0	0	0	1	1	1
No Training	0	0	0	2	0	0	1	1	0	0	0	0	0	0
Informal	2	1	2	1	2	2	0	0	0	0	0	1	1	1
Formal	2	2	1	0	1	0	1	0	1	1	1	0	0	0
Years	3.75	3.33	3.33	16.5	9	6	20	8	7	2	10	4	3	

views as necessary prerequisites to collect and verify the analysis. Similarly, HEART requires HTA before it can be conduted. Other links between methods are shown in lighter shading on Table 2.

Table 3 shows the uses to which the methods are put. The dark shading shows that questionnaires, interviews and observations are the principal methods used in data collection. The medium shading shows that simulators, computer simulation, interviews and repertory grids are used largely for design activities. It also shows that simulators are used for assessment activities and that repertory grids are used in validation activities. The light shading indicated occasional useage of other methods for a range of activities with no clear pattern.

3. GENERAL DISCUSSION

The appropriateness of the application of ergonomics methods to points in the design life cycle of products and devices is one of continuing debate. Obviously some methods depend upon the existence of a device to evaluate (such as observation, link analysis and layout analysis) whereas others do not (such as heuristics, checklists and repertory grids). An interesting picture is painted by the survey which asks what people use the methods for. The responses showed that four main areas of application were highlighted: data collection, design, assessment and validation activities. Table 4 summarises the methods used in these general areas. The findings agree with our assessment of interviews and repertory grids, i.e. that they are appropriate for most of the design stages.

In additon, some interesting links between methods came from the survey which are shown in Figure 1.

As shown, the interview is directly and indirectly linked to five other methods. This makes the interview an important design method. Given the concern about reliability and validity of the interview in other fields of research (Cook, 1988) we would caution users of this technique to ensure that they employ a semi-structured and situationally focused approach to the device evaluation interview.

	Event trees	Fault trees	Interviews	Observation	Questionnair	Rep grids	Protocols	HEART	PHEA	SHARP	SLIM	TAFEI	THERP
Event trees													
Fault trees	2												
Interviews	1												
Observation			7										
Questionnair	1		16	9									
Rep grids			5										
Protocols				1	1								
HEART	5	3	1										
PHEA			1										
SHARP			1										
SLIM			1										
TAFEI			1										
THERP	2	2	1										
Heuristics	1	1											
HTA	1		6	8		3		4	2	2	2	2	2
Layout			1	2	1								
Link			1	1									
TLA				1									
ARA			3	3									
Checklists													
Guidelines													
Simulation	1	1	3	2	3								
Focus groups													
Prototyping	1	1	4	1	1			1					
Simulators			1										
Walkthrough	1		1	1									

	Heuristics	HTA	Layout	Link	TLA	ARA	Checklists	Guidelines	Simulation	Focus group	Prototyping	Simulators
Event trees												
Fault trees												
Interviews												
Observation												
Questionnaire												
Rep grids												
Protocols												
HEART												
PHEA												
SHARP												
SLIM												
TAFEI												
THERP												
Heuristics												
HTA		1										
Layout		1										
Link		4	2									
TLA		3		1								
ARA		1										
Checklists							1					
Guidelines							1	2				
Simulation		3										
Focus groups												
Prototyping		4	1	1	1				2	2		
Simulators		2		1							4	
Walkthroughs		2	1	4	1					1	4	5

Key:
- N/A
- Heavy overlap
- Medium overlap
- Light overlap
- No overlap

Table 2. Reported links between methods

	Q'naires	Interviews	Observation	Rep grids	Checklists	Simulators	Simulation	HTA	Event/fault	HEART	Link	TLA	Protocols	TAKD
Data collection	**11**	**13**	**10**					1		1				
Attitudes	4			1										
Assessment	1	2	1		1	3	6	4	1		2	2	1	
Design	1		2			3	5	6	2		2	2		
Description									3					
Investigation	1	1	1		1				1	1		1	2	
HRA/HEI									1	3	3			
Analysis	4				1	3		1	3			1	4	1
Validation		1					3	4	1	1				
Mental models				2										
Training						1	1	1	1					

	ARA	Guidelines	Walkthrough	Prototyping	THEA	Focus group	Layout	THERP	STUDIO	Expert	Activity
Data collection						1	1				1
Attitudes						1	1				
Assessment	1	1	1	1	2	1	1	2	1	1	1
Design	1	5	5	2	5						
Description									1		
Investigation	1	1	1	1						1	1
HRA/HEI											
Analysis		1	1	3	5	1				1	
Validation										1	
Mental models											
Training											

Key:
Heavy usage
Medium usage
Light usage
No usage

Table 3. Reported applications of methods

Table 4. Application of ergonomic methods

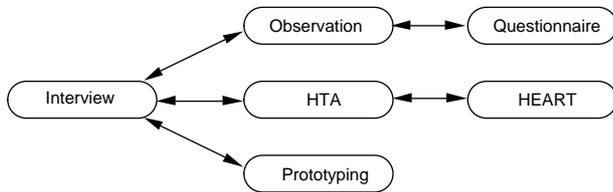| Application	Method(s)
Data Collection	Questionnaires, interviews and observation
Design	Simulators, computer simulation, interviews and repertory grids
Assessment	Simulators
Validation	Repertory grids

Figure 1. Links between methods

4. CONCLUSIONS

In conclusion, there is clearly little reported evidence in the literature of reliability or validity of ergonomics methods. This was confirmed by the survey. The patterns of usage suggest that there is no clear match of methods to applications, which presents a rather confusing picture when embarking upon an evaluation. Apart from a few clearly defined applications the pattern looks almost random. We would suggest ergonomists and designers would be well served by exploring the utility of other methods rather than always relying upon 3 or 4 of ones favourite approaches. However, it is an important goal of future research to establish the reliability and validity of ergonomics methods.

These data could provide the encouragement for designers to try alternative approaches.

REFERENCES

BABER, C. and MIRZA, M.G., 1998, 'Ergonomics and the evaluation of consumer products: surveys of evaluation practices' in N.A. Stanton (ed) *Human Factors in Consumer Products* (London: Taylor & Francis).

COOK, M., 1988, *Personnel Selection and Productivity* (Chichester: Wiley).

SALVENDY, G., 1997, *The Handbook of Human Factor Ergonomics*. (Chichester: Wiley).

STANTON, N.A. and BABER, C., 1996, Factors affecting the selection of methods and techniques prior to conducting a usability evaluation, in P.W. Jordan; B. Thomas; B.A. Weerdmeester & I.L. McClelland (eds) *Usability Evaluation in Industry*. (London: Taylor & Francis).

STANTON, N.A., 1998, *Human Factors in Consumer Products* (London: Taylor & Francis).

STANTON, N.A. and YOUNG, M.S., 1998, Is utility in the mind of the beholder? A review of ergonomics methods. *Applied Ergonomics*. 29 (1), 41–54.

WILSON, J., 1995, A framework and context for ergonomics methodology in J. Wilson and N. Corlett (eds) *Evaluation of Human Work*. 2nd Edition. (London: Taylor & Francis).

WILSON, J., and CORLETT, N., 1995, *Evaluation of Human Work*. 2nd Edition. (London: Taylor & Francis).

Task Analysis for Error Identification

C. Baber
University of Birmingham, UK

1. INTRODUCTION

In order to evaluate the design of a product, one often requires a physical working prototype. While this allows measurement of "real" user performance with the product, it also means that such evaluation is delayed until well into the design process. It would be useful to be able to take paper designs and submit them to "user testing" in order to determine the usability of the product. One way in which this can be done is through the application of error prediction techniques. In this entry, an error prediction technique is described which was developed specifically for product evaluation.

A given product will be capable of being in a finite number of states and will be capable of moving from its present state to a finite number of other states. A product can therefore be described in terms of states and transitions between states within the space of operation. In using a product, there will be a limited number of correct actions which can be performed at any one time. This will limit the number of actions that need to be taken into account. If we assume that products are used to perform tasks leading to the fulfillment of goals, then the definition of specific goals will limit the actions required. From this we can also propose that there will be situations in which a number of actions may be possible, but a specific goal may only require the performance of one action. If the product presents the opportunity to perform a range of actions not related to the goal, there will be a potential for error.

Task Analysis for Error Identification (TAFEI) rests upon three assumptions:

1. that human use of products will be goal directed, i.e. guided, at some level, by plans and following intended courses of action;
2. that goals will relate explicitly to, and be modified in light of, the information supplied to the user by a product;
3. that errors are the result of a cooperative endeavor between user and product.

2. CONDUCTING TAFEI ANALYSIS ON A VIDEO-CASSETTE RECORDER

As consumer products become more complicated, there will be an increase in the number of cognitive problems people have in not only operating, but also understanding the products. It is well known that people have difficulty operating videocassette recorders, and so it is fitting that this class of device be used to illustrate the use of TAFEI. This example (taken from Baber and Stanton 1994) shows the use of TAFEI to examine the ease with which programming could be performed. VCR programming gives the appearance of a logical task sequence. However, it requires several activity loops, which can lead to confusion.

The first phase of TAFEI analysis combines hierarchical task analysis (HTA) with state space diagrams (SSDs). The SSD is based on the supposed operation of a product — for instance, the sequence of states which the designer assumes will be followed to ensure efficient operation. It is important to note here that the SSD describes the states of the product and not user actions.

Each state in the diagram takes the form shown in Figure 1. The state is identified by a number and a brief title. Any notable
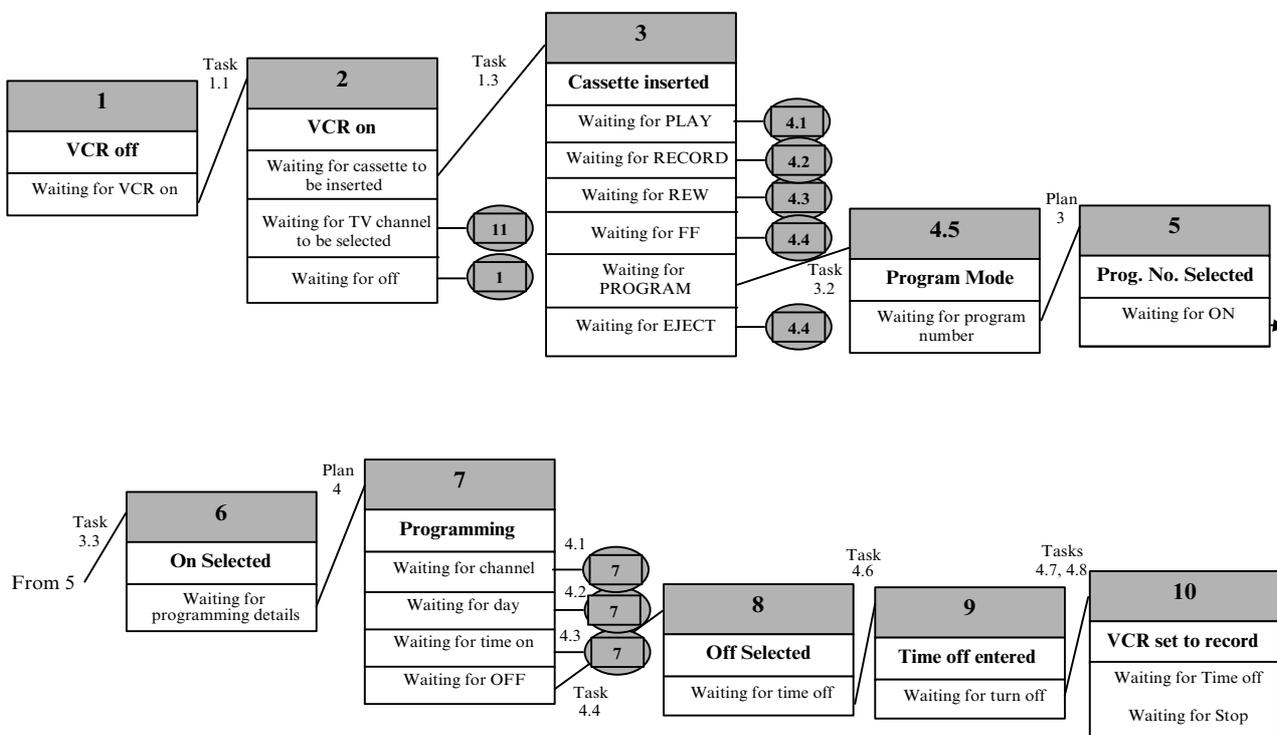

Figure 1. TAFEI diagram for programming a specific model of VCR

consequences of that state are also recorded. It is proposed that a product is capable of being in a finite number of states. By way of example, a kettle might be in one of five states: switched off, empty, full of water, switched on, boiling. Each state allows, or is "waiting for", a subsequent state, e.g. the kettle empty state is waiting for the kettle full state. Movement to any state other than the "legal" — i.e. designed for transition, such as pouring water from or switching on an empty kettle — would be considered an error on the part of the user.

The transition, or change, from one of these states to another requires action by either the human or the product. For example, the transition from kettle switched off to kettle switched on could be performed by the human pressing the on switch; the transition from kettle full of water to kettle boiling would require two steps: first, the human would switch the kettle on; second, the element in the kettle would heat up until the water boiled. It will be noted from this brief description that transitions between states are assumed to occur in sequences and that the defining feature of a particular sequence is the ultimate goal held by the user.

The SSD for the VCR was developed through careful study of the manufacturer's instruction manual. The SSD, then, represents the route through a set of actions which we assumed the designers had "designed into" their product in order to achieve a specified goal — for example, to enter details for recording a program.

HTA provides a normative description of (error free) human activity and is used as an explanation of transitions between states. In this way we can map human actions to transitions between states in the operation of a product. At present the mapping is a relatively simple process in which the transition between product states are assumed to occur as the result of either human or product actions. Where the transitions arise from human actions, we define the transition as a subgoal of product operation, and refer to the HTA for the "task" or "plan" to achieve that subgoal. For transitions requiring human action, the relevant task number or, where applicable, the plan number, derived from the HTA is attached to the transition lines in the SSD leading from one state to the next to produce a TAFEI diagram. We mapped user actions, defined by both plans and by single actions, onto product states represented in the SSD in order to indicate the means by which transitions occurred and developed a TAFEI diagram for VCR programming (see Figure 1).

3. PREDICTING HUMAN ERROR USING TAFEI

We term actions which will lead to a goal in a specific situation as "legal" in that they are allowed at that point in time, those which are permitted but will lead to outcomes other than the current goal as "illegal", and those actions which cannot be performed at that point in time as "impossible". TAFEI aims to elicit illegal actions within the closed loop of human product operation and allow the redesign of the product to make these actions impossible.

From the TAFEI diagram, we can suggest a number of illegal transitions. The definition of illegal actions is relatively straightforward. The analyst, faced with the TAFEI diagram asks whether it is conceivable that a transition can be made from one state to another state. (Note: we do not ask whether it is "sensible" to perform such a transition, only whether it is possible.) The question to ask is, can a user perform an action in state "x" which

will take them to state "z" rather than to state "y"? Some transitions will be impossible in that there will be no way in which an action can be performed to move between two states. However, other actions will be possible in the current design. In this way, we can demonstrate where actions can be performed which were not imagined or intended by the designers. The next step is to consider whether the transitions thus discovered can be accepted in product operation or whether they represent illegal activity. At present, this will be based on the intuitions and expectations of the designer, but we do not see this as a problem — after all, it is the designers who will have the clearest idea of how the product is designed/intended to be used. A transition matrix is used to define types of transition between states. Once the illegal transitions have been identified, it will be possible to begin to propose possible solutions. Furthermore, inclusion of by-products can indicate hazards which can be addressed by designers.

Entering the programming details for the VCR, state 7 in Figure 1, takes place in a poorly structured loop. After setting the timer to "program", and selecting "on", the user receives the prompt "CH". The user presses a "channel up/down" button to select the desired channel. Next, the user selects the day on which a program is to be recorded on. (Note: the "day" button must be pressed before the user can proceed; this means that if the desired program is being shown today, the user must go through the week before setting the product.) After selecting the desired channel and day, the user must then insert the start time. This is performed by selecting "set hours" and changing the time using "+" and "-" buttons. The action is repeated for the minutes, and the same process is used to set the finish time. However, after setting the start time, the user must wait for 5 seconds and then press the "off" button before setting the finish time. Once this sequence of activity has been performed, the user presses "set timer" again, and then presses the "time record" button. In addition to problems of reading the display, VCR users face difficulties in understanding the sequence of actions required. This problem is made worse when either no feedback or inappropriate feedback is given. While this interaction is largely user initiated, it is not always clear what the user should be doing.

A number of the illegal operations, such as switching off the recorder without stopping a tape playing or pressing play without inserting a tape, could be easily dealt with by considering the use of modes in the operation of the devices that would make specific actions impossible. Figure 1 also highlights the recursive transition (7—7) which required users to cycle through a sequence of data input. User reports suggested that this point was a major source of error, with users frequently failing to set the day for recording, for instance, setting the VCR to record on the day it was programmed rather than the day on which the desired television program was scheduled.

4. HOW RELIABLE IS TAFEI?

For any prediction method to be considered reliable, it must demonstrate a reasonable match with observed performance. Baber and Stanton (1996) compared predictions made using TAFEI with observed activity in the use of a ticket-vending machine. From the analysis of the TAFEI diagrams and the transition matrix, the following error types were predicted from the TAFEI analysis:
confuse next action

error in ticket selection
error in station selection
error in zone selection
error in money insertion
confuse mode
use wrong buttons
attempt to use closed machine
confuse return from cancel
attempt to use machine in waiting mode

In summary, TAFEI predicts 10 error types. From observation of users of the machine, we observed all of the predicted errors. We developed the following procedure to define concurrent validity:

(a) predicted errors from TAFEI = 10
(b) observed errors = 15
(c) errors predicted but not observed = 0
(d) errors observed but not predicted = 4
(e) errors defined as impossible = 4

concurrent validity of TAFEI

$$CV = \frac{(a/b) + [1 - (c/e)]}{2} = 0.83 \quad\quad [1]$$

If direct observation of human performance represents the "gold standard" for product evaluation, then obviously any technique which is proposed for this field should provide a reasonable match to "real" activity. We would suggest that TAFEI can produce a great deal of useful information for product evaluation; the fact that TAFEI produced around an 80% level of agreement with observed data suggests a sufficiently good relationship to warrant further consideration.

In conclusion it is proposed that human error prediction techniques, such as TAFEI, represent a reliable, cost-effective means of conducting product evaluation, and provide the advantage of allowing analysis during early stages of product development. We would also propose that such a structured technique could be both more reliable and rigorous than product evaluation checklists, which are commonly used in preference to observation as methods of speedy data collection (Mirza and Baber 1996).

REFERENCES

BABER, C. and MIRZA, M., 1998, Ergonomics and the evaluation of consumer products: surveys of evaluation practices. In N.A. Stanton (ed.) *Human Factors in Consumer Products* (London: Taylor and Francis), pp. 91–104.

BABER, C. and STANTON, N.A.,1994, Task analysis for error identification. *Ergonomics*, 37, 1923–42.

BABER, C. and STANTON. N.A., 1996, Human error identification techniques applied to public technology. *Applied Ergonomics*, 27, 119–31.

STANTON, N.A. and BABER, C.,1996, A systems approach to human error identification. *Safety Science*, 22, 215–28.

Task Analysis in Industry

H. Luczak and J. Stahl

Institute of Industrial Engineering and Ergonomics, Aachen University of Technology, Bergdriesch 27, D–52056 Aachen, Germany

1. INTRODUCTION

Task analysis is a crucial step in evaluating and designing systems. It provides insights into how users and systems interact and ensures that either the task is adapted to the person or, if feasible, measures can be derived to fit the person to the task.

2. FUNDAMENTALS OF TASK ANALYSIS

2.1 Etymology and definitions

The notion "task analysis" is composed of two words. The term "task" is a conceptual construct, so there are numerous ways of defining it. There are two major dimensions in which definitions differ: breadth of coverage and external versus intrinsic definitions (Fleishman and Quaintance 1984). The latter provides a fundamental difference in perspective. Tasks are viewed either in terms of work (usually in terms of observable behavior) or in terms of people (usually in terms of cognitive processes). In terms of work, a task may be defined as a set of actions with an objective, performed by a worker, and transforming inputs into outputs. From the people viewpoint, a task may be described as a series of goal-directed transactions controlled by "programs" that guide the operations of a human operator.

The term analysis, from its Greek origin, refers to the decomposition and separation of any material or abstract entity into its constituent elements. Task analysis can therefore be defined as a study that divides human work into component tasks and systematically and comprehensively describes these components in terms of their defining "elements" or aspects:

- action, activity or behavior
- sequence of subtasks
- object
- tools, aids, or equipment
- time
- frequency
- space
- environmental conditions
- goals
- cues
- information
- rules
- feedback
- cooperation and communication
- controls used
- errors or error proneness
- criticality
- importance
- required decisions
- required knowledge, skills and abilities (KSA)

Task analysis covers a range of techniques and methods. Depending on the goals and the areas applied, these techniques and methods focus on certain defining aspects and they neglect others.

2.2 Applications of task analysis

Task analysis is applied to various areas (Kirwan and Ainsworth 1992):

- Allocation of function (to machine or operator, extent of operator involvement in the control of a system, etc.)
- User interface design (information displays, controls, tools, equipment, etc.)
- System performance (productivity, reliability, limits, etc.)
- Safety (hazards to the operator, general system safety, human reliability assessment, incident and accident investigation)
- Quantitative and qualitative requirements in terms of KSA (number of staff, personnel selection, training measures, etc.)
- Work organization (allocation of responsibility, job descriptions, communication requirements, remuneration, etc.)

2.3 Task analysis process

The term "task analysis" might suggest that only analysis methods are used. However, the task analysis process usually comprises three steps: data collection; representation or description (and in some cases also simulation); and analysis. The process subsumes a variety of techniques, which are employed for data collection and representation or simulation, and methods which are based on theoretical models that allow interpretation of data in a broader framework and are used for analysis.

3. DATA COLLECTION TECHNIQUES

The type of data to be collected in a task analysis process depends on the intention and the scope of the analysis. Usually it is a subset of the defining aspects of tasks listed in Section 2.1. The data may be subjective or objective. Examples of subjective data collection techniques are questionnaires, checklists, interviews (structured or unstructured; single or group; initial, verification, or follow-up), verbal protocols, and the critical incident technique. Possible sources comprise operators or job incumbents, colleagues, superiors, and the analysts if they themselves are executing the tasks. Objective data collection techniques include any kind of continuous or sampled observation by an analyst or any kind of technically assisted access via perception (e.g., camera, tape recorder), viewing documents (e.g., job descriptions, blueprints, job cards, schedules, incident reports, diaries, training programs), and measurement (e.g., physiological variables, environmental conditions).

4. DATA REPRESENTATION TECHNIQUES

Data that has been collected must be represented in a meaningful way to aid the analysis, so representation techniques are often directly linked with analysis methods. Data representation techniques are either static or dynamic. There are a very wide range of static techniques, such as keyword protocols, tables, charts, event trees, and task hierarchies. The diversity can be illustrated by some examples. The critical incident technique (CIT) and failure mode and effects analysis (FMEA) simply represent the collected data textually in a list or table to be analyzed. Hierarchical task analysis (HTA) and event trees represent the data in treelike diagrams. The structured analysis and design technique (SADT) graphically represents an activity model and a data model. Nodes are denoted by activities, which are related to each other by input/output relations and control flow. Necessary resources are attributed to the activities.

Static data representation techniques merely provide a systematic structure of the tasks and their defining aspects, but dynamic representation techniques also describe the behavior of systems and may be used for simulation. They fall into one of three categories: state-oriented, event-oriented, or Petri-net-based. State-oriented representation techniques represent discrete states as nodes, and actions or events as transitions from one state to another. Generic techniques within this category are sequence diagrams and state diagrams. They have been adapted to create techniques with specific applicability to task analysis, such as object charts, ADV charts, task-object charts (TOCs), and the operator function model (OFM).

Event-oriented representation techniques represent events or actions as nodes and they represent states as edges. Tasks networks are widely used; they include techniques such as input/output block diagrams, process charts, flow diagrams, SAINT (system analysis of integrated networks of tasks), SLAM II (simulation language for alternative modeling), SIMAN (simulation analysis), and Simple++ (simulation in production, logistics and engineering implemented in C++). Other event-oriented representation techniques are time-based, e.g., Gantt diagrams and timeline analysis. Layout-based representations such as link analysis are also useful for systems design.

Petri nets are a synthesis of state-oriented and event-oriented graphs. In Petri nets, places represent states and transitions represent events. High-level nets are especially useful for task analysis. Examples are generic techniques such as place/transition nets (PT nets) or colored Petri nets (CPNs). Techniques have also been developed for specific purposes, such as role-function-action nets (RFA nets), which are useful in human-computer interaction, and expanded event-driven process chains (EEPCs), which are used to design corporate information systems (e.g., SAP/R3).

5. ANALYSIS METHODS

Analysis methods can be distinguished according to the breadth of coverage of the respective task definitions. This can be done following the seven structure-oriented levels of the ordering model by Luczak and Volpert (Luczak 1997). In this context, levels are not equally relevant, and analysis methods and their respective models may be grouped with respect to subtask, task, job, and group levels.

5.1 Subtask level

The subtask level comprises analysis methods that do not take into account the goal of a task. They can be distinguished as methods for motor processes and mental processes. Analysis methods for motor processes are therbligs, systems of predetermined times (SPT), the Ovako working posture analysis system (OWAS) and Ergon-Expert. Therbligs are 17 basic motion elements from which all industrial processes are composed and were identified by Gilbreth. SPT comprises methods such as work factor (WF) and methods time measurement (MTM), and consists of rules with which working times for limb operations can be determined, e.g., manual assembly. OWAS provides a matrix of basic work postures and is used to derive measures on the basis of relative frequency and stress of the respective postures. Ergon-Expert is a computer-based expert system to analyze lifting and material handling tasks.

Analysis methods for mental processes are based either on

stage models, resource models, or combinations thereof. In stage models, information is conceived as passing through a finite number of discrete stages (e.g., the Sternberg model). Resource models (e.g., the Wickens model) are based on the notion of availability of scarce processing resources and can be used to assess human performance limits. They are distinguished as single-resource and multiple-resource models. A combined model is the cognitive energetic linear stage model by Sanders. An application of SPT to mental processes is WF-Mento, which allows the analysis of time consumption of repetitive informational processes.

5.2 Task level

The task level is at the heart of task analysis. The underlying models of analysis methods on this level may be distinguished as single-task models, multiple-task models, and human error and reliability models. A few representative models of each category are named.

5.2.1 Single-task models

Single-task models comprise tasks such as manual control, tracking, and monitoring. Several models are based on applying control theory to human performance, such as the five-parameter quasilinear model and the crossover model, both by McRuer, and the optimal control model (OCM) by Kleinman, Baron, and Levison. They allow quite accurate prediction of task performance and control characteristics. To describe and predict human monitoring performance, specifically the way operators divide their attention among a number of instruments, the visual sampling models by Senders, Carbonell, and Smallwood may be used.

Information processing models include the model human processor (MHP) model, the goals, operators, methods, and selection rules (GOMS) model and their related models by Card, Moran, and Newell. MHP allows analysts to make general predictions about the time it will take to carry out tasks. It is based on the notion that information flows through discrete stages in order to be used and processed. GOMS provides a model of the necessary knowledge for successful execution of a task, which is assumed to be like a problem-solving process. Related models are the keystroke model and the unit tasks model.

5.2.2 Multiple-task models

Owing to increasing automation and use of advanced information technologies, there is growing interest in multitask performance, which has led to a wide variety of analysis methods and their respective models. Here are some of them:

- Predominantly aiming at optimizing system performance are simulation methods like SAINT, SLAM II, SIMAN, and the human operator simulator (HOS). The first three methods are based on task networks (Section 4), and HOS provides micromodels for human cognitive, perceptual, and psychomotor acts to predict timing and accuracy.
- Sheridan's supervisory control model is a qualitative model featuring a series of nested control loops incorporating these user tasks: plan, teach, monitor, intervene, and learn. It is especially useful in allocating functions to machines and operators.
- Rasmussen characterizes three kinds of human performance based on skills, rules, and knowledge (SRK model). From

this model, he developed a framework called the "decision ladder," which can be used to describe and analyze decision-making processes.

- Klein's recognition-primed decisions (RPD) model is a decision-making model that focuses on situation assessment rather than judgment of feasible alternatives. It distinguishes three cases: simple match, developing a course of action, and complex RPD strategy (situation evaluation).
- Cosimo (cognitive simulation model) by Cacciabue and Kjaer-Hansen is used to analyze decision-making processes in emergency situations. It comprises high-level decision making (HLDM) with higher cognitive processes like diagnosis and planning, and low-level decision making (LLDM) with predefined lines of action.
- Semiotic models are based on the work of Morris on signal theory and are used for the analysis and design of human-computer interaction. They distinguish three levels: pragmatic, semantic, and syntactic.
- The Siegel-Wolf model is based on the assumptions that operator loading is the basic element in effective human-machine system performance and that the variety of loading effects are compressed into one variable called stress. It can be applied to a wide variety of tasks and systems.
- Action regulation theory and HTA are both based on the notion that tasks can be described as hierarchies of goals, tasks, and operations, such as hierarchically coupled test-operate-test-exit (TOTE) units. HTA is quite generic in nature and application, whereas action regulation theory incorporates further concepts like the assignment of goal-action cycles to three levels (psychomotor, perceptive-conceptual, and intellectual) and is applied with the goal to design interesting and meaningful jobs.

5.2.3 Human error and reliability models

Although some of the methods mentioned above may also be used to model human error and reliability, there are dedicated techniques. Fault tree analysis combines hardware faults and human errors using AND/OR logic; it indicates potential weak links in system reliability.

- FMEA enables the analyst to consider the reliability of the operator and the consequences of errors for the system.
- Technique for human error rate prediction (THERP) is a method to quantitatively predict human error probabilities on the basis of human performance shaping models.
- Human cognitive reliability (HCR) by Hannaman and generic error-modeling system (GEMS) by Reason, both are based on Rasmussen's SRK model.

5.3 Job level

Task analysis methods at the job level are also known as job analysis methods. Some methods from the task level are also applied at the job level, with the additional aspect of considering the entire activity systems of single persons, i.e., their jobs as a

whole. Job analysis methods may be distinguished in four categories (Fleishman and Quaintance 1984):

- The behavior description approach emphasizes operators' overt behavior, i.e., what they actually do while performing a task. Well known is the position analysis questionnaire (PAQ) by McCormick, Jeanneret, and Mecham; this follows the stimulus-organism-response paradigm and additionally embraces behavioral adjustment in terms of cooperation and work context. It is quite general in nature and may be used for various applications, such as determining wages or required KSA. Many other methods rely on the PAQ for job description and expand it with additional models, e.g., Arbeitswissenschaftliches Erhebungsverfahren zur Tätigkeitsanalyse (AET), which translates as "ergonomic job analysis technique" and employs the stress-strain model.
- The behavior requirements approach catalogues behaviors that should be omitted or are assumed to be required in order to achieve criterion levels of performance. An example is Miller's task strategies approach, which is based on an information processing model.
- The ability requirements approach describes tasks in terms of the abilities that a given task requires of the individual performer. Well known is Fleishman's manual for ability requirements scales (MARS), which is based on a factor-analytical model of human abilities.
- The task characteristics approach is based on the definition of a task as a set of conditions that elicit performance. The task can be described in terms of intrinsic, objective properties. An example is Hackman and Oldham's job diagnostics survey (JDS), which employs a job characteristics model of work motivation based on five core job dimensions: skill variety, task identity, task significance, autonomy, and feedback from the job itself.

5.4 GROUP LEVEL

Although job analysis methods may also take into account the impact of group work on the individual, task analysis at the group level has an additional level of complexity, because communication and interaction among people and dependencies between their tasks must be considered. There are two distinct categories:

- Multioperator or crew models rely on very well-defined task descriptions (and are therefore mainly used in military settings) and address issues like global system performance (in terms of mission success or elapsed time), task organization, and crew size. An example is the Siegel-Wolf model (Section 5.2.2).
- Recently, in the context of computer-supported cooperative work (CSCW), models have been developed and are being developed that describe and analyze less defined tasks. Their issues are to understand cooperative processes like group work or collaborative learning and their goal is to design information and communication systems for their support. Models are taken from various backgrounds, e.g., coordination science.
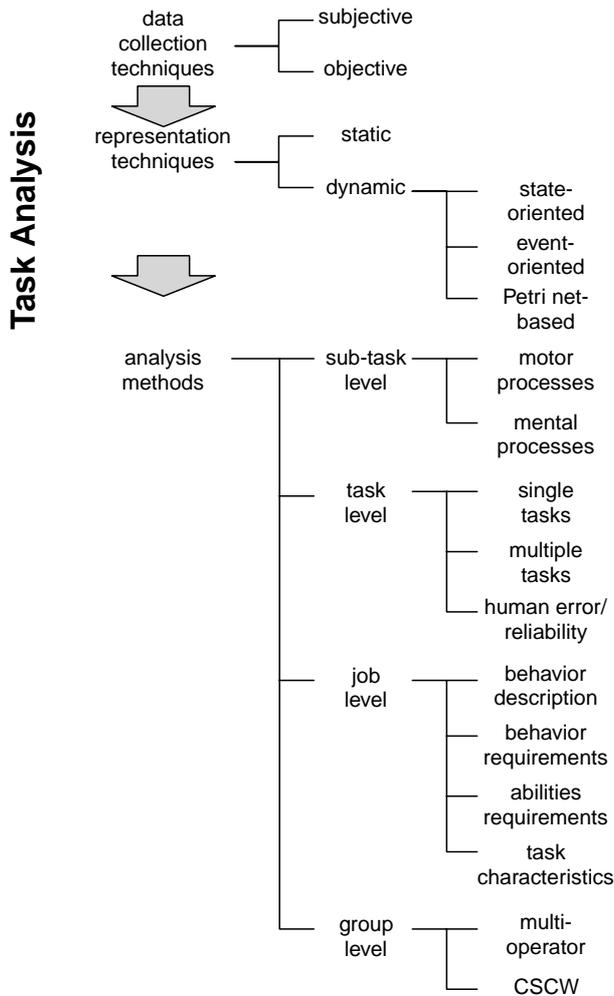
Task Analysis

data collection techniques
— subjective
— objective

⬇

representation techniques
— static
— dynamic
— state-oriented
— event-oriented
— Petri net-based

⬇

analysis methods
— sub-task level
— motor processes
— mental processes
— task level
— single tasks
— multiple tasks
— human error/reliability
— job level
— behavior description
— behavior requirements
— abilities requirements
— task characteristics
— group level
— multi-operator
— CSCW

Figure 1. Task analysis framework.

Contemplating past developments, the focus of task analysis is expanding to higher structural levels, especially collaborative work (figure 1). This will favor techniques and methods with wide applicability or generic nature.

REFERENCES

FLEISHMAN, E.A. and QUAINTANCE, M.K., 1984, *Taxonomies of Human Performance* (Orlando FL: Academic Press).

GAEL, S. (ed.), 1988, *The Job Analysis Handbook for Business, Industry, and Government* (New York: Wiley).

KIRWAN, B. and AINSWORTH L.K., 1992, *A Guide to Task Analysis* (London: Taylor & Francis).

LUCZAK, H., 1997, Task analysis. In Salvendy (ed.) *Handbook of Human Factors and Ergonomics* (New York: Wiley), pp. 340–416.

Work Stress Quantification and Evaluation using ErgoMOST™

J. R. Wilk

Senior Account Manager, H.B. Maynard and Company, Inc.

1. INTRODUCTION

There has been a rising number of ergonomic-related injuries over the past decade. This increase in work-related injuries has sparked aggressive auditing activities by the Occupational Safety and Health Administration (OSHA), and introduced the prospect of a national ergonomic standard. It is clear that ergonomics has become a subject that must concern everyone who performs or manages work.

To aid the management of these ergonomic problems, many companies have assigned the responsibility for their abatement to an individual or team within the organization. In cases where the engineering group or a cross-functional team has been given responsibility to investigate ergonomic concerns, many have employed the use of tools to help them. These tools are viewed as objective measures of ergonomic stress by these teams.

In order to provide an objective ergonomic evaluation tool, H. B. Maynard and Company, Inc. of Pittsburgh, Pennsylvania has developed ErgoMOST™, a software application that provides a direct link between ergonomic and methods analysis. ErgoMOST quantifies and evaluates the ergonomic stress associated with a method and pinpoints potential ergonomic concern. After redesigning a job, ErgoMOST can be used to simulate a method improvement or job rotation that reduces the operator's exposure to ergonomic stress.

This article will discuss how ErgoMOST functions as an essential piece of a comprehensive ergonomics program and as an important evaluation tool. It will demonstrate how ErgoMOST is used to detect ergonomic concerns and simulate and evaluate possible improvements to an operation. Finally, the link between basic methods improvement techniques and the use of ErgoMOST will be explained.

2. ERGOMOST AND AN OVERALL ERGONOMIC PLAN

Ergonomic evaluation is an important part of a company's overall ergonomic program. To do this evaluation, ErgoMOST provides the user with the means to quantify the ergonomic risk of operations. Ergonomic stress evaluation is one part of a comprehensive ergonomic program that should include the following elements:

- employee and workplace audit
- ergonomic stress evaluation
- ergonomic design of products, workplaces, methods
- environmental evaluation and improvement
- ergonomic program organization
- education and training program
- fitness and rehabilitation program
- reporting, feedback, and follow-up

These elements are essential to the successful implementation of ergonomic principles into the everyday operation of a facility. The following is a description of ErgoMOST and how it functions as the ergonomic stress evaluation part of an ergonomic program.

3. ERGOMOST AND ERGONOMIC STRESS EVALUATION

ErgoMOST allows the user to attach ergonomic analysis to work that follows a definable method. This method may reside in a work measurement or standards database or can be entered directly into the system. Given this method, the user applies ergonomic analysis to each method step. This analysis includes force, posture, grip, and vibration information. The force
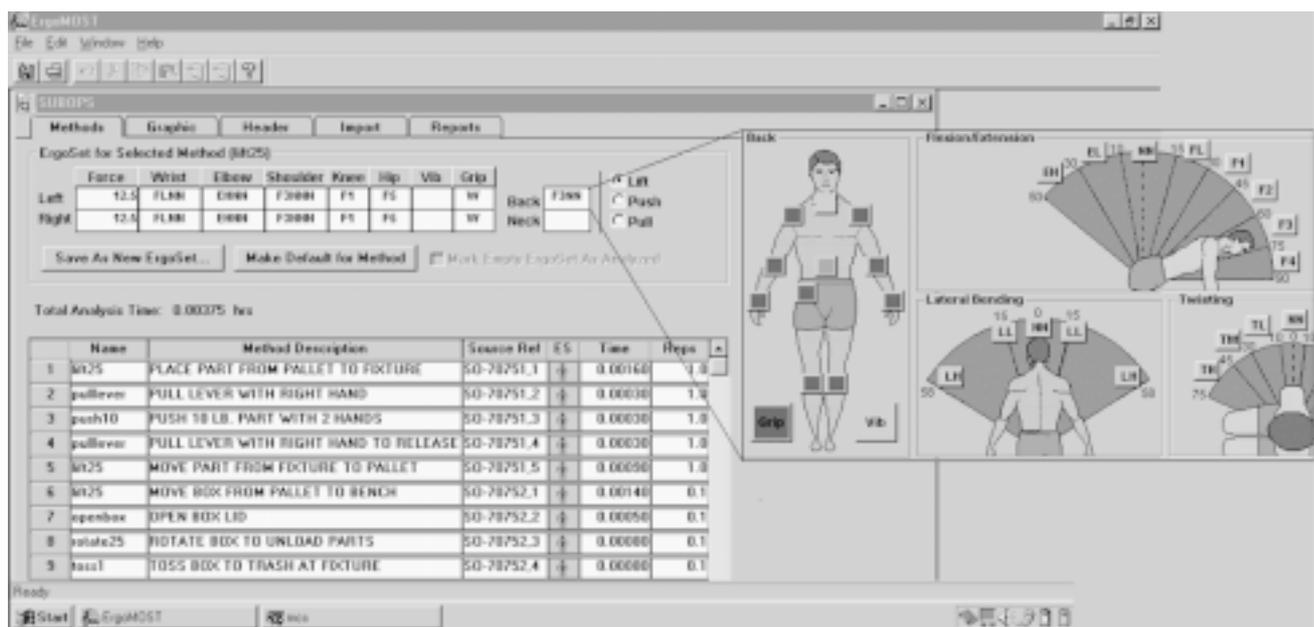

Figure 1: ErgoMOST methods screen with graphic input

Figure 2: Header information

requirement for each method is determined either by weighing the various objects handled or measuring the forces using a hand-held dynamometer, or other instrumentation. Next, the postures of various body members (wrists, elbows, shoulders, back, hips, neck, and knees) are entered for each step of a process. Then, the user classifies the grip that the operator is using in each method step as well as the vibration stress. This analysis is entered as an element called an ErgoSet™. Figure 1 shows the method definition for a job and the ErgoSet for step 1. The graphic in the inset is the posture input for the back.

After each step has an ErgoSet developed or assigned to it, job information is required, such as:
- Production period (hours in the shift)
- Profile of the operator (male or female and the percentile of the population)
- Product quantity produced in the shift

Figure 2 shows the header information with the job information that is required for the calculations.

The ErgoMOST system then evaluates this input to quantify the ergonomic stress of an operation by assigning an ergonomic stress index (ESI) for each body member in terms of force, posture, repetition, grip, and vibration stress. The evaluation criteria for the determination of the ESI values were developed in partnership between H.B. Maynard and Company, Inc. and university research partners. This approach to quantification of ergonomic risk has taken expertise and detailed modeling techniques and presented them to a user in simple terms. An ESI has a 1–5 rating system. These ratings indicate potential risk for each joint in the following manner:

A rating of 1 or 2 represents low ergonomic risk. A rating of 3 represents medium ergonomic risk, and a rating of 4 or 5 represents high ergonomic risk. Although all jobs should be monitored for feedback and improvement, operations with ESIs of four or five should be investigated for immediate ergonomic improvement. These ESI values have been developed and tested

Table 1: Ergonomic stress index table

1–2	Low Risk
3	Medium Risk
4–5	High Risk

to be in accord with current research and common ergonomic tools that are in use today, such as the NIOSH 1991 Lifting Guide.

Figure 3 shows the summary output for one analysis or job. In this case, the major areas of concern are force exposure to the shoulders as well as the posture of the elbows and back. This output reflects a job where there is lifting of heavy weight from a pallet on the floor, processing the part, and putting the finished object to the floor.

Measurement of ergonomic stress should not be considered as an end, but rather as a means for improving operations. Since absolute limits are not known and the variations of dealing with the human body are so diverse, ErgoMOST is recommended for use in a comparative sense. Method steps measuring the highest ergonomic stress levels are targeted for analysis to determine how they can be improved. Figure 4 shows a report that prioritizes the methods, which serves as a guide to making ergonomic improvement. In this case, the high force methods were the part handling steps of the operation. If the steps with the ESI of 4 can be redesigned to eliminate these high ESIs, then the job will be improved for the operator and ergonomic risk will be reduced.

After the initial ErgoMOST analysis has been completed, the ergonomic stressors contributing to the ESI ratings can be pinpointed. Finally, a redesigned solution is completed and simulated to evaluate the ergonomic improvement. Figure 5 shows the comparison of the original job and the redesigned method. The graph shows that the high ESIs for posture have been eliminated. In this case, implementing lift tables on either side of the working area so that the heavy weight could be presented in a better fashion reduced the fours and fives to threes or less.

Pinpointing problems in an operation is beneficial in identifying and documenting specific ergonomic problems of an operation in terms of force, posture, repetition, grip, and vibration. Documentation of problems and improvements is essential in having a complete evaluation of ergonomic stress. Simulating the method improvement allows the team to identify potential ergonomic stress before the operation is put into effect.

ErgoMOST can also analyze a group of processes. Since many ergonomic stressors are alleviated by job rotation, various combinations of tasks can be simulated and analyzed to assign a rating for a group of processes to determine the effects of leveling the peak stresses. ErgoMOST allows the engineer to specify exact quantities produced against each job in a rotation. After this information is determined, as we have seen earlier, the user can get feedback on the areas of ergonomic concern.

4. ERGOMOST AND METHODS IMPROVEMENT

By employing methods improvement techniques, which are a basic part of industrial engineering, one can see a great deal of ergonomic improvement in a short period of time. The following are examples of basic methods engineering principles. The ergonomic analyst should evaluate:
- *Workplace Heights*: These should be adjustable to accommodate the varying heights of operators.
- *Material Placement*: Simple arrangement of materials with consideration given to wrist, elbow and shoulder postures can result in greatly reduced stress. For example, placement of material in a container with an angled opening enables an operator to remove or replace objects from the container with minimal wrist deviation.
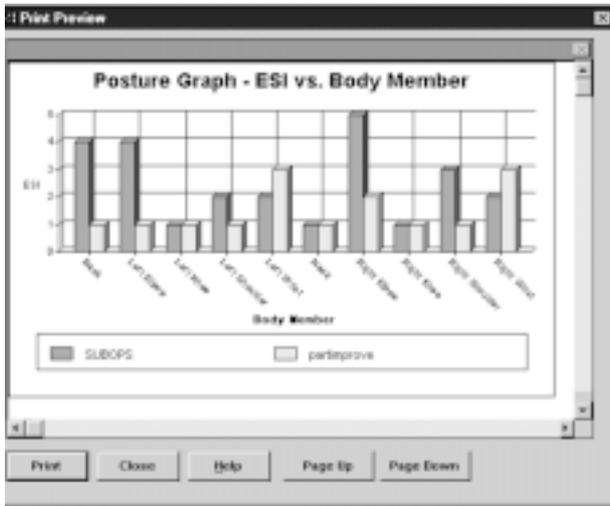
Figure 3: Summary output

Analysis Top Methods of Concern Report

H.B. Maynard and Co.

Header Information

Analysis Name:	SUBOPS		
Description:	PART ASSEMBLY - STATION 1		
Analyzed by:	MCS	Created:	1/11/99 00:00:47
		Last Updated:	4/9/99 00:00:30
Department	ASSEMBLY	Work Center	01-10
Operation No.	010	Material	STEEL
Operator Gender:	F, 50th		
Product Id:	FE2-34	MCS Data Type:	
Production Prd:	8.0	MCS Data Id:	
Product Qty:	1200	MCS Data Issue:	

Analysis	Step	Description	Source Reference	R. Shoulder Acute ESI
SUBOPS	1	PLACE PART FROM PALLET TO FIX	SO-70751,1	4
SUBOPS	3	PUSH 10 LB. PART WITH 2 HANDS	SO-70751,3	4
SUBOPS	5	MOVE PART FROM FIXTURE TO PA	SO-70751,5	4
SUBOPS	6	MOVE BOX FROM PALLET TO BENC	SO-70752,1	4
SUBOPS	9	TOSS BOX TO TRASH AT FIXTURE	SO-70752,4	3
SUBOPS	2	PULL LEVER WITH RIGHT HAND	SO-70751,2	2
SUBOPS	4	PULL LEVER WITH RIGHT HAND TO	SO-70751,4	2
SUBOPS	8	ROTATE BOX TO UNLOAD PARTS	SO-70752,3	2

Figure 4: Top methods of concern

- *Force Exposure*: Updating equipment to use modern actuating devices such as proximity switches rather than palm buttons can reduce forces that pose risk of injury to the operator.
- *Possibility for Job Rotation:* Combinations of jobs can be devised which level out the risk of certain high-stress operations because the operation is distributed between several employees rather than concentrated onto one worker.
- *Tooling Application and Design*: Obvious posture improvements can be seen and measured when a poorly designed tool is replaced with the proper tool or a specially designed tool that meets the requirements of its application.

5. CONCLUSION

ErgoMOST is designed to provide users with a tool that can be applied easily and objectively to quantify and evaluate ergonomic stress involved with an operation. The results are presented in an easy-to-use format. Using the results of the initial analysis, ErgoMOST should be used on a redesigned method to ensure that improvements are implemented on the floor, which provide safe and efficient processes for workers to perform.

Because of the present interest in the relatively new science of ergonomics, and the attention it has received during the recent years, it can be predicted that the knowledge of work-related stress will increase substantially in the next few years. As this knowledge grows, ErgoMOST will be refined as data becomes available, and its capabilities will increase to incorporate the data. More specifically, some of the features that are under consideration for inclusion in ErgoMOST are:

- energy expenditure
- job evaluation checklist

Figure 5: Comparison graph

- enhanced graphics
- ergonomic improvements database
- composite ratings for the entire body
- economic justification of ergonomic improvement

 For ergonomic evaluation to be successful, it is up to researchers to provide the data and guidelines, and industry to apply ergonomics through a properly structured ergonomic program that uses tools which provide useful solutions to the ergonomic problems.

REFERENCES

BHATTACHARYA *et al.*, 1996, *Occupational Ergonomics: Theory and Applications*, (New York: Marcel Dekker).

HODSON, W.K., 1992, *Maynard's Industrial Engineering Handbook*, 4th edn, (New York: McGraw-Hill).

KEYSERLING, W.M., 1992, Ergonomic considerations in job analysis, *Proceedings of the 1992 MOST ® User's Conference*, Pittsburgh, pp. 3–24.

KEYSERLING, W.M., ARMSTRONG, T.J. and PUNETT, L., 1991, Ergonomic job analysis: A structured approval for identifying risk factors associated with overexertion injuries and disorders, *Appl. Occ. Envir. Hyg.*

SANDERS, J., 1993, ErgoMOST as a tool for productivity improvement, *Proceedings of the 1993 MOST® User's Conference, Pittsburgh*, pp. 161–3.

US DEPARTMENT OF LABOR, 1990, Occupational safety and health administration. *Ergonomics Program Management Guidelines for Meatpacking Plants*. OSHA Publication No. 3108. Washington, DC.

ZANDIN, K.B., 1990, *MOST® Work Measurement Systems*, 2nd edn, (New York: Marcel Dekker).

Published Ergonomics Literature

T. Megaw

Ergonomics Information Analysis Centre, The University of Birmingham, UK

1. INTRODUCTION

Despite the accelerating introduction of electronic methods of information provision, printed copy still remains a primary source of ergonomics information and knowledge. The printed material available to ergonomists can be broadly divided into five sources: journals, books, proceedings, reports and standards. Journals contain articles or papers which have been refereed by experts in the reported topic area. This provides a good degree of confidence in the quality of the material. However, there can often be quite a delay, sometimes as long as three years, between the initial submission of a journal article and its publication. Books may either be written or edited by one or more authors. Some books provide a general coverage of all areas of ergonomics, while a majority restrict the areas they cover. Included under the heading of books are handbooks. Although, books are not refereed to the same degree as most journal material, they are usually written by experts in the relevant topic areas. Proceedings include the papers presented at symposia and conferences, often the annual conferences of the various ergonomics societies. The extent that conference papers have been peer-reviewed varies widely. It is rare for conference papers to be reviewed as thoroughly as journal papers and it is common for them not to have gone through any reviewing process at all. For this reason the quality of conference papers varies enormously. The main advantage of conference proceedings is that they provide the reader with a good impression of whatever research is being carried out in a particular topic area at a particular point in time. In this respect, one can judge "the flavor of the month". In some cases conference papers are published as books after undergoing some editorial process. A majority of ergonomics research laboratories and centers produce internal reports of their own work, including annual reports, which they make available, usually on request, to people outside their organizations, provided there are no issues of confidentiality. Such reports enable the readers to appreciate the areas in which the various ergonomics centers specialize as well as to get more detailed information on specific projects that could not be obtained from journals or proceedings. There are now numerous publicly available standards documents (International, European, and National) on several aspects of ergonomics.

The intention here is to provide a comprehensive list of the publicly available literature in respect of journals, books, and proceedings. Because of the multidisciplinary nature of ergonomics, it is naturally difficult to be confident on what should and what should not be included in the list. However, there can be no doubting the relevance of topics such as human–computer interaction, health and safety, and job design. Hence, these areas are well represented in the list.

It is important to emphasize that in compiling the list no judgements have been made on the quality of the publications. The extent that the material has been peer reviewed has already been mentioned. If there are any obvious omissions in the list, I can only apologize.

Finally, it is important to point out that increasingly both journals and conference proceedings are being offered in electronic formats to the extent that some proceedings are often no longer available in a traditional hardcopy format. Currently, many of the main publishers are offering their journals on-line, while proceedings are appearing both in on-line and CD-ROM formats.

2. JOURNALS

These have been divided into five groups:

Primary ergonomics journals

Applied Ergonomics (Amsterdam: Elsevier)

Ergonomics (London: Taylor & Francis)

Ergonomics Abstracts (London: Taylor & Francis)

Human Factors (Santa Monica, California: Human Factors and Ergonomics Society)

International Journal of Industrial Ergonomics (Amsterdam: Elsevier)

Theoretical Issues in Ergonomics Science (London: Taylor & Francis)

Other ergonomics journals

Ergonomia (Italy) (Milan: Moretti Vitali Editori)

Ergonomia (Poland) (Krakow: ORPAN)

Ergonomics Australia (Downer, ACT: The Ergonomics Society of Australia)

Ergonomics in Design (Santa Monica, California: Human Factors and Ergonomics Society)

Ergonomics SA (Grahamstown, SA: Ergonomics Society of South Africa)

Human Factors and Ergonomics in Manufacturing (New York: John Wiley)

International Journal of Cognitive Ergonomics (Mahwah, New Jersey: Lawrence Erlbaum)

International Journal of Occupational Safety and Ergonomics (Warsaw: Central Institute for Labor Protection)

Japanese Journal of Ergonomics (Tokyo: Business Center for Academic Societies Japan)

Journal of Human Ergology (Tokyo: Business Center for Academic Societies Japan)

Occupational Ergonomics (Amsterdam: IOS)

Tijdschrift voor Ergonomie (Nieuwegein: Nederlandse Vereniging voor Ergonomie)

Travail Humain (Paris: Presses Universitaires de France)

Zeitschrift fur Arbeitswissenschaft (Cologne: Otto Schmidt)

Human–computer interaction

ACM Transactions on Computer–Human Interaction (New York: Association for Computing Machinery)

Behaviour and Information Technology (London: Taylor & Francis)

Cognition, Technology & Work (Berlin: Springer)

Communications of the ACM (New York: The Association for Computing Machinery)

Computers in Human Behavior (Amsterdam: Elsevier)

CyberPsychology and Behavior (Larchmont, New York: Mary Ann Liebert)

Human–Computer Interaction (Mahwah, New Jersey: Lawrence Erlbaum)

Interacting with Computers (Amsterdam: Elsevier)

International Journal of Human–Computer Interaction (Mahwah, New Jersey: Lawrence Erlbaum)

International Journal of Human–Computer Studies (London: Academic)

Related health and safety journals

Accident Analysis and Prevention (Amsterdam: Elsevier)

International Archives of Occupational and Environmental Health (Berlin: Springer)

International Journal of Occupational Medicine and Environmental Health (Lodz: Nofer Institute of Occupational Medicine)

Journal of Loss Prevention in the Process Industries (Amsterdam: Elsevier)

Journal of Occupational and Environmental Medicine (Philadelphia, Pennsylvania: Lippincott Williams & Wilkins)

Journal of Safety Research (Itasca, Illinois: National Safety Council and Amsterdam: Elsevier)

Journal of Science of Labour (Kawasaki: Institute for Science of Labour)

Medicina del Lavoro (Fidenza: Casa Editrice Mattioli)

Medicina y Seguridad del Trabajo (Madrid: Instituto Nacional de Medicina Seguridad del Trabajo)

Occupational Medicine (London: Lippincott Williams & Wilkins)

Occupational and Environmental Medicine (London: BMJ Publishing)

Reliability Engineering and System Safety (Amsterdam: Elsevier)

Risk Analysis (New York: Plenum Press)

Safety Science (Amsterdam: Elsevier)

Scandinavian Journal of Work, Environment & Health (Helsinki: Finnish Institute of Occupational Health)

Work and Stress (London: Taylor & Francis)

Other related journals

Applied Cognitive Psychology (Chichester: Wiley)

Aviation, Space, and Environmental Medicine (Alexandria, Virginia: Aerospace Medical Association)

Clinical Biomechanics (Amsterdam: Elsevier)

Design Studies (Amsterdam: Elsevier)

Displays (Amsterdam: Elsevier)

European Journal of Applied Physiology and Occupational Physiology (Berlin: Springer)

IEEE Transactions on Professional Communication (New York: Institute of Electrical and Electronic Engineering)

IEEE Transactions on Systems, Man, and Cybernetics. Part A: Systems and Humans (New York: Institute of Electrical and Electronic Engineering)

Information Design Journal (Reading: IDJ)

Information Technology & People (Bradford: MCB University Press)

International Journal of Aviation Psychology (Mahwah, New Jersey: Lawrence Erlbaum)

International Journal of Lighting Research and Technology (London: Chartered Institution of Building Services Engineers)

Journal of Aging and Physical Activity (Champaign, Illinois: Human Kinetics)

Journal of Applied Psychology (Washington DC: American Psychological Association)

Journal of Biomechanics (Amsterdam: Elsevier)

Journal of Biological Rhythms (Thousand Oaks, California: Sage Science Press)

Journal of Experimental Psychology: Applied (Washington DC: American Psychological Association)

Journal of Human Performance in Extreme Environments (Dayton, Ohio: Society for Human Performance in Extreme Environments)

Journal of the Illuminating Engineering Society (New York: Illuminating Engineering Society of North America)

Journal of Occupational and Organizational Psychology (Leicester: British Psychological Society)

Journal of Organizational Behavior (Chichester: Wiley)

Medical Informatics and the Internet in Medicine (London: Taylor & Francis)

Military Psychology (Mahwah, New Jersey: Lawrence Erlbaum)

New Technology, Work and Employment (Oxford: Blackwell Publishers)

Ophthalmic and Physiological Optics (Amsterdam: Elsevier)

Perceptual and Motor Skills (Missoula, Montana: Perceptual and Motor Skills)

Personal Technologies (Berlin: Springer)

Presence (Cambridge, Massachusetts: MIT Press)

Proceedings of the Chartered Institution of Building Services Engineers Series A (London: Chartered Institution of Building Services Engineers)

Psychophysiology (Cambridge: Cambridge University Press)

Quarterly Journal of Experimental Psychology, Section A: Human Experimental Psychology (London: Taylor & Francis)

Spine (Philadelphia, Pennsylvania: Lippincott Williams & Wilkins)

Virtual Reality (Godalming, Surrey: Springer)

Visible Language (Providence, Rhode Island: Rhode Island School of Design)

3. BOOKS

These have been grouped under the headings shown below. Naturally, the headings are not mutually exclusive. Accordingly, the books have been classified under the heading which most closely reflects their primary contents.

General ergonomics

Human characteristics and performance

Anthropometry

Ageing and disability

Workplace, equipment, and product design

The physical environment

Display and control design

Human–computer interaction

Complex systems

Transport systems

Musculoskeletal injuries and manual materials handling

Health and safety

Human error and reliability

Job design and work organisation

Impact of information technology

Methods, techniques and measurement

GENERAL ERGONOMICS

ADAMS, J.A., 1989, *Human Factors Engineering* (New York: Macmillan) ISBN 0 02 300370 7.

ALEXANDER, D.C. and PULAT, B.M., 1985, *Industrial Ergonomics. A Practitioner's Guide* (Norcross, Georgia: Industrial Engineering & Management Press, Institute of Industrial Engineers) ISBN 0 89806 073 7.

ALEXANDER, D.C. (Ed), 1999, *Applied Ergonomics: Case Studies Volume 1* (Norcross, Georgia: Industrial Engineering & Management Press) ISBN 0 89806 214 4.

ALEXANDER, D.C. (Ed), 1999, *Applied Ergonomics: Case Studies Volume 2* (Norcross, Georgia: Industrial Engineering & Management Press) ISBN 0 89806 214 4.

ANDRE, T.S. and SCHOPPER, A.W. (eds), 1997, *Human Factors Engineering in System Design* (Wright-Patterson Air Force Base, Ohio: Crew System Ergonomics Information Analysis Center) no ISBN.

BHATTACHARYA, A. and McGLOTHLIN, J.D. (eds), 1996, *Occupational Ergonomics: Theory and Applications* (New York: Marcel Dekker) ISBN 0 8247 9419 2.

BRIDGER, R.S., 1995, *Introduction to Ergonomics* (New York: McGraw-Hill) ISBN 0 07 007741 X.

BRUNE, D., GERHARDSSON, G., CROCKFORD, G.W. and D'AURIA, D. (eds), 1997, *The Workplace*, 2 Volumes (Geneva: ILO and Oslo: Scandinavian Science Publisher) ISBN 82 91833 00 1.

BULLOCK, M.I. (ed.), 1990, *Ergonomics. The Physiotherapist in the Workplace* (Edinburgh: Churchill Livingstone) ISBN 0 443 03612 8.

BURGESS, J.H., 1989, *Human Factors in Industrial Design. The Designer's Companion* (Blue Ridge Summit, Pennsylvania: TAB Professional and Reference Books) ISBN 0 8306 3356 1.

BURKE, M., 1992, *Applied Ergonomics Handbook* (Boca Raton, Florida: Lewis Publishers) ISBN 0 87371 367 2.

CHAPANIS, A., 1965, *Man–Machine Engineering* (Belmont, California: Wadsworth), no ISBN.

CHAPANIS, A., 1996, *Human Factors in Systems Engineering* (New York: John Wiley) ISBN 0 471 13782 0.

CHAPANIS, A., 1999, *The Chapanis Chronicles* (Santa Barbara, California: Aegean Publishing) ISBN 0 9636178 9 3.

DUL, J. and WEERDMEESTER, B., 1993, *Ergonomics for Beginners. A Quick Reference Guide* (London: Taylor & Francis) ISBN 0 7484 0079 6.

Eastman Kodak Company, 1983, *Ergonomic Design for People at Work. Volume 1* (Belmont, California: Lifetime Learning Publications) ISBN 0 534 97962 9.

Eastman Kodak Company, 1986, *Ergonomic Design for People at Work. Volume 2* (New York: Van Nostrand Rheinhold) ISBN 0 442 22103 7.

FRASER, T.M., 1989, *The Worker at Work* (London: Taylor & Francis) ISBN 0 85066 481 0.

FRASER, T.M., 1992, *Fitness for Work* (London: Taylor & Francis) ISBN 0 85066 858 1.

FRASER, T.M., 1996, *Introduction to Industrial Ergonomics: A Textbook for Students and Managers* (Toronto: Wall & Emerson) ISBN 0 921332 46 7.

FRASER, T.M. and PITYN, P.J., 1994, *Work, Productivity, and Human Performance* (Springfield, Illinois: Charles C. Thomas) ISBN 0 398 05910 1.

GALER, I. (ed.), 1987, *Applied Ergonomics Handbook* (2nd Edition) (London: Butterworths) ISBN 0 408 00880 6.

GAY, K., 1986, *Ergonomics. Making Products and Places Fit People* (Hillside, New Jersey: Enslow) ISBN 0 89490 118 4.

GARNER, S., 1991, *Human Factors* (Oxford: Oxford University Press) ISBN 0 19 832783 8.

HANCOCK, P.A. (ed.), 1987, *Human Factors Psychology* (Amsterdam: North-Holland) ISBN 0 444 70319 5.

HELANDER, M., 1995, *A Guide to the Ergonomics of Manufacturing* (London: Taylor & Francis) ISBN 07484 0122 9.

HELANDER, M. and NAGAMACHI, M. (eds), 1992, *Design for Manufacturability: A Systems Approach to Concurrent Engineering and Ergonomics* (London: Taylor & Francis) ISBN 0 7484 0009 5.

HORRIGAN, K.R., O'SULLIVAN, J.J. and WHITING, J.F., 1987, *The NSCA Ergonomics Primer* (Brisbane: National Safety Council of Australia, Queensland Branch) ISBN 0 909670 19 6.

IVERGARD, T., 1982, *Information Ergonomics* (Bromley, Kent: Chartwell-Bratt) ISBN 0 86238 032 4.

JACOBS, K., 1999, *Ergonomics for Therapists* (2nd Edition) (Boston: Massachusetts Butterworth-Heinemann) ISBN 0 7506 7051 7.

KANTOWITZ, B.H. and SORKIN, R.D., 1983, *Human Factors: Understanding People–System Relationships* (Chichester: Wiley) ISBN 0 471 09595 X.

KARWOWSKI, W. and SALVENDY, G. (eds), 1998, *Ergonomics in Manufacturing: Raising Productivity Through Workplace Improvement* (Dearborn, Michigan: Society of Manufacturing Engineers, and Norcross, Georgia: Engineering & Management Press) ISBN 0 87263 485 X.

KARWOWSKI, W. and MARRAS, W.S., 1999, *The Occupational Ergonomics Handbook* (Boca Raton, Florida: CRC Press) ISBN 0 8493 2641 9.

KLEMMER, E.T. (ed.), 1989, *Ergonomics. Harness the Power of Human Factors in your Business* (Norwood, New Jersey: Ablex) ISBN 0 89391 559 9.

KOHN, J. P., 1997, *The Ergonomic Casebook: Real World Solutions* (Boca Raton, Florida: Lewis Publishers) ISBN 1 56670 269 0.

KRAGT, H. (ed.), 1992, *Enhancing Industrial Performance: Experiences of Integrating the Human Factor* (London: Taylor & Francis) ISBN 0 7484 0035 4.

KROEMER, K.H.E. and GRANDJEAN, E. (eds), 1997, *Fitting the Task to the Human: A Textbook of Occupational Ergonomics* (5th ed.) (London: Taylor & Francis) ISBN 0 7484 0665 4.

KROEMER, K.H.E., KROEMER, H.B. and KROEMER-ELBERT, K.E., 1994, *Ergonomics: How to Design for Ease and Efficiency* (Englewood Cliffs, New Jersey: Prentice Hall) ISBN 0 13 278359 2.

LIKER, J.K. and CHAFFIN, D.B., 1991, *Fitting Jobs to People—An Ergonomics Process: Job Improvement Guide* (Ann Arbor, Michigan: University of Michigan), no ISBN.

MAULE, H.G. and WEINER, J.S. (eds), 1981, *Design for Work and Use* (London: Taylor & Francis) ISBN 0 85066 208 7.

MEISTER, D., 1989, *Conceptual Aspects of Human Factors* (Baltimore: John Hopkins University Press) ISBN 0 8018 3742 4.

MEISTER, D., 1999, The *History of Human Factors and Ergonomics* (Mahwah, New Jersey: Lawrence Erlbaum) ISBN 0 8058 2768 4.

MUCKLER, F.A. (ed.), 1984, *Human Factors Review: 1984* (Santa Monica, California: Human Factors and Ergonomics Society), no ISBN.

MURRELL, H., 1969, *Ergonomics* (London: Chapman & Hall) ISBN 412 07800 7.

MURRELL, H., 1976, *Men and Machines* (London: Methuen) ISBN 0 416 82310 6.

NICKERSON, R.S., 1992, *Looking Ahead: Human Factors Challenges in a Changing World* (Hillsdale, New Jersey: Lawrence Erlbaum) ISBN 0 8058 1151 6.

NOULIN, M., 1992, *Ergonomie* (Paris: Techniplus) ISBN 2 85232 733 3.

OBORNE, D.J., 1995, *Ergonomics at Work* (3rd ed.) (Chichester: Wiley) ISBN 0 471 95235 4.

OBORNE, D.J. (ed.), 1987, *International Reviews of Ergonomics, Volume 1* (London: Taylor & Francis) ISBN 0 85066 359 8.

OBORNE, D.J. (ed.), 1988, *International Reviews of Ergonomics, Volume 2* (London: Taylor & Francis) ISBN 0 85066 408 X.

OBORNE, D.J. (ed.), 1989, *International Reviews of Ergonomics, Volume 3* (London: Taylor & Francis), no ISBN.

OBORNE, D. (ed.), 1995, *Ergonomics and Human Factors, Volumes 1 and 2* (Aldershot, Hampshire: Edward Elgar) ISBN 1 85278 757 0.

OBORNE, D.J., BRANTON, R., LEAL, F., SHIPLEY, P. and STEWART, T. (eds), 1993, *Person-Centred Ergonomics: A Brantonian View of Human Factors* (London: Taylor & Francis) ISBN 0 7484 0051 6.

OZOK, A.F. and SALVENDY, G. (eds), 1996, *Advances in Applied Ergonomics* (West Lafayette, Indiana: USA Publishing).

PELSMA, K.H. (ed.), 1987, *Ergonomics Sourcebook* (Lawrence, Kansas: Report Store) ISBN 0 916313 12 3.

PHEASANT, S., 1987, *Ergonomics - Standards and Guidelines for Designers*

(Milton Keynes, Bedford: British Standards Institution) ISBN 0 580 15391 6.

PHEASANT, S., 1991, *Ergonomics, Work and Health* (Basingstoke: Macmillan) ISBN 0 333 48998 5.

PROCTOR, R.W. and VAN ZANDT, T., 1994, *Human Factors in Simple and Complex Systems* (Boston, Massachusetts: Allyn and Bacon) ISBN 0 205 13999 X.

PULAT, B.M., 1992, *Fundamentals of Industrial Ergonomics* (Englewood Cliffs, New Jersey: Prentice Hall) ISBN 0 13 345364 2.

PULAT, B. and ALEXANDER, D.C. (eds), 1991, *Industrial Ergonomics: Case Studies* (Norcross, Georgia: Industrial Engineering and Management Press, IIE) ISBN 0 89806 099 0.

ROBERTS, D.L. and BECKER, W.J. (eds), 1991, *Human Factors: A Series of Quality Instructional Materials* (St Joseph, Michigan: American Society of Agricultural Engineers) ISBN 0 929355 18 0.

RUBINSTEIN, R. and HERSH, H., 1984, *The Human Factor. Designing Computer Systems for People* (Bedford, Massachusetts: Digital Press) ISBN 0 932376 44 4.

SALVENDY, G. (ed.), 1992, *Handbook of Industrial Engineering* (2nd ed.) (Chichester: Wiley) ISBN 0 471 50276 6.

SALVENDY, G. (ed.), 1997, *Handbook of Human Factors and Ergonomics* (2nd ed.) (New York: John Wiley) ISBN 0 471 11690 4.

SANDERS, M.S. and MCCORMICK, E.J., 1976, *Workbook for Human Factors in Engineering and Design* (Dubuque, Iowa: Kendall/Hunt Publishing) ISBN 0 8403 1624 0.

SANDERS, M.S. and MCCORMICK, E.J., 1992, *Human Factors in Engineering and Design* (7th ed.) (New York: McGraw-Hill) ISBN 0 07 112826 3.

SELAN, J.L. (ed.), 1994, *The Advanced Ergonomics Manual* (Dallas, Texas: Advanced Ergonomics, Inc), no ISBN.

SHEPHARD, R.J., 1974, *Men at Work: Applications of Ergonomics to Performance and Design* (Springfield, Illinois: Charles C. Thomas) ISBN 0 398 02965 2.

SINTSCHENKO, W.P., MUNIPOW, W.M., and SMOLJAN, G.L., 1976, *Ergonomische Grundlagen der Arbeitsorganisation* (Berlin: VEB Deutscher Verlag der Wissenschaften), no ISBN.

Standards Australia, 1994, *Ergonomics—The Human Factor: A Practical Approach to Work Systems Design* (Homebush, New South Wales: Standards Australia) ISBN 0 7262 9193 5.

STRAMLER, J.H., 1993, *The Dictionary for Human Factors/Ergonomics* (Boca Raton, Florida: CRC Press) ISBN 0 8493 4236 8.

Swedish Work Environment Fund, 1988, *Making the Job Easier. An Ergonomics Idea Book* (Chicago: National Safety Council), no ISBN.

TAYYARI, F. and SMITH, J.L., 1997, *Occupational Ergonomics: Principles and Applications* (London: Chapman & Hall) ISBN 0 412 58650 9.

VENTURINO, M. (ed.), 1990, *Selected Readings in Human Factors* (Santa Monica, California: Human Factors and Ergonomics Society) ISBN 0 945289 00 6.

WEIMER, J., 1993, *Handbook of Ergonomic and Human Factors Tables* (Englewood Cliffs, New Jersey: PTR Prentice Hall) ISBN 0 13 374174 5.

WEINER, J.S. and MAULE, H.G. (eds), 1977, *Human Factors in Work, Design and Production, Volume 1* (London: Taylor & Francis) ISBN 0 85066 076 9.

WICKENS, C.D., GORDON, S.E. and LIU, Y., 1998, *An Introduction to Human Factors Engineering* (New York: Longman:) ISBN 0 321 01229 1.

WOODSON, W.E., 1987, *Human Factors Reference Guide for Electronics and Computer Professionals* (New York: McGraw-Hill) ISBN 0 07 071766 4.

WOODSON, W.E., TILLMAN, B. and TILLMAN, P., 1992, *Human Factors*

Design Handbook (2nd ed.) (New York: McGraw-Hill) ISBN 0 07 071768 0.

ZINCHENKO, V. and MUNIPOV, V., 1989, *Fundamentals of Ergonomics* (Moscow: Progress Publishers) ISBN 5 01 001137 9.

Human characteristics and performance

ASTRAND, P.O. and RODAHL, K., 1986, *Textbook of Work Physiology* (3rd ed.) (New York: McGraw Hill) ISBN 0 07 100114 X.

BAKER, M.A. (ed.), 1987, *Sex Differences in Human Performance* (Chichester: Wiley) ISBN 0 471 91119 4.

BOFF, K.R., KAUFMAN, L. and THOMAS, J.P. (eds), 1986, *Handbook of Perception and Human Performance. Volume 1 Sensory Processes and Perception* (New York: Wiley) ISBN 0 471 88544 4.

BOFF, K.R., KAUFMAN, L. and THOMAS, J.P. (eds), 1986, *Handbook of Perception and Human Performance. Volume 2 Cognitive Processes and Performance* (New York: Wiley) ISBN 0 471 82957 9.

BOFF, K.R. and LINCOLN, J.E. (eds), 1988, *Engineering Data Compendium. Human Perception and Performance. Volumes 1, 2, and 3,* and *User's Guide* (Ohio: Harry G. Armstrong Aerospace Medical Research Laboratory, Wright-Patterson Air force Base), no ISBN.

CANNON-BOWERS, J.A. and SALAS E. (eds), 1998 *Making Decisions Under Stress: Implications for Individual and Team Training* (Washington, DC: National Academy Press) ISBN 0 55798 525 1.

CHAFFIN, D.B. ANDERSSON, G.B.J. and MARTIN, B.J., 1991, *Occupational Biomechanics* (3rd ed.) (New York: John Wiley) ISBN 0 471 24697 2.

DAMOS, D.L. (ed.), 1991, *Multiple-Task Performance* (London: Taylor & Francis) ISBN 0 85066 757 7.

DAVIES, D.R. and PARASURAMAN, R., 1982, *The Psychology of Vigilance* (London: Academic Press) ISBN 0 12 206180 2.

DENNIS, I. and TAPSFIELD, P. (eds), 1996, *Human Abilities: Their Nature and Measurement* (Mahwah, New Jersey: Lawrence Erlbaum) ISBN 0 8058 1800 6.

DRISKELL, J.E. and SALAS, E. (eds), 1996, *Stress and Human Performance* (Mahwah, New Jersey: Lawrence Erlbaum) ISBN 0 8058 1182 6.

ENDSLEY, M. R. and GARLAND, D. J., (Eds), 2000, *Situational Awareness Analysis and Measurement* (Mahwah, New Jersey: Lawrence Erlbaum Associates) ISBN 0 8058 3701 1.

ENGESTROM, Y. and MIDDLETON, D. (eds), 1996, *Cognition and Communication at Work* (Cambridge: Cambridge University Press) ISBN 0 521 44104 8.

FLEISHMAN, E.A. and QUAINTANCE, M.K., 1984, *Taxonomies of Human Performance. The Description of Human Tasks* (London: Academic Press) ISBN 0 12 260450 4.

FLIN, R., SALAS, E., STRUB, M. and MARTIN, L. (eds), 1999, *Decision Making Under Stress: Emerging Themes and Applications* (Aldershot, Hampshire: Ashgate) ISBN 0 291 39856 1.

GAL, R. and MANGELSDORFF, A.D. (eds), 1991, *Handbook of Military Psychology* (Chichester: Wiley) ISBN 0 471 92045 2.

GAWRON, V. J., 2000, *Human Performance Measures Handbook* (Mahwah, New Jersey: Lawrence Erlbaum Associates) ISBN 0 8058 2134 1.

GENTNER, D. and STEVENS, A.L. (eds), 1983, *Mental Models* (Hillsdale, New Jersey: Lawrence Erlbaum) ISBN 0 89859 242 9.

HALPERN, D.F., 1992, *Sex Differences in Cognitive Abilities* (2nd ed.) (Hillsdale, New Jersey: Lawrence Erlbaum) ISBN 0 8058 0845 0.

HANCOCK, P.A. (ed.), 1999, *Human Performance and Ergonomics: Perceptive and Cognitive Principles* (2nd ed.) (New York: Academic Press) ISBN 0 12 322735 6.

HANCOCK, P.A. and MESHKATI, N. (ed.), 1988, *Human Mental Workload* (Amsterdam: North-Holland) ISBN 0 444 70388 8.

HARTLEY, J. (ed.), 1980, *The Psychology of Written Communications* (London: Kogan Page) ISBN 0 85038 282 3.

HOLDING, D.H., 1989, *Human Skills* (2nd ed.) (Chichester: Wiley) ISBN 0 471 92076 2.

HUEY, B.M. and WICKENS, C.D. (eds), 1993, *Workload Transition: Implications for Individual and Team Performance* (Washington, DC: National Academy Press) ISBN 0 309 04796 X.

KELLER, C.M. and KELLER, J.D., 1996, *Cognition and Tool Use: The Blacksmith at Work* (Cambridge: Cambridge University Press) ISBN 0 521 55239 7.

KROEMER, K.H.E., KROEMER, H.J. and KROEMER-ELBERT, K.E., 1990, *Engineering Physiology. Bases of Human Factors/Ergonomics* (2nd ed.) (New York: Van Nostrand Reinhold) ISBN 0 442 00354 4.

LEPLAT, J., 1992, *L'Analyse du Travail en Psychologie Ergonomique* (Toulouse: Octares eds) ISBN 2 906769 05 3.

MACKIE, R.R. (ed.), 1977, *Vigilance* (New York: Plenum) ISBN 0 306 32883 6.

MCARDLE, J.J. and WOODCOCK, R.W. (eds), 1998, *Human Cognitive Abilities in Theory and Practice* (Mahwah, New Jersey: Lawrence Erlbaum) ISBN 0 8058 2717 X.

MATTHEWS, G., DAVIES, D. R., WESTERMAN, S. J. and STAMMERS, R. B., 2000, *Human Performance: Cognition, Stress and Individual Differences* (Hove, East Sussex: Psychology Press) ISBN 0 415 04407 03.

MONK, A. (ed.), 1991, *Sleep, Sleepiness and Performance* (Chichester: Wiley) ISBN 0 471 93002 4.

MANTOVANI, G., 1996, *New Communication Environments: From Everyday to Virtual* (London: Taylor & Francis) ISBN 0 7484 0396.

MONTGOMERY, H. and SVENSON, O. (eds), 1989, *Process and Structure in Human Decision Making* (Chichester: Wiley) ISBN 0 471 91977 2.

MORAY, N. (ed.), 1979, *Mental Workload: Its Theory and Measurement* (New York: Plenum Press) ISBN 0 306 40144 4.

NORTH, R.V., 1993, *Work and the Eye* (Oxford: Oxford University Press) ISBN 0 19 261885 7.

OZKAYA, N. and NORDIN, M., 1991, *Fundamentals of Biomechanics* (New York: Van Nostrand Reinhold) ISBN 0 442 00313 7.

PERVIN, L. A. and JOHN, O. P. (Eds), 1999, *Handbook of Personality: Theory and Research* (New York: The Guildford Press) ISBN 0 357230 483 7.

POULTON, E.C., 1974, *Tracking Skill and Manual Control* (New York: Academic Press) ISBN 0 12 563550 8.

PROCTOR, R.W. and DUTTA, A., 1995, *Skill Acquisition and Human Performance* (Thousand Oaks, California: SAGE Publications) ISBN 0 8039 5010 1.

RASMUSSEN, J., 1986, *Information Processing and Human–Machine Interaction* (Amsterdam: North-Holland) ISBN 0 444 00987 6.

RODAHL, K., 1989, *The Physiology of Work* (London: Taylor & Francis) ISBN 0 85066 483 7.

SCHMIDT, R.A. and LEE, T.D., 1999, *Motor Control and Learning: A Behavioral Emphasis* (3rd ed.) (Champaign, Illinois: Human Kinetics) ISBN 0 88011 484 3.

SHEPHARD, R.J., 1979, *Human Physiological Work Capacity* (Cambridge: Cambridge University Press) ISBN 0 521 21781 4.

SHEPHARD, R.J., 1994, *Aerobic Fitness and Health* (Champaign, Illinois: Human Kinetics Publishers) ISBN 0 87322 417 5.

SINGLETON, W.T. (ed.), 1978, *Analysis of Practical Skills* (Lancaster: MTP Press) ISBN 0 85200 199 1.

SINGLETON, W.T., 1982, *The Body at Work: Biological Ergonomics* (Cambridge: Cambridge University Press) ISBN 0 521 24087 5.

SINGLETON, W.T., 1989, *The Mind at Work. Psychological Ergonomics* (Cambridge: Cambridge University Press) ISBN 0 521 26579 7.

SMITH, A.P. and JONES, D.M. (eds), 1992, *Handbook of Human Performance, Volumes 1, 2 and 3* (Academic: London) Vol. 1 ISBN 0 12 650351 6; Vol. 2 ISBN 0 12 650352 4; Vol. 3 ISBN 0 12 650353 2.

SMITH, G. and ATCHISON, D.A., 1997, *The Eye and Visual Optical Instruments* (Cambridge: Cambridge University Press) ISBN 0 521 47820 0.

SWETS, J.A., 1996, *Signal Detection Theory and ROC Analysis in Psychology and Diagnostics: Collected Papers* (Mahwah, New Jersey: Lawrence Erlbaum) ISBN 0 8058 1834 0.

TOZEREN, A., 2000, *Human Body Dynamics: Classical Mechanics and Human Movement* (New York: Springer-Verlag) ISBN 0 387 98801 7.

VAN DER VEER, G.C., BAGNARA, S., and KEMPEN, G.A.M., 1992, *Cognitive Ergonomics: Contributions from Experimental Psychology* (Amsterdam: North-Holland) ISBN 0 444 89504 3.

WALKER, A.E., 1997, *The Menstrual Cycle* (London: Routledge) ISBN 0 415 16331 5.

WARR, P. (ed.), 1987, *Psychology at Work* (3rd ed.) (Harmondsworth, Middlesex: Penguin) ISBN 0 14 022654 0.

WATERHOUSE, J.M., MINORS, D.S., and WATERHOUSE, M.E., 1990, *Your Body Clock* (Oxford: Oxford University Press) ISBN 0 19 261982 9.

WEBB, W.B. (ed.), 1982, *Biological Rhythms, Sleep and Performance* (Chichester: Wiley) ISBN 0 471 10047 1.

WHITING, W.C. and ZERNICKE, R.F., 1998, *Biomechanics of Musculoskeletal Injury* (Champaign, Illinois: Human Kinetics) ISBN 0 87322 779 4.

WICKENS, C.D. and HOLLANDS, J.G., 2000, *Engineering Psychology and Human Performance* (3rd Edition) (Upper Saddle River, New Jersey: Prentice Hall) ISBN 0 321 04711 7.

WILKINS, A.J., 1995, *Visual Stress* (Oxford: Oxford University Press) ISBN 0 19 852174 X.

ZINCHENKO, V.P., LEONOVA, A.B., and STRELKOV, Y.K., 1985, *The Psychometrics of Fatigue* (London: Taylor & Francis) ISBN 0 85066 258 3.

ZSAMBOK, C.E. and KLEIN, G. (eds), 1997, *Naturalistic Decision Making* (Mahwah, New Jersey: Lawrence Erlbaum) ISBN 0 8058 1874 X.

Anthropometry

DIFFRIENT, N., TILLEY, A.R., and BARDAGJY, J.C., 1975, *Humanscale 1/2/3* (Cambridge, Mass.: MIT Press) ISBN 0 262 54027 4.

DIFFRIENT, N., TILLEY, A.R. and HARMAN, D., 1981, *Humanscale 4/5/ 6* (Cambridge, Mass.: MIT Press) ISBN 0 262 04059 X.

DIFFRIENT, N., TILLEY, A.R. and HARMAN, D., 1981, *Humanscale 7/8/ 9* (Cambridge, Mass.: MIT Press) ISBN 0 262 04059 X.

DREYFUSS, H., 1967, *The Measure of Man; Human Factors in Design* (2nd ed.) (New York: Whitney Library of Design), no ISBN.

KEMPER, H.C.G. (ed.), 1995, *The Amsterdam Growth Study* (Champaign, Illinois: Human Kinetics) ISBN 0 87322 507 4.

LOHMAN, T., ROCHE, A.F., and MARTORELL, R. (eds), 1988, *Anthropometric Standardization Reference Manual* (Champaign, Illinois: Human Kinetics Books) ISBN 0 87322 121 4.

LOHMAN, T.G., ROCHE, A.F., and MARTORELL, R. (eds), 1988, *Anthropometric Standardization Reference Manual* (Abridged ed.) (Champaign, Illinois: Human Kinetics Books) ISBN 0 87322 331 4.

National Aeronautics & Space Administration (NASA), 1978, *Anthropometric Source Book* (Washington, DC: NASA), no ISBN.

NORRIS, B. and WILSON, J.R., 1995, *Childata: The Handbook of Child Measurements and Capabilities* (London: Consumer Safety Unit, Department of Trade and Industry), no ISBN.

PANERO, J. and ZELNIK, M., 1979, *Human Dimensions and Interior Space* (London: Architectural Press) ISBN 0 8230 7271 1.

PARKER, J.F. and WEST, V.R. (eds), 1973, *Bioastronautics Data Book* (2nd ed.) (Washington, DC: NASA).

PEEBLES, L. and NORRIS, B., 1998, *Adultdata: The Handbook of Adult Anthropometric and Strength Measurements – Data for Design Safety* (London: Government Consumer Safety Research, Department of Trade and Industry), no ISBN.

PHEASANT, S., 1984, *Anthropometrics: An Introduction for Schools and Colleges* (London: British Standards Institution), no ISBN.

PHEASANT, S., 1996, *Bodyspace: Anthropometry, Ergonomics and Design* (2nd ed.) (London: Taylor & Francis) ISBN 0 7484 0326 4.

ROEBUCK, J.A., 1995, *Anthropometric Methods: Designing to Fit the Human Body for Human Factors* (Santa Monica, California: Human Factors and Ergonomics Society) ISBN 0 945289014.

ROEBUCK, J.A., KROEMER, K.H.E., and THOMSON, W.G., 1975, *Engineering Anthropometry Methods* (London: Wiley) ISBN 0 471 72975 2.

SCHMIDTKE, H. (ed.), 1984, *Ergonomic Data for Equipment Design* (New York: Plenum Press) ISBN 0 306 41893 2.

STEENBEKKERS, L.P.A., 1993, *Child Development, Design Implications and Accident Prevention* (Delft: Delft University Press) ISBN 90 6275 895 9.

TILLEY, A.R., 1993, *The Measure of Man and Woman* (New York: Whitney Library of Design) ISBN 0 8230 3031 8.

ULIJASZEK, S.J. and MASCIE-TAYLOR, C.G.N. (eds), 1994, *Anthropometry: The Individual and the Population* (Cambridge: Cambridge University Press) ISBN 0 521 41798 8.

Ageing and disability

CRAIK, F.I.M. and SALTHOUSE, T.A.. (Eds), 2000, *The Handbook af Aging and Cognition* (2nd Edition) (Mahwah, New Jersey: Lawrence erlbaum Associates) ISBN 0 8058 2966 0.

CULLEN, K. and ROBINSON, S., 1997, *Telecommunications for Older People and Disabled People in Europe: Preparing for the Information Society* (Amsterdam: IOS Press) ISBN 90 5199 346 3.

FISK, A.D. and ROGERS, W.A. (eds), 1997, *Handbook of Human Factors and the Older Adult* (San Diego, California: Academic Press) ISBN 0 12 257680 2.

EDWARDS, A.D.N. (ed.), 1995, *Extra-Ordinary Human–Computer Interaction. Interfaces for Users with Disabilities* (Cambridge: Cambridge University Press) ISBN 0 521 43413 0.

GOLDSMITH, S., 1984, *Designing for the Disabled* (3rd ed.) (London: RIBA Publications) ISBN 0 900630 50 7.

HJELMQUIST, E. and NILSSON, L.G. (ed.), 1986, *Communication and Handicap* (Amsterdam: North-Holland) ISBN 0 444 70034 X.

KING, T.W., 1999, *Assistive Technology:Essential Human Factors* (Boston, Massachusetts: Allyn and Bacon) ISBN 0 205 27126 2.

KUMAR, S. (ed.). 1997, *Perspectives in Rehabilitation Ergonomics* (London: Taylor & Francis) ISBN 0 7484 0673 5.

MARQUIE, J.C., CAU-BAREILLE, D.P., and VOLKOFF, S. (eds), 1998, *Working with Age* (London: Taylor & Francis) ISBN 0 7484 0785 5.

PERFECT, T.J. and MAYLOR, E.A., (Eds), 2000, *Models of Cognitive Aging* (Oxford: Oxford University Press) ISBN 0 19 852437 4.

POON, L.W., RUBIN, D.C., and WILSON, B.A. (eds), 1989, *Everyday Cognition in Adulthood and Late Life* (Cambridge: Cambridge University Press) ISBN 0 521 42860 2.

RASCHKO, B.B., 1982, *Housing Interiors for the Disabled and Elderly* (New York: Van Nostrand Reinhold) ISBN 0 442 22001 4.

ROGERS, W.A. (ed.), 1997, *Designing for an Aging Population: Ten Years of Human Factors/Ergonomics Research* (Santa Monica, California: Human Factors and Ergonomics Society) ISBN 0 945289 08 1.

ROSENTHAL, O. and PHILLIPS, R.H., 1997, *Coping with Color Blindness* (Garden City Park, New York: Avery) ISBN 0 89529 733 7.

SCHNEIDER, E.L. and ROWE, J.W. (eds), 1996, *Handbook of the Biology of Aging* (4th ed.) (San Diego, California: Academic Press) ISBN 0 12 627873 3.

SCHWARZ, N., PARK, D.C., KNAUPER, B., and SUDMAN, S. (eds), 1999, *Cognition, Aging, and Self-Reports* (Hove, East Sussex: Psychology Press) ISBN 1 56032 780 4.

SHEPHARD, R.J., 1997, *Aging, Physical Activity, and Health* (Champaign, Illinois: Human Kinetics) ISBN 0 87322 889 8.

SNEL, J. and CREMER, R. (eds), 1994, *Work and Aging: A European Perspective* (London: Taylor & Francis) ISBN 0 7484 0165 2.

SPIRDUSO, W.W., 1995, *Physical Dimensions of Aging* (Champaign, Illinois: Human Kinetics) ISBN 0 87322 323 3.

STEENBEKKERS, L.P.A. and VAN BEIJSTERVELDT, C.E.M. (eds), 1998, *Design-Relevant Characteristics of Ageing Users: Backgrounds and Guidelines for Product Innovation* (Delft: Delft University Press) ISBN 90 407 1709 5.

Workplace, equipment, and product design

ARONOFF, S. and KAPLAN, A., 1995, *Total Workplace Performance: Rethinking the Office Environment* (Ottawa: WDL Publications) ISBN 0 921804 95 4.

BERGMAN, E. (Ed), 2000, *Information Appliances and Beyond: Interaction Design for Consumer Products* (San Francisco, California: Morgan Kaufmann) ISBN 1 58860 600 9.

CACHA, C.A., 1999, *Ergonomics and Safety in Hand Tool Design* (Boca Raton, Florida: Lewis Publishers) ISBN 1 56670 308 5.

CORLETT, E.N. and CLARK, T.S., 1995, *The Ergonomics of Workspaces and Machines: A Design Manual* (2nd ed.) (London: Taylor & Francis) ISBN 0 7484 0320 5.

CUSHMAN, W.H. and ROSENBERG, D.J., 1991, *Human Factors in Product Design* (Amsterdam: Elsevier) ISBN 0 444 89031 9.

DAAMS, B.J., 1994, *Human Force Exertion in User-Product Interaction. Backgrounds for Design* (Delft: Delft University Press) ISBN 90 6275 995 5.

DRUIN, A. (Ed) 1999, *The Design of Childrens Technology* (San Francisco, California: Morgan Kaufmann) ISBN 1 55860 308 5.

GRANDJEAN, E. (ed.), 1969, *Proceedings of the Symposium on Sitting Posture, Sitzhaltung, Posture Assise* (London: Taylor & Francis) ISBN 85066 029 7.

GRANDJEAN, E., 1973, *Ergonomics of the Home* (London: Taylor & Francis) ISBN 0 85066 067 X.

GRANDJEAN, E., 1984, *Ergonomics and Health in Modern Offices* (London: Taylor & Francis) ISBN 0 85066 270 2.

GRANDJEAN, E., 1986, *Ergonomics in Computerised Offices* (London: Taylor & Francis) ISBN 0 85066 350 4.

GREEN, W.S. and JORDAN, P.W. (Eds), 1999, *Human Factors in Product Design: Current Practice and Future Trends* (London: Taylor & Francis) ISBN 0 748 40844 4.

HELANDER, M., 1981, *Human Factors Ergonomics for Building and Construction* (Chichester: Wiley) ISBN 0 471 05075 X.

HILTS, SR, 1994, *The Virtual Classroom* (Norwood, New Jersey: Ablex Publishing Corporation) ISBN 0 89391 928 4.

HUES, U., FORTE, W.B. and ROBINSON, S., 1990, *Telework: Towards the Elusive Office* (Chichester: Wiley) ISBN 0 471 92284 6.

IGBARIA, M. and TAN, M. (eds), 1998, *The Virtual Workplace* (Hershey, Pennsylvania: Idea Group Publishing) ISBN 1 878 28941 1.

INKELES, G. and SCHENCKE, I., (1994) *Ergonomic Living: How to Create a User Friendly Home and Office* (London: Simon & Schuster) ISBN 0 0209 3081 X.

IVERGARD, T., 1989, *Handbook of Control Room Design and Ergonomics* (London: Taylor & Francis) ISBN 0 85066 454 3.

JAMES, P. and NOAKES, T., 1994, *Hospital Architecture* (Harlow, Essex: Longman) ISBN 0 582 09389 9.

Jordan, P.W., 2000, *Designing Pleasurable Products: An Introduction to the New Human Factors* (London: Taylor & Francis) ISBN 0 748 40844 4.

KLEEMAN, W.B., 1991, *Interior Design of the Electronic Office: The Comfort and Productivity Payoff* (New York: Van Nostrand Reinhold) ISBN 0 442 00613 6.

KONZ, S., 1994, *Facility Design: Manufacturing Engineering* (2nd ed.) (Scottsdale, Arizona: Publishing Horizons) ISBN 0 942280 64 4.

KVALSETH, T.O. (ed.), 1983, *Ergonomics in Workstation Design* (London: Butterworths) ISBN 0 408 01253 6.

LINDQVIST, B., AHLBERG, E. and SKOGSBERG, L., 1986, *Ergonomic Tools in Our Time* (Stockholm: Atlas Copco) ISBN 91 7810 535 8.

Lockheed Missiles & Space Company, 1993, *Human Engineering Design Criteria for Military Systems, Equipment and Facilities: The MIL-STD-1472D Checklist* (Ohio: Distributed by CSERIAC, Wright-Patterson AFB), no ISBN.

LUEDER, R.K., 1986, *The Ergonomics Payoff: Designing the Electronic Office* (Toronto: Holt, Rinehart & Winston of Canada) ISBN 0 03 921998 4.

LUEDER, R. and NORO, K. (eds), 1994, *Hard Facts about Soft Machines: The Ergonomics of Seating* (London: Taylor & Francis) ISBN 0 85066 802 6.

MAHNKE, F.H., 1996, *Color, Environment, and Human Response* (New York: Van Nostrand Reinhold) ISBN 0 471 28667 2.

MANDAL, A.C., 1985, *The Seated Man. Homo Sedens* (Klampenborg, Denmark: Dafnia Publications) ISBN 87 982017 1 9.

MITAL, A. and KARWOWSKI, W. (eds), 1991. *Workspace, Equipment and Tool Design* (Amsterdam: Elsevier) ISBN 0 444 87441 0.

MORRIS, A. and DYER, H., 1998, *Human Aspects of Library Automation* (2nd ed.) (Aldershot, Hampshire: Gower) ISBN 0 566 07504 0.

MORGAN, C.T., COOK, J.S., CHAPANIS, A., and LUND, M.W. (eds), 1963, *Human Engineering Guide to Equipment Design* (New York: McGraw-Hill), no ISBN.

National Kitchen & Bath Association, 1996, *The Essential Kitchen Design Guide* (New York: John Wiley) ISBN 0 471 12672 1.

National Kitchen & Bath Association, 1997, *The Essential Bathroom Design Guide* (New York: John Wiley) ISBN 0 471 12673 X.

NORMAN, D.A., 1998, *The Psychology of Everyday Things* (2nd ed.) (London: MIT Press) ISBN 0 262 64037 6.

NORMAN, D.A., 1993, *Things That Make Us Smart* (Reading, Massachusetts: Addison-Wesley) ISBN 0 201 62695 0.

RAYFIELD, J.K., 1994, *The Office Interior Design Guide: An Introduction for Facilities Managers and Designers* (New York: John Wiley) ISBN 0 471 57286 1.

RUBIN, A. and ELDER, J., 1980, *Building for People* (Washington, DC: U.S. Government Printing Office), no ISBN.

ROOZENBURG, N.F.M. and EEKELS, J., 1995, *Product Design: Fundamentals and Methods* (Chichester: Wiley) ISBN 0 471 95465 9.

RUCK, N.C. (ed.), 1989, *Building Design and Human Performance* (New York: Van Nostrand Reinhold) ISBN 0 442 27847 0.

RUTTER, B.G. and DAINOFF, M.J., 1996, *The Ergonomic Office: Standards, Specifications and Design Guidelines* (St Louis, Missouri: Metaphase Publishing) ISBN 0 9643187 0 9..

SELAN, J.L. and SPENCER, S.P. (ed.), 1995, *The Ergonomic Products Report* (1996 ed.) (Dallas, Texas: Advanced Ergonomics), no ISBN.

STAARINK, H.A.M., 1995, *Sitting Posture, Comfort and Pressure: Assessing the Quality of Wheelchair Cushions* (Delft: Delft University Press) ISBN 90 407 1196 8.

STANTON, N. (ed.), 1998, *Human Factors in Consumer Products* (London: Taylor & Francis) ISBN 0 7484 0603 4.

STEIDL, R.E. and BRATTON, E.C., 1968, *Work in the Home* (New York: Wiley), no ISBN.

SUNDSTROM, E., 1986, *Work Places: The Psychology of the Physical Environment in Offices and Factories* (Cambridge: Cambridge University Press) ISBN 0 521 31947 1.

TEMPLER, J., 1992, *The Staircase: History and Theories* (Cambridge, Massachusetts: MIT Press) ISBN 0 262 20082 1.

TEMPLER, J., 1992, *The Staircase: Studies of Hazards, Falls, and Safer Design* (Cambridge, Massachusetts: MIT Press) ISBN 0 262 20083 X.

WIKLUND, M.E., 1995, *Medical Device and Equipment Design: Usability Engineering and Ergonomics* (Buffalo Grove, Illinois: Interpharm Press) ISBN 0 935184 69 4.

ZACHARKOW, D., 1988, *Posture: Sitting, Standing, Chair Design and Exercise* (Springfield, Illinois: Charles C. Thomas) ISBN 0 398 05418 5.

The physical environment

BOYCE, P.R., 1981, *Human Factors in Lighting* (London: Applied Science Publishers) ISBN 0 85334 912 6.

CASSIDY, T., 1997, *Environmental Psychology: Behaviour and Experience in Context* (London: Taylor & Francis) ISBN 0 86377 481 4.

Chartered Institution of Building Services Engineers, 1994, *Code for Interior Lighting 1994* (London: CIBSE) ISBN 0 900953 64 0.

CLARK, R.P. and EDHOLM, O.G., 1985, *Man and his Thermal Environment* (London: Edward Arnold) ISBN 0 7131 4445 9.

DUPUIS, H. and ZERLETT, G., 1986, *The Effects of Whole-Body Vibration* (Berlin: Springer) ISBN 0 387 16584 3.

FAHY, F.J. and WALKER, J.G. (Eds), 1998, *Fundamentals of Noise and Vibration* (London: E & FN Spon) ISBN 0 419 22700 8.

FANGER, P.O., 1970, *Thermal Comfort; Analysis and Applications in Environmental Engineering* (London: McGraw-Hill) ISBN 07 019915 9.

GRANDJEAN, E. and GILGEN, A., 1976, *Environmental Factors in Urban Planning* (London: Taylor & Francis) ISBN 0 85066 084 X.

GRIFFIN, M.J., 1990, *Handbook of Human Vibration* (London: Academic Press) ISBN 0 12 303040 4.

HAYMES, E.M. and WELLS, C.L., 1986, *Environment and Human Performance* (Champaign, Illinois: Human Kinetics Publishers) ISBN 0 87322 039 0.

HOLLIES, N.R.S. and GOLDMAN, R.F. (eds), 1977, *Clothing Comfort* (Ann Arbor, Michigan: Ann Arbor Science Publishers) ISBN 0 250 40159 2.

JONES, D.M. and CHAPMAN, A.J. (eds), 1984, *Noise and Society* (Chichester: Wiley) ISBN 0 471 90357 4.

KAUFMAN, J.E. and CHRISTENSEN, J.F. (eds), 1989, *IES Lighting Ready Reference* (2nd ed.) (New York: Illuminating Engineering Society of North America) ISBN 0 87995 030 7.

KRYTER, K.D., 1985, *The Effects of Noise on Man* (2nd ed.) (London: Academic Press) ISBN 0 12 42740 9.

KRYTER, K.D., 1994, *The Handbook of Hearing and the Effects of Noise: Physiology, Psychology, and Public Health* (San Diego, California: Academic Press) ISBN 0 12 427455 2.

LAING, A., DUFFY, F., JAUNZENS, D. and WILLIS, S., 1998, *New Environments for Working: The Re-Design of Offices and Environmental Systems for New Ways of Working* (London: E & FN Spon) ISBN 0 419 20990 5.

LOEB, M., 1986, *Noise and Human Efficiency* (Chichester: Wiley) ISBN 0 471 10287 3.

MEKJAVIC, I.B., BANISTER, E.W. and MORRISON, J.B. (eds), 1988, *Environmental Ergonomics* (London: Taylor & Francis) ISBN 0 85066 400 4.

NICOL, F., HUMPHREYS, M., SYKES, O. and ROAF, S. (Eds), 1995, *Standards for Thermal Comfort: Indoor Air Temperature Standards for the 21st Century* (London: E & FN Spon) ISBN 0 419 02420 2.

OBORNE, D.J. and GRUNEBERG, M.M. (eds), 1983, *The Physical Environment at Work* (Chichester: Wiley) ISBN 0 471 90123 7.

OLISHIFSKI, J.B. and HARFORD, E.R. (eds), 1975, *Industrial Noise and Hearing Conservation* (Chicago: National Safety Council) ISBN 0 87912 085 1.

O'REILLY, J.T., HAGAN, P., GOTS, R. and HEDGE, A., 1998, *Keeping Buildings Healthy: How to Monitor and Prevent Indoor Environmental Problems* (New York: John Wiley) ISBN 0 471 29228 1.

PANDOLF, K.B., SAWKA, M.N. and GONZALEZ, R.R. (eds), 1988, *Human Performance Physiology and Environmental Medicine at Terrestrial Extremes* (Indianapolis, Indiana: Benchmark Press) ISBN 0 936157 18 6.

PARSONS, K.C., 1993, *Human Thermal Environments* (London: Taylor & Francis) ISBN 0 7484 0041 9.

POULTON, E.C., 197O, *Environment and Human Efficiency* (Springfield, Illinois: Charles C. Thomas), no ISBN.

POULTON, E.C., 1979, *The Environment at Work* (Springfield, Illinois: Charles C. Thomas) ISBN 0 398 03848 1.

RAYMOND, S. and CUNLIFFE, R., 2000, *Tomorrow's Office Creating Effective and Humane Interiors* (London: E & FN Spon) ISBN 0 419 24400 X.

REA, M.S. (ed.), 1993, *Lighting Handbook: Reference and Application* (8th ed.) (New York: Illuminating Engineering Society of North America) ISBN 0 87995 102 8.

REASON, J.T. and BRAND, J.J., 1975, *Motion Sickness* (London: Academic Press) ISBN 0 12 584050 0.

RIVOLIER, J., GOLDSMITH, R., LUGG, D.J., and TAYLOR, A.J.W., 1988, *Man in the Antarctic* (London: Taylor & Francis) ISBN 0 85066 280 X.

ROBBINS, C.L., 1986, *Daylighting Design and Analysis* (New York: Van Nostrand Rheinhold) ISBN 0 442 27949 3.

ROSTRON, J. (ed.), 1997, *Sick Building Syndrome: Concepts, Issues and Practice* (London: E & FN Spon) ISBN 0 419 21530 1.

SAMUELOFF, S. and YOUSEF, M.K. (eds), 1987, *Adaptive Physiology to Stressful Environments* (Boca Raton, Florida: CRC Press) ISBN 0 8493 6458 2.

SLOAN, A.W., 1979, *Man in Extreme Environments* (Springfield, Illinois: Charles C. Thomas) ISBN 0 398 03941 0.

TREGENZA, P. and LOE, D., 1998, *The Design of Lighting* (London: E & FN Spon) ISBN 0 419 20440 7.

VISCHER, J.C., 1989, *Environmental Quality in Offices* (New York: Van Nostrand Reinhold) ISBN 0 442 20498 1.

WASSERMAN, D.E., 1987, *Human Aspects of Occupational Vibration* (Amsterdam: Elsevier) ISBN 0 444 42728 7.

YOULE, A. *et al.*, 1990, *The Thermal Environment* (Leeds: H & H Scientific Consultants) ISBN 0 905927 38 9.

Display and control design

BABER, C., 1997, *Beyond the Desktop: Designing and Using Interaction Devices* (San Diego, California: Academic Press) ISBN 0 12 069550 2.

BOSMAN, D. (ed.), 1989, *Display Engineering. Conditioning, Technologies, Applications* (Amsterdam: North-Holland) ISBN 0 444 87319 8.

CARD, S.K., MACKINLEY, J.D. and SHNEIDDERMAN, B. (eds), 1999, *Readings in Information Visualization: Using Vision to Think* (San Francisco, California: Morgan Kaufmann) ISBN 1 55860 533 9.

CARR, K. and ENGLAND, R. (eds), 1995, *Simulated and Virtual Realities: Elements of Perception* (London: Taylor & Francis) ISBN 0 7484 0129 6.

CARTER, R., DAY, B. and MEGGS, P., 1993, *Typographic Design: Form and Communication* (2nd ed.) (New York: Van Nostrand Reinhold) ISBN 0 442 00759 0.

COE, M., 1996, *Human Factors for Technical Communication* (New York: John Wiley) ISBN 0 471 03530 0.

DOUGLAS, S.A. and MITHAL, A.K., 1997, *The Ergonomics of Pointing Devices* (Berlin: Springer) ISBN 3 540 19986 1.

EARNSHAW, R.A., VINCE, J.A. and JONES, H. (eds), 1995, *Virtual Reality Applications* (London: Academic Press) ISBN 0 12 227755 4.

EASTERBY, R. and ZWAGA, H. (eds), 1984, *Information Design* (Chichester: Wiley) ISBN 0 471 10431 0.

EDWORTHY, J and ADAMS, A, 1996, *Warning Design: A Research Perspective* (London: Taylor & Francis) ISBN 0 7484 0467 8.

ELLIS, S.R., M.K. KAISER and A.J. GRUNWALD (eds), 1993, *Pictorial Communication in Virtual and Real Environments* (2nd ed.) (London: Taylor & Francis) ISBN 0 7484 0082 6.

FRASCARA, J., 1997, *User-Centred Graphic Design: Mass Communications and Social Change* (London: Taylor & Francis) ISBN 0 7484 0672 7.

GALITZ, W.O., 1981, *Handbook of Screen Format Design* (Wellesley, Massachusetts: QED Information Sciences) ISBN 0 89435 052 8.

GALITZ, W.O., 1994, *It's Time to Clean your Windows: Designing GUIs That Work* (New York: John Wiley) ISBN 0 471 60668 5.

GILMORE, W.E., GERTMAN, D.I., and BLACKMAN, H.S., 1989, *The User–Computer Interface in Process Control. A Human Factors Engineering Handbook* (Boston, Massachusetts: Academic Press) ISBN 0 12 283965 X.

GRANDJEAN, E. and VIGLIANI, E. (eds), 1980, *Ergonomic Aspects of Visual Display Terminals* (London: Taylor & Francis) ISBN 0 85066 211 7.

GUMP, D., 1993, *Virtual Reality Handbook: Products, Services and Resources* (Arlington, Virginia: Pasha Publications) ISBN 0 935453 56 3.

HARLING, P.A. and EDWARDS, A.D.N. (eds), 1997, *Progress in Gestural Interaction: Proceedings of Gesture Workshop '96*, 19 March 1996, University of York, UK (London: Springer) ISBN 3 540 76094 6.

HARTLEY, J. (ed.), 1992, *Technology and Writing: Readings in the Psychology of Written Communication* (London: Jessica Kingsley Publishers) ISBN 1 85302 097 4.

HORTON, W., 1994, *The Icon Book: Visual Symbols for Computer Systems and Documentation* (New York: John Wiley) ISBN 0 471 59901 8.

HORTON, W., 1991, *Illustrating Computer Documentation* (New York: John Wiley) ISBN 0 471 53845 0.

JACKSON, R., MACDONALD, L. and FREEMAN, K., 1994, *Computer Generated Colour: A Practical Guide to Presentation and Display* (Chichester: Wiley) ISBN 0 471 93378 3.

KALAWSKY, R.S., 1993, *The Science of Virtual Reality and Virtual Environments* (Wokingham, Berkshire: Addison-Wesley) ISBN 0 201 63171 7.

KOSSLYN, S.M., 1994, *Elements of Graph Design* (New York: W.H. Freeman) ISBN 0 7167 2362 X.

LEHTO, M.R. and MILLER, J.M., 1986, *Warnings. Volume I* (Ann Arbor, Michigan: Fuller Technical Publications) ISBN 0 940 537 00 1.

MACDONALD, L. and VINCE, J. (eds), 1994, *Interacting with Virtual Environments* (Chichester: Wiley) ISBN 0 471 93941 2.

MACDONALD, L.W. and LOWE, A.C. (eds), 1997, *Display Systems: Design and Applications* (Chichester: Wiley) ISBN 0 471 95870 0.

MACEACHREN, A.M., 1995, *How Maps Work: Representation, Visualization, and Design* (New York: Guilford Press) ISBN 0 89862 589 0.

MASHOUR, M., 1974, *Human Factors in Signalling Systems: Specific Applications to Railway Signalling* (Stockholm: Almqvist & Wiksell International and New York: Wiley) ISBN 0470 57465 8.

MELZER, J.E. and MOFFITT, K., 1997, *Head Mounted Displays: Designing for the User* (New York: McGraw-Hill) ISBN 0 07 041819 5.

MILLER, J.M. AND LEHTO, M.R., 1987, *Warnings. Volume II* (Ann Arbor, Michigan: Fuller Technical Publications) ISBN 0 940537 01 X.

National Research Council, 1983, *Video Displays, Work, and Vision* (Washington, DC: National Academy Press) ISBN 0 309 03388 8.

National Research Council, Committee on Human Factors, 1997, *Tactical Display for Soldiers: Human Factors Considerations* (Washington, DC: National Academy Press) ISBN 0 309 05638 1.

NEWMAN, R.L., 1995, *Head-Up Displays: Designing the Way Ahead* (Hampshire, Aldershot: Ashgate Publishing) ISBN 0 291 39811 1.

OKABAYASHI, S., 1996, *Visual Optics of Head-Up Displays (HUDs) in Automotive Applications* (Amsterdam: Gordon and Breach) ISBN 2 88449 033 7.

PETERS, G.A. and PETERS, B.J., 1999, *Warnings, Instructions, and Technical Communications* (Tucson, Arizona: Lawyers & Judges Publishing) ISBN 0 913875 61 9.

REYNOLDS, L. and BARRETT, S., 1981, *Signs and Guiding for Libraries* (London: Clive Bingley) ISBN 0 85157 312 6.

RIVLIN, C., LEWIS, R. and DAVIES-COOPER, R. (eds), 1990, *Guidelines for Screen Design* (Oxford: Blackwell Scientific Publications) ISBN 0 632 02686 3.

ROE, D.B. and WILPON, J.G. (eds), 1994, *Voice Communication Between Humans and Machines* (Washington, DC: National Academy Press) ISBN 0 309 04988 1.

RYAN, J.P., 1991, *Design of Warning Labels and Instructions* (New York: Van Nostrand Reinhold) ISBN 0 442 31953 3.

SASSOON, R., 1993, *Computers and Typography* (Oxford: Intellect) ISBN 1 871516 23 4.

SHERR, S. (ed.), 1988, *Input Devices* (Boston, Massachusetts: Academic Press) ISBN 0 12 639970 0.

SHERR, S., 1993, *Electronic Displays* (2nd ed.) (New York: Wiley) ISBN 0 471 63616 9.

SHURTLEFF, D.A., 1980, *How to Make Displays Legible* (La Miranda, California: Human Interface Design), no ISBN.

STANTON, N. (ed.), 1994, *Human Factors in Alarm Design* (London: Taylor & Francis) ISBN 0 7484 0109 1.

STOKES, A., WICKENS, C. and KITE, K., 1990, *Display Technology. Human Factors Concepts* (Warrendale, Pennsylvania: Society of Automotive Engineers) ISBN 0 89883 530 5.

THORELL, L.G. and SMITH, W.J., 1990, *Using Computer Color Effectively. An Illustrated Reference* (Englewood Cliffs, New Jersey: Prentice Hall) ISBN 0 13 939878 3.

TRAVIS, D., 1991, *Effective Color Displays* (London: Academic Press) ISBN 0 12 697690 2.

TUFTE, E.R., 1983, *The Visual Display of Quantitative Information* (Cheshire, Connecticut: Graphics Press), no ISBN.

TUFTE, E.R., 1990, *Envisioning Information* (Cheshire, Connecticut: Graphics Press), no ISBN.

TUFTE, E.R., 1997, *Visual Explanations: Images and Quantities, Evidence and Narrative* (Cheshire, Connecticut: Graphics Press) ISBN 0 9613921 2 6.

Van Der VAART, A.J.M., 1995, *Arm Movements in Operating Rotary Controls* (Delft: Delft University Press) ISBN 90 407 1131 3.

Van VLIET, J.C. (ed.), 1988, *Document Manipulation and Typography* (Cambridge: Cambridge University Press) ISBN 0 521 36294 6.

UMMELEN, N., 1997, *Procedural and Declarative Information in Software Manuals: Effects on Information Use, Task Performance and Knowledge* (Amsterdam: Rodopi) ISBN 90 420 0128 3.

VELGER, M., 1998, *Helmet-Mounted Displays and Sights* (Boston: Artech House) ISBN 0 89006 822 4.

WEINTRAUB, D.J. and ENSING, M., 1992, *Human Factors Issues in Head-Up Display Design: The Book of HUD* (Wright-Patterson Air Force Base, Ohio: Crew System Ergonomics Information Analysis Center), no ISBN.

WIDDEL, H., POST, D.L., GROSSMAN, J.D., and WALRAVEN, J. (eds), 1992, *Color in Electronic Displays* (New York: Plenum Press) ISBN 0 306 44191 8.

WILDBUR, P., 1989, *Information Graphics* (London: Trefoil Publications) ISBN 0 86294 110 5.

WOGALTER, M.S., DEJOY< D.M> and LAUGHERY, K.R., (Eds), 1999, *Warning and Risk Communication* (London: Taylor & Francis) ISBN 0 7484 0266 7.

YAZDANI, M. and BARKER, P. (Eds), 2000, *Iconic Communication* (Bristol: Intellect Books) ISBN 1 84150 016 X.

ZWAGA, H.J.G., BOERSEMA, T., and HOONHOUT, H.C.M. (eds), 1999, *Visual Information for Everyday Use: Design and Research Perspectives* (London: Taylor & Francis) ISBN 0 7484 0670 0.

Human–Computer interaction

ANDRIOLE, S. and ADELMAN, L., 1995, *Cognitive Systems Engineering for User-Computer Interface Design, Prototyping, and Evaluation* (Hillsdale, New Jersey: Lawrence Erlbaum) ISBN 0 8058 1244 X.

BADRE, A. and SHNEIDERMAN, B. (eds), 1982, *Directions in Human/Computer Interaction* (Norwood, New Jersey: Ablex Publishing Corporation) ISBN 0 89391 144 5.

BAECKER, R.M., GRUDIN, J. BUXTON, W.A.S. and GREENBERG, S. (eds), 1995, *Readings in Human–Computer Interaction. Toward the Year 2000* (2nd ed.) (San Francisco, California: Morgan Kaufmann) ISBN 1 55860 246 1.

BAECKER, R.M. and MARCUS, A., 1990, *Human Factors and Typography for More Readable Programs* (Reading, Massachusetts: Addison-Wesley) ISBN 0 201 10745 7.

BAILEY, R.W., 1989, *Human Performance Engineering. Using Human Factors/Ergonomics to Achieve Computer System Usability* (2nd ed.) (London: Prentice-Hall) ISBN 0 13 446048 0.

BARFIELD, L., 1993, *The User Interface: Concepts and Design* (Wokingham, Berkshire: Addison-Wesley) ISBN 0 201 54441 5.

BASS, L. and COUTAZ, J., 1991, *Developing Software for the User Interface* (Reading, Massachusetts: Addison-Wesley) ISBN 0 201 51046 4.

BASS, L. and DEWAN, P. (eds), 1993, *User Interface Software* (Chichester: Wiley) ISBN 0 471 93784 3.

BASS, L.J. and UNGER, C. (eds), 1996, *Engineering for Human–Computer Interaction* (London: Chapman & Hall) ISBN 0 412 72180 5.

BERRY, D. and HART, A. (eds), 1990, *Expert Systems: Human Issues* (London: Chapman & Hall) ISBN 0 412 37790 X.

BLATTNER, M.M. and DANNENBERG, R.B. (eds), 1992, *Multimedia Interface Design* (New York: ACM Press) ISBN 0 201 54981 6.

BODKER, S., 1991, *Through the Interface: A Human Activity Approach to*

User Interface Design (Hillsdale, New Jersey: Lawrence Erlbaum) ISBN 0 8058 0571 0.

BROWN, C.M., 1999, *Human–Computer Interaction Design Guidelines* (Exeter: Intellect) ISBN 1 871516 54 4.

CARD, S.K., MORAN, T.P. and NEWELL, A., 1983, *The Psychology of Human–Computer Interaction* (London: Lawrence Erlbaum) ISBN 0 89859 243 7.

CAREY, J.M., 1991, *Human Factors in Information Systems: An Organizational Perspective* (Norwood, New Jersey: Ablex) ISBN 0 89391 633 1.

CAREY, J.M. (ed.), 1988, *Human Factors in Management Information Systems* (Norwood, New Jersey: Ablex) ISBN 0 89391 448 7.

CAREY, J.M. (ed.), 1995, *Human Factors in Information Systems: Emerging Theoretical Bases* (Norwood, New Jersey: Ablex) ISBN 0 89391 940 3.

CAREY, J. M. (ed.), 1997, *Human Factors in Information Systems: The Relationship Between User Interface Design and Human Performance* (Greenwich, Connecticut: Ablex) ISBN 1 56750 286 5.

CARROLL, J.M. (ed.), 1991, *Designing Interaction. Psychology at the Human–Computer Interface* (Cambridge: Cambridge University Press) ISBN 0 521 40921 7.

CHATTY, S. and DEWAN, P. (Eds), 1998, *Engineering for Human-Computer Interaction* (Boston, Massachusetts: Kluwer Academic) ISBN 0 412 83520 7.

CONSTANTINE, L.L. and LOCKWOOD, L.A.D., 1999, *Software for Use: A Practical Guide to the Models and Methods of Usage-Centered Design* (Reading, Massachusetts: Addison-Wesley) ISBN 0 201 92478 1.

COX, K. and WALKER, D., 1993, *User–Interface Design* (2nd ed.) (New York: Prentice Hall) ISBN 0 13 952888 1.

DAMPER, R.I., HALL, W. and RICHARDS, J.W. (eds), 1994, *Multimedia Technologies and Future Applications* (London: Pentech Press) ISBN 0 7273 1320 7.

DAY, D.L. and KOVACS, D.K. (eds), 1996, *Computers, Communication and Mental Models* (London: Taylor & Francis) ISBN 0 7484 0543 7.

DEL GALDO, E.M. and NIELSEN, J. (eds), 1996, *International User Interfaces* (New York: Wiley) ISBN 0 471 14965 9.

DILLON, A., 1994, *Designing Usable Electronic Text: Ergonomic Aspects of Human Information Usage* (London: Taylor & Francis) ISBN 0 7484 0113 X.

DIX, A.J., FINLAY, J.E., ABOWD, G.D. and BEALE, R., 1998, *Human–Computer Interaction* (2nd ed.) (London: Prentice Hall Europe) ISBN 0 13 239864 8.

DOWNTON, A. (ed.), 1993 *Engineering the Human–Computer Interface* (Student ed.) (London: McGraw-Hill) ISBN 0 07 707727 X.

DUFFY, T.M., PALMER, J.E. and MEHLENBACHER, B., 1992, *Online Help: Design and Evaluation* (Norwood, New Jersey: Ablex Publishing Corporation) ISBN 0 89391 848 2.

EDMONDS, E. (ed.), 1992, *The Separable User Interface* (London: Academic Press) ISBN 0 12 232150 2.

FALZON, P. (ed.), 1990, *Cognitive Ergonomics: Understanding, Learning and Designing Human–Computer Interaction* (London: Academic Press) ISBN 0 12 248290 5.

FAULKNER, C., 1998, *The Essence of Human-Computer Interaction* (London: Prentice Hall) ISBN 0 13 751975 3.

FAULKNER, X., 2000, *Usability Engineering* (Basingstoke, Hampshire: Macmillan Press) ISBN 0 333 77321 7.

FORSYTHE, C., GROSE, E. and RATNER, J. (eds), 1998, *Human Factors and Web Development* (Mahwah, New Jersey: Lawrence Erlbaum) ISBN 0 8058 2824 9.

FOWLER, S., 1998, *GUI Design Handbook* (New York: McGraw-Hill) ISBN 0 07 059274 8.

GALITZ, W.O., 1996, *The Essential Guide to User Interface Design: An Introduction to GUI Design Principles and Techniques* (New York: Wiley) ISBN 0 471 15755 4.

GREEN, T.R.G., PAYNE, S.J., and VAN DER VEER, G.G., 1983, *The Psychology of Computer Use* (London: Academic Press) ISBN 0 12 297420 4.

GREENBAUM, J. and KYNG, M. (eds), 1991, *Design at Work: Cooperative Design of Computer Systems* (Hillsdale, New Jersey: Lawrence Erlbaum) ISBN 0 8058 0612 1.

GREENBERG, S., 1993, *The Computer User as Toolsmith* (Cambridge: Cambridge University Press) ISBN 0 521 40430 4.

GREIF, I. (ed.), 1988, *Computer-Supported Cooperative Work: A Book of Readings* (San Mateo, California: Morgan Kaufmann Publishers) ISBN 0 93461 3575.

GUINDON, R. (ed.), 1988, *Cognitive Science and Its Applications for Human–Computer Interaction* (Hillsdale, New Jersey: Lawrence Erlbaum) ISBN 0 89859 884 2.

HANCOCK, P.A. and CHIGNELL, M.H. (eds), 1989, *Intelligent Interfaces. Theory, Research and Design* (Amsterdam: North-Holland) ISBN 0 444 87313 9.

HARRISON, M. and THIMBLEBY, H. (eds), 1990, *Formal Methods in Human–Computer Interaction* (Cambridge: Cambridge University Press) ISBN 0 521 37202 X.

HARTSON, H.R. (ed.), 1985, *Advances in Human–Computer Interaction. Volume 1* (Norwood, New Jersey: Ablex Publishing) ISBN 0 89391 244 1.

HARTSON, H.R. and HIX, D. (eds), 1988, *Advances in Human–Computer Interaction. Volume 2* (Norwood, New Jersey: Ablex Publishing) ISBN 0 89391 428 2.

HARTSON, H.R. and HIX, D. (eds), 1992, *Advances in Human–Computer Interaction. Volume 3* (Norwood, New Jersey: Ablex Publishing) ISBN 0 89391 751 6.

HARTSON, H.R. and HIX, D. (eds), 1993, *Advances in Human–Computer Interaction. Volume 4* (Norwood, New Jersey: Ablex Publishing) ISBN 0 89391 934 9.

HELANDER, M., LANDAUER, T.K. and PRABHU, P.V. (eds), 1997, *Handbook of Human–Computer Interaction* (2nd ed.) (Amsterdam: Elsevier) ISBN 0 444 81862 6.

HICKS, R. and ESSINGER, J., 1991, *Making Computers More Human. Designing for Human–Computer Interaction* (Oxford: Elsevier Advanced Technology) ISBN 1 85617 057 8.

HIX, D. and HARTSON, H.R., 1993, *Developing User Interfaces. Ensuring Usability Through Product and Process* (New York: John Wiley) ISBN 0 471 57813 4.

HOC, J.M., CACCIABUE, P.C., and HOLLNAGEL, E., 1995, *Expertise and Technology: Cognition and Human–Computer Cooperation* (Hillsdale, New Jersey: Lawrence Erlbaum) ISBN 0 8058 1511 2.

HUTT, A., 1993, *User Interface* (New York: Prentice Hall) ISBN 0 13 630591 1.

JOHNSON, J., 2000, *GUI Bloopers: Don's and Do's for Software Developers and Web Designers* (San Francisco: Morgan Kaufmann) ISBN 1 55860 582 7.

JOHNSON, P., 1992, *Human–Computer Interaction. Psychology, Task Analysis and Software Engineering* (London: McGraw-Hill) ISBN 0 07 707235 9.

KAJLER, N. (ed.), 1998, *Computer–Human Interaction in Symbolic Computation* (Vienna: Springer-Verlag) ISBN 3 211 82843 5.

KEARSLEY, G., 1988, *Online Help Systems: Design and Implementation* (Norwood, New Jersey: Ablex Publishing) ISBN 0 89391 472 X.

KEARSLEY, G., 1994, *Public Access Systems: Bringing Computer Power to the People* (Norwood, New Jersey: Ablex Publishing) ISBN 0 89391 947 0.

KLINGER, A. (ed.), 1991, *Human–Machine Interactive Systems* (New York: Plenum Press) ISBN 0 306 43758 9.

KLIX, F., STREITZ, N.A., WAERN, Y. and WANDKE, H. (eds), 1989, *Man-Computer Interaction Research - Macinter-II* (Amsterdam: North-Holland) ISBN 0 444 87336 8.

KLIX, F. and WANDKE, H. (eds), 1986, *Man–Computer Interaction Research: Macinter-I* (Amsterdam: North-Holland) ISBN 0 444 87910 2.

KOLSKI, C., 1993, *Ingenierie des Interfaces Homme–Machine: Conception et Evaluation* (Paris: Hermes) ISBN 2 86601 377 8.

LANDAUER, T.K., 1995, *The Trouble with Computers: Usefulness, Usability, and Productivity* (Cambridge, Massachusetts: MIT Press) ISBN 0 262 12186 7.

LANSDALE, M.W. and ORMEROD, T.C., 1994, *Understanding Interfaces: A Handbook of Human–Computer Dialogue* (London: Academic Press) ISBN 0 12 528390 3.

LAUREL, B. (ed.), 1990, *The Art of Human–Computer Interface Design* (Reading, Massachusetts: Addison-Wesley) ISBN 0 201 51797 3.

LAUREL, B., 1991, *Computers as Theatre* (Reading, Massachusetts: Addison-Wesley) ISBN 0 201 51048 0.

LIFE, M.A., NARBOROUGH-HALL, C.S. and HAMILTON, W.I. (eds), 1990, *Simulation and the User Interface* (London: Taylor & Francis) ISBN 0 85066 803 4.

LONG, J. and WHITEFIELD, A. (eds), 1989, *Cognitive Ergonomics and Human–Computer Interaction* (Cambridge: Cambridge University Press) ISBN 0 521 37179 1.

MACAULAY, L., 1995, *Human–Computer Interaction for Software Designers* (London: International Thomson Computer Press) ISBN 1 85032 177 9.

MADDIX, F., 1990, *Human–Computer Interaction* (New York: Ellis Horwood) ISBN 0 13 446220 3.

MANDEL, T., 1997, *The Elements of User Interface Design* (New York: John Wiley) ISBN 0 471 16267 1.

MAYHEW, D.J., 1992, *Principles and Guidelines in Software User Interface Design* (Englewood Cliffs, New Jersey: Prentice Hall) ISBN 0 13 721929 6.

MAYHEW, D.J., 1999, *The Usability Engineering Lifecycle: A Practitioner's Handbook for User Interface Design* (San Francisco, California: Morgan Kaufmann) ISBN 0 55860 561 4.

MCALEESE, R. (ed.), 1999, *Hypertext: Theory into Practice* (2nd ed.) (Exeter, Devon: Intellect Books) ISBN 1 871516 28 5.

MEDYCKYJ-SCOTT, D. and HEARNSHAW, H.M. (eds), 1993, *Human Factors in Geographical Information Systems* (London: Belhaven Press) ISBN 1 85293 262 7.

MONK, A., 1984, *Fundamentals of Human–Computer Interaction* (London: Academic Press) ISBN 0 12 504580 8.

MONK, A., WRIGHT, P., HABER, J. and DAVENPORT, L., 1993, *Improving Your Human–Computer Interface: A Practical Technique* (New York: Prentice Hall) ISBN 0 13 010034 X.

MONK, A.F. and GILBERT, G.N. (eds), 1995, *Perspectives in HCI: Diverse Approaches* (London: Academic Press) ISBN 0 12 504575 1.

MULLET, K. and SANO, D., 1995, *Designing Visual Interfaces: Communication Oriented Techniques* (Englewood Cliffs, New Jersey: Sunsoft Press, Prentice Hall) ISBN 0 13 303389 9.

MYERS, B.A., 1988, *Creating User Interfaces by Demonstration* (Boston, Massachusetts: Academic Press) ISBN 0 12 512305 1.

NeXTSTEP COMPUTER INC., 1992, *NeXTSTEP User Interface Guidelines* (Reading, Massachusetts: Addison-Wesley) ISBN 0 201 63250 0.

NICKERSON, R.S., 1986, *Using Computers: The Human Factors of Information Systems* (Cambridge, Massachusetts: MIT Press) ISBN 0 262 14040 3.

NIELSEN, J. (ed.), 1989, *Coordinating User Interfaces for Consistency* (Boston, Massachusetts: Academic Press) ISBN 0 12 518400 X.

NIELSEN, J. (ed.), 1990, *Designing User Interfaces for International Use* (Amsterdam: Elsevier) ISBN 0 444 88428 9.

NIELSEN, J. (ed.), 1995, *Advances in Human–Computer Interaction, Volume 5* (Norwood, New Jersey: Ablex) ISBN 1 56750 199 0.

NORMAN, D.A., 1998, *The Invisible Computer* (Cambridge, Massachusetts: MIT Press) ISBN 0 262 14065 9.

NORMAN, D.A. and DRAPER, S.W. (eds), 1986, *User Centered System Design. New Perspectives on Human–Computer Interaction* (Hillsdale, New Jersey: Lawrence Erlbaum) ISBN 0 89859 781 1.

NORMAN, K.L., 1991, *The Psychology of Menu Selection* (Norwood, New Jersey: Ablex) ISBN 0 89391 553 X.

NOYES, J.M. and COOK, M. (eds), 1999, *Interface Technology: The Leading Edge* (Baldock, Hertfordshire: Research Studies Press) ISBN 0 86380 233 8.

OLSEN, D.R., 1992, *User Interface Management Systems: Models and Algorithms* (San Mateo, California: Morgan Kaufmann) ISBN 1 55860 220 8.

OPPERMANN, R. (ed.), 1994, *Adaptive User Support* (Hillsdale, New Jersey: Lawrence Erlbaum) ISBN 0 8058 1655 0.

ORAVEC, J.A., 1996, *Virtual Individuals, Virtual Groups: Human Dimensions of Groupware and Computer Networking* (Cambridge: Cambridge University Press) ISBN 0 521 45493 X.

PERLMAN, G., GREEN, G.K. and WOGALTER. M.S. (eds), 1995, *Human Factors Perspectives on Human–Computer Interaction* (Santa Monica, California: Human Factors and Ergonomics Society) ISBN 0 945289 05 7.

PETERSON, H.E. and SCHNEIDER, W. (eds), 1986, *Human–Computer Communications in Health Care* (Amsterdam: North-Holland) ISBN 0 444 87880 7.

PFLEGER, S., GONCALVES, J. and VARGHESE, K. (eds), 1995, *Advances in Human–Computer Interaction: Human Comfort and Security* (Berlin: Springer) ISBN 3 540 60145 7.

POWELL, J.E., 1990, *Designing User Interfaces* (San Marcos, California: Microtrend Books) ISBN 0 915391 40 6.

PREECE, J., 2000, *Online Communities: Designing Usability, Supporting Socialbility* (Chichester: Wiley) ISBN 0 471 80599 8.

PREECE, J., ROGERS, Y., SHARP, H., BENYON, D., HOLLAND, S and CAREY, T., 1994, *Human–Computer Interaction* (Wokingham, Berkshire: Addison-Wesley) ISBN 0 201 62769 8.

RASMUSSEN, J., ANDERSEN, H.B. and BERNSEN, N.O., 1991, *Human–Computer Interaction. Research Directions in Cognitive Science: European Perspectives Volume 3* (Hillsdale, New Jersey: Lawrence Erlbaum) ISBN 0 86377 113 0.

RUBIN, T., 1988, *User Interface Design for Computer Systems* (Chichester: Ellis Horwood) ISBN 0 7458 0299 0.

RUDISILL, M., LEWIS, C., POLSON, P.G. and MCKAY, T.D. (eds), 1996, *Human–Computer Interface Design: Success Stories, Emerging Methods, and Real-World Context* (San Francisco, California: Morgan Kaufmann) ISBN 1 55860 310 7.

SALVENDY, G. (ed.), 1987, *Cognitive Engineering in the Design of Human–Computer Interaction and Expert Systems* (Amsterdam: Elsevier) ISBN 0 444 42848 8.

SALVENDY, G. and SMITH, M.J. (eds) 1989, *Designing and Using Human–Computer Interfaces and Knowledge Based Systems* (Amsterdam: Elsevier) ISBN 0 444 88078 X.

SALVENDY, G. and SMITH, M.J. (eds) 1993, *Human–Computer Interaction: Software and Hardware Interfaces* (Amsterdam: Elsevier) ISBN 0 444 89540 X.

SALVENDY, G., SMITH, M.J. and KOUBEK, R.J. (eds) 1997, *Design of Computing Systems: Cognitive Considerations* (Amsterdam: Elsevier) ISBN 0 444 82183 X.

SCHNEIDER-HUFSCHMIDT, M., KUHME, T. and MALINOWSKI, U. (eds), 1993, *Adaptive User Interfaces: Principles and Practice* (Amsterdam: North-Holland) ISBN 0 444 81545 7.

SCOTT, D., 1991, *Human–Computer Interaction: A Cognitive Ergonomics Approach* (New York: Ellis Horwood) ISBN 0 13 395500 1.

SHACKEL, B. and RICHARDSON, S. (eds), 1991, *Human Factors for Informatics Usability* (Cambridge: Cambridge University Press) ISBN 0 521 36570 8.

SHNEIDERMAN, B. (ed.), 1993, *Sparks of Innovation in Human–Computer Interaction* (Norwood, New Jersey: Ablex Publishing Corporation) ISBN 1 56750 079 X.

SHNEIDERMAN, B., 1998, *Designing the User Interface* (3rd ed.) (Reading, Massachusetts: Addison-Wesley) ISBN 0 201 69497 2.

SHORROCK, B., 1988, *Systems Design and HCI: A Practical Handbook* (Wilmslow, Cheshire: Sigma Press) ISBN 1 85058 101 0.

SIME, M.E. and COOMBS, M.J., 1983, *Designing for Human–Computer Communication* (London: Academic Press) ISBN 0 12 643820 X.

SMITH, A., 1997, *Human–Computer Factors: A Study of Users and Information Systems* (London: McGraw-Hill) ISBN 0 07 709261 9.

SMITH, H.T. and GREEN, T.R.G., 1980, *Human Interaction with Computers* (London: Academic Press) ISBN 0 12 652852 7.–

SMITH, M.J. and SALVENDY, G. (eds), 1993, *Human–Computer Interaction: Applications and Case Studies* (Amsterdam: Elsevier) ISBN 0 444 89540 X.

SMITH, M.J., SALVENDY, G. and KOUBEK, R.J. (eds), 1997, *Design of Computing Systems: Social and Ergonomic Considerations* (Amsterdam: Elsevier) ISBN 0 444 82183 X.

SUCHMAN, L.A., 1987, *Plans and Situated Actions. The Problem of Human–Machine Communication* (Cambridge: Cambridge University Press) ISBN 0 521 33137 4.

SULLIVAN, J.W. and TYLER, S.W. (eds), 1991, *Intelligent User Interfaces* (Reading, Massachusetts: Addison-Wesley) ISBN 0 201 50305 0.

SUTCLIFFE, A., 1988, *Human–Computer Interface Design* (Basingstoke, Hampshire: Macmillan Education) ISBN 0 333 42898 6.

SUTCLIFFE, A., ZIEGLER, J. and JOHNSON, P. (eds), 1998, *Designing Effective and Usable Multimedia Systems* (Boston, Massachusetts: Kluwer Academic) ISBN 0 412 84270 X.

TAUBER, M.J. and ACKERMANN, D. (eds), 1991, *Mental Models and Human–Computer Interaction 2* (Amsterdam: North-Holland) ISBN 0 444 88602 8.

THIMBLEBY, H., 1990, *User Interface Design* (New York: ACM Press) ISBN 0 201 41618 2.

THOMAS, J.C. and SCHNEIDER, M.L. (eds), 1984, *Human Factors in Computer Systems* (Norwood, New Jersey: Ablex Publishing Corporation) ISBN 0 89391 146 1.

THOMAS, P.J. (ed.), 1995, *The Social and Interactional Dimensions of Human–Computer Interfaces* (Cambridge: Cambridge University Press) ISBN 0 521 45302 X.

TOGNAZZINI, B., 1992, *TOG on Interface* (Reading, Massachusetts: Addison-Wesley) ISBN 0 201 60842 1.

TREU, S., 1994, *User Interface Design: A Structured Approach* (New York: Plenum Press) ISBN 0 306 44681 2.

TREU, S., 1994, *User Interface Evaluation: A Structured Approach* (New York: Plenum Press) ISBN 0 306 44746 0.

Van Der VEER, G.C., GREEN, T.R.G., HOC, J.M. and MURRAY, D.M. (eds), 1988, *Working with Computers: Theory Versus Outcome* (London: Academic) ISBN 0 12 711705 9.

Van Der VEER, G.C. and MULDER, G. (eds), 1988, *Human–Computer Interaction: Psychonomic Aspects* (Berlin: Springer-Verlag) ISBN 0 387 18901 7.

Van VLIET, J.C. (ed.), 1986, *Text Processing and Document Manipulation* (Cambridge: Cambridge University Press) ISBN 0 521 32592 7.

VASKE, J.J. and GRANTHAM, C.E., 1990, *Socializing the Human–Computer Environment* (Norwood, New Jersey: Ablex) ISBN 0 89391 471 1.

VASSILIOU, Y. (ed.), 1984, *Human Factors and Interactive Computer Systems* (Norwood, New Jersey: Ablex Publishing Corporation) ISBN 0 89391 182 8.

WEINSCHENK, S. and BARKER, D.T., 2000, *Designing Effective Speech Interfaces* (New York: John Wiley) ISBN 0 471 37545 4.

WEINSCHENK, S., JAMAR, P., and YEO, S.C., 1997, *GUI Design Essentials* (New York: Wiley) ISBN 0 471 17549 8.

WINOGRAD, T. and FLORES, F., 1986, *Understanding Computers and Cognition* (Norwood, New Jersey: Ablex Publishing Corporation) ISBN 0 89391 050 3.

WOOD, L.E. (ed.), 1998, *User Interface Design: Bridging the Gap from User Requirements to Design* (Boca Raton, Florida: CRC Press) ISBN 0 8493 3125 0.

ZETIE, C., 1995, *Practical User Interface Design: Making GUIs Work* (London: McGraw-Hill) ISBN 0 07 709167 1.

Complex systems

BEEVIS, D. (Ed), 1999, *Analysis Techniques for Human-Machine Systems Design* (Ohio: Human Systems Infromation Analysis Center, Wright-Patterson Air Force Base) No ISBN.

DENNIS, A., and WIXON, B.H., 2000, *Systems Analysis and Design* (New York: John Wiley) ISBN 0 471 24100 8.

EDWARDS, E. and LEES, F.P., 1973, *Man and Computer in Process Control* (London: Institution of Chemical Engineers), no ISBN.

FLACH, J., HANCOCK, P., CAIRD, J. and VICENTE, K. (eds), 1995, *Global Perspectives on the Ecology of Human–Machine Systems, Volume 1* (Hillsdale, New Jersey: Lawrence Erlbaum) ISBN 0 8058 1381 0.

GARCIA, R.R., 1988, *Human Factors in Systems Development* (Manchester: NCC Publications) ISBN 0 85012 723 8.

GILL, K.S. (ed.), 1996, *Human Machine Symbiosis: The Foundations of Human-Centred Systems Design* (Berlin: Springer) ISBN 3 540 76024 5.

Institution of Chemical Engineers & Ergonomics Society, 1984, *Ergonomics Problems in Process Operations* (Oxford: Pergamon) ISBN 0 08 030282 3.

MAJCHRZAK, A., 1988, *The Human Side of Factory Automation* (San Francisco, California: Jossey-Bass) ISBN 1 55542 050 8.

MCMILLAN, G.R., BEEVIS, D., SALAS, E., STRUB, M.H., SUTTON, R., and VAN BREDA, L. (eds), 1989, *Applications of Human Performance Models to System Design* (New York: Plenum Press) ISBN 0 306 43242 0.

MILL, R.C. (ed.), 1992, *Human Factors in Process Operations* (Rugby, Warwickshire: Institution of Chemical Engineers) ISBN 0 85295 294 5.

MOULOUA, M. and KOONCE, J.M. (eds), 1997, *Human–Automation Interaction: Research and Practice* (Mahwah, New Jersey: Lawrence Erlbaum) ISBN 0 8058 2841 9.

NEWMAN, W.M. and LAMMING, M.G., 1995, *Interactive System Design* (Wokingham, Berkshire: Addison-Wesley) ISBN 0 201 63162 8.

NTUEN, C.A. and PARK, E.H., 1996, *Human Interaction with Complex Systems: Conceptual Principles and Design Practice* (Boston, Massachusetts: Kluwer Academic Publishers) ISBN 0 7923 9779 7.

PARASURAMAN, R. and MOULOUA, M. (eds), 1996, *Automation and Human Performance: Theory and Applications* (Mahwah, New Jersey: Lawrence Erlbaum) ISBN 0 8058 1616 X.

RAHIMI, M. and KARWOWSKI, W. (eds), 1992, *Human-Robot Interaction* (London: Taylor & Francis) ISBN 0 85066 809 3.

RASMUSSEN, J., BREHMER, B., and LEPLAT, J. (eds), 1991, *Distributed Decision Making. Cognitive Models for Cooperative Work* (Chichester: Wiley) ISBN 0 471 92828 3.

RASMUSSEN, J., PEJTERSEN, A.M., and GOODSTEIN, L.P., 1994, *Cognitive Systems Engineering* (New York: Wiley) ISBN 0 471 01198 3.

RASMUSSEN, J. and ROUSE, B. (eds), 1981, *Human Detection and Diagnosis of System Failures* (New York: Plenum Press) ISBN 0 306 40744 2.

RIJNSDORP, J.E., 1991, *Integrated Process Control and Automation* (Amsterdam: Elsevier) ISBN 0 444 88128 X.

ROUSE, W.B. (ed.), 1984, *Advances in Man–Machine Systems Research, Volume 1* (Greenwich, Conn.: JAI Press) ISBN 0 89232 404 X.

ROUSE, W.B. (ed.), 1987, *Advances in Man–Machine Systems Research, Volume 3* (Greenwich, Conn.: JAI Press) ISBN 0 89232 659 X.

ROUSE, W.B. (ed.), 1988, *Advances in Man–Machine Systems Research, Volume 4* (Greenwich, Conn.: JAI Press) ISBN 0 89232 753 7.

ROUSE, W.B. (ed.), 1989, *Advances in Man–Machine Systems Research, Volume 5* (Greenwich, Conn.: JAI Press) ISBN 1 55938 011 X.

ROUSE, W.B. (ed.), 1993, *Human/Technology Interaction in Complex Systems, Volume 6* (Greenwich, Conn.: JAI Press) ISBN 1 55938 192 2.

ROUSE, W.B. (ed.), 1995, *Human/Technology Interaction in Complex Systems, Volume 7* (Greenwich, Conn.: JAI Press) ISBN 1 55938 843 9.

ROUSE, W.B. (ed.), 1996, *Human/Technology Interaction in Complex Systems, Volume 8* (Greenwich, Conn.: JAI Press) ISBN 0 7623 0013 2.

SALAS, E. (ed.), 1999, *Human/Technology Interaction in Complex Systems, Volume 9* (Greenwich, Connecticut: JAI Press) ISBN 0 7623 01775 5.

SCERBO, M.W. and MOULOUA, M. (eds), *Automation Technology and Human Performance* (Mahwah, New Jersey: Lawrence Erlbaum) ISBN 0 8058 3135 5.

SHERIDAN, T.B., 1992, *Telerobotics, Automation, and Human Supervisory Control* (Cambridge, Mass.: MIT Press) ISBN 0 262 19316 7.

SHERIDAN, T.B. and FERRELL, W.R., 1974, *Man–Machine Systems: Information, Control, and Decision Models of Human Performance* (Cambridge, Mass.: MIT Press) ISBN 0 262 19118 0.

SHERIDAN, T.B. and JOHANNSEN, G., 1976, *Monitoring Behavior and Supervisory Control* (New York: Plenum Press) ISBN 0 306 32881 X.

SHERIDAN, T.B. and VAN LUNTEREN, T. (eds), 1997, *Perspectives on the Human Controller: Essays in Honor of Henk G. Stassen* (Mahwah, New Jersey: Lawrence Erlbaum) ISBN 0 8058 2190 2.

SIGRIMIS, N. and GROUMPOS, P. (eds), 1999, *Control Applications and Ergonomics in Agriculture* (Amsterdam: Elsevier) ISBN 0 08 043229 8.

SINGLETON, W.T., 1974, *Man-Machine Systems* (Harmondsworth, Middlesex: Penguin Education), no ISBN.

SUTTON, R., 1990, *Modelling Human Operators in Control System Design* (Taunton, Somerset: Research Studies Press) ISBN 0 86380 109 9.

WEIR, G.R.S. and ALTY, J.L. (eds), 1991, *Human–Computer Interaction and Complex Systems* (London: Academic Press) ISBN 0 12 742660 4.

WISE, J.A., HOPKIN, V.D. and STAGER, P. (eds), 1993, *Verification and Validation of Complex Systems: Human Factors Issues* (Berlin: Springer-Verlag) ISBN 0 387 56574 4.

Transport systems

BALDWIN, R. (ed.), 1999, *Developing the Future Aviation System* (Aldershot, Hampshire: Ashgate) ISBN 0 291 39843 X.

BARFIELD, W. and DINGUS, T.A. (eds), 1998, *Human Factors in Intelligent Transportation Systems* (Mahwah, New Jersey: Lawrence Erlbaum) ISBN 0 8058 1434 5.

BILLINGS, C.E., 1997, *Aviation Automation: The Search for A Human-Centered Approach* (Mahwah, New Jersey: Lawrence Erlbaum) ISBN 0 8058 2127 9.

CUSHING, S., 1994, *Fatal Words: Communication Clashes and Aircraft Crashes* (Chicago: University of Chicago Press) ISBN 0 226 13200 5.

DEITZ, S.R. and THOMS, W.E. (eds), 1992, *Pilots, Personality, and Performance: Human Behavior and Stress in the Skies* (Westport, Connecticut: Quorum Books).

EDWARDS, M. and EDWARDS, E., 1990, *The Aircraft Cabin: Managing the Human Factors* (Aldershot, Hampshire: Gower Technical) ISBN 0 566 09056 2.

FARMER, E. (ed.), 1991, *Stress and Error in Aviation* (Aldershot, Hampshire: Avebury Technical) ISBN 1 85628 169 8.

GALOTTI, V.P., 1997, *The Future Air Navigation System (Fans): Communication Navigation Surveillance Air Traffic Management* (Aldershot, Hampshire: Avebury Aviation) ISBN 0 291 39833 2.

GARLAND, D.J., WISE, J.A. and HOPKIN, V.D. (eds), 1999, *Handbook of Aviation Human Factors* (Mahwah, New Jersey: Lawrence Erlbaum) ISBN 0 8058 1680 1.

GREEN, R.G., MUIR, H., JAMES, M., GRADWELL, D., and GREEN, R.L., 1996, *Human Factors for Pilots* (2nd ed.) (Aldershot, Hampshire: Avebury Technical) ISBN 0 291 39827 8.

HARTLEY, L. (ed.), 1995, *Fatigue and Driving: Driver Impairment, Driver Fatigue and Driving Simulation* (London: Taylor & Francis) ISBN 0 7484 0262 4.

HAWKINS, F.H., 1993, *Human Factors in Flight* (2nd ed.) (Aldershot, Hampshire: Gower Technical Press) ISBN 1 85742 134 5.

HAYWARD, B.J. and LOWE, A.R. (eds), 1996, *Applied Aviation Psychology: Achievement, Change and Challenge* (Aldershot, Hampshire: Avebury Aviation) ISBN 0 29139 840 5.

HOPKIN, V.D., 1995, *Human Factors in Air Traffic Control* (London: Taylor & Francis) ISBN 0 7484 0357 4.

HUNT, G.J.F. (ed.), 1997, *Designing Instruction for Human Factors Training in Aviation* (Aldershot, Hampshire: Avebury Aviation) ISBN 0 291 39800 6.

ISAAC, A.R. and RUITENBERG, B., 1999, *Air Traffic Control: Human Performance factors* (Aldershot, Hampshire: Ashgate Publishing) ISBN 0 291 39854 5.

JENSEN, R.S. (ed.), 1989, *Aviation Psychology* (Aldershot, Hampshire: Gower Technical) ISBN 0 291 39778 6.

JENSEN, R.S., 1995, *Pilot Judgement and Crew Resource Management* (Aldershot, Hampshire: Ashgate Publishing) ISBN 0 291 39804 9.

JOHNSTON, N., FULLER, R., and MCDONALD, N. (eds), 1995, *Aviation Psychology: Training and Selection* (Aldershot, Hampshire: Ashgate Publishing) ISBN 0 291 39819 7.

JOHNSTON, N., MCDONALD, N. and FULLER, R. (eds), 1994, *Aviation Psychology in Practice* (Aldershot, Hampshire: Avebury Technical) ISBN 0 291 39808 1.

MAURINO, D.E., REASON, J., JOHNSTON, N., and LEE, R.B., 1995, *Beyond Aviation Human Factors* (Aldershot, Hampshire: Ashgate Publishing) ISBN 0 291 39822 7.

MCDONALD, N., 1984, *Fatigue, Safety and the Truck Driver* (London: Taylor & Francis) ISBN 0 85066 207 9.

MCDONALD, N., JOHNSTON, N., and FULLER, R. (eds), 1995, *Applications of Psychology to the Aviation System* (Aldershot, Hampshire: Ashgate Publishing) ISBN 0 291 39818 9.

MICHON, J.A. (ed.), 1993, *Generic Intelligent Driver Support* (London: Taylor & Francis) ISBN 0 7484 0069 9.

National Research Council, Committee on the Effect of Smaller Crews on Maritime Safety, 1990, *Crew Size and Maritime Safety* (Washington, DC: National Academy Press) ISBN 0 309 04375 1.

NOY, T.I. (ed.), 1997, *Ergonomics and Safety of Intelligent Driver Interfaces* (Mahwah, New Jersey: Lawrence Erlbaum) ISBN 0 8058 1956 8.

OBORNE, D.J. and LEVIS, J.A. (eds), 1980, *Human Factors in Transport Research* (London: Academic Press) Vol.1 ISBN 0 12 523 8010; Vol.2 ISBN 0 12 523 8029.

O'HANLON, J.F. and De GIER, J.J. (eds), 1986, *Drugs and Driving* (London: Taylor & Francis) ISBN 0 85066 290 7.

O'HARE, D. and ROSCOE, S., 1990, *Flightdeck Performance. The Human Factor* (Ames, Iowa: Iowa State University Press) ISBN 0 8138 0161 3.

O'NEIL, H.F. and ANDREWS, D.H. (Eds), 2000, *Aircrew Training and Assessment* (Mahwah, New Jersey: Lawrence Erlbaum Associates) ISBN 0 8058 2977 6.

PARKES, A.M. and FRANZEN, S. (eds), 1993, *Driving Future Vehicles* (London: Taylor & Francis) ISBN 0 7484 0042 7.

PAUWELUSSEN, J.P. and PACEJKA, H.B. (eds), 1995, *Smart Vehicles* (Lisse, The Netherlands: Swets & Zeitlinger) ISBN 90 265 1456 5.

PEACOCK, B. and KARWOWSKI, W. (eds), 1993, *Automotive Ergonomics* (London: Taylor & Francis) ISBN 0 7484 0005 2.

ROSCOE, S.N., 1980, *Aviation Psychology* (Oxford: Oxford University Press) ISBN 0 8138 1925 3.

SATCHELL, P., 1993, *Cockpit Monitoring and Alerting Systems* (Aldershot, Hampshire: Ashgate) ISBN 1 85742 109 4.

SATCHELL, P., 1998, *Innovation and Automation* (Aldershot, Hampshire: Ashgate Publishing) ISBN 1 84014 315 0.

SEAMSTER, T.L., REDDING, R.E., and KAEMPF, G.L., 1997, *Applied Cognitive Task Analysis in Aviation* (Aldershot, Hampshire: Avebury Aviation) ISBN 0 291 39830 8.

SLOAN, S.J. and COOPER, C.L., 1986, *Pilots Under Stress* (London: Routledge & Kegan Paul) ISBN 0 7102 0479 5.

SMOLENSKY, M.W. and STEIN, E.S. (eds), 1998, *Human Factors in Air Traffic Control* (San Diego, California: Academic Press) ISBN 0 12 653010 6.

STOKES, A. and KITE, K., 1994, *Flight Stress: Stress, Fatigue, and Performance in Aviation* (Aldershot, Hampshire: Ashgate) ISBN 0 291 39806 5.

VON HOLST, H., NYGREN, A., and THORD, R. (eds), 1997, *Transportation, Traffic Safety and Health: The New Mobility* (Berlin: Springer) ISBN 3 540 62524 0.

WICKENS, C.D., MAVOR, A.S., and MCGEE, J.P. (eds), 1997, *Flight to the Future: Human Factors in Air Traffic Control* (Washington, DC: National Academy Press) ISBN 0 309 05637 3.

WICKENS, C.D., MAVOR, A.S., PARASURAMAN, R., and MCGEE, J.P. (eds), 1998, *The Future of Air Traffic Control: Human Operators and Automation* (Washington, DC: National Academy Press) ISBN 0 309 06412 0.

WIENER, E.L. and NAGEL, D.C. (eds), 1988, *Human Factors in Aviation* (San Diego, California: Academic Press) ISBN 0 12 750030 8.

Musculoskeletal injuries and manual materials handling

APTS, D.W., 1992, *Back Injury Prevention Handbook* (Boca Raton, Florida: Lewis Publishers) ISBN 0 87371 419 9.

ARKSEY, H., 1998, *RSI and the Experts: The Construction of Medical Knowledge* (London: Taylor & Francis) ISBN 1 85728 813 0.

AYOUB, M.M. and MITAL, A., 1989, *Manual Materials Handling* (London: Taylor & Francis) ISBN 0 85066 383 0.

DEMBE, A.T., 1996, *Occupation and Disease: How Social Factors Affect the Conception of Work-Related Disorders* (New Haven, Connecticut: Yale University Press) ISBN 0 300 06436 5.

ERDIL, M. and DICKERSON, O.B. (eds), 1997, *Cumulative Trauma Disorders: Prevention, Evaluation, and Treatment* (New York: Van Nostrand Reinhold) ISBN 0 442 01074 5.

FEYER, A.M. and WILLIAMSON, A. (eds), 1998, *Occupational Injury: Risk, Prevention and Intervention* (London: Taylor & Francis) ISBN 0 7484 0647 6.

KARWOWSKI, W., WOGALTER, M.S., and DEMPSEY, P.G. (eds), 1997, *Ergonomics and Musculoskeletal Disorders: Research on Manual Handling 1983–1996* (Santa Monica, California: Human Factors and Ergonomics Society) ISBN 0 945289 09 X.

KENDZIOR, R.J., 1999, *Slip-and-Fall Prevention Made Easy: A Coomprehensive Guide to Preventing Accidents* (Rockville, Maryland: Government Institutes) ISBN 0 86587 664 9.

KHALIL, T.M., ABDEL-MOTY, E.M., ROSOMOFF, R.S., and ROSOMOFF, H.L., 1993, *Ergonomics in Back Pain: A Guide to Prevention and Rehabilitation* (New York: Van Nostrand Reinhold) ISBN 0 442 01375 2.

KROEMER, K.H.E., 1997, *Ergonomic Design of Material Handling Systems* (Boca Raton, Florida: Lewis Publishers) ISBN 1 56670 224 0.

KUMAR, S. (ed.), 1999, *Biomechanics in Ergonomics* (London: Taylor & Francis) 0 7484 0704 9.

KUORINKA, I. et al., 1995, *Work Related Musculoskeletal Disorders (WMSDS): A Reference Book for Prevention* (London: Taylor & Francis) ISBN 0 7484 0132 6.

LLOYD, P., TARLING, C., TROUP, J.D.G. and WRIGHT, B., 1987, *The Handling of Patients: A Guide for Nurses* (2nd ed.) (Teddington, Middlesex: Back Pain Association) ISBN 0 950 7726 2 3.

London Hazards Centre, 1997, *RSI Hazards Handbook* (London : London Hazards Centre Trust) ISBN 0 948974 14 1.

MARRAS, W.S., KARWOWSKI, W., SMITH, J.L. and PACHOLSKI, L. (eds), 1993, *The Ergonomics of Manual Work* (London: Taylor & Francis) ISBN 0 7484 0060 5.

MAYER, T.J., GATCHEL, R.J. and POLATI, P.B. (Eds), 2000, *Occupational Musculoskeletal Disorders Function, Outcomes, and Evidence* (Philadelphia: Lippincott Williams & Wilkins) ISBN 0 309 06397 3.

MCIVOR, D., 1987, *Overuse Injuries: The RSI Phenomenon* (Melbourne: Pitman Publishing) ISBN 0 85896 374 4.

MITAL, A., NICHOLSON, A.S. and AYOUB, M.M., 1997, *A Guide to Manual Materials Handling* (2nd ed.) (London: Taylor & Francis) ISBN 0 7484 0728 6.

MOON, S.D. and SAUTER, S.L. (eds), 1996, *Beyond Biomechanics: Psychosocial Aspects of Musculoskeletal Disorders in Office Work* (London: Taylor & Francis) ISBN 0 7484 0322 1.

MOORE, J.S. and GARG, A. (eds), 1992, *Ergonomics: Low-Back Pain, Carpal Tunnel Syndrome, and Upper Extremity Disorders in the Workplace, Occupational Medicine State of the Art Reviews Volume 7 Number 4* (Philadelphia, Pennsylvania: Hanley & Belfus) ISBN 1 56053 095 2.

National Back Pain Association, 1997, *The Guide to Handling of Patients* (4th ed.) (Teddington, Middlesex National Back Pain Association) ISBN 0 9530582 0 4.

National Research Council, 1999, *Work-Related Musculoskeletal Disorders: Report, Workshop Summary, and Workshop Papers* (Washington D.C.: National Academy Press) ISBN 0 309 06397 3.

NELSON, M. (ed.), 1983, *Low Back Pain and Industrial and Social Disablement* (Teddington, Middlesex: National Back Pain Association) ISBN 0 9507726 1 5.

NORDIN, M, ANDERSSON, G.B.J. and POPE, M.H., 1997, *Musculoskeletal Disorders in the Workplace: Principles and Practice* (St Louis, Missouri: Mosby) ISBN 0 8016 7984 2.

PARKER, K.G. and IMBUS, H.R., 1992, *Cumulative Trauma Disorders* (Boca Raton, Florida: Lewis Publishers) ISBN 0 87371 322 2.

PEATE, W.F. and LUNDA, K.A., 1997, *Cumulative Trauma Disorders: A Practical Guide to Prevention and Control* (Rockville, Maryland: Government Institutes) ISBN 0 86587 553 7.

PUTZ-ANDERSON, V. (ed.), 1988, *Cumulative Trauma Disorders. A Manual for Musculoskeletal Diseases of the Upper Limbs* (London: Taylor & Francis) ISBN 0 85066 405 5.

RING, L., 1993, *Back*s (Auckland, New Zealand: UNIPRINT) ISBN 0 473 01832 2.

ROWE, M.L., 1983, *Backache at Work* (Fairport, New York: Perinton Press), no ISBN.

WIESEL, S.W., FEFFER, H.L., BORENSTEIN, D.G., and ROTHMAN, R.H., 1989, *Industrial Low Back Pain. A Comprehensive Approach* (Charlottesville, Virginia: Michie) ISBN 0 87473 449 5.

WILLS, G. and COOPER, C.L., 1988, *Pressure Sensitive. Popular Musicians Under Stress* (London: SAGE Publications) ISBN 0 8039 8142 2.

Health and safety

CANTER, D. (ed.), 1990, *Fires and Human Behaviour* (2nd ed.) (London: David Fulton) ISBN 1 85346 105 9.

COOPER, D., 1998, *Improving Safety Culture: A Practical Guide* (Chichester: Wiley) ISBN 0 471 95821 2.

ERICKSON, P.A., *Practical Guide to Occupational Health and Safety* (San Diego, California: Academic Press) ISBN 0 12 240570 6.

FLIN, R., 1996, *Sitting in the Hot Seat: Leaders and Teams for Critical Incident Management* (Chichester: Wiley) ISBN 0 471 95796 8.

GLENDON, A.I. and MCKENNA, E.F., 1995, *Human Safety and Risk Management* (London: Chapman & Hall) ISBN 0 412 40250 5.

HALE, A. and BARAM, M., 1998, *Safety Management: The Challenge of Change* (Amsterdam: Elsevier) ISBN 0 08 043075 9.

HALE, A.R. and GLENDON, A.I., 1987, *Individual Behaviour in the Control of Danger* (Amsterdam: Elsevier) ISBN 0 444 42838 0.

HAMMER, W. and PRICE, D., 2001, *Occupational Safety Management and Engineering* (5th Edition) (Upper Saddle River, New Jersey: Prentice Hall) ISBN 0 13 896515 3.

HANSEN, D.J. (ed.), 1991, *The Work Environment. Volume 1: Occupational Health Fundamentals* (Chelsea, Michigan: Lewis Publishers) ISBN 0 87371 303 6.

HANSEN, D.J. (ed.), 1991, *The Work Environment. Volume 3: Indoor Health Hazards* (Boca Raton, Florida: Lewis Publishers) ISBN 0 87371 393 1.

HARRINGTON, J.M. and GILL, F.S., 1992, *Occupational Health* (Oxford: Blackwell Scientific Publications) ISBN 0 632 03189 1.

HOYOS, C.G. and ZIMOLONG, B., 1988, *Occupational Safety and Accident Prevention. Behavioral Strategies and Methods* (Amsterdam: Elsevier) ISBN 0 444 70478 7.

HURRELL, J.J., MURPHY, L.R., SAUTER, S.L. and COOPER, C.L. (eds), 1988, *Occupational Stress: Issues and Developments in Research* (New York: Taylor & Francis) ISBN 0 85066 418 7.

KARASEK, R. and THEORELL, T., 1990, *Healthy Work: Stress, Productivity, and the Reconstruction of Working Life* (New York: Basic Books) ISBN 0 465 02896 9.

KENEN, R.H., 1993, *Reproductive Hazards in the Workplace: Mending Jobs, Managing Pregnancies* (New York: Harrington Park Press) ISBN 1 56023 006 1.

KERR, J., GRIFFITHS, A. and COX, T. (eds), 1996, *Workplace Health, Employee Fitness and Exercise* (London: Taylor & Francis) ISBN 0 7484 0143 1.

LANCASTER, J., 1996, *Engineering Catastrophes: Causes and Effects of Major Accidents* (Cambridge: Abington Publishing) ISBN 1 85573 204 1.

LEES, F.P., 1996, *Loss Prevention in the Process Industries* (2nd ed.) (Oxford: Butterworth-Heinemann) ISBN 0 7506 1547 8.

MATTILA, M. and KARWOWSKI, W. (eds), 1992, *Computer Applications in Ergonomics, Occupational Safety and Health* (Amsterdam: North-Holland) ISBN 0 444 89605 8.

MISUMI, J., WILPERT, B. and MILLER, R. (eds), 1998, *Nuclear Safety: A Human Factors Perspective* (London: Taylor & Francis) ISBN 0 7484 0818 5.

MITAL, A. and KARWOWSKI, W., 1988, *Ergonomics and Rehabilitation* (London: Taylor & Francis) ISBN 0 85066 456 X.

NEAL, A.C. and WRIGHT, F.B. (eds), 1992, *The European Communities' Health and Safety Legislation* (London: Chapman & Hall) ISBN 0 412 46690 2.

NICHOLSON, A.S. and RIDD, J.E. (eds), 1988, *Health, Safety and Ergonomics* (London: Butterworths) ISBN 0 408 02386 4.

NORO, K. (ed.), 1987, *Occupational Health and Safety in Automation and Robotics* (London: Taylor & Francis) ISBN 0 85066 351 2.

OXENBURGH, M., 1991, *Increasing Productivity and Profit Through Health and Safety* (North Ryde, NSW, Australia: CCH International) ISBN 1 86264 264 8.

PANTRY, S. (ed.), 1995, *Occupational Health* (London: Chapman & Hall) ISBN 0 412 60410 8.

PEARCE, B.G. (ed.), 1984, *Health Hazards of VDTs?* (Chichester: Wiley) ISBN 0 471 90065 6.

QUICK, J.C., MURPHY, L.R., and HURRELL, J.J. (eds), 1992, *Stress and Well-being at Work: Assessments and Interventions for Occupational Mental Health* (Washington, DC: American Psychological Association) ISBN 1 55798 175 2.

QUICK, J.C., QUICK, J.D., NELSON, D.L., and HURRELL, J.J., 1997, *Preventive Stress Management in Organizations* (Washington, DC: American Psychological Association) ISBN 1 55798 432 8.

REDMILL, F. and RAJAN, J. (eds), 1997, *Human Factors in Safety-Critical Systems* (Oxford: Butterworth-Heinemann) ISBN 0 7506 2715 8.

RIDLEY, J. and CHANNING, J. (eds), 1999, *Safety at Work* (5th ed.) (Oxford: Butterworth-Heinemann) ISBN 0 7506 4018 9.

RODAHL, K., 1993, *Stress Monitoring in the Workplace* (Boca Raton, Florida: Lewis Publishers) ISBN 0 87371 655 8.

SAUTER, S.L., DAINOFF, M.J. and SMITH, M.J. (eds), 1990, *Promoting Health and Productivity in the Computerized Office: Models of Successful Ergonomic Interventions* (London: Taylor & Francis) ISBN 0 85066 756 9.

SAUTER, S.L., HURRELL, J.J., and COOPER, C.L. (eds), 1989, *Job Control and Worker Health* (Chichester: Wiley) ISBN 0 471 92355 9.

SAUTER, S.L. and MURPHY, L.R. (eds), 1995, *Organizational Risk Factors for Job Stress* (Washington, DC: American Psychological Association) ISBN 1 55798 297 X.

SCALET, E.A., 1987, *VDT Health and Safety. Issues and Solutions* (Lawrence, Kansas: Ergosyst Associates) ISBN 0 916313 13 1.

SCHAUFELI, W. and ENZMANN, D., 1998, *The Burnout Companion to Study and Practice: A Critical Analysis* (London: Taylor & Francis) ISBN 0 7484 0698 0.

SINGLETON, W.T. and DIRKX, J. (eds), 1991, *Ergonomics, Health and Safety: Perspectives for the Nineties* (Leuven: Leuven University Press:) ISBN 90 6186 467 4.

SMITH, M.J. and SALVENDY, G. (eds), 1989, *Work with Computers: Organizational, Management, Stress and Health Aspects* (Amsterdam: Elsevier) ISBN 0 444 88077 1.

STANTON, N. (ed.), 1996, *Human Factors in Nuclear Safety* (London: Taylor & Francis) ISBN 0 7484 0166 0.

STRANKS, J., 1994, *Human Factors and Safety* (London: Pitman Publishing) ISBN 0 273 60440 6.

SULOWSKI, A.C. (ed.), 1991, *Fundamentals of Fall Protection* (Toronto: International Society for Fall Protection) ISBN 0 921952 01 5.

SUTHERLAND, V.J. and COOPER, C.L., 1990, *Understanding Stress: A Psychological Perspective for Health Professionals* (2nd ed.) (London: Chapman & Hall) ISBN 0 412 33930 7.

VAN STEEN, J. (ed.), 1996, *Safety Performance Measurement* (Rugby, Warwickshire: Institution of Chemical Engineers) ISBN 0 85295 382 8.

WARING, A., 1996, *Safety Management Systems* (London: Chapman & Hall) ISBN 0 412 71910 X.

WEISS, S.M., FIELDING, J.E. and BAUM, A. (eds), 1991, *Health at Work* (Hillsdale, New Jersey: Lawrence Erlbaum) ISBN 0 8058 0770 5.

WILLIAMS, T.A., 1988, *Computers, Work and Health: A Socio-Technical Approach* (London: Taylor & Francis) ISBN 0 85066 453 5.

WOODSON, W.E., 1998, *Human Factors Engineering for Forensic and Safety Specialists: Improper Design Can Lead to Product Mis-Use and Personal Injury* (Tucson, Arizona: Lawyers & Judges Publishing) ISBN 0 913875 40 6.

Human error and reliability

BOGNER, M.S. (ed.), 1994, *Human Error in Medicine* (Hillsdale, New Jersey: Lawrence Erlbaum) ISBN 0 8058 1385 3.

BROADBENT, D.E., REASON, J. and BADDELEY, A. (eds), 1990, *Human Factors in Hazardous Situations* (Oxford: Clarendon) ISBN 0 19 852191 X.

COX, S. and COX, T., 1996, *Safety, Systems and People* (Oxford: Butterworth-Heinemann) ISBN 0 7506 2089 7.

DHILLON, B.S., 1986, *Human Reliability with Human Factors* (Oxford: Pergamon) ISBN 0 08 033981 6.

DOUGHERTY, E.M. and FRAGOLA, J.R., 1988, *Human Reliability Analysis. A Systems Engineering Approach with Nuclear Power Plant Applications* (New York: Wiley) ISBN 0 471 60614 6.

DRURY, C.G. and FOX, J.G. (eds), 1975, *Human Reliability in Quality Control* (London: Taylor & Francis) ISBN 0 85066 088 2.

GERTMAN, D.I. and BLACKMAN, H.S., 1994, *Human Reliability and Safety Analysis Data Handbook* (New York: John Wiley) ISBN 0 471 59110 6.

GOODSTEIN, L.P., ANDERSON, H.B. and OLSEN, S.E. (eds), 1988, *Tasks, Errors and Mental Models* (London: Taylor & Francis) ISBN 0 85066 401 2.

HARMS-RINGDAHL, L., 1993, *Safety Analysis: Principles and Practice in Occupational Safety* (Barking, Essex: Elsevier Science).

HOLLNAGEL, E., 1993, *Human Reliability Analysis: Context and Control* (London: Academic Press) ISBN 0 12 352658 2.

HOLLNAGEL, E., 1998, *Cognitive Reliability and Error Analysis Method: CREAM* (Amsterdam: Elsevier) ISBN 0 08 042848 7.

KIRWAN, B., 1994, *A Guide to Practical Human Reliability Assessment* (London: Taylor & Francis) ISBN 0 7484 0111 3.

KLETZ, T., 1991, *An Engineer's View of Human Error* (2nd ed.) (Rugby, Warwickshire: Institution of Chemical Engineers) ISBN 0 85295 265 1.

KLETZ, T., 1993, *Lessons from Disaster* (Rugby, Warwickshire: Institution of Chemical Engineers) ISBN 0 85295 307 0.

KLETZ, T., CHUNG, P., BROOMFIELD, E. and SHEN-ORR, C., 1995, *Computer Control and Human Error* (Rugby, Warwickshire: Institution of Chemical Engineers) ISBN 0 85295 362 3.

PARK, K.S., 1987, *Human Reliability. Analysis, Prediction, and Prevention of Human Errors* (Amsterdam: Elsevier) ISBN 0 444 42727 9.

PERROW, C., 1984, *Normal Accidents. Living with High-Risk Technologies* (New York: Basic Books) ISBN 0 465 05142 1.

PETERSEN, D., 1996, *Human Error Reduction and Safety Management* (3rd ed.) (New York: Van Nostrand Reinhold) ISBN 0 442 02183 6.

PITBLADO, R. and TURNEY, R. (eds), 1996, *Risk Assessment in the Process Industries* (2nd ed.) (Rugby, Warwickshire: Institution of Chemical Engineers) ISBN 0 85295 323 2.

RASMUSSEN, J., DUNCAN, K. and LEPLAT, J. (eds), 1987, *New Technology and Human Error* (Chichester: Wiley) ISBN 0 471 91044 9.

REASON, J., 1990, *Human Error* (Cambridge: Cambridge University Press) ISBN 0 521 31419 4.

REASON, J., 1997, *Managing the Risks of Organizational Accidents* (Aldershot, Hampshire: Ashgate Publishing) ISBN 1 84014 105 0.

SAYERS, B.A. (ed.), 1988, *Human Factors and Decision Making: Their Influence on Safety and Reliability* (London: Elsevier Applied Science) ISBN 1 85166 289 8.

SENDERS, J.W. and MORAY, N.P., 1991, *Human Error: Cause, Prediction, and Reduction* (Hillsdale, New Jersey: Lawrence Erlbaum) ISBN 0 89859 598 3.

SILVERMAN, B.G., 1992, *Critiquing Human Error: A Knowledge Based Human–Computer Collaboration Approach* (London: Academic Press) ISBN 0 12 643740 8.

SINGLETON, W.T. and HOVDEN, J. (eds), 1994, *Risk and Decisions* (2nd ed.) (Chichester: Wiley) ISBN 0 471 95224 9.

SWAIN, A.D., 1989, *Comparative Evaluation of Methods for Human Reliability Analysis* (Cologne: Gesellschaft fur Reaktorsicherheit (GRS) mbH) ISBN 3 923875 21 5.

TRIMPOP, R.M., 1994, *The Psychology of Risk Taking Behavior* (Amsterdam: North-Holland) ISBN 0 444 89961 8.

VINCENT, C. and DE MOL, B., (Eds), 2000, *Safety in Medicine* (Amsterdam: Pergamon) ISBN 0 08 043656 0.

VISCUSI, W.K. and MAGAT, W.A., 1987, *Learning about Risks. Consumer and Worker Responses to Hazard Information* (Cambridge, Massachusetts: Harvard University Press) ISBN 0 674 51915 9.

WILPERT, B. and QVALE, T. (eds), 1993, *Reliability and Safety in Hazardous Work Systems* (Hove, Sussex: Lawrence Erlbaum) ISBN 0 86377 309 5.

WOODS, D.D., JOHANNESEN, L.J., COOK, R.I., and SARTER, N.B., 1994, *Behind Human Error: Cognitive Systems, Computers, and Hindsight* (Ohio: Wright-Patterson AFB, CSERIAC), no ISBN.

YATES, J.F. (ed.), 1992, *Risk-Taking Behavior* (Chichester: Wiley) ISBN 0 471 92250 1.

Job design and work organization

CARROLL, J.M. (Ed), 1998, *Minimalism Beyond the Nurnberg Funnel* (Cambridge, Massachusetts: The MIT Press) ISBN 0 262 03249 X.

CHMIEL, N., 1998, *Jobs, Technology and People* (London: Routledge) ISBN 0 415 15817 6.

COLQUHOUN, W.P. and RUTENFRANZ, J., 1980, *Studies of Shiftwork* (London: Taylor & Francis) ISBN 0 85066 210 9.

COLQUHOUN, W.P., COSTA, G., FOLKARD, S., and KNAUTH, P. (eds), 1996, *Shiftwork: Problems and Solutions* (Frankfurt am Main: Peter Lang) ISBN 3 631 49133 6.

DAVIS, L.E. and TAYLOR, J.C., 1979, *Design of Jobs* (2nd ed.) (Santa Monica, California: Goodyear Publishing) ISBN 0 87620 219 0.

DI MARTINO, V. and CORLETT, N. (eds), 1998, *Work Organization and Ergonomics* (Geneva: International Labour Office) ISBN 92 2 109518 5.

EASON, K., 1988, *Information Technology and Organisational Change* (London: Taylor & Francis) ISBN 0 85066 388 1.

FOLKARD, S. and MONK, T.H. (eds), 1985, *Hours of Work: Temporal Factors in Work Scheduling* (Chichester: Wiley) ISBN 0 471 10524 4.

GROTH-MARNAT, G., 1997, *Handbook of Psychological Assessment* (3rd ed.) (New York: John Wiley) ISBN 0 471 05220 5.

HACKMAN, J.R. and OLDHAM, G.R., 1980, *Work Redesign* (London: Addison-Wesley) ISBN 0 201 02779 8.

HAIDER, M., KOLLER, M., and CERVINKA, R. (eds), 1986, *Night and Shiftwork: Long-term Effects and Their Prevention* (Frankfurt am Main: Peter Lang) ISBN 3 8204 8753 0.

HASLEGRAVE, C.M., WILSON, J.R., CORLETT, E.N., and MANENICA, I. (eds), 1990, *Work Design in Practice* (London: Taylor & Francis) ISBN 0 85066 770 4.

KARWOWSKI, W. and SALVENDY, G. (eds), 1994, *Organization and Management of Advanced Manufacturing* (New York: John Wiley) ISBN 0 471 55508 8.

KIDD, P. and KARWOWSKI, W. (eds), 1994, *Advances in Agile Manufacturing: Integrating Technology, Organization and People* (Amsterdam: IOS) ISBN 90 5199 176 2.

KOSLOWSKI, M., 1998, *Modeling the Stress–Strain Relationship in Work Settings* (London: Routledge) ISBN 0 415 15320 4..

KOUBEK, R.J. and KARWOWSKI, W. (eds), 1996, *Manufacturing Agility and Hybrid Automation* (Louisville, Kentucky: IEA Press) ISBN 0 9653395 0 5.

LUND, R.T., BISHOP, A.B., NEWMAN, A.E., and SALZMAN, H., 1993, *Designed to Work: Production Systems and People* (Englewood Cliffs, New Jersey: PTR Prentice Hall) ISBN 0 13 203944 3.

MONK, T.H. and FOLKARD, S., 1992, *Making Shift Work Tolerable* (London: Taylor & Francis) ISBN 0 85066 822 0.

MOORE-EDE, M., 1993, *The 24-Hour Society: The Risks, Costs and Challenges of a World That Never Stops* (London: Judy Piatkus) ISBN 0 7499 1255 3.

MORRISON, J.E. (ed.), 1991, *Training for Performance* (Chichester: Wiley) ISBN 0 471 92248 X.

MURRELL, H., 1976, *Motivation at Work* (London: Methuen) ISBN 0 416 84090 6.

OBORNE, D.J., 1985, *Computers at Work. A Behavioural Approach* (Chichester: Wiley) ISBN 0 471 90410 4.

O'NEILL, M., 1998, *Ergonomic Design for Organizational Effectiveness* (Boca Raton, Florida: Lewis Publishers) ISBN 1 56670 252 6.

PARKER, S. and WALL, T., 1998, *Job and Work Design: Organizing Work to Promote Well-Being and Effectiveness* (Thousand Oaks, California: SAGE Publications) ISBN 0 7619 0420 4.

PASMORE, W.A., 1988, *Designing Effective Organizations: The Sociotechnical Systems Perspective* (New York: John Wiley) ISBN 0 471 88785 4.

PATRICK, J., 1992, *Training: Research and Practice* (London: Academic Press) ISBN 0 12 546660 9.

PATRICK, J. and DUNCAN, K.D. (eds), 1988, *Training, Human Decision Making and Control* (Amsterdam: North-Holland) ISBN 0 444 70381 0.

SALVENDY, G. and KARWOWSKI, W. (eds), 1994, *Design of Work and Development of Personnel in Advanced Manufacturing* (New York: John Wiley) ISBN 0 471 59447 4.

SALVENDY, G., SAUTER, S.L. and HURRELL, J.J. (eds), 1987, *Social, Ergonomic and Stress Aspects of Work with Computers* (Amsterdam: Elsevier) ISBN 0 444 42847 X.

SALVENDY, G. and SMITH, M.J. (eds), 1981, *Machine Pacing and Occupational Stress* (London: Taylor & Francis) ISBN 0 85066 225 7.

SEIDEL, R.J. and CHATELIER, P.R. (eds), 1993, *Advanced Technologies Applied to Training Design* (New York: Plenum Press) ISBN 0 306 44308 2.

SEIDEL, R.J. and CHATELIER, P.R. (eds), 1997, *Virtual Reality, Training's Future? Perspectives on Virtual Reality and Related Emerging Technologies* (New York: Plenum Press) ISBN 0 306 45486 6.

SHOTTON, M.A., 1989, *Computer Addiction? A Study of Computer Dependency* (London: Taylor & Francis) ISBN 0 85066 796 8.

SWEZEY, R.W. and SALAS, E. (eds), 1992. *Teams: Their Training and Performance* (Norwood, New Jersey: Ablex) ISBN 0 89391 852 0.

WALLACE, A.P. (ed.), 1982, *Women in the Workplace* (Boston, Massachusetts: Auburn House) ISBN 0 86569 069 3.

WIENER, E.L., KANKI, B.G. & HELMREICH, R.L. (eds), 1993, *Cockpit Resource Management* (San Diego, California: Academic Press) ISBN 0 12 750025 1.

ZEIDNER, J. (ed.), 1986, *Human Productivity Enhancement. Training and Human Factors in Systems Design, Volume 1* (New York: Praeger) ISBN 0 275 92130 1.

Impact of information technology

AGRE, P.E. and ROTENBERG, M. (eds), 1997, *Technology and Privacy: The New Landscape* (London: MIT Press) ISBN 0 262 51101 0.

ANNETT, J. and STANTON, N.A. (Eds), 2000, *Task Analysis* (London: Taylor & Francis) ISBN 0 7484 0906 8.

BACKS, R.W. and BOUCSEIN, W. (Eds), 2000, *Engineering Psychophysiology: Issues and Applications* (Mahwah, New Jersey: Lawrence Erlbaum Associates) ISBN 0 8058 2453 7.

BAINBRIDGE, L. and RUIZ QUINTANILLA, S.A. (eds), 1989, *Developing Skills with Information Technology* (Chichester: Wiley) ISBN 0 471 92396 6.

BROSNAN, M.J., 1998, *Technophobia: The Psychological Impact of Information Technology* (London: Routledge) ISBN 0 415 13597 4.

DOWNEY, G.L., 1998, *The Machine in Me* (London: Routledge) ISBN 0 415 92022 1.

FARMER, E., VAN ROOIJ, J., RIEMERSMA, J., JORNA, P. and MORAAL, J., 1999, *Handbook of Simulator-Based Training* (Aldershot, Hampshire: Ashgate Publishing) ISBN 0 7546 1187 6.

FRIEDMAN, B. (ed.), 1997, *Human Values and the Design of Computer Technology* (Cambridge: Cambridge University Press) ISBN 1 57586 080 5.

KATSIKIDES, S. and ORANGE, G., 1998, *International Perspectives on Information Systems: A Social and Organisational Dimension* (Aldershot, Hampshire: Ashgate Publishing) ISBN 1 84014 508 0.

KATZ, Y.J., MILLIN, D., and OFFIR, B. (eds), 1996, *The Impact of*

Information Technology: From Practice to Curriculum (London: Chapman & Hall) ISBN 0 412 78470 X.

LUPTON, T.L. (ed.), 1986, *Human Factors: Man, Machine and New Technology* (Kempston, Bedford: IFS Publications) ISBN 0 948507 22 5.

MITAL, A., KILBOM, A. and KUMAR, S. (Eds), 2000, *Ergonomis Guidelines and Problem Solving* (Amsterdam: Elsevier Science) ISBN 0 08 043643 9.

National Academy of Engineering and National Research Council, Commission on Behavioral and Social Sciences and Education, 1991, *People and Technology in the Workplace* (Washington, DC: National Academy Press) ISBN 0 309 04583 5.

National Computing Centre, 1986, *The Impact of Office Technology* (Manchester: NCC Publications) ISBN 0 85012 504 9.

National Electronics Council, 1983, *Human Factors and Information Technology* (London: National Electronics Council) ISBN 0 9508590 0 1.

NIEBEL, B.W. and FREIVALDS, A., 1999, *Methods, Standards, and Work Design* (10th Edition) (Boston, Massachusetts: WCB/McGraw-Hill) ISBN 0 256 19507 29.

OSKAMP, S. and SPACAPAN, S. (eds), 1990, *People's Reactions to Technology in Factories, Offices, and Aerospace* (Newbury Park, California: SAGE Publications) ISBN 0 8039 3853 5.

PHILLIPS, C.A., 2000, *Human Factors Engineering* (New York: John Wiley) ISBN 0 471 24089 3.

ROCHLIN, G.I., 1995, *Trapped in the Net: The Unanticipated Consequences of Computerization* (Princeton, New Jersey: Princeton University Press) ISBN 0 691 00247 9.

ROSENBERG, R.S., 1992, *The Social Impact of Computers* (Boston, Massachusetts: Academic Press) ISBN 0 12 597130 3.

ROWE, C., 1986, *People and Chips: The Human Implications of Information Technology* (London: Paradigm Publishing) ISBN 0 948825 00 6.

SAGE, A.P. and ROUSE, W.B. (Eds), 1999, *Handbook of Systems Engineering and Management* (New York: John Wiley) ISBN 0 471 15405 9.

SCHRAAGEN, J.M., CHIPMAN, J.F. and SHALIN, V.L. (Eds), 2000, *Cognitive Task Analysis* (Mahwah, New Jersey: Lawrence Erlbaum Associates) ISBN 0 8058 3383 8.

STANTON, N.A. and M.S. YOUNG, M.S., 1999, *A Guide to Methodology in Ergonomics: Designing for Human Use* (London: Taylor & Francis) ISBN 0 7484 0703 0.

SZEWCZAK, E. and KHOSROWPOUR, M. (eds), 1996, *The Human Side of Information Technology Management* (Harrisburg, Pennsylvania: Idea Group Publishing) ISBN 1 878289 33 0.

WAERN, Y. (ed.), 1998, *Co-operative Process Management: Cognition and Information Technology* (London: Taylor & Francis) ISBN 0 7484 0713 8.

Methods, techniques and measurement

ADLER, P.S. and WINOGRAD, T.A. (eds), 1992, *Usability: Turning Technologies into Tools* (New York: Oxford University Press) ISBN 0 19 507510 2.

BIAS, R.G. and MAYHEW, D.J. (eds), 1994, *Cost-Justifying Usability* (Boston, Massachusetts: Academic Press) ISBN 0 12 095810.

BEYER, H. and HOLTZBLATT, K., 1997, *Contextual Design: Defining Customer-Centered Systems* (San Francisco: Morgan Kaufmann) ISBN 1 55860 411 1.

BORG, G., 1998, *Borg's Perceived Exertion and Pain Scales* (Champaign, Illinois: Human Kinetics) ISBN 0 88011 623 4.

BORG, G. and OTTOSON, D. (eds), 1986, *The Perception of Exertion in*

Physical Work (Basingstoke, Hampshire: Macmillan Press) ISBN 0 333 42306 2.

BRANNICK, M.T., SALAS, E. and PRINCE, C. (eds), 1997, *Team Performance Assessment and Measurement: Theory, Methods, and Applications* (Mahwah, New Jersey: Lawrence Erlbaum) ISBN 0 8058 2687 4.

BURKE, M., 1998, *Ergonomics Toolkit: Practical Applications* (Gaithersburg, Maryland: Aspern) ISBN 0 8342 1006 1..

CARROLL, J.M. (ed.), 1995, *Scenario-Based Design: Envisioning Work and Technology in System Development* (New York: John Wiley) ISBN 0 471 07659 7.

CORLETT, E.N., WILSON, J. and MANENICA, I. (eds), 1986, *The Ergonomics of Working Postures: Models, Methods and Cases* (London: Taylor & Francis) ISBN 0 85066 338 5.

DIAPER, D. (ed.), 1989, *Knowledge Elicitation: Principles, Techniques and Applications* (Chichester: Ellis Horwood) ISBN 0 7458 0451 9.

DIAPER, D. (ed.), 1989, *Task Analysis for Human–Computer Interaction* (Chichester: Ellis Horwood) ISBN 0 7458 0721 6.

ERICSSON, K.A. and SIMON, H.A., 1993, *Protocol Analysis: Verbal Reports as Data* (Revised ed.) (Cambridge, Massachusetts: MIT Press) ISBN 0 262 55023 7.

GALER, M., HARKER, S. and ZIEGLER, J. (eds), 1992, *Methods and Tools in User-Centred Design for Information Technology* (Amsterdam: North-Holland) ISBN 0 444 89301 6.

GARLAND, D.J. and ENDSLEY, M.R. (eds), 1995, *Experimental Analysis and Measurement of Situation Awareness* (Daytona Beach, Florida: Embry-Riddle Aeronautical University Press), no ISBN.

GROSS, C.M., 1996, *The Right Fit: The Power of Ergonomics as a Competitive Strategy* (Portland, Oregon: Productivity Press) ISBN 1 56327 111 7.

GUTKIN, T.B. and WISE, S.L. (eds), 1991, *The Computer and the Decision-Making Process* (Hillsdale, New Jersey: Lawrence Erlbaum) ISBN 0 8058 0610 5.

HACKOS, J.T. and REDISH, J.C., 1998, *User and Task Analysis for Interface Design* (New York: John Wiley) ISBN 0 471 17831 4.

HARRIGAN, J.E., 1987, *Human Factors Research: Methods and Applications for Architects and Interior Designers* (Amsterdam: Elsevier) ISBN 0 444 42746 5.

HASHIMOTO, K., KOGI, K. and GRANDJEAN, E. (eds), 1971, *Methodology in Human Fatigue Assessment* (London: Taylor & Francis) ISBN 0 85066 049 1.

HORROCKS, I., 1999, *Constructing the User Interface with Statecharts* (Harlow, Essex: Addison-Wesley) ISBN 0 201 34278 2.

HORTON, W.K., 1994, *Designing and Writing Online Documentation. Hypermedia for Self-Supporting Products* (2nd ed.) (New York: John Wiley) ISBN 0 471 30635 5.

HUTT, A., 1993, *Usability* (New York: Prentice Hall) ISBN 0 13 630930 5.

JONES, J.C., 1992, *Design Methods* (2nd ed.) (New York: Van Nostrand Reinhold) ISBN 0 442 01182 2.

JORDAN, P.W., 1998, *An Introduction to Usability* (London: Taylor & Francis) ISBN 0 7484 0726 6.

JORDAN, P.W., THOMAS, B., WEERDMEESTER, B.A. and MCCLELLAND, I.L. (eds), 1996, *Usability Evaluation in Industry* (London: Taylor & Francis) ISBN 0 7484 0460 0.

KARWOWSKI, W., GENAIDY, A.M. and ASFOUR, S.S. (eds), 1990, *Computer-Aided Ergonomics. A Researcher's Guide* (London: Taylor & Francis) ISBN 0 85066 753 4.

KARWOWSKI, W. and MITAL, A. (eds), 1986, *Applications of Fuzzy Set Theory in Human Factors* (Amsterdam: Elsevier) ISBN 0 444 42723 6.

KIRAKOWSKI, J. and CORBETT, M., 1990, *Effective Methodology for the Study of HCI* (Amsterdam: North-Holland) ISBN 0 444 88447 5.

KIRWAN, B. and AINSWORTH, L.K. (eds), 1992, *A Guide to Task Analysis* (London: Taylor & Francis) ISBN 0 7484 0058 3.

KOHN, J.P. (ed.), 1998, *Ergonomic Process Management: A Blueprint for Quality and Compliance* (Boca Raton, Florida: Lewis Publishers) ISBN 1 56670 226 7.

KUMAR, S. and MITAL, A. (eds), 1996, *Electromyography in Ergonomics* (London: Taylor & Francis) ISBN 0 7484 0130 X.

LANDAU, K. and ROHMERT, W. (eds), 1989, *Recent Developments in Job Analysis* (London: Taylor & Francis) ISBN 0 85066 790 9.

LIM, K.Y. and LONG, J., 1994, *The MUSE Method for Usability Engineering* (Cambridge: Cambridge University Press) ISBN 0 521 47494 9.

LINDGAARD, G., 1994, *Usability Testing and System Evaluation: A Guide for Designing Useful Computer Systems* (London: Chapman & Hall) ISBN 0 412 46100 5.

MAJCHRZAK, A., CHANG, T.C., BARFIELD, W., EBERTS, R., and SALVENDY, G., 1987, *Human Aspects of Computer-Aided Design* (London: Taylor & Francis) ISBN 0 85066 364 4.

MARSH, J.P., GORAYSKA, B. and MEY, J.L. (eds), *Human Interfaces: Questions of Method and Practice in Cognitive Technology* (Amsterdam: North-Holland) ISBN 0 444 82874 5.

MCGRATH, R.A., 1991, *Where's the Manual. Preparing and Producing the Software User's Manual* (New York: Van Nostrand Reinhold) ISBN 0 442 00425 7.

MCGRAW, K.L. and HARBISON, K., 1997, *User-Centered Requirements: The Scenario-Based Engineering Process* (Mahwah, New Jersey: Lawrence Erlbaum) ISBN 0 8058 2065 5.

MEISTER, D., 1971, *Human Factors: Theory and Practice* (London: Wiley) ISBN 0 471 59190 4.

MEISTER, D., 1976, *Behavioral Foundations of System Development* (New York: John Wiley) ISBN 0 471 59195 5.

MEISTER, D., 1985, *Behavioral Analysis and Measurement Methods* (Chichester: Wiley) ISBN 0 471 89640 3.

MEISTER, D., 1986, *Human Factors Testing and Evaluation* (Amsterdam: Elsevier) ISBN 0 444 42701 5.

MEISTER, D., 1991, *Psychology of System Design* (Amsterdam: Elsevier) ISBN 0 444 88378 9.

MORAAL, J. and KRAISS, K.F. (eds), 1981, *Manned Systems Design, Methods, Equipment, and Applications* (New York: Plenum Press) ISBN 0 306 40804 X.

MORAN, T.P. and CARROLL, J.M. (eds), 1996, *Design Rationale: Concepts, Techniques, and Use* (Mahwah, New Jersey: Lawrence Erlbaum) ISBN 0 8058 1566 X.

NIELSEN, J., 1993, *Usability Engineering* (Boston, Massachusetts: Academic Press) ISBN 0 12 518405 0.

NIELSEN, J. and MACK, R.L. (eds), 1994, *Usability Inspection Methods* (New York: John Wiley) ISBN 0 471 01877 5.

NOBLE, B.J. and ROBERTSON, R.J., 1996, *Perceived Exertion* (Champaign, Illinois: Human Kinetics) ISBN 0 88011 508 4.

NORO, K. and IMADA, A. (eds), 1991, *Participatory Ergonomics* (London: Taylor & Francis) ISBN 0 85066 382 2.

NOYES, J. and BABER, C., 1999, *User Centred Design of Systems* (Berlin: Springer) ISBN 3 540 76007 5.

O'BRIEN, T.G. and CHARLTON, S.G. (eds), 1996, *Handbook of Human Factors Testing and Evaluation* (Mahwah, New Jersey: Lawrence Erlbaum) ISBN 0 8058 1725 5.

OPPENHEIM, A.N., 1992, *Questionnaire Design, Interviewing and Attitude Measurement* (2nd ed.) (London: Pinter Publishers) ISBN 1 85567 044 5.

PALANQUE, P. and PATERNO, F. (eds), 1998, *Formal Methods in Human–Computer Interaction* (Berlin: Springer) ISBN 3 540 76158 6.

PREECE. J. (ed.), 1993, *A Guide to Usability: Human Factors in Computing* (Wokingham, Berkshire: Addison-Wesley) ISBN 0 201 62768 X.

RAVDEN, S. and JOHNSON, G., 1989, *Evaluating Usability of Human–Computer Interfaces: A Practical Method* (Chichester: Ellis Horwood) ISBN 0 7458 0614 7.

ROHMERT, W. and LANDAU, K., 1983, *A New Technique for Job Analysis* (London: Taylor & Francis) ISBN 0 85066 251 6.

ROUSE, W.B., 1991, *Design for Success* (New York: John Wiley) ISBN 0 471 52483 2.

ROUSE, W.B. and BOFF, K.R. (eds), 1987, *System Design. Behavioral Perspectives on Designers, Tools, and Organizations* (New York: North-Holland) ISBN 0 444 01230 3.

RUBIN, J., 1994, *Handbook of Usability Testing: How to Plan, Design, and Conduct Effective Tests* (New York: John Wiley) ISBN 0 471 59403 2.

SALVENDY, G. and SEYMOUR, W.D., 1973, *Prediction and Development of Industrial Work Performance* (London: Wiley) ISBN 0 471 75080 8.

SCHULER, D. and NAMIOKA, A. (eds), 1993, *Participatory Design: Principles and Practices* (Hillsdale, New Jersey: Lawrence Erlbaum) ISBN 0 8058 0952 X.

SIDES, C.H., 1992, *How to Write and Present Technical Information* (2nd ed.) (Cambridge: Cambridge University Press) ISBN 0 521 43861 6.

SIMPSON, H. and CASEY, S.M., 1988, *Developing Effective User Documentation. A Human Factors Approach* (New York: McGraw-Hill) ISBN 0 07 057336 0.

SINGLETON, W.T., FOX, J.G. and WHITFIELD, D. (eds), 1971, *Measurement of Man at Work: An Appraisal of Physiological and Psychological Criteria in Man–Machine Systems* (London: Taylor & Francis) ISBN 0 85066 041 6.

SINGLETON, W.T. and SPURGEON, P. (eds), 1975, *Measurement of Human Resources* (London: Taylor & Francis) ISBN 0 85066 068 8.

SMITH, W J., 1996, *ISO and ANSI Ergonomic Standards for Computer Products: A Guide to Implementation and Compliance* (Upper Saddle River, New Jersey: Prentice Hall PTR) ISBN 0 13 151119 X.

TRENNER, L. and BAWA, J. (eds), 1998, *The Politics of Usability: A Practical Guide to Designing Usable Systems in Industry* (Berlin: Springer-Verlag) ISBN 3 540 76181 0.

VAN DER SCHAAF, T.W., LUCAS, D.A. and HALE, A.R. (eds), 1991, *Near Miss Reporting as a Safety Tool* (Oxford: Butterworth-Heinemann) ISBN 0 7506 1178 2.

VAN SOMEREN, M.W., BARNARD, Y.F. and SANDBERG, J.A.C., 1994, *The Think Aloud Method: A Practical Guide to Modelling Cognitive Processes* (London: Academic Press) ISBN 0 12 714270 3.

VENDA, V.F. and VENDA, Y.V., 1995, *Dynamics in Ergonomics, Psychology, and Decisions: Introduction to Ergodynamics* (Norwood, New Jersey: Ablex) ISBN 1 56750 129 X.

WEIMER, J., 1995, *Research Techniques in Human Engineering* (Englewood Cliffs, New Jersey: Prentice Hall PTR) ISBN 0 13 097072 7.

WHITAKER, K., 1995, *A Guide to Publishing User Manuals* (New York: John Wiley) ISBN 0 471 11846 X.

WIKLUND, M.E. (ed.), 1994, *Usability in Practice: How Companies Develop User-Friendly Products* (Boston, Massachusetts: AP Professional) ISBN 0 12 751250 0.

WILSON, J.R. and CORLETT, E.N. (eds), 1995, *Evaluation of Human Work: A Practical Ergonomics Methodology* (2nd ed.) (London: Taylor & Francis) ISBN 0 7484 0084 2.

WILSON, J.R., CORLETT, E.N. and MANENICA, I. (eds), 1987, *New Methods in Applied Ergonomics* (London: Taylor & Francis) ISBN 0 85066 384 9.

ZIMMERMAN, C.M. and CAMPBELL, J.J., 1988, *Fundamentals of Procedure Writing* (2nd ed.) (London: Kogan Page) ISBN 1 85091 772 8.

4. PROCEEDINGS

Advances in Industrial Ergonomics and Safety (London: Taylor & Francis).

Advances in Occupational Ergonomics and Safety (Amsterdam: IOS).

Analysis, Design and Evaluation of Man-Machine Systems (Oxford: Pergamon).

Australasian Computer Human Interaction Conference: OZCHI Annual Conference (Piscataway, New Jersey: IEEE Computer Society).

Aviation Psychology International Symposium (Columbus, Ohio: Ohio State University).

Cognitive Science Approaches to Process Control Conference (various publishers).

Engineering Psychology and Cognitive Ergonomics (Aldershot, Hampshire: Ashgate Publishing).

Ergonomics for Global Quality and Productivity (Hong Kong: Hong Kong University of Science and Technology).

Enhanced Safety of Vehicles International Technical Conference (Washington DC: National Highway Traffic Safety Administration).

Ergonomics Society Annual Conference, UK (London: Taylor & Francis).

Ergonomics Society of Australia Annual Conference (Downer, ACT: Ergonomics Society of Australia).

Ergonomics Society of New Zealand Conference (Palmerston North: New Zealand Ergonomics Society).

Ergonomics Society of South Africa Annual Conference (Ergonomics Society of South Africa).

European Conference on Cognitive Ergonomics (ECCE) (various publishers).

Human Aspects of Advanced Manufacturing and Hybrid Automation (Amsterdam: Elsevier, and Santa Monica, California: IEA Press).

Human–Computer Interaction International Conference (Amsterdam: Elsevier).

Human Decision Making and Manual Control Conference (various publishers).

Human Factors Association of Canada Annual Conference (Mississauga, Ontario: Human Factors Association of Canada).

Human Factors in Computing Systems Conference (Amsterdam: North-Holland, and Reading, Massachusetts: Addison-Wesley).

Human Factors and Ergonomics Society Annual Meeting (Santa Monica, California: Human Factors and Ergonomics Society).

Human Factors in Manufacturing International Conference (various publishers).

Human Factors in Organizational Design and Management (Amsterdam: Elsevier).

Human Factors in Telecommunications International Symposium (various publishers).

IEEE Conference on Human Factors and Power Plants (New York: Institute of Electrical and Electronic Engineers).

IFIP International Conference on Human–Computer Interaction: INTERACT (Amsterdam: Elsevier, and London: Chapman & Hall).

Interface, HFES Consumer Product Technical Group Conference (Santa Monica, California: Human Factors and Ergonomics Society).

International Ergonomics Association Triennial Congress (various publishers).

Pan-Pacific Conference (various publishers).

People and Computers, British Computer Society Human Interaction Specialist Group (Cambridge: Cambridge University Press).

PSAM: International Conference on Probabilistic Safety Assessment and Management (various publishers).

RESNA Annual Conference Association for the Advancement of Rehabilitation Technology (Arlington, Virginia: RESNA Press).

SAFE Association Annual Symposium (Nashville, Tennessee: SAFE Association).

Silicon Valley Ergonomics Conference - ErgoCon (San Jose, California: Silicon Valley Ergonomics Institute, San Jose State University).

Society of Information Display Symposium (Santa Ana, California: Society for Information Display).

Trends in Ergonomics/Human Factors (Amsterdam: Elsevier).

Vision in Vehicles (Amsterdam: Elsevier).

Visual Search (London: Taylor & Francis).

Wearable Computers International Symposium (Piscataway, New Jersey: IEEE Computer Society).

Work with Display Units (Amsterdam: Elsevier).

Bibliography

W. Karwowski

Center for Industrial Ergonomics, Department of Industrial Engineering, University of Louisville, USA

AARAS, A., 1987, Postural load and the development of musculoskeletal illness, *Scandinavian Journal of Rehabilitation Medicine*, **18**, 6–35.

AARAS, A., 1994, Relationship between trapezius load and the incidence of musculoskeletal illness in the neck and shoulder, *International Journal of Industrial Ergonomics*, **14**, 341–8.

AARAS, A., 1994, The impact of ergonomic intervention on individual health and corporate prosperity in a telecommunications environment, *Ergonomics*, **37**, 1679–96.

AARAS, A., VEIEROD, M.B., LARSEN, S., ORTENGREN, R. and RO, O., 1996, Reproducibility and stability of normalized EMG measurements on musculus trapezius, *Ergonomics*, **39**, 171–85.

ADAMS, J.A., 1989, *Human Factors Engineering* (New York: Macmillan).

ALEXANDER, D.C., 1986, *The Practice and Management of Industrial Ergonomics* (Englewood Cliffs: Prentice-Hall).

ALLREAD, W.G., MARRAS, W.S. and PARNIANPOUR, M., 1996, Trunk Kinematics of one-handed lifting and effects of asymmetry and load weight, *Ergonomics*, **39**, 322–34.

AMAR, J., 1920, *The Human Motor* (London: George Routledge & Sons).

American National Standards for Human Factors Engineering of Visual Display Terminal Workstations (Santa Monica: Human Factors Society, 1988).

ANDERSSON B.J.G. and MARRAS W.S., 1996, The Lumbar Spine. In J.N. Weinstein and S.W. Wiesel (eds), *Occupational Biomechanics*, 2nd edn (Philadelphia: W.B. Saunders).

ANDERSSON, B.J.G., ORTENGREN, R., NACHEMSON, A. and ELFSTROM, G., 1974, Lumbar disc pressure and myoelectric back muscle activity during sitting, *Scandinavian Journal of Rehabilitation Medicine*, **6**, 104–14.

ANKRUM, D.R. and NEMETH, K.J.,. 1995, Posture, comfort and monitor placement, *Ergonomics in Design*, **April**, 7–9.

ASHFORD, F.C., 1969, *The Aesthetics of Engineering Design* (London: Business Books).

AYOUB, M.M. and DEMPSEY, P.G., 1999, The psychophysical approach to manual materials handling task design, *Ergonomics*, **42**, 17–31.

AYOUB, M.M. and MITAL, A., 1989, *Manual Materials Handling* (London: Taylor & Francis).

BADDELEY, A.D., 1986, *Working Memory* (Oxford: Clarendon).

BADLER, N.I., PHILLIPS, C.B. and WEBBER, B.L., 1993, *Simulating Humans: Computer Graphics Animation and Control* (New York: Oxford University Press).

BAILEY, R.W., 1982, *Human Performance Engineering: A Guide for System Designers* (Englewood Cliffs: Prentice-Hall).

BAILEY, R.W., 1983, *Human Error in Computer Systems* (Englewood Cliffs: Prentice-Hall).

BAILEY, R.W., 1989, *Human Performance Engineering: Using Human Factors/Ergonomics to Achieve Computer System Usability* (Englewood Cliffs: Prentice Hall).

BAINBRIDGE, L., 1983, Ironies of automation, *Automatica*, **19**, 775–9.

BAIRD, J.C. and NOMA, E., 1978, *Fundamentals of Scaling and Psycophysics* (New York: Wiley).

BAKER, C.A. (ed.), 1965, *Visual Capabilities in the Space Environment* (Oxford: Pergamon).

BANDURA, A., 1977, *Social Learning Theory* (Englewood Cliffs: Prentice Hall).

BARON, R.A., 1994, The physical environment of work settings: effects on task performance, interpersonal relations, and job satisfaction. In B.M. Staw and L.L. Cummings (eds), *Research in Organizational Behavior*, vol. 16 (Greenwich: JAI Press), 1–46.

BATESON, G., 1987, *Steps to An Ecology of Mind* (Northvale: Jason Aronson).

BEACH, L.R. and MITCHELL, T.R., 1987, Image theory: principles, goals and plans in decision making, *Acta Psychologica*, **66**, 201–20.

BECK, J., HOPE, B. and ROSENFELD, A. (eds), 1983, *Human and Machine Vision* (New York: Academic Press).

BENDIX, T., 1984, Seated trunk postures at various seat inclinations, seat heights, and table heights, *Human Factors*, **26**, 695–703.

BENDIX, T. and JESSEN, F., 1986, Wrist support during typing — a controlled, electromyographic study, *Applied Ergonomics*, **17**, 162–8.

BENDIX, T., POULSEN, V., KLAUSEN, K. and JENSEN, C.V., 1996, What does a backrest actually do to the lumbar spine? *Ergonomics*, **39**, 533–42.

BENDIX, T., WINKEL, J. and JESSEN, F., 1985, Comparison of office chairs with fixed forwards or backward inclining, or tiltable seats, *European Journal of Applied Physiology*, **54**, 378–85.

BENNETT, E.M., DEGAN, J. and SPIEGEL, J. (eds), 1963, *Human Factors in Technology* (New York: McGraw-Hill).

BERGQVIST, U., WOLGAST, E., MILSSON, B. and VOSS, M., 1995, Musculoskeletal disorders among visual display terminal workers: individual, ergonomic, and work organizational factors, *Ergonomics*, **38**, 763–76.

BERGQVIST, U., WOLGAST, E., MILSSON, B. and VOSS, M., 1995, The influence of VDT work on musculoskeletal disorders, *Ergonomics*, **38**, 754–62.

BERNARD, B.P. and PUTZ-ANDERSON, V., 1997, *Musculoskeletal Disorders and Workplace Factors: A Critical Review of Epidemiological Evidence for Work-related Musculoskeletal Disorders of the Neck, Upper Extremity, and Low Back*. No. 97–141 (Atlanta: DHHS (NIOSH)).

BERNECKER, C.A., BRAINARD, G.C., FERNSLER, F.J., ROLLAG, M.D., LONG, R.R., TIERNEY, S. and GADDY, J.R., 1994, Biological effects of architectural lighting and their associated energy utilization, *Journal of the Illuminating Engineering Society*, **23**, 31–9.

BERNECKER, C.A., SANDERS, P.A. and MISTRICK, R.G., 1995, Reflected luminance patterns in visual display units, *Journal of the Illuminating Engineering Society*, **24**, 8–18.

BIRKBECK, M.Q. and BEER, T.C., 1975, Occupation in relation to carpal tunnel syndrome. *Rheumatology and Rehabilitation*, **14**, 218–20.

BOFF, K.R. and LINCOLN, J.E. (eds), 1988, *Engineering Data Compendium — Human Perception and Performance* (Wright Patterson Air Force Base, OH: Harry G. Armstrong Aerospace Medical Research Laboratory).

BOFF, K.R., KAUFMAN, L. and THOMAS, J.P. (eds), 1986, *Handbook of Perception and Human Performance. 1. Sensory Processes and Perception* (New York: Wiley).

BOFF, K.R., KAUFMAN, L. and THOMAS, J.P. (eds), 1986, *Handbook of Perception and Human Performance. 11. Cognitive Processes and Perception* (New York: Wiley).

BOLES, D.B., 1997, Multiple resource contributions to training. In *Proceedings of the Human Factors and Ergonomic Society 41st Annual Meeting* (Santa Monica: Human Factors and Ergonomic Society), 1176–9.

BOLES, D.B. and LAW, M.B., 1998, A simultaneous task comparison of differentiated and undifferentiated hemispheric resource theories, *Journal of Experimental Psychology: Human Perception and Performance*, **24**, 204–15.

BOLES, D.B. and WICKENS, C.D., 1987, Display formatting in information integration and nonintegration tasks, *Human Factors*, **29**, 395–406.

BOLMAN, L.G. and TERRENCE, E.D., 1994, *Reframing Organizations* (San Francisco: Jossey-Bass).

BONGERS, P.M., DE WINTER, C.R., KOMPIER, M.J. and HILDERBRANDT, V.H., 1993, Psychosocial factors and work and musculoskeletal disease, *Scandinavian Journal of Work, Environment and Health*, **19**, 297–312.

BOOHER, H.R. (ed.), 1990, *Manprint: An Approach to Systems Integration* (New York: van Nostrand Reinhold).

BOURDIEU, P., 1977, *Outline of a Theory Of Practice*. Studies in Social Anthropology 16 (Cambridge: Cambridge University Press).

BOYCE, P.R., 1981, *Human Factors in Lighting* (London: Applied Science).

BRADLEY, G., 1989, *Computers and the Psychosocial Work Environment* (London: Taylor & Francis).

BRADLEY, G. and HENDRICK, H.W. (eds), 1994, *Human Factors in Organizational Design and Managsement — IV* (Amsterdam: North-Holland).

BRIDGER, R.S., 1988, Postural adaptations to a sloping chair and work surface, *Human Factors*, **30**, 237–47.

BRIDGER, R.S., 1995, *Introduction to Ergonomics* (New York: McGraw-Hill).

BRIDGER, R.S., SPARTO, P. and MARRAS, W.S., 1998, Spade design, lumbar motions risk of low-back injury and digging posture, *Occupational Ergonomics*, **1**, 157–72.

BROADBENT, D.E., 1958, *Perception and Communication* (London: Pergamon).

BROADBENT, D.E., 1971, *Decision and Stress* (New York: Academic Press).

BRÖDNER, P. and KARWOWSKI, W. (eds), 1992, *Ergonomics of Hybrid Automated Systems III* (Amsterdam: Elsevier).

BROOKS, L.R., 1968, Spatial and verbal components of the act of recall, *Canadian Journal of Psychology*, **22**, 349–68.

BROWN, O., Jr and HENDRICK, H.W. (eds), 1986, *Human Factors in Organizational Design and Management — II* (Amsterdam: North-Holland).

BROWN, O., Jr and HENDRICK, H.W. (eds), 1996, *Human Factors in Organizational Design and Management — V* (Amsterdam: North-Holland).

BRUNSWIK, E., 1956, *Perception and the Representative Design of Experiments* (Berkley: University of California Press).

BULLINGER, H.J., 1994, *Ergonomie* (Stuttgart: Teubner).

BURGESS, J.H., 1989, *Human Factors in Industrial Design: The Designer's Companion* (Blue Ridge Summit: TAB Professional and Reference Books).

BURKE, M., 1993, *Ergonomics Analysis and Problem Solving* (Boca Raton: Lewis).

CAKIR, A., HART, D.J. and STEWART, D.F.M., 1978, *Research into the Effects of Video display Workplaces on the Physical and Psychological Function of Person* (Bonn: Federal Ministry for Work and Social Order).

CARAYON, P., 1993, Job design and job stress in office workers, *Ergonomics*, **36**, 463–77.

CARAYON, P., 1994, Stressful jobs and non-stressful jobs: a cluster analysis of office jobs, *Ergonomics*, **37**, 311–23.

CARD, S.K., MORAN, T.P. and NEWELL, A., 1983, *The Psychology of Human-Computer Interaction* (Hillsdale: Lawrence Erlbaum).

CERNEA, M., 1995, Social organization and development anthropology, *Human Organization*, **54**, 340–52.

CHAFFIN, D.B., ANDERSSON, G. and MARTIN, B. (eds), 1999, *Occupational Biomechanics* (New York: Wiley).

CHAFFIN, D.B. and PARK, K.S., 1973, A longitudinal study of low-back pain as associated with occupational weight lifting factors. *American Industrial Hygiene Association Journal*, **34**, 513–25.

CHAPANIS, A., 1956, *The Design and Conduct of Human Engineering Studies* (San Diego: San Diego State College Foundaiton).

CHAPANIS, A., 1959, *Research Techniques in Human Engineering* (Baltimore: Johns Hopkins University Press).

CHAPANIS, A., 1965, *Man-Machine Engineering* (Belmont: Wadsworth).

CHAPANIS, A. (ed.), 1975, *Ethnic Variables in Human Factors Engineering* (Baltimore: Johns Hopkins University Press).

CHAPANIS, A., 1996, *Human Factors in Systems Engineering* (New York: Wiley).

CHAPANIS, A., GARNER, W.R. and MORGAN, C.T., 1949, *Applied Experimental Psychology: Human factors in Engineering Design* (New York: Wiley).

CHRISLIP, D.D. and LARSON, C.E., 1994, *Collaborative Leadership* (San Francisco: Jossey-Bass).

CLARK, T.S. and CORLETT, E.N., 1995, *The Ergonomics of Workspace and Machines: A Design Manual* (Washington, DC: Taylor & Francis).

CLEGG, C., RAVDEN, S., CORBERRT, M. and JOHNSON, S., 1989, Allocating functions in computer integrated manufacturing: a review and new method, *Behavior and Information Technology*, **8**, 175–90.

CLEVENSON, S.A., 1978, *Effect of Vibration Duration on Human Discomfort* (Washington, DC: NASA).

COHEN, A.L., 1997, *Elements of Ergonomics Programs: A Primer Based on Workplace Evaluations of Musculoskeletal Disorders*. No. 97–117 (Atlanta: DHHS (NIOSH)).

COLQUHOUN, W.P. (ed.), 1971, *Biological Rhythms and Human Performance* (London: Academic Press).

CONGLETON, J.J., AYOUB, M.M. and SMITH, J.L., 1985, The design and evaluation of the neutral posture chair for surgeons, *Human Factors*, **27**, 589–600.

CORLETT, E.N. and BISHOP, R.P., 1976, A technique for assessing postural discomfort, *Ergonomics*, **19**, 175–82.

CORLETT, E.N. and RICHARDSON, J. (eds), 1981, *Stress, Work Design and Productivity* (Chichester: Wiley).

COURTNEY, T.K. and CLANCY, E.A., 1998, A descriptive study of US OSHA penalties and inspection frequency for musculoskeletal disorders in the workplace, *American Industrial Hygiene Association Journal,* *59, 563–71*.

CRAIK, K.J.W., 1947, The theory of the human operator in control systems, Man as an element in a control system, *British Journal of Psychology*, **38**, 142–8.

CUSHMAN, W.H. and ROSENBERG, D.J., 1991, *Human Factors in Product Design* (Amsterdam: Elsevier).

DAINOFF, M.J., 1982, Occupational stress factors in visual display terminal (VDT) operation: a review of empirical research, *Behavior and Information Technology*, **1**, 141–76.

DAMON, A., STOUDT, H.W. and McFARLAND, R.A., 1966, *The Human Body in Equipment Design* (Cambridge, MA: Harvard University Press).

DAS, B. and KARWOWSKI, W. (eds), 1997, *Advances in Occupational Ergonomics and Safety 1* (Amsterdam: IOS Press).

DAVIS, K.G., KIRKING, B., GAUDES, L., JAY, W. and MARRAS, W.S., 1997, The influence of load moment and visual perception on psychophysically determined acceptable lifting weight, *Human Factors*, **39**, 312–22.

DAVIS, K.G., MARRAS, W.S. and WATERS, T.R., 1997, Evaluation of spinal loading during lowering and lifting, *Clinical Biomechanics*, **13**, 141–52.

Bibliography

DAVIS, L.E., 1982, Organization design. In G. Salvendy (ed.), *Handbook of Industrial Engineering* (New York: Wiley).

DEGREENE, K., 1973, *Sociotechnical Systems* (Englewood Cliffs: Prentice-Hall).

DELLEMAN, N.J., 1999, *Working Postures: Prediction and Evaluation* (Amstelveen: Nico J. Delleman).

DEMING, W.E., 1986, *Out of the Crisis* (Cambridge, MA: MIT Center for Advanced Engineering Study).

DEMPSEY, P.G. and AYOUB, M.M., 1996, The influence of gender, grasp type, pinch width and wrist position on sustained pinch strength, *International Journal of Industrial Ergonomics*, **17**, 259–73.

DEMPSEY, P., AYOUB, M.M., BERNARD, T.M., ENDSLEY, M.R., KARWOWSKI, W., LIN, C.J. and SMITH, J.L., 1996, Ergonomic investigation of letter-carrier satchels: Part 1 field study, *Applied Ergonomics*, **27**, 303–13.

DESANCTIS, G., 1984, Computer graphics as decision aids: Directions for research, *Decision Sciences*, **15**, 463–87.

DHILLON, B.S., 1986, *Human Reliability: With Human Factors* (New York: Pergamon).

DICKINSON, K., CAMPION, A.F, FOSTER, A.F, NEWMAN, S.J., O'ROURKE, A.M.T. and THOMAS, P.G., 1992, Questionnaire development: an examination of the Nordic Musculoskeletal Questionnaire, *Applied Ergonomics*, **23**, 197–201.

DINER, D.B. and FENDER, D.H., 1993, *Human Engineering in Stereoscopic Viewing Devices* (New York: Plenum).

DONALDSON, S. and DONALDSON, M., 1996, Carpal Tunnel Syndrome: a multi-etiological problem, *Physical Therapy Products*, **March**, 22–4.

DREYFUSS, H., 1960, *The Measure of Man: Human Factors in Design* (New York: Whitney Library of Design).

DREYFUSS, H., 1967, *Designing for People* (New York: Paragraphic).

DUL, J. and WEERDMEESTER, B.A., 1993, *Ergonomics for Beginners: A Quick Reference Guide* (London: Taylor & Francis).

DURSO, F.T., NICKERSON, R.S., SCHVANEVELDT, R.W., DUMAIS, S.T., LINDSAY, D.S. and CHI, M.T.H. (eds), 1999, *Handbook of Applied Cognition* (Chichester: Wiley).

DYER, H. and MORRIS, A., 1990, *Human Aspects of Library Automation* (Brookfield: Gower).

EASTMAN KODAK CO., 1983, *Ergonomic Design for People at Work*, vol. 1 (New York: van Nostrand Reinhold).

EASTMAN KODAK CO., 1986, *Ergonomic Design for People at Work*, vol. 2 (New York: van Nostrand Reinhold).

EDWARDS, R., 1988, Hypotheses of peripheral and central mechanisms underlying occupational muscle pain and injury, *European Journal of Physiology*, **57**, 275–81.

EHRICH, R.W. and WILLIGES, R.C. (eds), 1986, *Human-Computer Dialogue Design* (Amsterdam: Elsevier).

EKBERG, K., KARLSSON, M. and AXELSON, O., 1995, Cross-sectional study of risk factors for symptoms in the neck and shoulder area, *Ergonomics*, **38**, 971–80.

EMERY, F.E. and TRIST, E.L., 1960, Sociotechnical Systems. In C.W. Chruman and M. Verhulst (eds), *Management Science* (Oxford: Pergamon).

EMERY, F.E. and TRIST, E.L., 1965, The causal nature of organizational environments, *Human Relations*, **18**, 21–32.

ENDSLEY, M.R., 1995, Measurement of situation awareness in dynamic systems, *Human Factors*, **37**, 65–84.

ENDSLEY, M.R., 1995, Toward a theory of situation awareness in dynamic systems, *Human FActors*, **37**, 32–64.

ENDSLEY, M.R. and KIRIS, E.O., 1995, The out-of-the-loop performance problema nd level of control in automation, *Human Factors*, **37**, 381–94.

ERDIL, M. and DICKERSON, O.B. (eds), 1997, *Cumulative Trauma Disorders — Prevention, Evaluation, and Treatment* (New York: van Nostrand Reinhold).

ERICSON, M.O., 1989, Spinal shrinkage with three different types of chair whilst performing video display unit work, *International Journal of Industrial Ergonomics*, **3**, 177–83.

EVANS, G.W., KARWOWSKI, W. and WILHELM, M.R. (eds), 1989, *Application of Fuzzy Set Methodologies in Industrial Engineering* (Amsterdam: Elsevier Science).

FATHALLAH, F.A., MARRAS, W.S. and PARNIANPOUR, M., 1997, The effect of complex dynamic lifting and lowering characteristics on trunk muscles recruitment, *Journal of Occupational Rehabilitation*, **7**, 121–38.

FATHALLAH, F.A., MARRAS, W.S. and PARNIANPOUR, M., 1998, The role of complex, simultaneous trunk motions in the risk of occupationally-related low back disorders, *Spine*, **23**, 1035–42.

FATHALLAH, F.A., MARRAS, W.S., PARNIANPOUR, M. and GRANATA, K.P., 1997, A method for measuring external spinal loads during whole-body free-dynamic lifting, *Journal of Biomechanics*, **30**, 975–8.

FAUCETT, J. and REMPEL, D., 1994, VDT-related musculoskeletal symptoms: interactions between work posture and psychosocial work factors, *American Journal of Industrial Medicine*, **26**, 597–612.

FAUCETT, J. and REMPEL, D., 1996, Musculoskeletal symptoms related to video terminal use, *AAOHN Journal*, **44**, 33–9.

FEIGENBAUM, A.V., 1961, *Total Quality Control* (New York: McGraw-Hill).

FERGUSON, S.A. and MARRAS, W.S., 1997, A review of low back disorder assessment methods and risk factors, *Clinical Biomechanics*, **12**, 211–26.

FERGUSON, S.A., MARRAS, W.S. and WATERS, T.R., 1992, Quantification of back motion during asymmetric lifting, *Ergonomics*, **35**, 845–59.

FERNSTROM, E.A.C. and ERICSON, M.O., 1996, Upper-arm elevation during office work, *Ergonomics*, **39**, 1221–30.

FISK, A.D. and ROGERS, W.A. (eds), 1997, *Handbook of Human Factors and the Older Adult* (San Diego: Academic Press).

FITTS, P.M. (ed.), 1959, *Notes and Selected Readings on Human Engineering Concepts and Theory* (Ann Arbor: University of Michigan).

FITTS, P.M. and POSNER, M.I., 1969, *Human Performance* (Belmont: Brooks Cole).

FLACH, J.M., 1990, The ecology of human-machine systems 1: Introduction, *Ecological Psychology*, **2**, 191–205.

FLACH, J.M., HANCOCK, P.A., CAIRD, J.K. and VICENTE, K.J. (eds), 1992, *An Ecological Approach to Human Machine Systems II: Local Applications* (Hillsdale: Erlbaum).

FLURSCHEIM, C.H. (ed.), 1983, *Industrial Design in Engineering: A Marriage of Techniques* (Berlin: Springer).

FOGEL, L.J., 1963, *Biotechnology: Concepts and Applications* (New Jersey: Prentice-Hall).

FOLKARD, S. and MONK. T.H. (eds), 1985, *Hours of Work* (New York: Wiley).

FRANUS, E.A. (ed.), 1991, *Connective Networks in Ergonomics* (Amsterdam: Elsevier).

FRASER, T.M., 1989, *The Worker at Work* (London: Taylor & Francis).

FRIELING, E. and SONNTAG, K., 1999, *Lehrbuch Arbeitspsychologie* (Bern: Aufl).

FRYMOYER, J.W. and GORDON, S.L. (eds), 1988, *New Perspective on Low Back Pain* (Illinois: American Academy of Orthopedic Surgeons).

GARGNê, R.M., 1962, *Psychological Principles in System Development* (New York: Holt, Rinehart & Winston).

GAGNê, R., BRIGGS, L. and WAGNER, R., 1988, *Principles of Instructional Design* (New York: Holt Rinehart and Winston).

GALLAGHER, S., HAMRICK, C.A., LOVE, A.C. and MARRAS, W.S., 1994, Dynamic biomechanical modeling of symmetric and asymmetric lifting tasks in restricted postures, *Ergonomics*, **37**, 1289–1310.

GENAIDY, A. and KARWOWSKI, W., 1993, The effects of body movements on perceived joint discomfort ratings in sitting and standing postures, *Ergonomics*, **36**, 785–92.

GENAIDY, A., KARWOWSKI, W. and CHRISTENSEN, D., 1999, Principles of work system performance optimization: a business ergonomics approach, *Human Factors and Ergonomics in Manufacturing*, **9**, 105–28.

GENAIDY, A., KARWOWSKI, W., CHRISTENSEN, D., VOGIATZIS, C. and PRINS, A., 1998, What is 'heavy'?, *Ergonomics*, **41**, 320–432.

GENAIDY, A., KARWOWSKI, W., GUO, L., HIDALGO, J. and GARBUTT, G., 1992, Physical training: a tool for increasing work tolerance limits of employees engaged in manual handling tasks, *Ergonomics*, **36**, 1081–1102.

GERARD, M., JONES, S.K., SMITH, L.A., THOMAS, R.E. and WANG, T., 1994, An ergonomic evaluation of the Kinesis ergonomic computer keyboard, *Ergonomics*, **37**, 1661–8.

GERTMAN, D.I. and BLACKMAN, H.S., 1994, *Human Reliability and Safety Analysis Data Handbook* (New York: Wiley).

GIBSON, J.J., 1966, *The Senses Considered as Perceptual Systems* (Westport: Greenwood).

GIBSON, J.J., 1986, *The Ecological Approach to Visual Perception* (London: Lawrence Erlbaum).

GIESE, F., 1925, *Handbuch der Arbeitswissenschaft* (Halle).

GILBRETH, F.B., 1911, *Motion Study* (New York: van Nostrand).

GILL, K.S. (ed.), 1996, *Human-Machine Symbiosis: The Foundations of Human-Centred Systems Design* (Berlin: Springer).

GOETTL, B.P. and WICKENS, C.D., 1989, Multiple resources versus information integration. *Proceedings of the Human Factors Society*, **33**, 1454–8.

GOLDSTEIN, I.L., 1993, *Training in Organizations* (Belmont: Wadsworth).

GOODSTEIN, L.P., ANDERSEN, H.B. and OLSEN, S.E. (eds), 1988, *Tasks, Errors and Mental Models* (London: Taylor & Francis).

GORDON, S., 1994, *Systematic Training Program Design: Maximizing and Minimizing Liability* (Englewood Cliffs: Prentice Hall).

GOULD, J.D., 1966, Tribute to Karl U. Smith (1907–1994), *American Psychologist*, **51**, 977.

GRAF, M., GUGGENBUHL, U. and KRUEGER, H., 1993, Investigations on the effects of seat shape and slope on posture comfort and back muscle activity, *International Journal of Industrial Ergonomics*, **12**, 91–103.

GRAF, M., GUGGENBUHL, U. and KRUEGER, H., 1995, An assessment of seated activity and postures at five workplaces, *International Journal of Industrial Ergonomics*, **15**, 81–90.

GRAHAM, J.H. (ed.), 1991, *Safety, Reliability and Human Factors in Robotic Systems* (New York: van Nostrand).

GRANATA, K.P. and MARRAS, W.S., 1993, An EMG-assisted model of loads on the lumbar spine during asymmetric trunk extensions, *Journal of Biomechanics*, **26**, 1429–38.

GRANATA, K.P. and MARRAS, W.S., 1995, An EMG-assisted model of trunk loading during free-dynamic lifting, *Journal of Biomechanics*, **28**, 1309–17.

GRANATA, K.P. and MARRAS, W.S., 1999, Relation between spinal load factors and the high risk probability of occupational low back disorder, *Ergonomics*, **42**, 1187–99.

GRANDJEAN, E., 1973, *Ergonomics of the Home* (London: Taylor & Francis).

GRANDJEAN, E., (ed.), 1984, *Ergonomics and Health in Modern Offices* (London: Taylor & Francis).

GRANDJEAN, E., 1987, *Ergonomics in Computerized Offices* (London: Taylor & Francis).

GRANDJEAN, E., 1988, *Fitting the Tasks to the Man: A Textbook of Occupational Ergonomics* (London: Taylor & Francis).

GRANDJEAN, E. and VIGLIANI, E. (eds), 1980, *Ergonomic Aspects of Visual Display Terminals* (London: Taylor & Francis).

GREEK, D.C., 1959, *Check List of Human Engineering Design Principles* (North American Aviation, Inc., Missile Division).

GREEN, D.M. and SWETS, J., 1974, *Signal Detection Theory and Psychophysics* (New York: Plenum).

GRIECO, A., MOLTENI, G., PICCOLI, B. and OCCHIPINTI, E. (eds), 1995, *Work With Display Units '94* (Amsterdam: Elsevier).

GROBELNY, J., KARWOWSKI, W. and ZURADA, J., 1995, Application of fuzzy-based linguistic patterns for the assessment of computer screen design quality, *International Journal of Human-Computer Interaction*, **7**, 193–212.

GRÖNQVIST, R., 1999, Slips and falls. In S. Kumar (ed.), *Biomechanics in Ergonomics* (London: Taylor & Francis), 351–75.

GROSS, C.M., 1996, *The Right Fit: The Power of Ergonomics as a Competitive Strategy* (Portland: Productivity).

GUMPERZ, J.J., 1982, Interethnic communication. In *Discourse Strategies* (Cambridge: Cambridge University Press).

HACKER, W., 1998, *Allgemeine Arbeitspsychologie, Psychische Regulation von ArbeitstŇtigkeiten* (Bern).

HACKMAN, J.R., 1990, *Groups That Work* (San Francisco: Jossey-Bass).

HACKMAN, J.R. and OLDHAM, G.R., 1980, *Work Redesign* (New York: Addison-Wesley).

HACKSTEIN, R., 1977, *Arbeitswissenschaft im Umri&Greek;b*, Bd. 1: *Gegenstand und RechtsverhŇltnisse*, Bd. 2 (Essen: Grundlagen & Anwendung).

HAGBERG, M. and SUNDELIN, G., 1986, Discomfort and load on the upper trapezius muscle when operating a word procesor, *Ergonomics*, **29**, 1637–45.

HAMILTON, N., 1996, Source document position as it affects head position and neck muscle tension, *Ergonomics*, **39**, 593–610.

HANCOCK, P.A. (ed.), 1987, *Human Factors Psychology* (Amsterdam: North-Holland).

HANCOCK, P.A., 1997, *The Future of Human-Machine Systems* (Eden Prairie: BANTA Information Services Group).

HANCOCK, P.A., CHIGNELL, M.H. and LOWENTHAL, A., 1985, An adaptive human-machine system. *Proceedings of the IEEE Conference on Systems, Man and Cybernetics*, **15**, 627–9.

HANCOCK, P.A. and PARASURAMAN, R., 1992, Human factors and safety in the design of intelligent vehicle-highway systems, *Journal of Safety Research*, **23**, 181–98.

HANNUM, W. and HANSEN, C., 1992, *Instructional Systems Development in Large Organizations* (Englewood Cliffs: Educational Technology).

HARRIS, D.H. and CHANEY, F.B., 1969, *Human Factors in Quality Assurance* (New York: Wiley).

HART, S.G. and STAVELAND, L.E., 1988, Development of NASA-TLX (Task Load Index): results of empirical and theoretical research. In

P.A. Hancock and N. Meshkati (eds), *Human Mental Workload* (North-Holland: Elsevier Science), 139–83.

HASHEMI, L., WEBSTER, B.S., CLANCY, E.A., 1998, Trends in disability duration and cost for workers' compensation LBP claims (1988–1996), *Journal of Occupational and Environmental Medicine*, **40**, 1110–19.

HASHEMI, L., WEBSTER, B.S., CLANCY, E.A. and COURTNEY, T.K., 1998, Length of disability and cost of work-related musculoskeletal disorders of the upper extremity, *Journal of Occupational and Environmental Medicine*, **40**, 261–9.

HASLEGRAVE, C.M., 1994, What do we mean by a 'working posture'?, *Ergonomics*, **37**, 781–99.

HEDGE, A. and POWERS, J., 1995, Wrist postures while keyboarding: effects of a negative slope keyboard system and full motion forearm supports, *Ergonomics*, **38**, 508–17.

HELANDER, M. (ed.), 1981, *Human FActors/Ergonomics for Building and Construction* (New York: Wiley).

HELANDER, M., 1995, *A Guide to Ergonomics of Manufacturing* (London: Taylor & Francis).

HELANDER, M. and NAGAMACHI, M. (eds), 1992, *Design for Manufacturability: A Systems Approach to Concurrent Engineering and Ergonomics* (London: Taylor & Francis).

HELANDER, M.G., ZHANG, L. and MICHEL, D., 1995, Ergonomics of ergonomic chairs: a study of the adjustability features, *Ergonomics*, **38**, 2007–29.

HENDRICK, H.W., 1986, Macroergonomics: a conceptual model for integrating human factors with organizational design. In O. Brown Jr and H.W. Hendrick (eds), *Human Factors in Organizational Design and Management — II* (Amsterdam: North-Holland).

HENDRICK, H.W., 1987, Organizational design. In G. Salvendy (ed.), *Handbook of Human Factors* (New York: John Wiley and sons), 470–94.

HENDRICK, H.W., 1996, Human factors in ODAM: an historical perspective. In O. Brown Jr and H.W. Hendrick (eds) *Human Factors in Organizational Design and Management — V* (Amsterdam: North-Holland).

HENDRICK, H.W., 1997, *Good Ergonomics is Good Economics* (Santa Monica: Human Factors and Ergonomics Society).

HENDRICK, H.W. and BROWN, O., Jr (eds), 1984, *Human Factors in Organizational Design and Management* (Amsterdam: North Holland).

HETTINGER, T., 1970, *Angewandte Ergonomie, Arbeitsphysiologische und arbeitsmedizinische Probleme der Betriebspraxis* (Frechen).

HIDALGO, J., GENAIDY, A., KARWOWSKI, W., CHRISTENSEN, D., HUSTON, R. and STAMBOUGH, J., 1995, A cross-validation of the NIOSH limits for manual lifting. *Ergonomics*, **38**, 2455–64.

HIDALGO, J., GENAIDY, A., KARWOWSKI, W., CHRISTENSEN, D., HUSTON, R. and STAMBOUGH, J., 1997, A comprehensive lifting model: beyond the NIOSH lifting equation, *Ergonomics*, **40**, 916–27.

HILBORN, E.H., 1975, *Human Factors Experiments for Data Link: Final Report* (Washington, DC: US Department of Transportation, Federal Aviation Administration, Systems Research and Development Service).

HOLLNAGEL, E., 1993, *Human Reliability Analysis: Context and Control* (London: Academic Press).

HOWARD, R.A., 1980, An assessment of decision analysis, *Operations Research*, **28**, 4–27.

HOWELL, W.C. and GOLDSTEIN, I.L., 1971, *Engineering Psychology: Current Perspectives in Research* (New York: Appleton-Century-Crofts).

HSIANG, S.M., BROGMUS, G.E. and COURTNEY, T.K., 1997, Low back pain and lifting technique- a review, *International Journal of Industrial Ergonomics*, **19**, 59–74.

HUCHINGSON, R.D., 1981, *New Horizons for Human Factors in Design* (New York: McGraw-Hill).

HUNTER, T.A., 1992, *Engineering Design for Safety* (New York: McGraw-Hill).

HUNTING, W., GRANDJEAN, E. and MAEDA, K., 1980, Constrained postures in accounting machine operators, *Applied Ergonomics*, **11**, 145–9.

HUNTING, W., LAUBLI, T. and GRANDJEAN, E., 1981, Postural and visual loads at VDT workplaces 1. Constrained postures, *Ergonomics*, **24**, 917–31.

ILLIMUNIATING ENGINEERING SOCIETY OF NORTH AMERICA, 1999, *IESNA Lighting Handbook*, 9th edn (New York: IESNA).

INSTITUTE FOR APPLIED EXPERIMENTAL PSYCHOLOGY, 1952, *Handbook of Human Engineering Data* (Medford: Tufts University).

INSTRUMENT SOCIETY OF AMERICA, 1985, *Human Engineering for Control Centers: Recommended Practice* (Research Triangle Park: Instrument Society of America).

INTERNATIONAL LABOUR OFFICE, 1996, *Ergonomics Checkpoints* (Geneva: ILO).

IRVINE, C.H., SNOOK, S.H. and SPARSHATT, J.H., 1990, Stairway risers and treads: acceptable and preferred dimensions. *Applied Ergonomics*, **21**, 215–55.

ISO, 1998, *Ergonomic Requirements for Office Work with Visual Display Terminals (VDTs) Part 5: Workstation Layout and Postural Requirements*. ISO 9241–5 (Geneva: IOS).

IVERGÅRD, T., 1989, *Handbook of Control Room Design and Ergonomics* (London: Taylor & Francis).

JARVINEN, J. and KARWOWSKI, W., 1995, Analysis of self-reported accidents attributed to advanced manufacturing systems, *International Journal of Human Factors in Manufacturing*, **5**, 251–66.

JASCHINSKI-KRUZA, W., 1990, On the preferred viewing distances to screen and document at VDU workplaces, *Ergonomics*, **33**, 1055–63.

JAVITZ, A.E. (ed.), 1956, *Human Engineering in Equipment Design* (New York: Sage).

JOHNSEN, E.G. and CORLISS, W.R., 1971, *Human Factors Applications in Teleoperator Design and Operation* (New York: Wiley-Interscience).

JOHNSON, A.T., 1991, *Biomechanics and Exercise Physiology* (New York: Wiley).

JOHNSON-LAIRD, P.N., 1994, Mental models and probabilistic thinking, *Cognition*, **50**, 189–209.

JOINT ARMY-NAVY-AIR FORCE STEERING COMMITTEE, 1972, *Human Engineering Guide to Equipment Design* (New York: Wiley).

JONES, S., 1997, *The Archaeology of Ethnicity* (London: Routledge).

JONSSON, B., 1978, Kinesiology, *Contemporary Clinical Neurophysiology*, **34** 417–28.

JONSSON, B., 1988, The static component in muscle work, *European Journal of Physiology*, **57**, 305–10.

JORGENSEN, K., FALLENTIN, N., KROGH-LUND, C. and JENSEN, B., 1988, Electromyography and fatigue during prolonged, low-level static contractions, *European Journal of Physiology*, **57**, 316–21.

JORGENSEN, M.J., DAVIS, K.G., KIRKING, B.C., LEWIS, K.E. and MARRAS, W.S., 1999, Significance of biomechanical and physiological variables during the determination of maximum acceptable weight of lift, *Ergonomics*, **42**, 1216–32.

JOYCE, M. and WALLERSTEINER, U., 1989, *Ergonomics: Humanizing the Automated Office* (Cincinatti: South-Western).

KAHNEMAN, D., 1973, *Attention and Effort* (Englewood Cliffs: Prentice Hall).

KAHNEHMAN, D., SLOVIC, P. and TVERSKY, A. (eds), 1982, *Judgement Under Uncertainty: Heuristics and Biases* (Cambridge: Cambridge University Press).

KANTOWITZ, B.H. and SORKIN, R.D., 1983, *Human Factors: Understanding People-System Relationships* (New York: Wiley).

KANTOWITZ, B.H. and SORKIN, R.D., 1987, Allocation of function. In G. Salvendy (ed.), *Handbook of Human Factors* (New York: Wiley).

KAPLAN, M. and SCHWARTZ, S. (eds), 1975, *Human Judgement and Decision Processes* (New York: Academic Press).

KARWOWSKI, W., 1991, Complexity, fuzziness and ergonomic incompatibility issues in the control of dynamic work environments, *Ergonomics*, **34**, 671–86.

KARWOWSKI, W., 1992, The human world of fuzziness, human entropy, and the need for general fuzzy systems theory, *Journal of Japan Society for Fuzzy Theory and Systems*, **4**, 591–609.

KARWOWSKI, W. and AYOUB, M.M., 1984, Fuzzy modeling of stresses in manual lifting tasks, *Ergonomics*, **27**, 641–9.

KARWOWSKI, W., EBERTS, R., SALVENDY, G. and NOLAND, S., 1994, The effects of computer interface design on human postural dynamics, *Ergonomics*, **37**, 703–24.

KARWOWSKI, W., GENAIDY, A.M. and ASFOUR, S.S. (eds), 1990, *Computer-Aided Ergonomics* (Bristol: Taylor & Francis).

KARWOWSKI, W., KOSIBA, E., BENABDALLAH, S. and SALVENDY, G., 1990, A framework for development of fuzzy GOMS model for human-computer interaction, *International Journal of Human-Computer Interaction*, **2**, 287–305.

KARWOWSKI, W., LEE, W.G., JAMALDIN, B., GADDIE, P. and JANG, R., 1999, Beyond psychophysics: a need for cognitive modeling approach to setting limits in manual lifting tasks, *Ergonomics*, **42**, 40–60.

KARWOWSKI, W. and MARRAS, W.S. (eds), 1999, *The Occupational Ergonomics Handbook* (Boca Raton: CRC Press).

KARWOWSKI, W. and MITAL, A. (eds), 1986, *Applications of Fuzzy Set Theory in Human Factors* (Amsterdam: Elsevier Science).

KARWOWSKI, W., OSTASZEWSKI, K. and ZURADA, J., 1992, Applications of the catastrophe theory in modeling the risk of low back injury in manual lifting tasks, *La Travail Humain*, **55**, 259–75.

KARWOWSKI, W. and SALVENDY, G. (eds), 1998, *Ergonomics in Manufacturing: Raising Productivity Through Workplace Improvement* (Michigan: SME and IIE Presses).

KARWOWSKI, W. and SALVENDY, G. (eds), 1999, *Design of Work and Organization* (New York: Wiley).

KEEGAN, J.J., 1953, Alterations of the lumbar curve related to posture and seating, *Journal of Bone and Joint Surgery*, **35-A**, 589–603.

KENNEDY, K.W., 1978, *Reach Capability of Men and Women: A Three-dimensional Analysis*. AMRL-TR-77-50 (Wright-Patterson Air Force Base, OH: Aerospace Medical Research Laboratory, Aerospace Medical Division, Air Force Systems Command).

KEPPEL, G., 1982, *Design and Analysis: A Researcher's Handbook* (Englewood Cliffs: Prentice-Hall).

KIDD, P.T. and KARWOWSKI, W. (eds), 1994, *Advances in Agile Manufacturing* (Amsterdam: IOS Press).

KIM, J.Y. and MARRAS, W.S., 1987, Quantitative trunk muscle electromyography during lifting at different speeds, *International Journal of Industrial Ergonomics*, **1**, 219–29.

KIM, J.Y., STUART-BUTTLE, C. and MARRAS, W.S., 1994, The effects of mats on back and leg fatigue, *Applied Ergonomics*, **25**, 29–34.

KIRA, A., 1976, *The Bathroom* (New York: Viking).

KIRKPATRICK, D., 1979, Techniques for evaluating training programs, *Training and Development Journal*, **31**, 9–12.

KIRLIK, A., MILLER, R.A. and JAGACINSKI, R.J., 1993, Supervisory control in a dynamic and uncertain environment: a process model of skilled human-environment interaction, *IEEE Transactions on Systems, Man, and Cybernetics*, **23**, 929–52.

KIRWAN, B., 1994, *A Guide to Practical Human Reliability Assessment* (Philadelphia: Taylor & Francis).

KIRWIN, A., 1999, Expansion of community modeling: a case study in Park Hill, Denver, Colorado, *High Plains Applied Anthropologist*, **19**, 133–43.

KLEIN, G.A. and CALDERWOOD, R., 1991, Decision models: some lessons from the field, *IEEE Transactions on Systems, Man, and Cybernetics*, **21**, 1018–26.

KLEIN, G.A., ORASANU, J., CALDERWOOD, R. and ZSAMBOK, C.E. (eds), 1993, *Decision Making in Action: Models and Methods* (Norwood: Ablex).

KLEINER, B.M., 1996, Macroergonomics: lessons learned from large scale change efforts industry, government, and academia. In O. Brown and H.W. Hendrick (eds), *Human Factors in Organizational Design and Management V* (Amsterdam: North-Holland), 483–8.

KLEINER, B.M., 1997, An integrative framework for measuring and evaluating information management performance, *International Journal of Computers and Industrial Engineering*, **32**, 545–55.

KLING, J.W. and RIGGS, L.A. (eds), 1971, *Woodworth and Schlosberg's Experimental Psychology* (New York: Holt, Rinehart & Winston).

KNAVE, B.G., WIBOM, R.I., VOSS, M., HEDSTROM, L.D. and BERGQVIST, U., 1985, Work with video display terminals among office employees, *Scandinavian Journal of Work: Environmental Health*, **11**, 457–66.

KNIRK, F.G. and GUSTAFSON, K.L., 1986, *Instructional Technology: A Systematic Approach to Education* (New York: Holt, Rinehart & Winston).

KONZ, S. and JOHNSON, S., 2000, *Work Design: Industrial Ergonomics* (Scottsdale: Holcomb Hathaway).

KOPFSTEIN-PENK, A., 1994, *The Healthy Guitar: Finding a Guitar that Reduces Injury Risk and Improves Your Playing* (Saint Louis: MMB Music).

KORADECKA, D. (ed.), 1997, *Work Safety and Ergonomics* (Warsaw: Central Institute of Labour Protection). [in Polish]

KORZYBSKI, A., 1921, *Manhood of Humanity: The Science and Art of Human Engineering* (New York: E.P. Dutton).

KOUBEK, R. and KARWOWSKI, W. (eds), 1996, *Human Factors in Agile Manufacturing* (Santa Monica: IEA Press).

KRAISS, M., 1976, *Introduction to Human Engineering* (Cologne: TöV).

KROEMER, K.H.E. and GRANDJEAN, E., 1997, *Fitting the Task to the Man: A Textbook of Occupational Ergonomics* (Philadelphia: Taylor & Francis).

KROEMER, K.H.E., KROEMER, H.J. and KROEMER-ELBERT, K.E., 1990, *Engineering Physiology: Bases of Human Factors/Ergonomics* (New York: van Nostrand Reinhold).

KROEMER, K.H.E., KROEMER, H.J. and KROEMER-ELBERT, K.E., 1994, *Ergonomics: How to Design for Ease and Efficiency* (Englewood Cliffs: Prentice-Hall).

KROEMER, K.H.E., MARRAS, W.S., MCGLOTHLIN, J.D., MCINTYRE, D.R. and NORDIN, M., 1990, On the measurement of human strength, *International Journal of Industrial Ergonomics*, **6**, 199–210.

KUMAR, S. and MITAL, A. (eds), 1996, *Electromyography in Ergonomics* (London: Taylor & Francis).

KUMASHIRO, M. and MEGAW, E.D. (eds), 1991, *Towards Human Work:*

Solutions to Problems in Occupational Health and Safety (Bristol: Taylor & Francis).

KUORINKA, I. and FORCIER, L. (eds), 1995, *Work Related Musculoskeletal Disorders (WMSDs): A Reference Book for Prevention* (London: Taylor & Francis).

KUORINKA, I., JONSSON, B., KILLBOM, A., VINTERBERG, H., BIERING-SORENSON, F., ANDERSON, G. and JORGENSEN, K., 1987, Standardized Nordic Questionnaires for the analysis of musculoskeletal symptoms, *Applied Ergonomics*, **18**, 233–7.

KVÅLSETH, T.O., 1983, *Ergonomics of Workstation Design* (London: Butterworths).

LANDAU, K. and STÜBLER, E., 1992, *Die Arbeit im Dienstleistungsbetrieb* (Stuttgart: Ulmer).

LANG, R. and HELLPACH, W., 1922, *Gruppenfabrikation* (Berlin).

LAUGHERY, K.R., 1999, Modeling human performance during system design. In *Human/Technology Interaction in Complex Systems* (Greenwich: JAI Press), 147–73.

LAURIG, W., 1990, *GrundzÅge der Ergonomie, Erkenntnisse und Prinzipien* (Berlin: Aufl).

LAVENDER, S.A., MIRKA, G.A., SCHOENMARKLIN, R.W., SOMMERICH, C.M., SUDHAKER, L.R. and MARRAS, W.S., 1989, An investigation into the effects of preview and symmetry on trunk loading, *Human Factors*, **31**, 101–15.

LAVINE, N.D., PETERS, S.D. and LAUGHERY, K.R., 1996, *Methodology for Predicting and Applying Human Response to Environmental Stressors*. DNA-TR-95-116 (Alexandria: Defense Special Weapons Agency).

LEAMON, T.B. and DEMPSEY, P.G., 1995, The unusual congruence between subjective evaluations and losses associated with inadequate hand tool design, *International Journal of Industrial Ergonomics*, **16**, 23–8.

LECLERCQ, S., 1999, The prevention of slipping accidents: a review and discussion of work related to the methodology of measuring slip resistance, *Safety Science*, **31**, 95–125.

LEDGARD, H., SINGER, A. and WHITESIDE, J., 1981, *Directions in Human Factors for Interactive Systems* (New York: Springer).

LEE, G.C.H. (ed.), 1999, *Advances in Occupational Ergonomics and Safety* (Amsterdam: IOS Press).

LEHMANN, G., 1962, *Praktische Arbeitsphysiologie* (Stuttgart: Aufl).

LELONG, C., DREVET, R., CHEVALLIER, R. and PHELPS, X., 1988, Spinal biomechanics and the sitting position, *Revue de Rhumatisme*, **55**, 375–80.

LIN, M., RADWIN, R.G. and SNOOK, S.H., 1997, A single metric for quantifying biomechanical stress in repetitive motions and exertions, *Ergonomics*, **40**, 543–58.

LINTON, S.J. and KAMWENDO, K., 1989, Risk factors in psychosocial work environment for neck and shoulder pain in secretaries, *Journal of Occupational Medicine*, **31**, 609–13.

LIPMANN, O., 1932, *Lehrbuch der Arbeitswissenschaft* (Jena).

LOE, D.L., MANSFIELD, K.P. and ROWLANDS, E., 1994, Appearance of lit environment and its relevance in lighting design: experimental study, *Lighting Research and Technology*, **26**, 119–33.

LUCZAK, H., 1998, *Arbeitswissenschaft* (Berlin: Aufl).

LUEDER, R. and NORO, K. (eds), *Hard Facts About Soft Machines — The Ergonomics of Seating* (London: Taylor & Francis).

LUND, R.L., BISHOP, A.B., NEWMAN, A.E. and SALZMAN, H., 1993, *Designed to Work: Production Systems and People* (Englewood Cliffs: Prentice-Hall).

LUPTON, T. (ed.), 1986, *Human Factors: Man, Machine, and New Technology* (Berlin: Springer).

MacKINNON, S.E. and NOVAK, C.B., 1994, Clinical commentary: pathogenesis of cumulative trauma disorder, *Journal of Hand Surgery*, **19-A**, 873–83.

MacKWORTH, N.H., 1970, *Vigilance and Attention* (Baltimore: Pergamon).

MacLEOD, D., 1995, *The Ergonomics Edge: Improving Safety, Quality, and Productivity* (New York: van Nostrand Reinhold).

MAINZER, K., 1999, *Thinking in Complexity: The Complex Dynamics of Matter, Mind and Mankind* (Berlin: Springer).

MAJHCHRZAK, A., 1988, *The Human Side of Factory Automation* (San Francisco: Jossey-Bass).

MANDAL, A.C., 1976, Work-chair with tilting seat, *Ergonomics*, **19**, 157–64.

MANDAL, A.C., 1981, The seated man (*Homo sedans*), *Applied Ergonomics*, **12**, 19–26.

MANGUM, B.W. and HILL, J.E. (eds), 1977, *Thermal Analysis — Human Comfort — Indoor Environments* (Washington, DC: Department of Commerce, National Bureau of Standards).

MAREK, T., NOWOROL, C. and KARWOWSKI, W., 1988, Mental fatigue at work and pain perception, *Work and Stress*, **2**, 133–7.

MAREK, T., NOWOROL, C., WOS, H., KARWOWSKI, W. and HAMIGA, K., 1992, Muscular loading and subjective ratings of muscular tension experienced by novices when typing with standard and split-design computer keyboards, *International Journal of Human-Computer Interaction*, **4**, 387–94.

MARGOLIS, B.L. and KROES, W.H. (eds), 1975, *The Human Side of Accident Prevention; Psychological Concepts and Principles which Bear on Industrial Safety* (Springfield: C.C. Thomas).

MARK, L.S., WARM, J.S. and HUSTON, R.L. (eds), 1987, *Ergonomics and Human Factors: Recent Research* (New York: Springer).

MARQUIE, J.C., CAU-BARELLE, D.P. and VOLKOFF, S. (eds), 1998, *Working with Age* (London: Taylor & Francis).

MARQUIE, J.C., THON, B. and BARACAT, B., 1994, Age influence on attitudes of office workers faced with new computerized technologies, *Applied Ergonomics*, **25**, 130–42.

MARRAS, W.S., DAVIS, K.G., KIRKING, B.C. and GRANATA, K.P., 1998, Spine loading and trunk kinematics during team lifting, *Ergonomics*, **42**, 1258–73.

MARRAS, W.S., FINE, L.J., FERGUSON, S.A. and WATERS, T.R., 1999, The effectiveness of commonly used lifting assessment methods to identify industrial jobs associated with elevated risk of low-back disorders, *Ergonomics*, **42**, 229–45.

MARRAS, W.S. and GRANATA, K.P., 1997, An EMG-assisted model of trunk lateral bending, *Journal of Biomechanics*, **30**, 697–703.

MARRAS, W.S., GRANATA, K.P., DAVIS, K.G., ALLREAD, W.G. and JORGENSEN, M.J., 1997, Spine loading and probability of low back disorder risk as a function of box location on a pallet, *International Journal of Human Factors in Manufacturing*, **7**, 323–36.

MARRAS, W.S., KING, A.I. and JOYNT, R.L., 1984, Measurements of loads on the lumbar spine under isometric and isokinetic conditions, *Spine*, **9**, 176–88.

MARRAS, W.S. and LAVENDER, S.A., 1991, The effects of method of use, tool design, and roof height on trunk muscle activities during underground scaling bar use, *Ergonomics*, **34**, 221–32.

MARRAS, W.S. and MIRKA, G.A., 1989, Trunk strength during asymmetric trunk motion, *Human Factors*, **31**, 667–77.

MARRAS, W.S., RANGARAJULU, S.L. and LAVENDER, S.A., 1987, Trunk loading and expectation, *Ergonomics*, **30**, 551–62.

MARRAS, W.S. and REILLY, C.H., 1988, Networks of internal trunk loading activities under controlled trunk motion conditions, *Spine*, **13**, 661–7.

MARRAS, W.S. and SCHOENMARKLIN, R.W., 1993, Wrist motion in industry, *Ergonomics*, **36**, 341–51.

MARRAS, W.S., SCHOENMARKLIN, R.W., GREENSPAN, G.J. and LEHMAN, K.R., 1995, Quantification of wrist motions during scanning, *Human Factors*, **37**, 412–23.

MARRAS, W.S. and SOMMERICH, C.M., 1991, A three-dimensional motion model of loads on the lumbar spine. I. Model Structure, *Human Factors*, **33**, 123–37.

MARRAS, W.S. and SOMMERICH, C.M., 1991, A three-dimensional motion model of loads on the lumbar spine. II. Model validation, *Human Factors*, **33**, 139–49.

MATTILA, M. and KARWOWSKI, W. (eds), 1992, *Computer Applications in Ergonomics, Occupational Safety and Health* (Amsterdam: North-Holland).

MATTILA, M., KARWOWSKI, W. and VIKKI, M., 1993, Analysis of working postures in hammering tasks at building construction sites using the computerized OWAS-method, *Applied Ergonomics*, **24**, 405–12.

MAYO, E., 1933, *The Human Problems of an Industrial Civilization* (New York: Macmillan).

MCCORMICK, E.J., 1964, *Human Factors Engineering* (New York: McGraw-Hill).

MCFARLAND, R.A., 1953, *Human Body Size and Capabilities in the Design and Operation of Vehicular Equipment* (Boston: Harvard School of Public Health).

MCFARLAND, R.A. and MOSELEY, A.L., 1954, *Human Factors in Highway Transport Safety* (Boston: Harvard School of Public Health).

MCMILLAN, G.R., BEEVIS, D., SALAS, E., STRUB, M.H., SUTTON, R. and VAN BREDA, L., 1989, *Applications of Human Performance Models to System Design* (New York: Plenum).

MEISTER, D., 1971, *Human Factors: Theory and Practice* (New York: Wiley-Interscience).

MEISTER, D., 1976, *Behavioral Foundations of System Development* (New York: Wiley).

MEISTER,, D., 1985, *Behavioral Analysis and Measurement Methods* (New York: Wiley).

MEISTER, D., 1989, *Conceptual Aspects of Human Factors* (Baltimore: Johns Hopkins University Press).

MEISTER, D., 1991, *Psychology of System Design* (Amsterdam: Elsevier).

MEISTER, D., 1997, *The Practice of Ergonomics* (Bellingham: BCPE).

MEISTER, D., 1999, *The History of Human Factors Ergonomics* (Mahwah: Erlbaum).

MEISTER, D. and RABIDEAU, G.F., 1965, *Human Factors Evaluation in System Development* (New York: Wiley).

MICHEL, D.P., 1994, Two types of chairs on stature change and comfort for individuals with healthy and herniated discs, *Ergonomics*, **37**, 1231–44.

MILLER, R.B., 1953, *A Method for Man-machine Task Analysis*. WADC Technical Report 53–137 (Wright-Patterson Air Force Base, OH: Wright Air Development Center).

MIRKA, G.A. and MARRAS, W.S., 1990, Lumbar motion response to a constant load velocity lift, *Human Factors*, **32**, 493–501.

MITAL, A. and KARWOWSKI, W. (eds), 1988, *Ergonomics in Rehabilitation* (Philadelphia: Taylor & Francis).

MITAL, A. and KARWOWSKI, W., 1991, *Workspace, Equipment and Tool Design* (Amsterdam: Elsevier).

MITAL, A., NICHOLSON, A.S. and AYOUB, M.M., 1993, *A Guide to Manual Materials Handling* (London: Taylor & Francis).

MONK, A. (ed.), 1984, *Fundamentals of Human-Computer Interaction* (London: Academic Press).

MOORE, D.J., 1996, *Manager's Guide to Workplace Ergonomics* (Madison: Business & Legal Reports).

MOORHEAD, G. and GRIFFIN, R.W. (eds), 1998, *Organizational Behavior: Managing People and Organizations* (New York: Houghton Mifflin).

MORAAL, J. and KRAISS, K.-F. (eds), 1981, *Manned Systems Design: Methods, Equipment, and Applications* (New York: Plenum).

MORAY, N., 1979, *Mental Workload* (New York: Plenum).

MÜNSTERBERG, H., 1914, *GrundzÅge einer Psychotechnik* (Leipzig).

MURRELL, K.F.H., 1965, *Ergonomics: Man in His Working Environment* (London: Chapman & Hall).

MURRELL, K.F.H., 1965, *Human Performance in Industry* (New York: Reinhold).

MURTHY, G., KAHAN, N.J., HARGENS, A.R. and REMPEL, D.M., 1997, Forearm muscle oxygenation decreases with low levels of voluntary contraction, *Journal of Orthopedic Research*, **15**, 507–11.

NAG, P.K. (ed.), 1996, *Ergonomics and Work Design* (New Delhi: New Age International Press).

NERTNEY, R.J. and BULLOCK, M.G., 1976, *Human Factors in Design* (Washington, DC: US Energy Research and Development Administration, Division of Operational Safety).

NEWELL, A. and SIMON, H.A., 1972, *Human Problem Solving* (Englewood Cliffs: Prentice-Hall).

NEWELL, A. and SIMON, H.A., 1990, *Unified Theories of Cognition* (Cambridge, MA: Harvard University Press).

NICHOLSON, A.S. and RIDD, J.E. (ed.), 1988, *Health, Safety, and Ergonomics* (London: Butterworths).

NICKERSON, R.S., 1986, *Using Computers: The Human Factors of Information Systems* (Cambridge, MA: MIT Press).

NIEBEL, B. and FREIVALDS, A., 1999, *Methods, Standards, and Work Design* (Boston: McGraw-Hill).

NIKERSON, R. (ed.), 1995, *Emerging Needs and Opportunities for Human Factors Research* (Washington, DC: National Academy Press).

NISBET, R., 1962, *Community and Power* (New York: Oxford University Press).

NORDIN, M., ANDERSON, G. and POPE, M. (eds), 1997, *Musculoskeletal Disorders in the Workplace: Principles and Practice* (St Louis: Mosby-Year Book).

NORMAN, D.A., 1988, *The Psychology of Everyday Things* (New York: Basic).

NORMAN, D.A. and BOBROW, D., 1975, On data-limited and resource-limited processing, *Cognitive Psychology*, **7**, 44–60.

NORMAN, D.A. and DRAPER, S.W. (eds), 1986, *User Centered System Design: New Perspectives on Human- Computer Interaction* (Hillsdale: Lawrence Erlbaum).

NORO, K. and BROWN, O., Jr (eds), 1990, *Human Factors in Organizational Design and Management — III* (Amsterdam: North Holland).

NORO, K. and IMADA, A.S. (ed.), 1991, *Participatory Ergonomics* (London: Taylor & Francis).

NORTH, R.A. and RILEY, V.A., 1989, W/INDEX: a predictive model of operator workload. In G.R. McMillan, D. Beevis, E. Salas, M.H. Strub, R. Sutton and L. van Breda (eds), *Applications of Human Performance Models to System Design* (New York: Plenum), 81–9.

NOY, Y.I. (ed.), 1997, *Ergonomics and Safety of Intelligent Driver Interfaces* (Mahwah: Lawrence Erlbaum).

O'BRIEN, T.G. and CHARLTON, S.G. (eds), 1996, *Handbook of Human Factors Testing and Evaluation* (Mahwah: Lawrence Erlbaum).

O'HARA, J.M., BROWN, W.S., STUBLER, W.F., WACHTEL, J.A. and PERSENSKY, J.J., 1995, *Human-System Interface Design Review*

Guideline: Draft Report for Comment. NUREG-0700 Rev. 1, (Washington, DC: US Nuclear Regulatory Commission).

O'NEILL, M.J., 1998, *Ergonomic Design for Organizational Effectiveness* (New York: Lewis).

OBORNE, D.J., 1995, *Ergonomics at Work: Human Factors in Design and Development* (Chichester: Wiley).

OBORNE, D.J., BRANTON, R., LEAL, F., SHIPLEY, P. and STEWART, T. (eds), 1993, *Person-Centered Ergonomics: A Brantonian View of Human Factors* (Washington, DC: Taylor & Francis).

OBORNE, D.J. and GRUNEBERG, M.M. (eds), 1983, *The Physical Environment at Work* (Chichester: Wiley).

OSTROM, L.T., 1993, *Creating the Ergonomically Sound Workplace* (San Francisco: Jossey-Bass).

OWENS, D.A. and WOLF-KELLY, K., 1987, Near work, visual fatigue, and variations of oculomotor tonus, *Investigative Ophthalmology and Visual Science*, **28**, 743–9.

PACE, R.W., 1983, *Organizational Communication, Foundations for Human Resource Development* (Englewood Cliffs: Prentice Hall).

PACHOLSKI, L. and MARCINKOWSKI, J.S. (eds), 1999, *Certification and Accreditation of Ergonomics* (Poznan: Labour Protection and Work Safety Education, Put Institute of Management Engineering).

PAPANEK, V.J., 1971, *Design for the Real World: Human Ecology and Social Change* (Toronto: Bantam).

PARASURAM, R., 1987, Human-computer monitoring, *Human Factors*, **29**, 695–706.

PARASURAMAN, R., MOLLOY, R. and SINGH, I.L., 1993, Performance consequences of automation induced complacency, *International Journal of Aviation Psychology*, **3**, 1–23.

PARASURAMAN, R. and MOULOUA, M. (eds), 1995, *Automation and Human Performance: Theory and Applications* (Hillsdale: Lawrence Erlbaum).

PARASURAMAN, R. and RILEY, V.A., 1997, Humans and automation: use, misuse, disuse, abuse, *Human Factors*, **39**, 230–53.

PARKER, K.G. and IMBUS, H.R., 1992, *Cumulative Trauma Disorders: Current Issues and Ergonomic Solutions: A Systems Approach* (Boca Raton: Lewis).

PARSONS, K.C., 1993, *Human Thermal Environments* (London: Taylor & Francis).

PASCARELLI, E.F. and KELLA, J.J., 1993, Soft tissue injuries related to use of the computer keyboard, *Journal of Occupational Medicine*, **35**, 522–31.

PASMORE, W.A., 1988, *Designing Effective Organizations: The Sociotechnical Systems Perspective* (New York: Wiley).

PASMORE, W.A. and SHERWOOD, J.J. (eds), 1978, *Sociotechnical Systems: A Sourcebook* (La Jolla: University Associates).

PEACOCK, B. and KARWOWSKI, W. (eds), 1993, *Automative Ergonomics* (London: Taylor & Francis).

PELSMA, K.H. (ed.), 1987, *Ergonomics Sourcebook: A Guide to Human Factors Information* (Lawrence: Report Store).

PEPER, E., TAYLOR, W., PIERCE, A. and TIBBETS, V., 1994, *Prevent Computer User Injury with Biofeedback* (San Francisco: San Francisco State University).

PERROW, C., 1999, *Normal Accidents: Living with High Risk Technologies* (Princeton: Princeton University Press).

PETROSKI, H., 1985, *To Engineer is Human: The Role of Failure in Successful Design* (New York: St Martin's).

PHEASANT, S., 1986, *Bodyspace: Anthropometry, Ergonomics and Design* (Philadelphia: Taylor & Francis).

PHEASANT, S., 1991, *Ergonomics, Work, and Health* (Gaithersburg: Aspen).

PIAGET, J., 1968, *On the Development of Memory and Identity*. Heinz Werner Lectures, 1967 (Worcester: Clark University Press).

PITZ, G.F. and SACHS, N.J., 1984, Judgment and decision: theory and application, *Annual Review of Psychology*, **35**, 139–63.

PROCTOR, R.W. and ZANDT, T.V., 1994, *Human Factors in Simple and Complex Systems* (Boston: Allyn & Bacon).

PULAT, B.M., 1997, *Fundamentals of Industrial Ergonomics* (ILlinois: Waveland).

PULAT, B.M. and ALEXANDER, D.C. (eds), 1991, *Industrial Ergonomics: Case Studies* (Norcross: Industrial Engineering and Management Press).

PUTZ-ANDERSON, V., 1988, *Cumulative Trauma Disorders: A Manual for Musculoskeletal Diseases of the Upper Limbs* (New York: Taylor & Francis).

QÉINNEC, Y. and DANIELLOU, F. (eds), 1991, *Designing for Everyone: Proceedings of the Eleventh Congress of the International Ergonomics Association* (London: Taylor & Francis).

RAHIMI, M. and KARWOWSKI, W. (eds), 1992, *Human-Robot Interaction* (London: Taylor & Francis).

RAHMAN, M.M. and SPRIGLE, S., 1997, Design and selection guidelines for accessible consumer product controls, *Assistive Technology*, **9**, 3–14.

RAIFFA, H., 1968, *Decision Analysis: Introductory Lectures on Choices Under Uncertainty* (New York: Addison-Wesley).

RASMUSSEN, J., 1983, Skills, rules, and knowledge; signals, signs, and symbols, and other distinctions in human performance models, *IEEE Transactions on Systems, Man, and Cybernetics*, **13**, 257–66.

RASMUSSEN, J., 1985, The role of hierarchical knowledge representation in decision making and system management, *IEEE Transactions on Systems, Man, and Cybernetics*, **15**, 234–43.

RASMUSSEN, J., 1986, *Information Processing and Human Machine Interaction: An Approach to Cognitive Engineering* (Amsterdam: Elsevier Science/North-Holland).

RASMUSSEN, J., DUNCAN, K. and LEPLAT, J. (eds), 1987, *New Technology and Human Error* (London: Wiley).

RASMUSSEN, J. and ROUSE, W. (eds), 1981, *Human Detection and Diagnosis of System Failures* (New York: Plenum).

REA, M.S., 1991, Solving the problem of VDT reflections, *Progressive Architecture*, **October**, 35–40.

REA, M.S. and OUELLETTE, M.J., 1991, Relative visual performance: a basis for application, *Lighting Research and Technology*, **23**, 135–44.

REASON, J.T., 1990, *Human Error* (Cambridge: Cambridge University Press).

ROBERTS, D.L. and BECKER, W.J. (eds), 1991, *Human Factors: A Series of Quality Instructural Materials* (St Joseph: American Society of Agricultural Engineers).

ROBERTS, S.L. and FALKENBURG, S.A., 1992, *Biomechanics: Problem Solving for Functional Activity* (St Louis: Mosby-Year Book).

ROBERTSON, M.M. and RAHIMI, M., 1990, A systems analysis for implementing video display terminals, *IEEE Transactions Engineering Management*, **37**, 55–61.

ROBERTSON, M.M., ROBINSON, M., O'NEIL, M.J. and SLES, J., 1998, Measuring the impact of work environment change programs: a systems approach. In *Proceedings of the Human Factors and Ergonomics Society 42nd Annual Meeting* (Santa Monica), 984–8.

RODAHL, K., 1989, *The Physiology of Work* (London: Taylor & Francis).

ROEBUCK, J.A. Jr, KROEMER, K.H.E. and THOMSON, W.G., 1975, *Engineering Anthropometry Methods* (New York: Wiley).

ROGERS, E.M., 1995, *Diffusion of Innovations* (New York: Free Press).

ROHMERT, W. and LANDAU, K., 1983, *A New Technique for Job Analysis* (London: Taylor & Francis).

ROHMERT, W. and RUTENFRANZ, J., 1983, *Praktische Arbeitsphysiologie* (Stuttgart: Aufl).

ROSE, M.J., 1991, Keyboard operating posture and actuation force: implications for muscle over-use, *Applied Ergonomics*, **22**, 198–203.

ROSENBROCK, H.H. (ed.), 1989, *Designing Human-centered Technology: A Cross-disciplinary Project in Computer-aided Manufacturing* (London: Springer).

ROSNER, J. (ed.), 1968, *Ergonomics* (Warsaw: Book & Knowledge). [in Polish]

ROUSE, W.B., 1977, Human-computer interaction in multitask situations, *IEEE Transactions on Systems, Man, and Cybernetics*, **7**.

ROUSE, W.B., 1988, Adaptive aiding for human/computer control, *Human Factors*, **30**, 431–43.

ROUSE, W.B., 1991, *Design for Success: A Human-centred Approach to Designing Successful Products and Systems* (New York: Wiley).

RUBIN, A.I., 1976, *Energy Conservation in Buildings: A Human Factors/ System Viewpoint* (Washington, DC: National Bureau of Standards, Institute for Applied Technology).

RUBINSTEIN, R. and HERSH, H.M., 1984, *The Human Factor: Designing Computer Systems for People* (Burlington: Digital).

SAGAN, S.D., 1993, *The Limits of Safety: Organization, Accidents and Nuclear Weapons* (Princeton: Princeton University Press).

SAGE, A.P. (ed.), 1987, *System Design for Human Interaction* (New York: IEEE Press).

SALVENDY, G. (ed.), 1992, *Handbook of Industrial Engineering* (New York: Wiley).

SALVENDY, G. (ed.), 1997, *Handbook of Human Factors and Ergonomics* (New York: Wiley-Interscience).

SALVENDY, G. and KARWOWSKI, W. (eds), 1994, *Design of Work and Development of Personnel in Advanced Manufacturing* (New York: Wiley).

SALVENDY, G. and SMITH, M.J. (eds), 1981, *Machine Pacing and Occupational Stress* (London: Taylor & Francis).

SANDERS, M.S. and MCCORMICK, E.J., 1992, *Human Factors in Engineering and Design*, 7th edn (New York: McGraw-Hill).

SAUTER, S.L., DAINOFF, M.J. and SMITH, M.J., 1990, *Promoting Health and Productivity in the Computerized Office: Models of Successful Ergonomics Interventions* (London: Taylor & Francis).

SAUTER, S.L. and MURPHY, L.R. (eds), 1995, *Organizational Risk Factors for Job Stress* (Washington, DC: APA).

SAYERS, B.A. (ed.), 1988, *Human Factors and Decision Making: Their Influence on Safety and Reliability* (London: Elsevier).

SCHMIDTKE, H., 1976, *Ergonomische Bewertung von Arbeitssystemen* (Munich).

SCHMIDTKE, H. (ed.), 1984, *Ergonomic Data for Equipment Design* (New York: Plenum).

SCHMIDTKE, H., 1993, *Ergonomie* (Munich: Aufl).

SCHULDT, K., 1988, On neck muscle activity and load reduction in sitting postures, *Scandinavian Journal of Rehabilitation Medicine*, **19**, 18–47.

SCOTT, P.A., BRIDGER, R.S. and CHARTERIS, J. (eds), 1998, *Global Ergonomics* (Amsterdam: Elsevier).

SEJERSTED, O.M. and WESTGAARD, R.H., 1988, Occupational muscle pain and injury: scientific challenge, *European Journal of Physiology*, **57**, 271–4.

SHERIDAN, T.B., 1992, *Telerobotics, Automation, and Human Supervisory Control* (Cambridge, MA: MIT Press).

SHERIDAN, T.B. and FERRELL, W.A., 1974, *Man-Machine Systems: Information, Control, and Decision Models of Human Performance* (Cambridge: MIT Press).

SIEGEL, A.I. and WOLF, J.A., 1969, *Man-Machine Simulation Models* (New York: Wiley-Interscience).

SILVERMAN, M., 1962, *Human Engineering: An ASTIA Report Bibliography* (Arlington: Defense Documentation Center).

SILVERSTEIN, B.A., FINE, L.J. and ARMSTRONG, T.J., 1986, Hand wrist cumulative trauma disorders in industry, *British Journal of Industrial Medicine*, **43**, 1179–84.

SIMON, H.A., 1981, *The Sciences of the Artificial* (Cambridge, MA: MIT Press).

SINGLETON, W.T., 1972, *Introduction to Ergonomics* (Geneva: WHO).

SINGLETON, W.T., 1982, *The Body at Work: Biological Ergonomics* (Cambridge: Cambridge University Press).

SINGLETON, W.T., 1989, *The Mind at Work: Psychological Ergonomics* (Cambridge: Cambridge University Press).

SINGLETON, W.T., WHITFIELD, D. and EASTERBY, R.S., 1967, *The Human Operator in Control Systems* (London: Taylor & Francis).

SMITH, K.U., 1962, *Delayed Sensory Feedback and Behavior* (Philadelphia: W. B. Saunders).

SMITH, K.U., 1965, *Behavior Organization and Work* (Madison: College Printing).

SMITH, K.U., 1987, Origins of human factors science, *Human Factors Society Bulletin*, **30**, 1–3.

SMITH, K.U., 1988, Origins of ergonomics and the International Ergonomics Association, *Human Factors Society Bulletin*, **31**, 2–5.

SMITH, K.U. and HENRY, J.P., 1967. Cybernetic foundations for rehabilitation, *American Journal of Physical Medicine*, **46**, 379–467.

SMITH, K.U. and SMITH, M.F., 1973, *Psychology: An Introduction to Behavior Science* (Boston: Little Brown).

SMITH, K.U. and SMITH, W.M., 1958, *The Behavior of Man: Introduction to Psychology* (New York: Holt, Rinehart & Winston).

SMITH, K.U. and SMITH, W.M., 1962, *Perception and Motion: An Analysis of Space-Structured Behavior* (Philadelphia: W. B. Saunders).

SMITH, M. and SAINTFORT, P., 1989, A balance theory of job design for stress reduction, *International Journal of Industrial Ergonomics*, **4**, 67–79.

SNOOK, S.H., 1978, The design of manual handling tasks, *Ergonomics*, **21**, 963–85.

SNOOK, S.H. and CIRIELLO, V.M., 1991, The design of manual handling tasks: revised tables of maximum acceptable weights and forces. *Ergonomics*, **34**, 1197–1213.

SNOOK, S.H., CIRIELLO, V.M. and WEBSTER, B.S., 1999, Maximum Acceptable Forces for Repetitive Wrist Extension with a Pinch Grip. *International Journal of Industrial Ergonomics*, **24**, 579–90.

SNOOK, S.H., FVAILLANCOURT, D.R., CIRIELLO, V.M. and WEBSTER, B.S., 1997, Maximum Acceptable Forces for Repetitive Ulnar Deviation of the Wrist. *American Industrial Hygiene Association Journal*, **58**, 509–17.

SNOOK, S.H., VAILLANCOURT, D.R., CIRIELLO, V.M. and WEBSTER, B.S., 1995, Psychophysical studies of repetitive wrist flexion and extension. *Ergonomics*, **38**, 1488–1507.

SOMMERICH, C.M. and MARRAS, W.S., 1992, Temporal patterns of trunk muscle activity throughout a dynamic, asymmetric lifting motion, *Human Factors*, **34**, 215–30.

SOMMERICH, C.M., MARRAS, W.S. and PARNIANPOUR, M., 1996, A

quantitative description of typing biomechanics, *Journal of Occupational Rehabilitation*, **6**, 33–55.

SOMMERICH, C.M., MCGLOTHLIN, J.D. and MARRAS, W.S., 1993, Occupational risk factors associated with soft tissue disorders of the shoulder: a review of recent investigations in the literature, *Ergonomics*, **36**, 697–717.

STACK, B., 1988, *Repetitive Strain Injury: Prevention and Rehabilitation* (Tasmania: Tasmania State Department of Health Services).

STANTON, N. (ed.), 1994, *Human Factors in Alarm Design* (London: Taylor & Francis).

STEPHEN, P., 1988, *Bodyspace: Anthropometry, Ergonomics, and Design* (London: Taylor & Francis).

STRAMLER, J.H. Jr, 1993, *The Dictionary for Human Factors/Ergonomics* (Boca Raton: CRC Press).

STRANDBERG, L., 1983, Ergonomics applied to slipping accidents. In T.O. Kvålseth (ed.), *Ergonomics of Workstation Design* (London: Butterworths), 201–28.

SUH, N.P., 1990, *The Principles of Design* (New York: Oxford University Press).

TAKALA, E., LAMMI, S., NIEMINEN, H. and VIIKARI-JUNTURA, E., 1993, Electromyographic changes in the static holding test of the arm, *International Journal of Industrial Ergonomics*, **12**, 85–90.

TAYLOR, J.C. and FELTON, D.F., 1993, *Performance by Design* (Englewood Cliffs: Prentice-Hall).

TAYYARI, F., BURFORD, C.L. and RAMSEY, J.D., 1989, Guidelines for the use of systematic workload estimation, *International Journal of Industrial Ergonomics*, **4**, 61–5.

TAYYARI, F. and SMITH, J.L., 1997, *Occupational Ergonomics: Principles and Applications* (London: Chapman & Hall).

TAYYARI, F., SMITH, J.L., BURFORD, C.L. and RAMSEY, J.D., 1989, Design criteria for microclimate cooling systems, *Ergonomics*, **32**, 1247–50.

THACKRAY, R.I. and TOUCHSTONE, R.M., 1989, Detection efficiency on an air traffic control monitoring task with and without computer aiding, *Aviation, Space, and Environmental Medicine*, **60**, 744–8.

TICHAUER, E.R., 1978, *The Biomechanical Basis of Ergonomics: Anatomy Applied to the Design of Work Situations* (New York: Wiley).

TITTIRANONDA, P., REMPEL, D., ARMSTRONG, T. and BURASTERO, S., 1999, Effect of four computer keyboards in computer users with upper extremity musculoskeletal disorders, *American Journal of Industrial Medicine*, **35**, 647–61.

TODA, M., 1962, The design of a fungus-eater: a model of human behavior in an unsophisticated environment, *Behavioral Science*, **7**, 164–83.

TOUGAS, G. and NORDIN, M.C., 1987, Seat features recommendations for workstations, *Applied Ergonomics*, **18**, 207–10.

TREASTER, D. and MARRAS, W.S., 1987, Measurement of seat and back pressure distributions, *Human Factors*, **29**, 563–75.

TRIST, E.L. and BAMFORTH, K.W., 1951, Some social and psychological consequences of the longwall method of coal-getting, *Human Relations*, **4**, 3–38.

TRIST, E.L., HIGGIN, G.W., MURRAY, H. and POLLOCK, A.B., 1963, *Organizational Choice* (London: Tavistock).

TSANG, P.S. and VELAZQUEZ, V.L., 1996, Diagnosticity and multidimensional subjective workload ratings, *Ergonomics*, **39**, 358–81.

TULGA, M.K. and SHERIDAN, T.B., 1980, Dynamic decisions and work load in multitask supervisory control, *IEEE Transactions on Systems, Man, and Cybernetics*, **10**.

TVERSKY, A. and KAHNEMAN, D., 1974, Judgment under uncertainty: heuristics and biases, *Science*, **185**, 1124–31.

ULICH, E., 1998, *Arbeitspsychologie* (Zurich: Auflage).

ULVILA, J.W. and BROWN, R.V., 1982, Decision analysis comes of age, *Harvard Business Review*, **60**, 130–41.

US ARMY, 1991, *1988 Anthropometeric Survey of US Army Personnel: Pilot Summary Statistics*. Technical report TR-91/040 (Natick: Research, Development and Engineering Center).

VAILLANCOURT, D.R. and SNOOK, S.H., 1995, A review of machine guarding recommendations. *Applied Ergonomics*, **26**, 141–5.

VAN COTT, H.P. and KINKADE, R.G. (eds), 1972, *Human Engineering Guide to Equipment Design* (Washington, DC: Department of Defense, US Government Printing Office).

VAN GIGCH, J.F., 1974, *Applied General System Theory* (New York: Harper & Row).

VASSILIOU, Y. (ed.), 1984, *Human Factors and Interactive Computer Systems* (Norwood: Ablex).

VEITCH, J.A. and NEWSHAM, G.R., 1998, Determinants of lighting quality I: State of the science, *Journal of the Illuminating Engineering Society*, **27**, 92–106.

VEITCH, J.A. and NEWSHAM, G.R., 1998, Lighting quality and energy-efficiency effects on task performance, mood, health, satisfaction and comfort, *Journal of the Illuminating Engineering Society*, **27**, 107–29.

VENDA, V.F. and VENDA, Y.V., 1995, *Dynamics in Ergonomics, Psychology and Decisions: Introduction to Ergodynamics* (Norwood: Ablex).

VENTURINO, M. (ed.), 1990, *Selected Readings in Human Factors* (Santa Monica: Human Factors Society).

VERBEEK, J., 1991, The use of adjustable furniture: evaluation of an instruction programme for office workers, *Applied Ergonomics*, **22**, 179–84.

VICENTE, K.J., 1990, A few implications of the ecological approach to human factors, *Human Factors Society Bulletin*, **33**, 1–4.

VICENTE, K.J., 1999, *Coignitive Work Analysis: Toward Safe, Productive, and Health Computer-based Work* (Mahwah: Erlbaum).

VICENTE, K.J. and RASMUSSEN, J., 1990, The ecology of human-machine systems II: Mediating 'direct perception' in complex work domains, *Ecological Psychology*, **2**, 207–50.

VICENTE, K.J. and RASMUSSEN, J., 1992, Ecological interface design: theoretical foundations, *IEEE Transactions on Systems, Man, and Cybernetics*, **22**, 589–606.

VIDULICH, M.A., 1989, The use of judgment matrices in subjective workload assessment: the Subjective WORkload Dominance (SWORD) technique. In *Proceedings of the Human Factors Society 33rd Annual Meeting* (Santa Monica: Human Factors Society), 1406–10.

VILLANUEVA, M., SOTOYAMA, M., JONAI, H., TAKEUSHI, Y. and SAITO, S., 1996, Adjustments of posture and viewing parameters of the eye to changes in the screen height of the visual display terminal, *Ergonomics*, **39**, 933–45.

VINK, P., KONINGSVELD, E.A. and DHONDT, S. (eds), 1998, *Human Factors in Organizational Design and Management – VI* (Amsterdam: North Holland).

VOLINN, E., 1999, Do workplace interventions prevent low-back disorders? If so, Why? A methodologic commentary, *Ergonomics*, **42**, 258–72.

VOLINN, E., VAN KOEVERING, D., LOESER, J.D., 1991, Back sprain in industry: the role of socioeconomic factors in chronicity, *Spine*, **16**, 542–8.

WALL, T.D., CLEGG, C.W. and KEMP, N.J. (eds), 1987, *The Human Side of Advanced Manufacturing Technology* (Chichester: Wiley).

WEBB ASSOCIATES, 1978, *Anthropometric Source Book*, vol. 1. NASA reference publication 1023 (Washington, DC: NASA, Scientific and Information Service).

WEIMER, J., 1993, *Handbook of Ergonomic and Human Factors Tables* (Englewood Cliffs: Prentice Hall).

WEIMER, J., 1995, *Research Techniques in Human Engineering* (Englewood Cliffs: Prentice-Hall).

WESTGAARD, R.H., 1988, Measurement and evaluation of postural load in occupational work situations, *European Journal of Physiology*, **57**, 291–304.

WICKENS, C.D., GORDON, S.E. and LIU, Y., 1998, *An Introduction to Human Factors Engineering* (New York: Addison-Wesley Longman).

WICKENS, C.D. and HOLLANDS, J.G., 2000, *Engineering Psychology and Human Performance* (New Jersey: Prentice-Hall).

WICKENS, C.D. and HUEY, B.M., 1993, *Workload Transition* (Washington, DC: National Academy Press).

WICKENS, C.D. and LIU, Y., 1988, Codes and modalities in multiple resources: a success and a qualification, *Human Factors*, **30**, 599–616.

WICKENS, C.D., SANDRY, D.L. and VIDULICH, M., 1983, Compatibility and resource competition between modalities of input, central processing, and output, *Human Factors*, **25**, 227–48.

WICKENS, C.D., VIDULICH, M. and SANDRY-GARZA, D., 1984, Principles of S-C-R compatibility with spatial and verbal tasks: the role of display-control location and voice-interactive display-control interfacing, *Human Factors*, **26**, 533–43.

WEINER, E.L., 1988, Cockpit automation. In E.L. Wiener and D.C. Nagel (eds), *Human Factors in Aviation* (San Diego: Academic Press), 433–59.

WIENER, E.L. and CURRY, R.E., 1980, Flight-deck automation: promises and problems, *Ergonomics*, **23**, 995–1011.

WILKINS, A.J., 1986, Intermittent illumination from visual display units and fluorescent lighting affects movements of the eyes across text, *Human Factors*, **28**, 75–81.

WILLIAMS, T.A., 1989, *Computers, Work, and Health: A Socio-technical Approach* (London: Taylor & Francis).

WILSON, J.R. and CORLETT, E.N. (eds), 1995, *Evaluation of Human Work: A Practical Ergonomics Methodology* (Bristol: Taylor & Francis).

WILSON, J.R., CORLETT, E.N. and MANENICA, I. (eds), 1987, *New Methods in Applied Ergonomics: The Proceedings of the Second International Occupational Ergonomics Symposium* (London: Taylor & Francis).

WOODSON, W.E., 1987a, *Human Factors Reference Guide for Electronics and Computer Professionals* (New York: McGraw-Hill).

WOODSON, W.E., 1987b, *Human Factors Reference Guide for Process Plants* (New York: McGraw-Hill).

WOODSON, W.E., TILLMAN, B. and TILLMAN, P., 1992, *Human Factors Design Handbook: Information and Guidelines for the Design of Systems, Facilities, Equipment, and Products for Human Use* (New York: McGraw-Hill).

YATES, F.A., 1966, *The Art of Memory* (London: Routledge & Kegan Paul).

YOUNG, L., 1969, On adaptive manual control, *IEEE Transactions on Man–Machine Systems*, **10**, 292–331.

ZEITLIN, L.R. and FINKLEMAN, J.M., 1975, Research note: subsidiary task techniques of digit generation and digit recall as indirect measures of operator loading, *Human Factors*, **17**, 218–20.

ZINK, K.J., 1994, *WettbewerbsfÑhigkeit durch innovative Strukturen und Konzepte, Festschrift zum 80* (Munich: Geburtstag von GÅnter RÅhl).

ZINK, K.J. (ed.), 1996, *Business Excellence through TQM* (Basingstoke: Gower).

ZINK, K.J. (ed.), 1997, *Succesful TQM — Inside Stories from European Quality Award Winners* (Basingstoke: Gower).

ZINK, K.J., 1999, Safety and quality issues as part of a holistic (i.e., sociotechnological) approach, *International Journal of Occupational Safety and Ergonomics*, **5**, 279–90.

ZURADA, J., KARWOWSKI, W. and MARRAS, W.S., 1997, A neural network-based system for classification of industrial jobs with respect to the risk of low back disorders, *Applied Ergonomics*, **28**, 49–58.